

Annotation of *Helicobacter pylori* at Locus Tags HP1538, HP1260, HP1010 and HP0828

Luke George, Emily Gersitz, Grace Gominiak, Elisabeth Horvath, and Laura O'Donnell

Attica High School- Attica, NY and The Western New York Genetics in Research and Health Care Partnership



Abstract

A group of genes from the microorganism *Helicobacter pylori* were annotated using the collaborative genome annotation website GENI-ACT. The GenBank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, and cellular localization data. The GenBank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the computer database.

Introduction

Helicobacter pylori is a gram-negative, spiral shaped bacteria that can enter and live in your digestive tract. After being infected for many years, they can cause ulcers in the stomach and small intestine. For some, being infected with *H. pylori* can lead to stomach cancer. An infection with this bacteria is very common, with about two-thirds of the world's population having it in their bodies. However, for most of the population, *H. pylori* doesn't cause ulcers or any other symptoms, such as bloating, burping, nausea, vomiting, and unconditional weight loss. As more and more of the world is receiving sanitation and clean water, *H. pylori* is becoming less and less popular.

Once *H. pylori* enters your body, it uses its flagella to move through the mucus lining and attack the epithelial cells of your stomach. When it has done enough damage, the acid in your stomach can get through the lining, leading to ulcers. These ulcers can cause bleeding, infection, and the inability for food to travel through your digestive tract. *H. pylori* is most common in places with bad water and sewage systems, but can still be spread through food, water, utensils, and bodily fluids from infected people. Most are infected during childhood, but it can still be contracted in adulthood. The germs can even live in your body for years before any symptoms start. Ulcers caused by *H. pylori* will need treatment, such as antibiotics to kill the bacteria.

Colored scanning electron micrograph of *H. pylori* on surface of gastric cells
web.stanford.edu

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Helicobacter pylori* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3- Structure-Based Evidence	TIGRFam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?

Results

Gene Locus: HP1538		
Basic Information	DNA Coordinates	complement(1615878..1616735)
	DNA Length	858 nt
	Amino Acid Sequence Length	285 aa
Sequence-based Similarity	COGS	COG2857- Cytochrome C
	T-Coffee	Very well conserved until the very end
	Web-Logo	Well conserved until position 105, then not conserved at all
Structure-based Evidence	TIGRFAMs	No significant Hits
	PFAMs	pfam00034 cytochrome c
	HMM Logo Key Residues	C12, C15, H16, Y87
	PDB	cytochrome C
Cellular Localization	Transmembrane Helices	2
	Signal Peptide	none
	PSORTb Final Prediction	Periplasmic

Helicobacter pylori HP1538:

COG, PFAM, and PDB data supports HP1538 to be a cytochrome C protein which is an electron acceptor in the electron transport chain portion of cellular respiration. Cellular localization predicts that this protein has two transmembrane helices and functions within the periplasm.

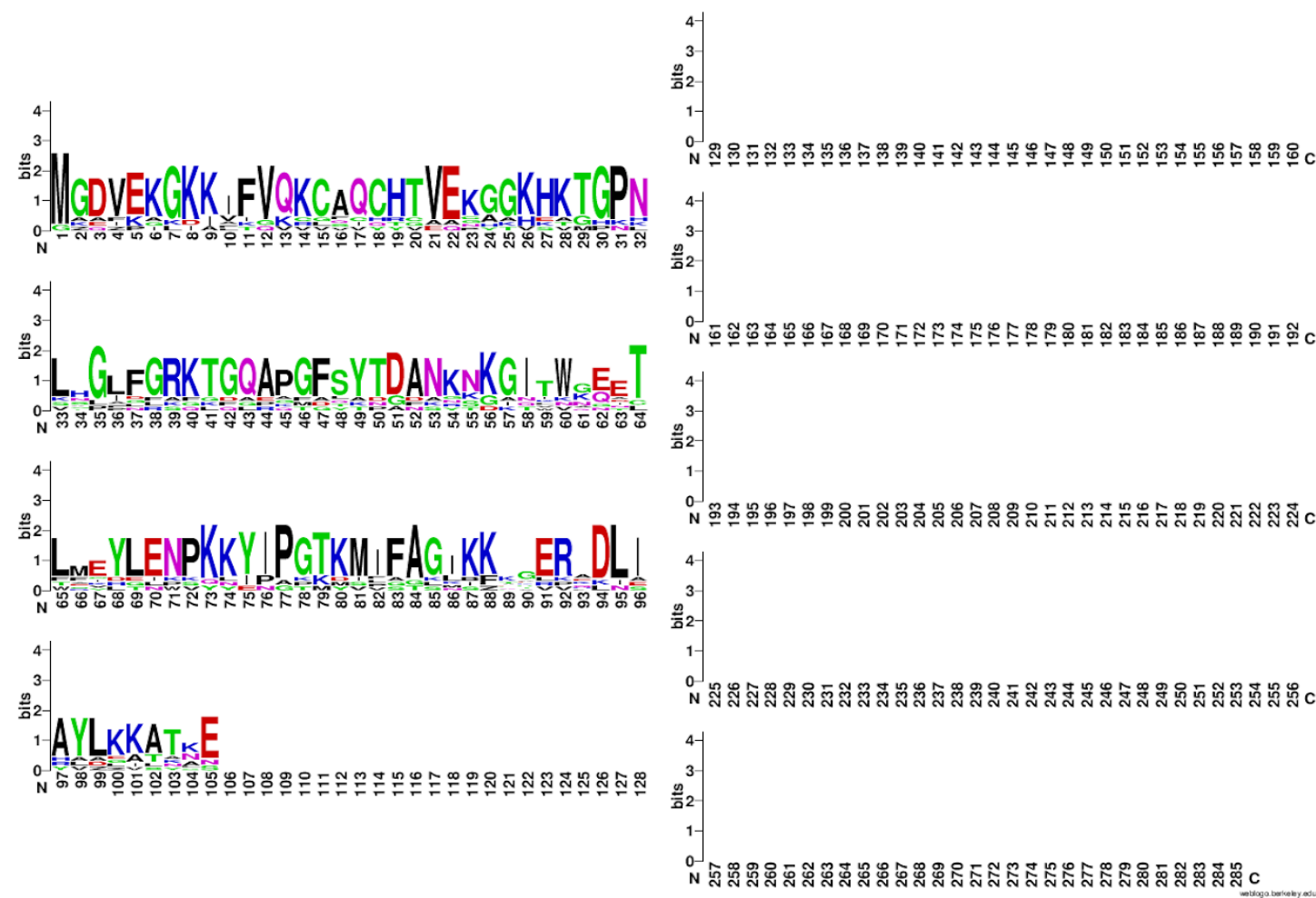


Figure 1 – WebLogo data for HP1538 shows the protein to be well conserved until position 105 and then is not conserved at all until end.

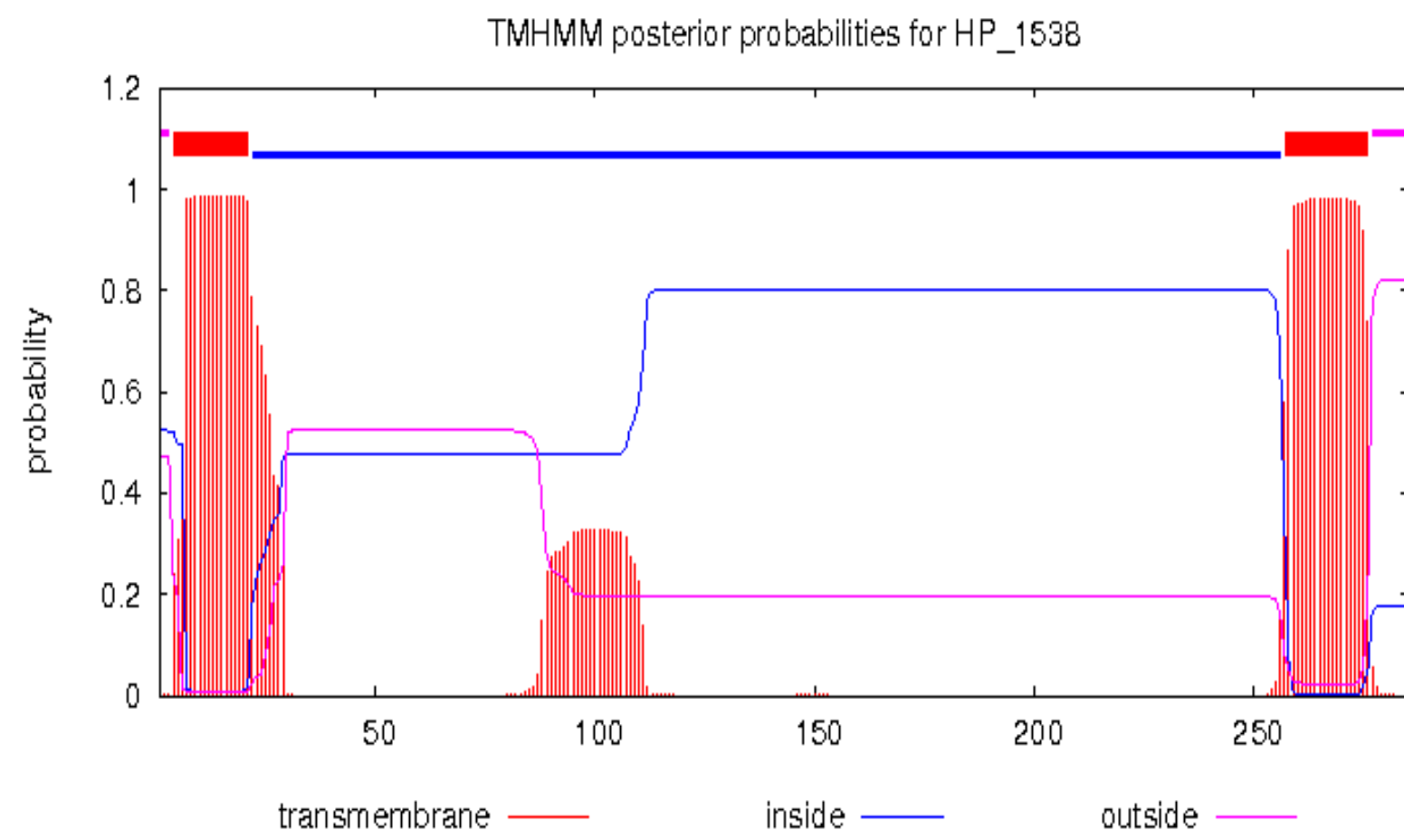


Figure 2 – TMHMM data for HP1538: The cellular localization data shows the presence of 2 transmembrane helices.

Gene Locus: HP1260		
Basic Information	DNA Coordinates	1333813..1334214
	DNA Length	402 nt
	Amino Acid Sequence Length	133 aa
Sequence-based Similarity	COGS	COG0838: NuoA NADH:ubiquinone oxidoreductase subunit 3 (chain A) [Energy production and conversion];
	T-Coffee	Well conserved until the very end
	Web-Logo	Fairly well conserved until the end
Structure-based Evidence	TIGRFAMs	No significant TIGRFAM hits
	PFAMs	pfam00507, Oxidored_q4
	HMM Logo Key Residues	Y24, G27, F42, F49, F80, Y90, Y99
	PDB	photosynthetic complex I
Cellular Localization	Transmembrane Helices	3
	Signal Peptide	none
	PSORTb Final Prediction	Cytoplasmic Membrane

Helicobacter pylori HP1260:

COG, PFAM, and PDB data supports HP1260 to be an enzyme that works with NADH and NADPH along with other electron transporters that are part of the cell respiration process involving the electron transport chain. Cellular localization data predicts this protein to have three transmembrane helices.

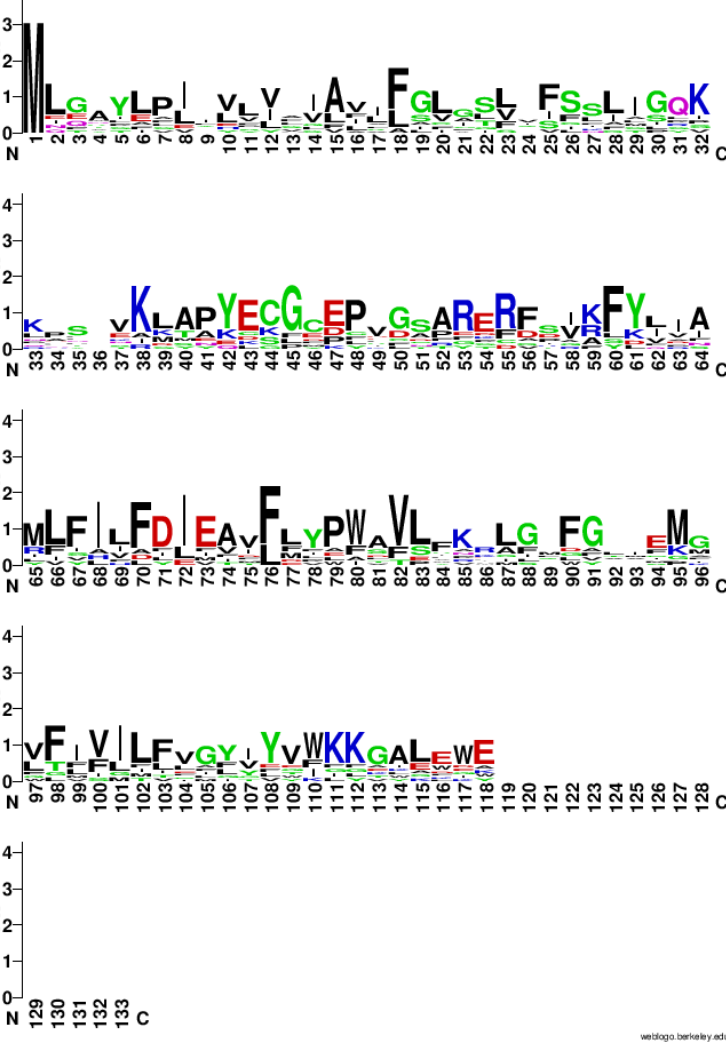


Figure 3 – WebLogo data for HP1260 shows the protein to be fairly well conserved throughout

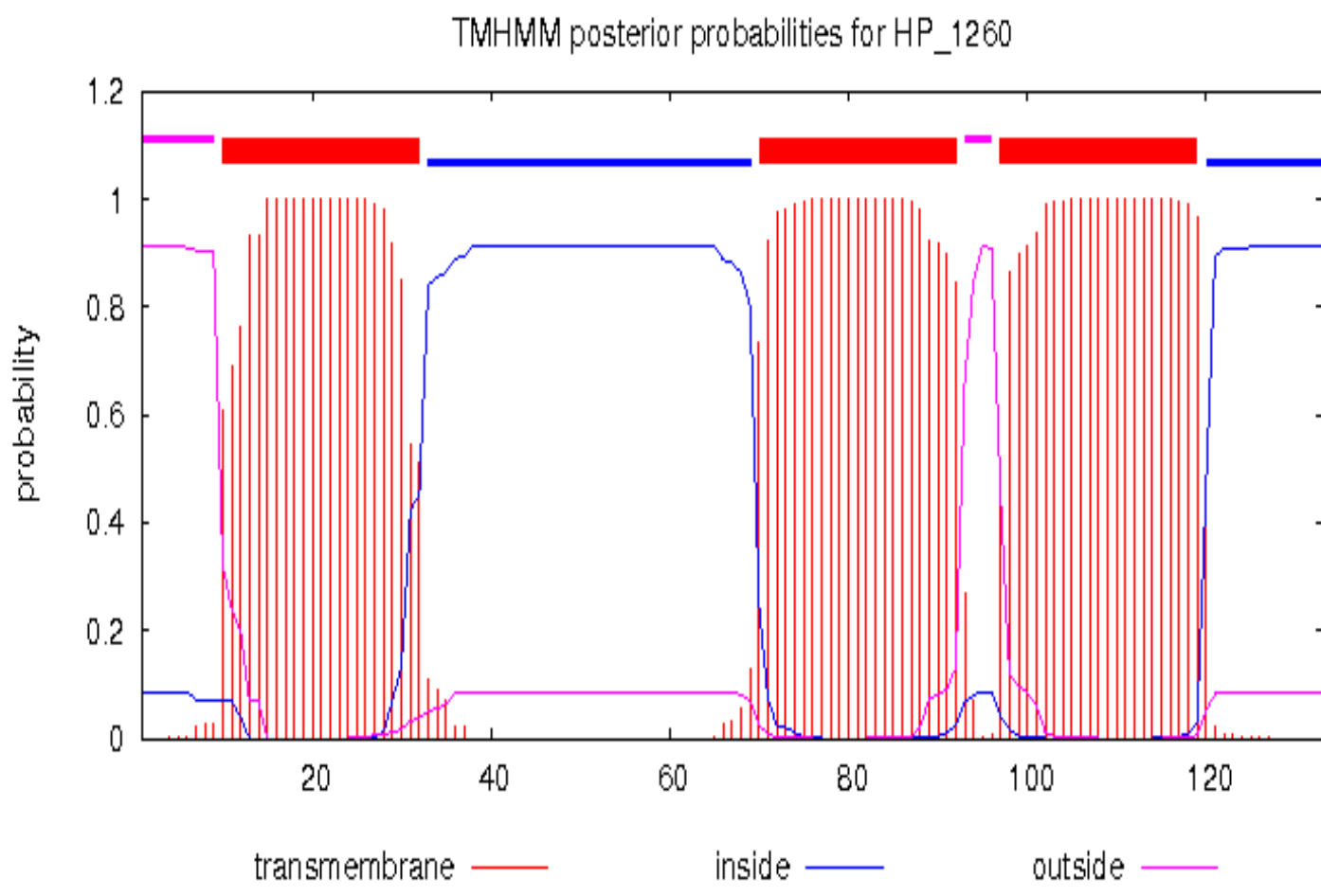


Figure 4 – TMHMM data for HP1260: The cellular localization data shows the presence of 3 transmembrane helices.

Gene Locus: HP1010		
Basic Information	DNA Coordinates	1072429..1074456
	DNA Length	2028 nt
	Amino Acid Sequence Length	675 aa
Sequence-based Similarity	COGS	No significant COG hits
	T-Coffee	Well conserved until the very end
	Web-Logo	Not well conserved
Structure-based Evidence	TIGRFAMs	No significant TIGRFAM hits
	PFAMs	No significant PFAM hits
	HMM Logo Key Residues	No key residues
	PDB	Polyphosphate Kinase
Cellular Localization	Transmembrane Helices	None
	Signal Peptide	None
	PSORTb Final Prediction	Cytoplasmic Membrane

Helicobacter pylori HP1010:

There is little evidence to confirm the function of HP1010, however PDB data suggests that it may be a polyphosphate kinase which catalyzes the formation of ADP from ATP. This protein has an undetermined location due to conflicting data. The TMHMM and Phobius results are predicting that this protein may be found in the cytoplasm, while PSORT-B is predicting that the protein is located in the cytoplasmic membrane. This may be occurring because the formation of ADP and polyphosphate from ATP can be happening in multiple locations where energy is needed.

Gene Locus: HP0828		
Basic Information	DNA Coordinates	complement(879963..880643)
	DNA Length	681 NT
	Amino Acid Sequence Length	226 AA
Sequence-based Similarity	COGS	COG0356FoF1-type ATP synthase, membrane subunit a [Energy production and conversion
	T-Coffee	Well conserved until the very end
	Web-Logo	Fairly well conserved throughout
Structure-based Evidence	TIGRFAMs	TIGR01131: ATP_synt_6_or_A: ATP synthase subunit 6
	PFAM	pfam00119: ATP synthase A chain
	HMM Logo Key Residues	R147, N151, A154, G155, Q198, F202, L205, Y209
	PDB	ATP Synthase
Cellular Localization	Transmembrane Helices	6
	Signal Peptide	none
	PSORTb Final Prediction	cytoplasmic membrane

Helicobacter pylori HP0828:

COG, TIGRFAM, PFAM, and PDB data supports HP0828 to be the enzyme ATP synthase which synthesizes ATP from ADP for energy. Cellular localization data predicts that HP0828 has 6 transmembrane helices, residing in the cytoplasmic side of the cell membrane.

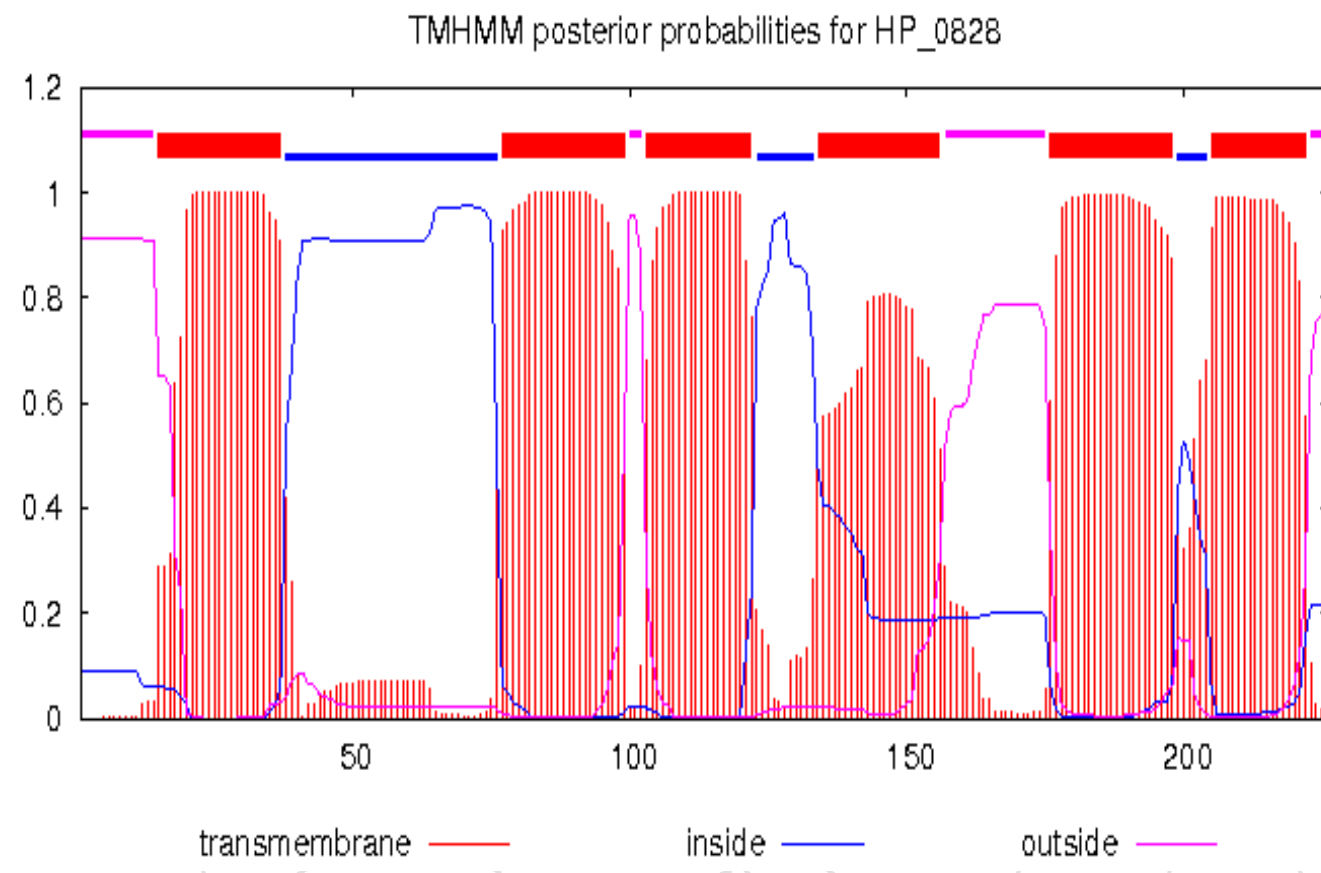
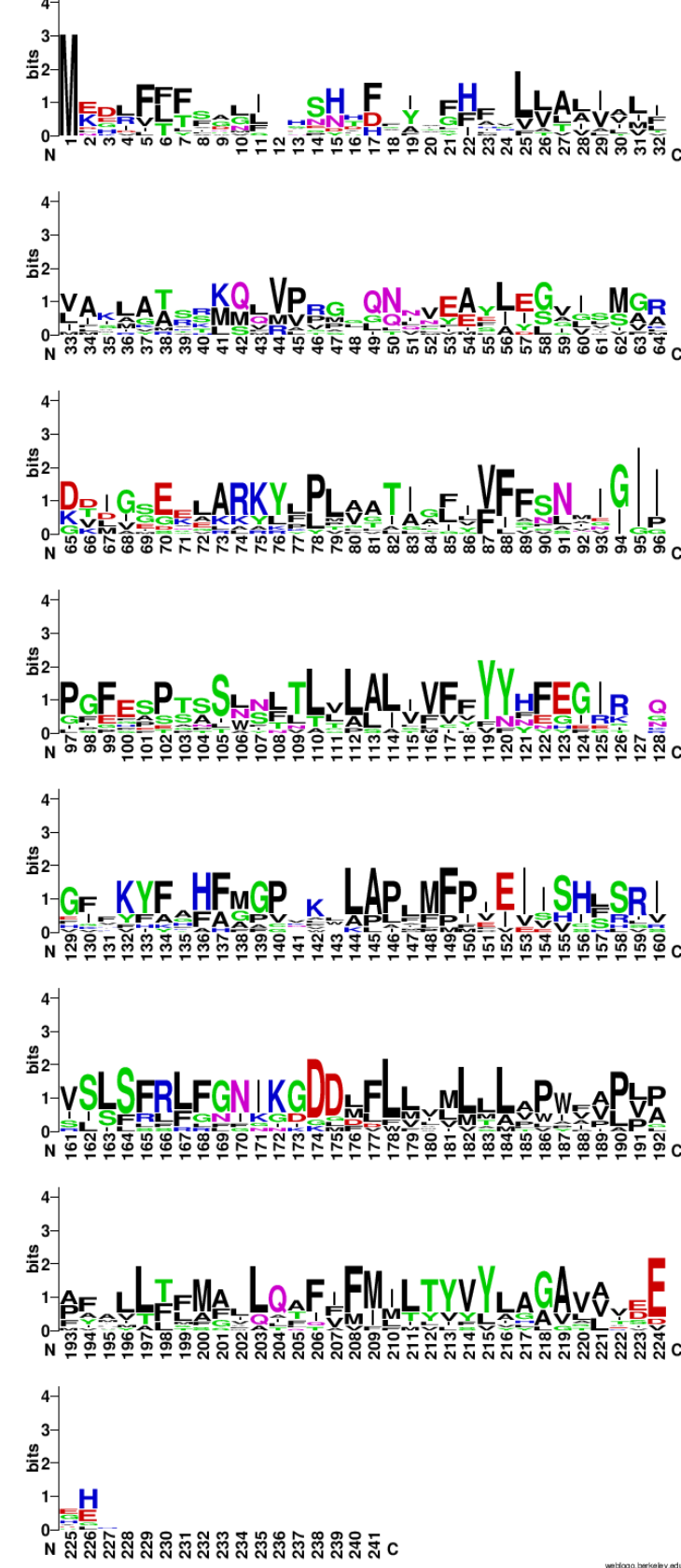


Figure 5 (above) – TMHMM data for HP0828: The cellular localization data shows the presence of 6 transmembrane helices.

Figure 3 (to the left) – WebLogo data for HP0828 shows the protein to be fairly well conserved throughout

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the computer database. All proteins annotated seem to be involved with the electron transport chain for energy in *H. pylori*.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
HP1538	Cytochrome C protein	Cytochrome C protein
HP1260	NuoA: NADH:ubiquinone oxidoreductase subunit 3	NuoA: NADH:ubiquinone oxidoreductase subunit 3
HP1010	Polyphosphate kinase	Polyphosphate kinase
HP0828	ATP Synthase	ATP Synthase

References

H. pylori Bacteria Infection: Symptoms, Diagnosis, Treatment, Prevention. (n.d.). Retrieved from <https://www.webmd.com/digestive-disorders/h-pylori-helicobacter-pylori#1>

Helicobacter pylori (*H. pylori*) infection. (2017, May 17). Retrieved from <https://www.mayoclinic.org/diseases-conditions/h-pylori/diagnosis-treatment/drc-20356177>

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu