Winter 2-6-2016

Bioinformatics Comparison of *M. ruber* Mrub_2507 to *E. coli* pdxK/b1636 and *M. ruber* Mrub_2888 to *E. coli* pdxH/b1638 to Determine the Orthologous Nature

Adam Bernardi
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College, Rock Island Illinois

Bioinformatics Comparison of *M. ruber* Mrub_2507 to *E. coli pdxK*/b1636 and *M. ruber* Mrub_2888 to *E. coli pdxH*/b1638 to Determine the Orthologous Nature

Adam Bernardi

Introduction

Meiothermus ruber (*M. ruber*), a gram-negative bacteria, is a member of the Deinococcus-Thermus phylum. The Deinococcus-Thermus phylum is found on a very poorly studied branch of the Tree of Life (Tindall *et al.*, 2010). Studying bacteria, like *M. ruber*, from these poorly studied branches is very important is vital to increasing our understanding of life. These studies often yield novel discoveries that expand or understanding about an organism's role in different biological processes including energy production, bioremediation, the global nutrient cycle, and pathogenesis. These discoveries also increase our understanding and knowledge of the diversity of life and allow us to draw connections between different organisms expanding the understanding of the evolutionary descent of organisms. (Phylogenetic Diversity). To study organisms like *M. ruber* a model organism is needed, that is why *Escherichia coli* (*E. coli*) was chosen as a model organism. *E. coli* is a great model organism to use for comparisons because it has been widely studied and the complete genome has been sequenced. It is also believed that *E. coli* has genes and proteins that are orthologs to genes and proteins in other species, such as *M. ruber*.

Vitamin B_6 biosynthesis and salvage is vital for the proper function of living organisms (Yang *et al.*, 1998). A main product of the vitamin B_6 salvage pathway is pyridoxal 5'phosphate (PLP) which is the active form of vitamin B_6 (Allgood *et al.* 1993). PLP is an essential cofactor that is necessary for the proper running of over 100 cellular enzymes. Further increasing the

importance of PLP, many of these enzymes are involved in amino acid synthesis, as well as being a coenzyme involved in neurotransmitter synthesis, steroid-receptor interaction modulation, and immune regulation. (di Salvo *et al.*, 2003). Studies of the vitamin B_6 salvage pathway in *E. coli* and it has been found that PLP can be synthesized through the utilization of pyridoxine (PN), pyridoxamine (PM), or pyridoxal (PL) (Figure 1) (Yang *et al.*, 1996). The phosphorylation of PL, PN, and PM to PLP, PNP, and PMP respectively is catalyzed by the enzymes PL kinase, PN kinase, and PM kinase encoded by the *pdxK*/b2418 and *pdxY*/b1636 genes. These two genes appear to be paralogs and provide very similar functions, however, previous research has found that *pdxK* encodes for a kinase that can act as a PL/PN/PM kinase, although its PN kinase activity is much higher while it has a lower level of PL kinase activity, whereas *pdxY* encodes for PL kinase that expresses low levels of PN kinase activity (Yang *et al.*, 1998). After the formation of PNP and PMP the enzymes PNP oxidase and PMP oxidase, encoded for by the gene *pdxH*/b1638, oxidize PNP and PMP, respectively, to form PLP (di Salvo *et al.*, 2003).

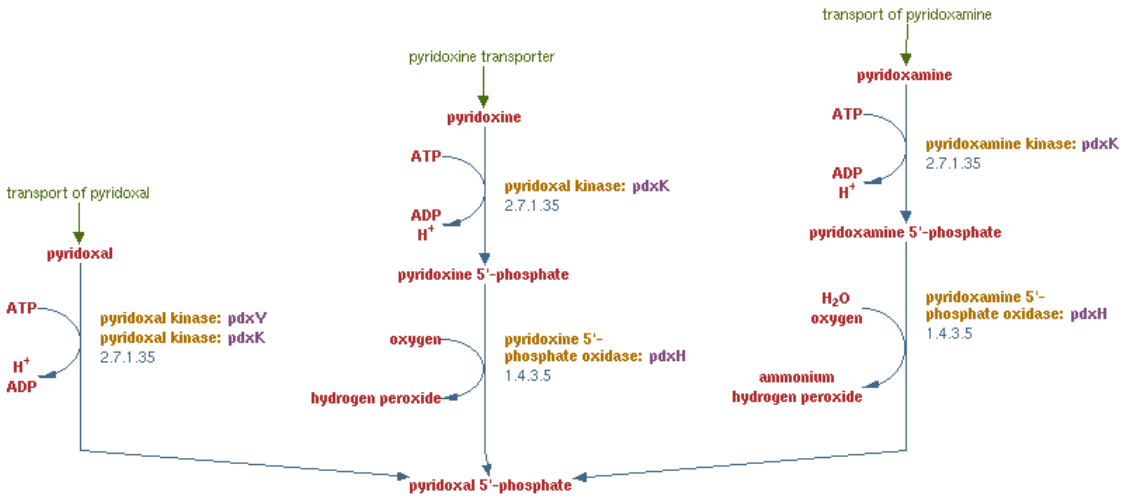


Figure 1. Pyridoxal 5'-phosphate salvage pathway for *E. coli*. EcoCyc *E. coli* Database (http://ecocyc.org/) generated PLP salvage pathways.

Bioinformatics tools allow for the analysis of the genome, amino acid sequence, and function of proteins in different species. Bioinformatics is the branch of science that allows for the analysis, comparison, and interpretation of biological data. The possible orthologous nature of proteins in different species can be determined through the use of these tools. The use of these tools greatly improves the study of living organisms because the understanding and knowledge of biological processes in different organisms increases. The almost instant feedback provided by these tools on the differences between sequences makes identifying mutants easily observable. The continued use and refinement of these tools could lead to the possibility of individualized medicine, leading to improved healthcare (Pujari).

This system in *E. coli* was chosen to be compared to a possible similar system in *M. ruber*. *E. coli* PdxK and PdxY were compared against the *M. ruber* genome because PdxK and PdxY play a role in the vitamin B_6 salvage pathway (Yang *et al.*, 1998). However, after a comparison BLAST search of *E. coli* PdxK/b2418 and PdxY/b1636 against Mrub_2507 it was determined that PdxY was the more likely identity of the possible ortholog in *M. ruber* because the comparison of PdxY to the *M. ruber* genome yielded a score of 236 bits and E-value of 1e-76 and percent identity of 43% (Table 1), whereas the comparison of PdxK to *M. ruber* yielded a score of 130 bits and E-value of 4e-36 and percent identity of 32%. Complete annotation was conducted for PdxY and PdxK, yielding almost identical results and consistently smaller bit scores and larger E-values. Therefore, only the results of the comparison of *E. coli* PdxY to Mrub_2507 were included in detail while only essential data will be presented comparing *E. coli* PdxK/b2418 to Mrub_2507. To compare the genes of *E. coli* and *M. ruber* the bioinformatics tools used were BLAST, CCD, T-Coffee, WebLogo, TMHMM, SignalP, LipoP, PSORT-B, Phobius, TIGRFAM, Pfam, PDB, IMG/EDU, KEGG, and MetaCyc. The proposed *M. ruber*

genes, Mrub_2507 (PdxY) and Mrub_2888 (PdxH), were hypothesized to be orthologous to *E. coli* PdxY/b1636 and PdxH/b1638, respectively.

Methods

The bioinformatics tools outlined in the GENI-ACT lab notebook (http://www.geni-act.org/education/main/) were used to compare Mrub_2507 to *E. coli* PdxY/b1636 and Mrub_2888 to *E. coli* PdxH/b1638, with the following changes. To generate the Vitamin B_6 salvage pathway EcoCyc, not MetaCyc was used for *E. coli*. The first BLAST performed for *E. coli* was against *M. ruber* to determine the possible presence of an orthologous gene in *M. ruber*. When the T-Coffee analysis was run fifteen sequences were used instead of just ten.

Results

Table 1: *E. coli* PdxY/b1636 and Pdxk/b2418 are paralogs and *E. coli* PdxY/b1636 and Mrub_2507 are orthologs

Evidence Collected	PdxY/b1636	PdxK/b2418	Mrub_2507
BLAST *E. coli* against *M. ruber*	Score = 236 E-value = 1e-76 % identity = 43%	Score = 130 E-value = 4e-36 % identity = 32%	
KEGG pathway	Vitamin B6 Metabolism		
E.C. number	E.C.2.7.1.35 Pyridoxal kinase		
LipoP (cellular location)	Cytoplasm		
PSORT-B (cellular location and score)	Cytoplasm (score:9.97)	Cytoplasm (score:9.26)	Cytoplasm (score:9.97)
CDD (COG category)	COG2240 E = 1.17e-146 pdxK (Pyridoxal/pyridoxamine/ pyridoxamine kinase)	COG2240 E = 2.03e-117 pdxK (Pyridoxal/pyridoxamine/ pyridoxamine kinase)	COG2240 E = 1.40e-114 pdxK (Pyridoxal/pyridoxamine/ pyridoxamine kinase)
TIGRFAM – protein family	TIGR00687 Score = 499.9 E = 4.3e-147 pdxK: pyridoxal kinase	TIGR00687 Score = 131.1 E = 4.6e-36 pdxK: pyridoxal kinase	TIGR00687 Score = 411.7 E = 1.5e-120 pdxK: pyridoxal kinase
Pfam – protein family	PF08543 Phosphomethylpyrimidine kinase	PF08543 Phosphomethylpyrimidine kinase	PF08543 Phosphomethylpyrimidine kinase

	Score = 57.4, E = 1.2e-15	Score = 89.4, E = 2e-25	Score = 54.9, E = 7e-15
PDB	1TD2 E = 1.53744e-158 pdxY	2DDM E = 5.89994e-163 Pyridoxal kinase	1VI9 E = 1.87369e-62 Pyridoxamine kinase

Using the bioinformatics tools outlined in the GENI-ACT lab notebook the analysis of information regarding the protein sequence alignments, cellular location, and protein family and cellular pathways they are involved in was conducted.

The amino acid sequence alignment of *E. coli* PdxY and Mrub_2507 were initially compared through a BLAST comparison yielding a score of 236 bits and E-value of 1e-76 and percent identity of 43% (Table 1). The Pfam analysis of the sequences yielded a pairwise alignment (Figure 2), showing several conserved amino acids in both sequences. The conserved amino acids are G85 and G83, D115 and D113, P116 and P114, V117 and V115, P148 and P146, N149 and N147, E152 and E 150, L156 and L154, V186 and V180, G227 and G222, G229 and G224, and A253 and A244 in Mrub_2507 and *E. coli* PdxY respectively.

Sequence A

Sequence B

Sequence A

Sequence B

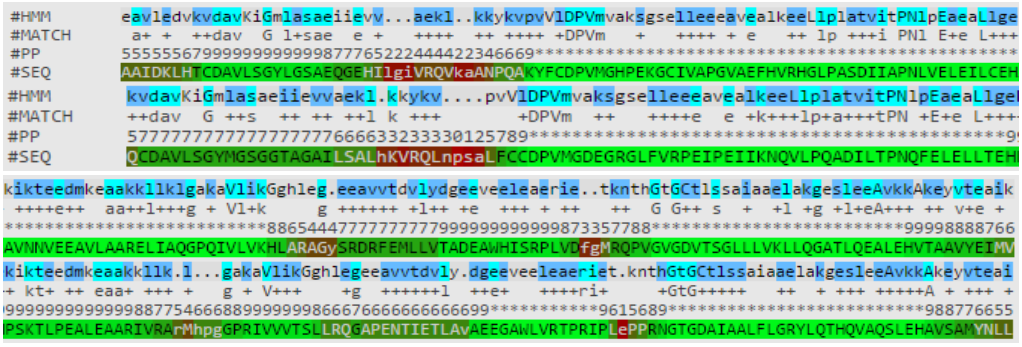


Figure 2. *E. coli* PdxY and Mrub_2507 contain similar conserved amino acids. Sequence A = *E. coli* PdxY/b1636; Sequence B = Mrub_2507. #HMM = consensus protein sequence; #SEQ = bacteria amino acid sequence; #MATCH = match between bacteria and consensus sequence. Search Pfam (http://pfam.xfam.org/search) produced pairwise sequence alignment.

Next the cellular pathways and protein family of Mrub_2507 and *E. coli* PdxY were analyzed. KEGG was used to confirm the presence of Mrub_2507 and *E. coli* PdxY in the vitamin B_6 metabolism pathway (Figure 3). The E.C. number and protein names (Table 1) were found using the KEGG pathway and they were found to be E.C.2.7.1.35 and pyridoxal kinase for

both Mrub_2507 and *E.coli* PdxY. The CDD analysis (Table 1) of Mrub_2507 and *E. coli* PdxY were returned the same COG number, COG2240, with significant E-values, 1.17e-146 for *E. coli* PdxY and 1.40e-114 for Mrub_2507. The COG analysis returned that both genes were identified as PdxK pyridoxal/pyridoxine/pyridoxamine kinase. Pfam analysis of the Mrub_2507 and *E. coli* PdxY also returned results for the genes that were identical (Table 1). Both Mrub_2507 and *E. coli* PdxY returned Pfam numbers of PF08543, both were from the same family phosphomethylpyrimidine kinase, and both returned relatively large bit scores and small E-values, 54.9 and 7e-15 for Mrub_2507 and 57.4 and 1.2e-15 for *E. coli* PdxY. TIGRFAM analysis of Mrub_2507 and *E. coli* pdxY (Table 1) returned the same TIGRFAM number, TIGR00687, with accompanying large bit scores and significant E-values, 411.7 and 1.5e-120 for Mrub_2507 and 499.9 and 4.3e-147 for *E. coli* PdxY, and the same name PdxK: pyridoxal kinase. PDB analysis yielded different PDB codes, 1VI9 for Mrub_2507 and 1TD2 for *E. coli* PdxY. The name returned for Mrub_2507 was pyridoxamine kinase and for *E. coli* PdxY it was PdxY, and the E-values for these were significant, 1.87369e-62 for Mrub_2507 and 1.53744e-158 for *E. coli* PdxY. A comparison of the gene context (Figure 4) using the IMG/EDU database confirmed that *E. coli pdxY* are found in an operon that also contains other *E. coli* gene of interest *E. coli pdxH* (Yang *et al.*, 1998). Mrub_2507 was found to have genes near it on the chromosome, however, without further study it cannot be said for certain whether or not the Mrub_2507 gene is contained within an operon.

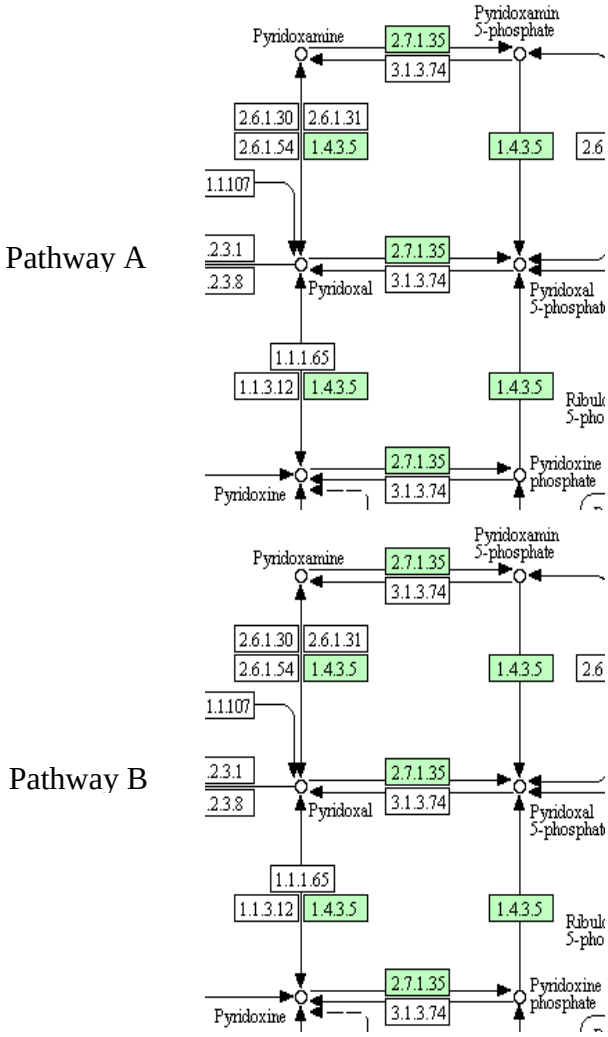


Figure 3. Vitamin B_6 Metabolism pathways for *E. coli* and *M. ruber* shows the presence of PdxY (E.C.2.7.1.35) and PdxH (E.C.1.4.3.5). Pathway A = *E. coli*, Pathway B = *M. ruber*. Proteins highlighted green produced by bacteria. KEGG PATHWAY Database (http://www.genome.jp/kegg/pathway.html) created vitamin B_6 metabolism pathways.

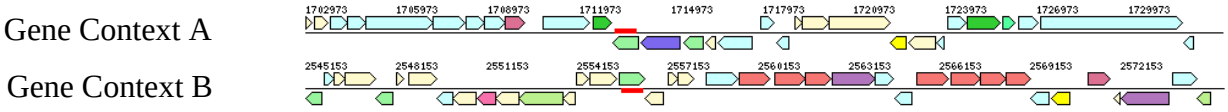
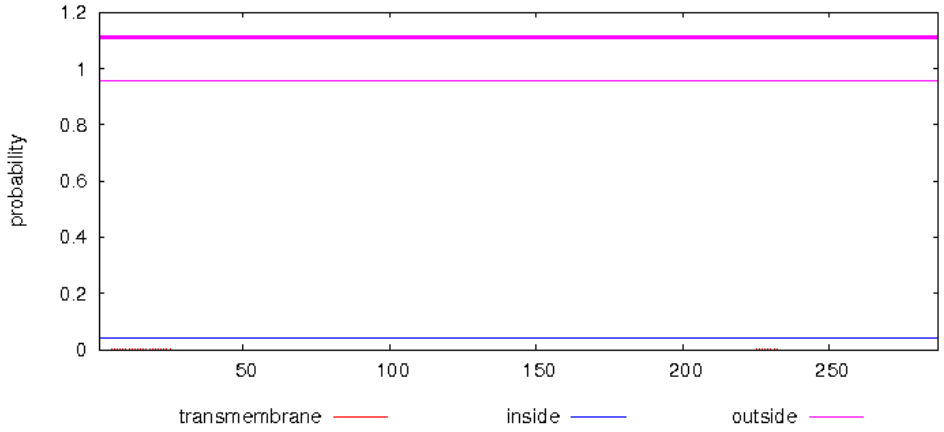


Figure 4. *E. coli* PdxY is part of an operon and Mrub_2507 may be in an operon. Gene Context A = *E. coli* PdxY/b1636; Gene Context B = Mrub_2507. Red line above or below sequence indicates desired gene. Search IMG/EDU (http://img.jgi.doe.gov/cgi-bin/edu/main.cgi) produced gene context.

The cellular location of Mrub_2507 and *E. coli* PdxY were analyzed. The TMHMM analysis (Figure 5) was used and the hydropathy plots returned for both Mrub_2507 and *E. coli* PdxY indicated that no transmembrane helices were present in either protein. The LipoP

prediction (Table 1) was that both Mrub_2507 and *E. coli* PdxY were cytoplasmic proteins. PSORT-B was then used to also predict the cellular location of the protein and this analysis returned a cytoplasmic localization score of 9.97 for both Mrub_2507 and *E. coli* PdxY. SignalP analysis (Figure 6) also indicated Mrub_2507 and *E. coli* PdxY are both located in the cytoplasm of the cell. Along with the prediction of Mrub_2507 and *E. coli* PdxY being cytoplasmic SignalP also returned a D-score of 0.231 for Mrub_2507 and 0.215 for *E. coli* PdxY, indicating the proteins are not signaling peptides. This also indicates both proteins are most likely cytoplasmic because most signaling proteins will be membranous or extracellular.

Panel A



Panel B

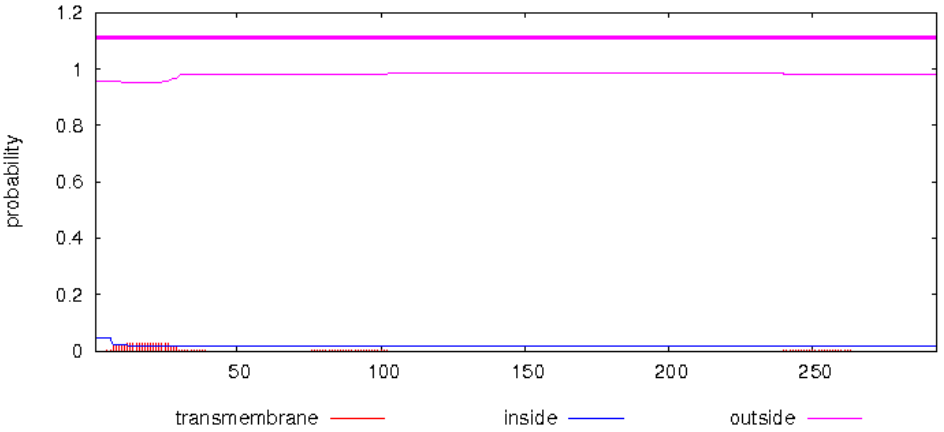
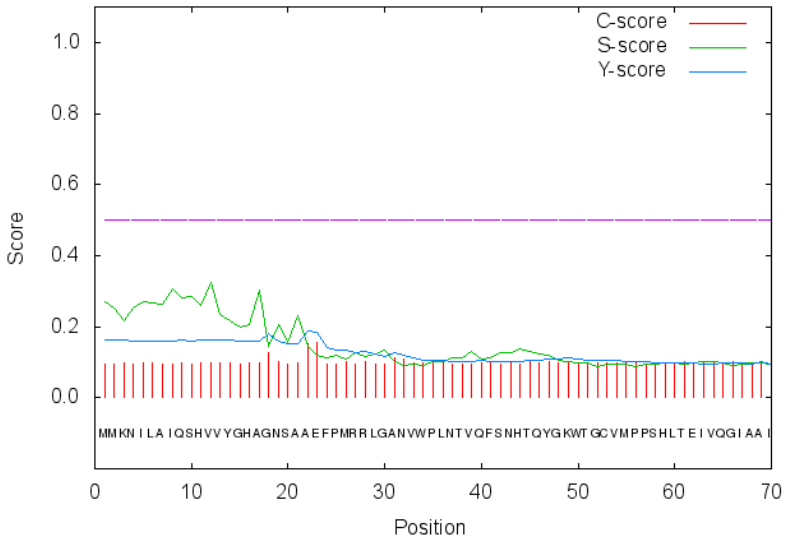


Figure 5. *E. coli* PdxY and Mrub_2507 do not contain TMH regions; enzyme predicted to be in the cytoplasm. Panel A = *E. coli* PdxY/b1636; Panel B = Mrub_2507. TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) generated hydropathy plots.

Panel A



Panel B

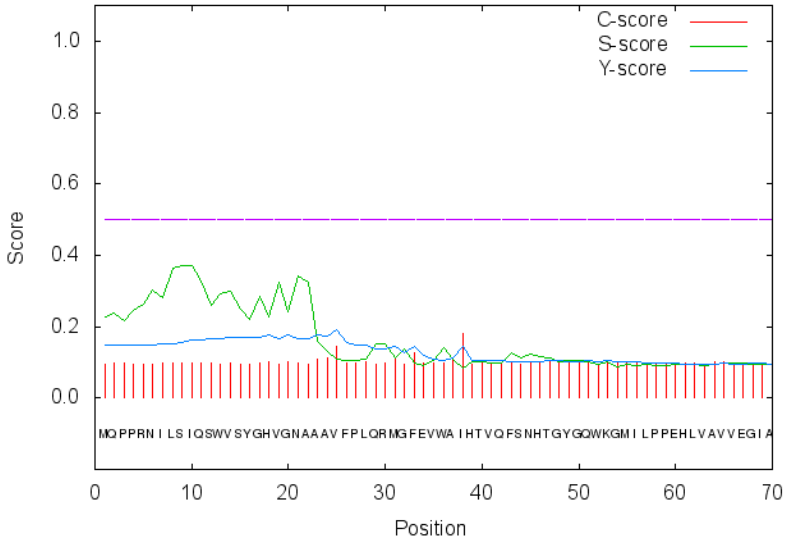


Figure 6. *E. coli* PdxY and Mrub_2507 are not signaling proteins. *E. coli* PdxY D-score = 0.215; Mrub_2507 D-score = 0.231. Panel A = *E. coli* PdxY/b1636; Panel B = Mrub_2507. SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP) generated hydropathy plot.

The cellular pathways and protein family of Mrub_2507 and *E. coli* PdxK were analyzed. KEGG was used to confirm the presence of Mrub_2507 and *E. coli* PdxK in the vitamin B_6 metabolism pathway (Figure 3). The E.C. number and protein names (Table 1) were found using the KEGG pathway and were found to be E.C.2.7.1.35 and pyridoxal kinase for both Mrub_2507 and *E.coli* PdxK. CDD analysis (Table 1) of *E. coli* PdxK returned the same COG number, COG2240, with a significant E-values, 2.03e-117 for *E. coli* PdxK, however, larger than the E-

value returned for PdxY. The COG analysis identified the protein as PdxK pyridoxal/pyridoxine/pyridoxamine kinase. Pfam analysis of *E. coli* PdxK also returned results for the genes that were identical to Mrub_2507 and PdxY (Table 1). *E. coli* PdxK returned Pfam numbers of PF08543, from the family phosphomethylpyrimidine kinase, and a relatively large bit score and small E-valuem 89.4 and 2e-25, although the bit score was smaller and the E-value was larger than those of PdxY. TIGRFAM analysis of *E. coli* pdxK (Table 1) returned the same TIGRFAM number, TIGR00687, as PdxY and Mrub_2507 with a large bit score and significant E-value, 131.1 and 4.6e-36 for *E. coli* PdxK, both less significant than the PdxY analysis, but still significant. The same name PdxK: pyridoxal kinase was also returned. A different PDB code was returned through PDB analysis, 2DDM, for *E. coli* PdxK, with the name pyridoxal kinase, and the E-value 5.89994e-163 which was about the same as that of PdxY.

The cellular location *E. coli* PdxK was analyzed. The TMHMM analysis (not pictured) was conducted and the hydropathy plots returned for *E. coli* PdxK indicated no transmembrane helices were present. The LipoP prediction (Table 1) for *E. coli* Pdxk was that it was a cytoplasmic protein. PSORT-B predicted the cellular location of the protein with a cytoplasmic localization score of 9.26 for PdxK. SignalP analysis (not pictured) also indicated *E. coli* PdxK is located in the cytoplasm of the cell. SignalP also returned a D-score of 0.141 *E. coli* PdxK, indicating the protein is also not a signaling peptide. This also indicates both proteins are most likely cytoplasmic because most signaling proteins will be membranous or extracellular.

Table 2: *E. coli* PdxH/b1638 and Mrub_2888 are orthologs

Evidence Collected	***E. coli***	***M. ruber***
BLAST *E. coli* against *M. ruber*	Score = 182 E-value = 2e-57 % identity = 45%	
KEGG pathway	Vitamin B6 Metabolism	
E.C. number	E.C.1.4.3.5 Pyridoxal 5'-phosphate synthase	

LipoP (cellular location)	Cytoplasm	
PSORT-B (cellular location and score)	Cytoplasm (score:9.97)	Cytoplasm (score:9.26)
CDD (COG category)	COG0259 E = 6.75e-130 pdxH (Pyridoxine/pyridoxamine 5'-phosphate oxidase)	COG0259 E = 6.90e-108 pdxH (Pyridoxine/pyridoxamine 5'-phosphate oxidase)
TIGRFAM – protein family	TIGR00558 Score = 447.4, E = 2.8e-131 pdxH: pyridoxamine 5'-phosphate oxidase	TIGR00558 Score = 440.6, E = 3.1e-129 pdxH: pyridoxamine 5'-phosphate oxidase
Pfam – protein family	PF08543 Pyridoxamine 5'phosphate oxidase Score = 87.2, E = 5.6e-25	PF01243 Pyridoxamine 5'phosphate oxidase Score = 86.9, E = 7e-25
PDB	1G76 E = 1.32317e-158 pyridoxine 5'-phosphate oxidase	1G76 E = 3.52844e-46 pyridoxine 5'-phosphate oxidase

The amino acid sequence alignment of *E. coli* PdxH and Mrub_2888 were initially compared through a blast comparison yielding a score of 182 bits and E-value of 2e-57 and percent identity of 45% (Table 2). Pfam analysis of the sequences yielded a pairwise alignment (Figure 7), revealing several conserved amino acids in both sequences. The conserved amino acids are T50 and T58, G55 and G63, P57 and P65, and G109 and G117 in Mrub_2888 and *E. coli* PdxH respectively.

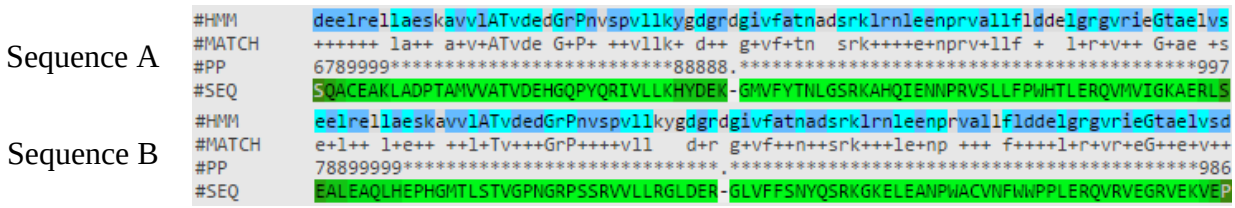


Figure 7. *E. coli* PdxH/b1638 and Mrub_2888 contain similar conserved amino acids. Sequence A = *E. coli* PdxH/b1638; Sequence B = Mrub_2888. #HMM = consensus protein sequence; #SEQ = bacteria amino acid sequence; #MATCH = match between bacteria and consensus sequence. Search Pfam (http://pfam.xfam.org/search) produced pairwise sequence alignment.

Next the cellular pathways and protein family of Mrub_2888 and *E. coli* PdxH were analyzed. KEGG was used to confirm the presence of Mrub_2888 and *E. coli* PdxH in the vitamin B_6 metabolism pathway (Figure 3). The E.C. number and protein names (Table 2) were

found using the KEGG pathway and they were found to be E.C.1.4.3.5 and pyridoxal kinase for both Mrub_2888 and *E.coli* PdxH. The CDD analysis (Table 1) returned the same COG number, COG0259, for Mrub_2888 and *E. coli* PdxH with significant E-values, 6.75e-130 for *E. coli* PdxH and 6.90e-108 for Mrub_2888, and both genes were identified as PdxH pyridoxine/pyridoxamine 5'-phosphate oxidase. Pfam analysis of the Mrub_288 and *E. coli* PdxH also returned results for the genes that were identical (Table 1). Both Mrub_2888 and *E. coli* PdxH returned the same Pfam number, PF01243, both were also from the pyridoxamine 5'phosphate oxidase family, and both returned large bit scores and small E-values, 86.9 and 7e-25 for Mrub_2888 and 87.2 5.6e-25 for *E. coli* PdxH. TIGRFAM analysis of Mrub_2888 and *E. coli* PdxH (Table 1) returned the same TIGRFAM number, TIGR00558, and the same name PdxH: pyridoxamine 5'-phosphate oxidase with large bit scores and small E-values, 440.6 3.1e-129 for Mrub_2888 and 447.4 and 2.8e-131 for *E. coli* PdxH. The same PDB code, 1G769, was returned from the PDB analysis of Mrub_2888 and *E. coli* PdxH. The name returned for Mrub_2888 and *E. coli* PdxH was pyridoxine 5'-phosphate oxidase, and the E-values for these were significant, 3.52844e-46 for Mrub_2888 and 1.32317e-158 for *E. coli* PdxY. A comparison of the gene context (Figure 8) was conducted using IMG/EDU which found that Mrub_2888 is not in an operon while *E. coli* PdxH was confirmed to be found in an operon that also contains *E. coli* PdxY (Yang *et al.*, 1998).

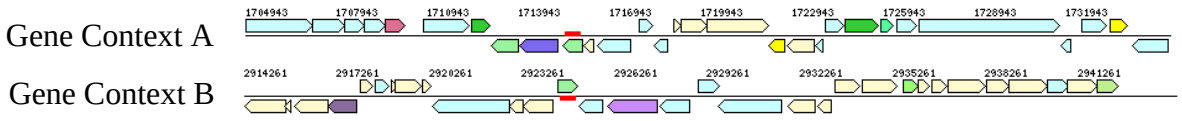


Figure 8. *E. coli* PdxH is found in an operon while Mrub_2888 is not. Gene Context A = *E. coli* PdxH/b1638; Gene Context B = Mrub_2888. Red line above or below sequence indicates desired gene. Search IMG/EDU (http://img.jgi.doe.gov/cgi-bin/edu/main.cgi) produced gene context.

The cellular location of Mrub_2888 and *E. coli* PdxH were analyzed. TMHMM analysis (Figure 9) was conducted and the hydropathy plots returned for both Mrub_2888 and *E. coli*

PdxH indicated that no transmembrane helices were present in either protein. LipoP predictions (Table 2) for Mrub_2888 and *E. coli* PdxH revealed the proteins were cytoplasmic. PSORT-B was used to predict the cellular location of the proteins and this returned a cytoplasmic localization score of 9.26 for Mrub_2888 and 9.97 for *E. coli* PdxH. SignalP analysis (Figure 10) also indicated Mrub_2888 and *E. coli* PdxH are both cytoplasmic proteins. SignalP also returned Mrub_2888 and *E. coli* PdxH D-scores of 0.111 for Mrub_2888 and 0.105 for *E. coli* PdxH, indicating the proteins are not signaling peptides. Since most signaling proteins are membranous or extracellular the low D-scores also indicate Mrub_2888 and *E. coli* PdxH are cytoplasmic proteins.

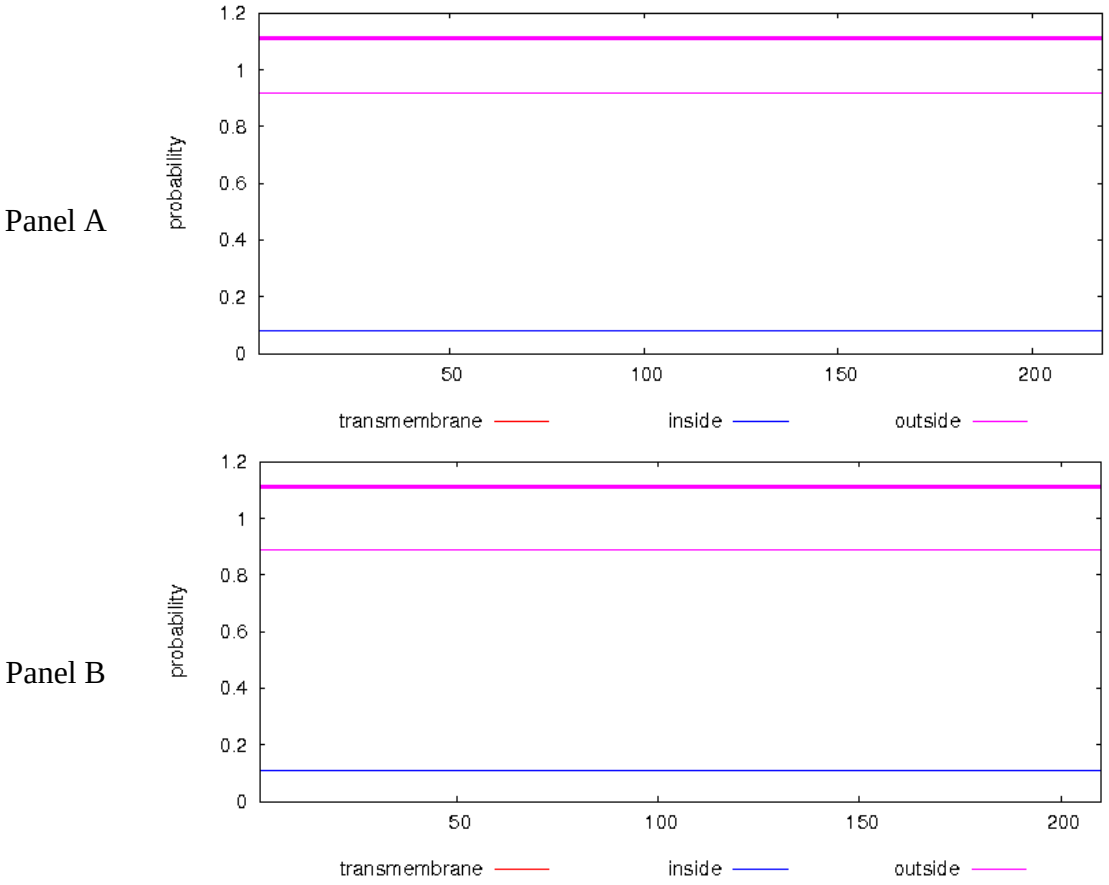
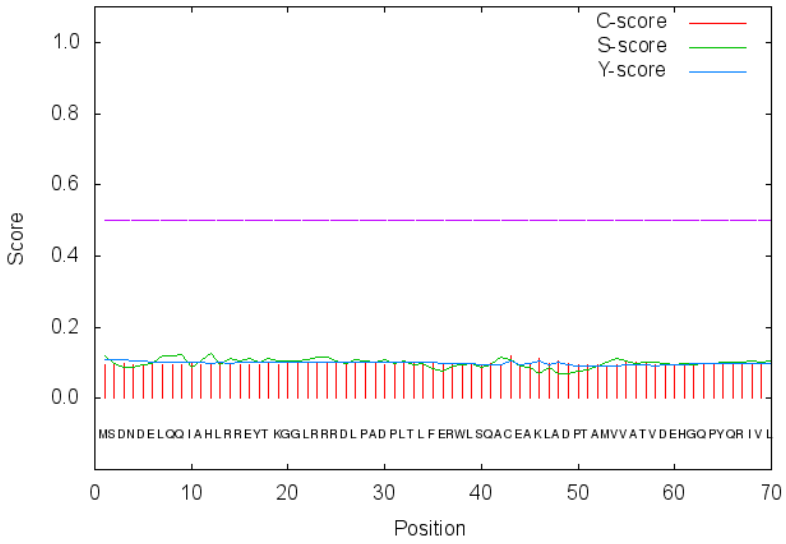


Figure 9. *E. coli* PdxH and Mrub_2888 do not contain TMH regions; enzyme predicted to be in the cytoplasm. Panel A = *E. coli* PdxH/b1638; Panel B = Mrub_2888. TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) generated hydropathy plots.

Panel A



Panel B

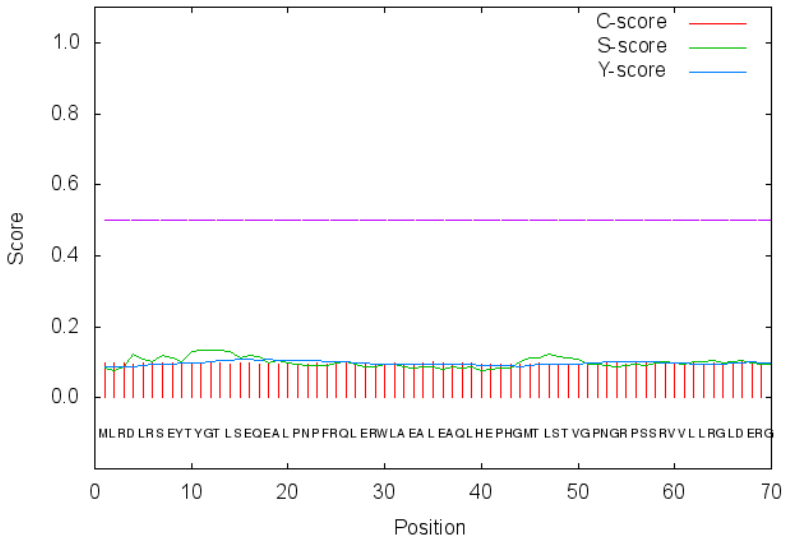


Figure 10. *E. coli* PdxH and Mrub_2888 are not signaling proteins. *E. coli* PdxH D-score = 0.105; Mrub_2888 D-score = 0.111. Panel A = *E. coli* PdxH/b1638; Panel B = Mrub_2888. SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP) generated hydropathy plot.

Conclusion

Based on the data presented above, the hypothesis that Mrub_2507 has an ortholog in *E. coli* PdxY/b1636 was supported. The first result that confirmed this was the high bit score and low E-value that the BLAST of *E. coli* PdxY/b1636 against Mrub_2507 returned. Also, the matching conserved amino acids in the Pfam pairwise alignment between Mrub_2507 and *E. coli* PdxY helped confirm this hypothesis. The identical E.C. number, TIGRFAM number and name,

and Pfam number also further confirm the orthologous nature of Mrub_2507 and *E. coli* PdxY. The COG number returned for Mrub_2507 and *E. coli* PdxY were the same, however, the COG name returned was PdxK pyridoxal/pyridoxine/pyridoxamine kinase. This could be explained by the similarities between *E. coli* PdxY/b1636 and PdxK/b2418 because although *E. coli pdxY* mainly encodes PL kinase with low levels of PN kinase activity *E. coli pdxK* encodes a kinase that acts as a PL/PN/PM kinase, although it favors PN kinase activity over PL kinase activity. These similar functions are the result of similarities in the amino acid sequences of *E. coli* PdxY and PdxK, and since *pdxY* primarily encodes a PL kinase it is possible that the COG results for PdxY and Mrub_2507 viewed these proteins as PdxK since it also encodes for a PL kinase (Yang *et al.,* 1998). It is for this reason that the COG identification of PdxK does not weaken the argument that Mrub_2507 and *E. coli* PdxY are orthologous. Pfam returned the name, phosphomethylpyrimidine kinase for Mrub_2507 and *E. coli* PdxY, a kinase in the thiamine pyrophosphate (TPP) synthesis pathway, a completely different pathway from the vitamin B_6 metabolism pathway. However, further investigation on the Pfam database found that both phosphomethylpyrimidine kinase and PL kinase are members of the ribokinase superfamily of enzymes, so it is possible that since Pfam investigates protein families that the similarities between these kinases caused the product of Mrub_2507 and *E. coli pdxY* to be phosphomethylpyrimidine kinase instead of PL kinase. The PDB analysis also returned different PDB codes for Mrub_2507 and *E. coli* PdxY, however, the names given to Mrub_2507, pyridoxamine kinase is just the gene product of *E. coli pdxY* which the name PdxY. The final piece of evidence were the determination that Mrub_2507 and *E. coli* PdxY had no transmembrane helices, are cytoplasmic, and are not signaling proteins by the TMHMM, LipoP, PSORT-B, and SignalP analyses.

The data presented above also supported the hypothesis that Mrub_2888 is orthologous to *E. coli* PdxH/b1638. The first result that confirmed this was the high bit score and low E-value returned by the BLAST search of *E. coli* PdxH/b1638 against Mrub_2888. The matching conserved amino acids in the Pfam pairwise alignment between Mrub_2888 and *E. coli* pdxH further confirmed this orthologous nature. The identical E.C. number, COG, TIGRFAM, Pfam, and PDB names and numbers also confirmed the orthologous nature of Mrub_2888 and *E. coli* PdxH. The final piece of evidence were the determination that Mrub_2888 and *E. coli* PdxH had not transmembrane helices, are cytoplasmic, and are not signaling proteins by the TMHMM, LipoP, PSORT-B, and SignalP analyses.

Through the use of various bioinformatics tools, strong evidence was obtained to support the hypotheses that Mrub_2507 is orthologous to *E. coli* PdxY/b1636 and Mrub_2888 is orthologous to *E. coli* PdxH/b1638.

References

Allgood V.E., Oakley R.H., Cidlowski J.A.. 1993. Modulation of vitamin B_6 of glucocorticoid receptor-mediated gene expression requires transcription factors in addition to the glucocorticoid receptor. J. Biol. Chem. 268(28):20870-20876.

di Salvo ML, Safo MK, Musayev FN, Bossa F, Schirch V. 2003. Structure and mechanism of Escherichia coli pyridoxine 5'-phosphate oxidase. Biochim. Biophys. Acta. 1647:76-82.

Phylogenetic Diversity. [Internet]. Joint Genome Institute; [accessed 17 Dec 2015]. Available from: http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/.

Pujari, S. "Bioinformatics: An useful essay on bioinformatics & biotechnology" [Internet]. Your Article Library; [2015 Dec 2015]. Available from:

http://www.yourarticlelibrary.com/essay/bioinformatics-an-useful-essay-on-bioinformaticsbiotechnology/29374/.

Tindall BJ, Sikorski J, Lucas S, Goltsman E, Copeland A, Del Rio TG, Nolan M, Tice H, Cheng JF, Han C, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Ovchinnikova G, Pati A, Fähnrich R, Goodwin L, Chen A, Palaniappan K, Land M, Hauser L, Chang YL, Jefferies CD, Rohde M, Göker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Lapidus A. 2010. Complete genome sequence of Meiothermus ruber type strain. Stand Genomic Sci. (3): 26-36.

Yang, Y., G. Zhao, and M. E. Winkler. 1996. Identification of the pdxK gene that encodes pyridoxine (vitamin B6) kinase in Escherichia coli K-12. FEMS Microbiol. Lett. 141:89–95.

Yang Y, Tsui H-CT, Man T-K, Winkler ME. 1998. Identification and function of the pdxY gene, which encodes a novel pyridoxal kinase involved in the salvage pathway of pyridoxal 5'-phosphate biosynthesis in Escherichia coli K-12. J Bacteriol. 180(7):1814-1821.