Characterization of the Mutant ProP Protein in *Salmonella enterica* serovar Typhimurium

by

Faez Patrick Hamcho

A Thesis

Submitted to the Honor's Department

Florida Southern College

In Partial Fulfillment of the Requirements

for the Honor's Program

April 2021

Acknowledgments

Thank you Dr. Gasper for guiding me on the research project for the last 2 years. I have learned so much and I believe being able to talk about this research opened doors to research programs and even medical school.

Abstract

Mechanisms invoked by organisms to combat osmotic stress are ubiquitous. The information gained from analyzing osmotic adaptations can be broadly applied, to creating drought-resistant crops, understanding disease pathologies, and fighting bacterial infections. *Salmonella enterica* serovar Typhimurium (*Salmonella typhimurium*, hereafter) has a transport protein called ProP that undergoes post-translational modifications that allow it to uptake osmoprotectants in the face of osmotic stress. The nature of these post-translational modifications is not well understood since the protein has not been crystallized. The aim of my study was to further characterize 6 *Salmonella typhimurium* strains that have a mutant ProP protein that has conformational changes that mimic post-translationally modified wild-type proteins. I did so by running a sequence alignment between *Salmonella typhimurium* and 22 related species to gain an understanding of the importance of the regions of ProP needed for function and I modeled the mutated amino acid sequences on a related protein to gain a better understanding of how the mutations affect the confirmation of the protein. The regions where the mutations occurred in the mutants were found to play a significant role in ProP's function based on their conservation among 22 ProP orthologs, and the majority of the mutations could significantly affect ProP's function in a way that likely mimics the wildtype.

Table of contents

Introduction	5
Methodology	8
Results	10
Discussion	15
Conclusion	17
Reference list	18
Supplementary Data	19

Introduction

All organisms have to withstand osmotic stress. Many of the mechanisms used to withstand this stress are similar between plants, humans, and bacteria. The ubiquitously shared order of events entail, osmotic sensors recognizing the disturbance in the environment, signal transduction, and appropriate cellular response such as genetic regulations and molecular mechanisms. When the water concentration outside of a cell is lower than the concentration inside the cell, the cell is in a state of hypertonic stress that leads to cell shrinkage, macromolecular crowding, cell cycle arrest, DNA damage, oxidative stress, protein carbonylation, protein degradation, downregulated protein transcription, translation, and finally apoptosis. To combat this, cells will upregulate their transporter translocation mechanisms, osmoprotectant influx, osmoprotectant synthesis, cytoskeletal remodeling, chaperone proteins, and antioxidant proteins (Brocker, Thompson, & Vasiliou, 2012).

The larger the information database is on osmotic regulation, the better, since most organisms share the same mechanisms. In plants, for example, there are certain alleles and molecular mechanisms that species commonly exposed to drought have developed that allow for them to survive. Given that plants cannot move in the face of abiotic stressors, it is no surprise that they have retained key osmoregulatory genes through evolution. By studying the genomes of drought-resistant plants, key alleles can be revealed that can be implemented in the production of drought-resistant crops (Lefebvre , Poormohammad Kiani, & Durand-Tardif, 2009).

In addition to the prominent impacts that studying osmotic adaptations can have on crop production, studying these mechanisms can yield dramatic impacts on treatments for human disease. Many human diseases have been linked to hyperosmotic stress on the tissues where the disease occurs. When these tissues are exposed to stress, regulatory pathways are triggered and proinflammatory cytokines are released, leading to both systematic and acute inflammation. Hypertonic stress has been implicated in eye disease, diabetes, inflammatory bowel disease, liver disease, cardiovascular disease, and cancer. By learning more about osmoregulatory mechanisms,

both effective treatments and models can be used to relieve the symptoms and prevent these diseases (Brocker, Thompson, & Vasiliou, 2012).

Learning more about the osmoregulatory mechanisms of *Salmonella typhimurium* can add to the osmoregulatory database. In the face of hyperosmotic stress, *Salmonella typhimurium* accumulates compatible solutes like proline through *de novo synthe*is or it synthesizes more transport proteins to increase the influx of proline. There are 3 known transport proteins for proline. They are ProP, ProU, and PutT. The proline taken in by PutP is used for nitrogen fixation, so it plays no role in osmoregulation. Further, ProP and ProU play a role in osmoregulation, and both experience osmotically induced transcription. Moreover, ProP is of even more interest because it is post-translationally modified in the face of osmotic stress, and these modifications lead to a 20-fold increase in proline intake (Gasper , 2012).

ProP has not yet been crystallized. The types of post-translational modifications that occur to it, and the conformational changes that ensue, are unknown. Most of what is known about ProP is through site-directed mutagenesis protocols that replace amino acids with cysteines that can be tagged. Moreover, the current models of ProP are made through protein modeling software that find the most similar proteins based on the amino acid sequence (Poolman, Spitzer, & Wood, 2004).

ProP is a 500 residue integral membrane protein that is part of the major facilitator super family. The protein uses the proton motor force to pump solutes like proline and glycine betaine into the cell. In addition to being a transporter, the protein is an osmotic sensor.

In a past study done by Dr. Gasper, a series of different mutagenesis protocols were performed on *Salmonella typhimurium* and a positive selection for auxotrophs that could grow in the presence of glycine betaine antagonism was done. These mutants were sequenced and the locations on ProP where the mutations occurred were all localized. All of the mutations were in regions involved in the formation of the transport pore. This research was done to address the knowledge gap regarding the type of post-translational modifications that occur. Since the mutations all led to increased proline uptake by ProP, the hopes were that these mutations mimic the post-translational modifications that

occur to ProP in osmotically stressful environments. And a simple way of demonstrating this can be seen in **supplementary 1**, where the structural changes associated between an amino acid being phosphorylated is similar to the amino acid being mutated to an ionizable acidic amino acid-like aspartic acid.

The goal of my study was to further characterize these mutated *Salmonella typhimurium* strains. Originally, I was going to measure growth rates of the mutants and wildtype strains with proline analogs, to further gauge the effects of these mutations. Because of the pandemic, my project was adjusted to a more bio-informatic approach. I visualized the effects of these mutations on 3d modeling programs and analyzed the conservation of amino acids in 22 ProP orthologs in the regions where the mutations were made to gauge the importance of these regions.

Methodology

Using the mutant strains, I was going to prepare growth curves to assess how they respond to proline analogs as compared to the wild type. This could have told us more about the mutations and possibly the post-translational ProP mechanisms. In preparing for this project, I cultured the 6 mutant strains of *Salmonella typhimurium* plus a wild-type strain. I then grew them in 3 different environments, M63+Proline, M63+NaCl, and M63+NaCl+Proline. The growth rates were then measured using a Biotek plate reader. The point of this experiment was to ensure that growth rates could be indirectly measured through the plate reader system, and this was done by comparing the growth rates measured from this experiment to a past experiment with a more established growth rate protocol. Moreover, because of COVID, the project took a bio-informatic direction.

The first part of my bio-informatic project involved analyzing the amino acid sequences around the regions where the mutations were located in each of the mutant strains and comparing these sequences to 22 related species to gain an idea of the importance of the regions where the mutations occurred. The first step involved creating a list of closely related *Salmonella typhimurium* relatives, which were gram-negative, and then looking to more distantly related gram-positive relatives, all of which contained ProP orthologs and can be found in **Supplementary 2**. From there each species of bacteria's ProP sequence was acquired through the Patric database (**Supplementary 2**). When more than one result appeared, the result with the closest number of amino acid residues to the *Salmonella typhimurium* ortholog was picked. From there, each of the 22 sequences was input into the Clustal Omega sequence alignment program. Then, the exact region on *Salmonella typhimurium* where a mutation was made, was localized, and I isolated an 11 residue sequence from this region consisting of 5 residues to the left of where the mutation occurred and 5 residues to the right of the mutation from both *Salmonella typhimurium* and the 22 ProP orthologs.

Next, I input the 11 sequence segments of each species into excel and used the column stats function to identify how often the same residue found in *Salmonella typhimurium* appeared in each of the 22 ProP orthologs. From there, I had 11 percentage values from each residue in the sequence and averaged them for each of the 6 regions I looked at. Then, I was able to see which region of the 6

regions I looked at was most conserved. I then compared the entire region's conservation value to the exact location of the residue number where the mutation was made, to look for any significant differences.

For the next part of my project, I visualized the 6 mutant ProP sequences and wild-type sequences to look for any visual changes to the structure of the protein. I did this by inputting the sequences into the Swiss modeling program. From there I was able to compare the wild-type protein to the mutated proteins.

Results

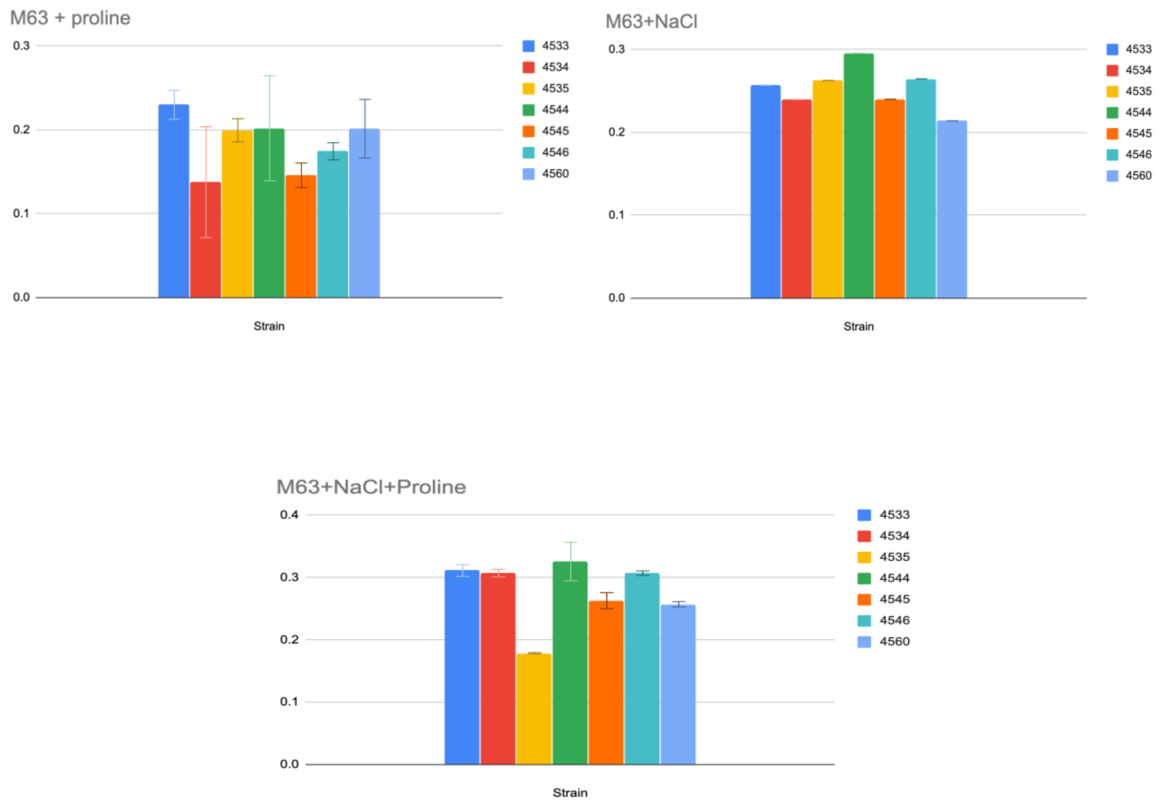


Figure 1: The growth charts of the 6 mutant strains and wild-type *Salmonella typhimurium* in M63+Proline, M63+NaCl, and M63+NaCl+Proline.These results told us that the BioTek method of measuring the growth rate was viable. This meant that we could move on with growing the strains with proline analogs.

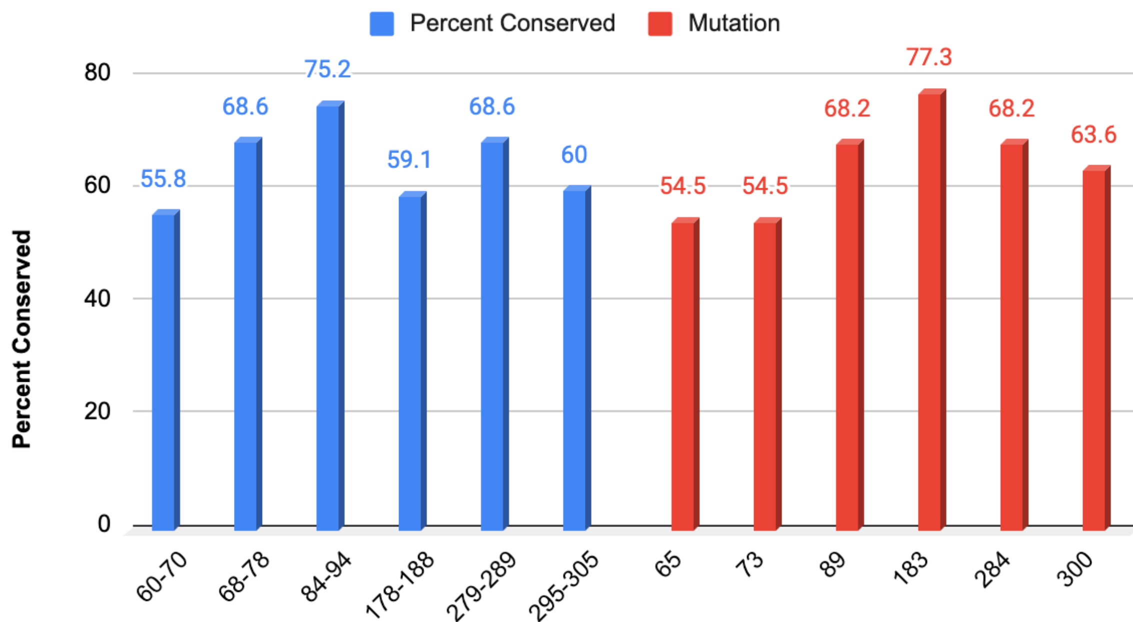
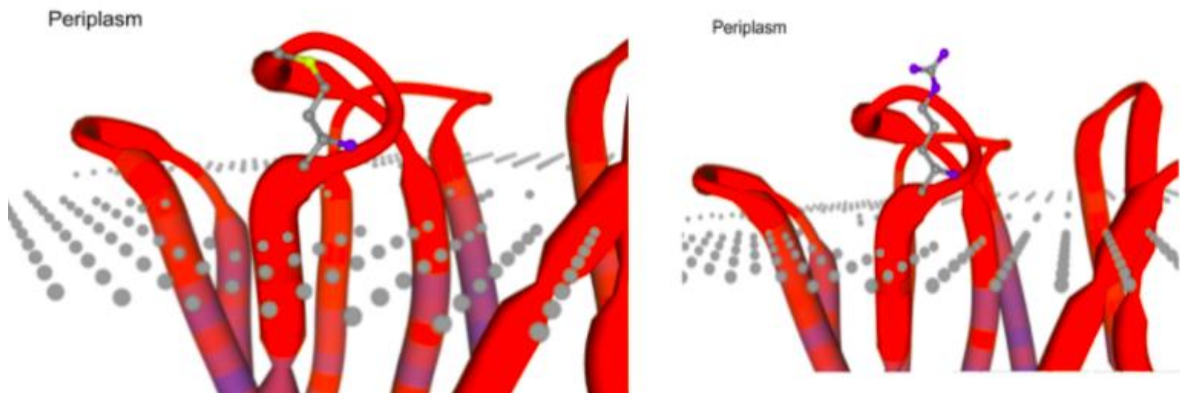
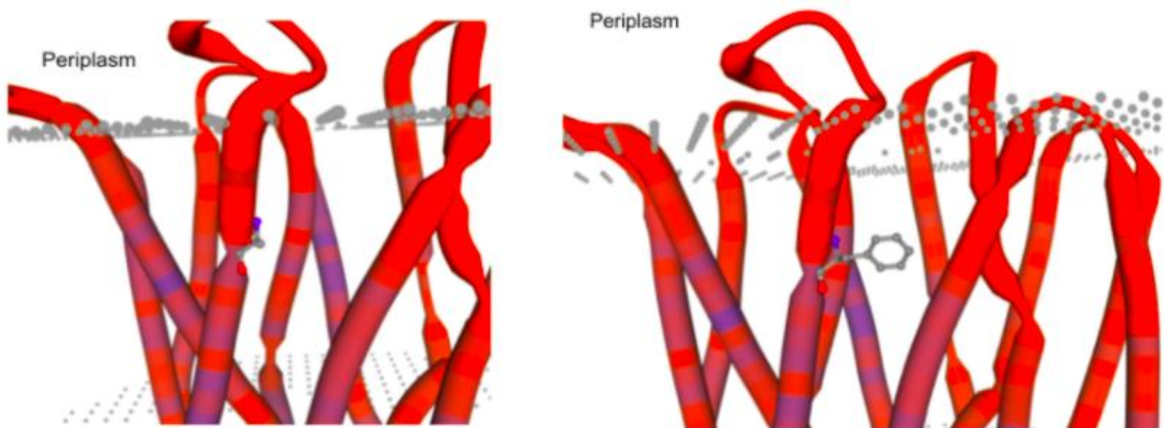


Figure 2: 6 regions of 11 amino acid residues were isolated from 22 related ProP orthologs. Each region was around the residue where the mutation occurred (**Supplementary 2**). The percentage of times the specific amino acid residue in *Salmonella typhimurium* repeated in the 22 orthologs was calculated and then averaged with the entire region. Additionally, the percentage of times the specific residue where the mutation occurred on *Salmonella typhimurium* appeared in the orthologs was analyzed. Each of the 6 regions had an average conservation percentage over 50%, and the trend was very similar in the specific residues.

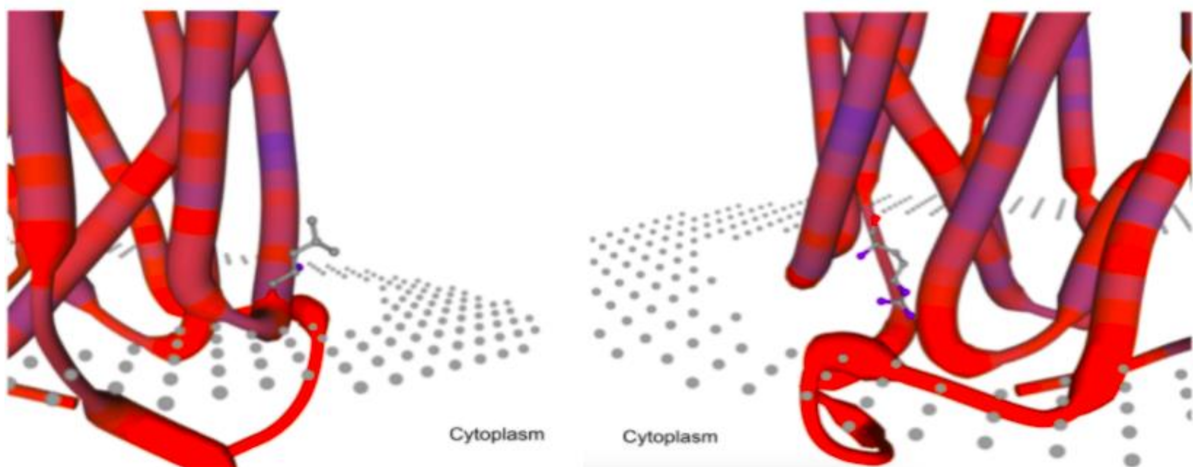
Mutation: M65R

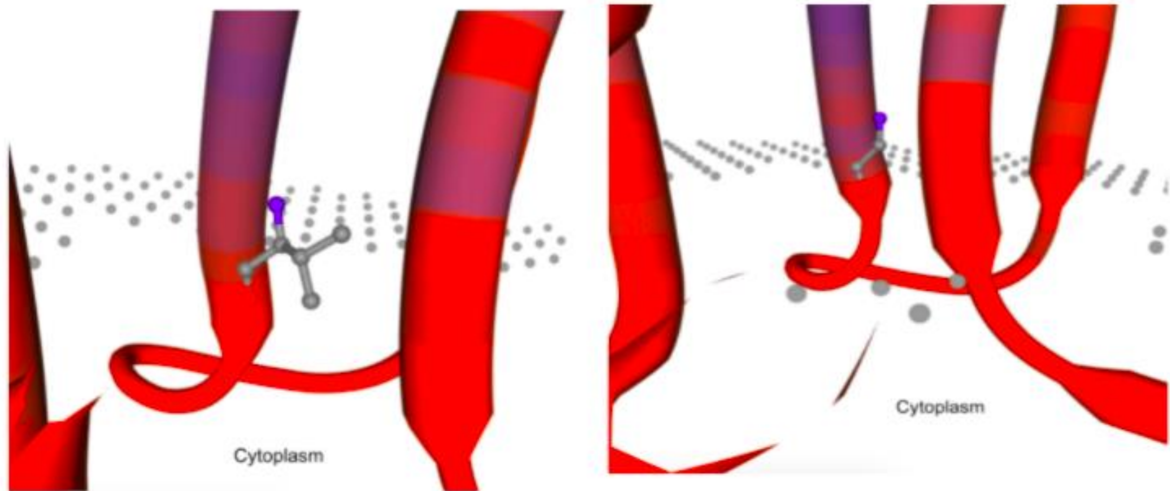


Mutation: S73F



Mutation: L89R





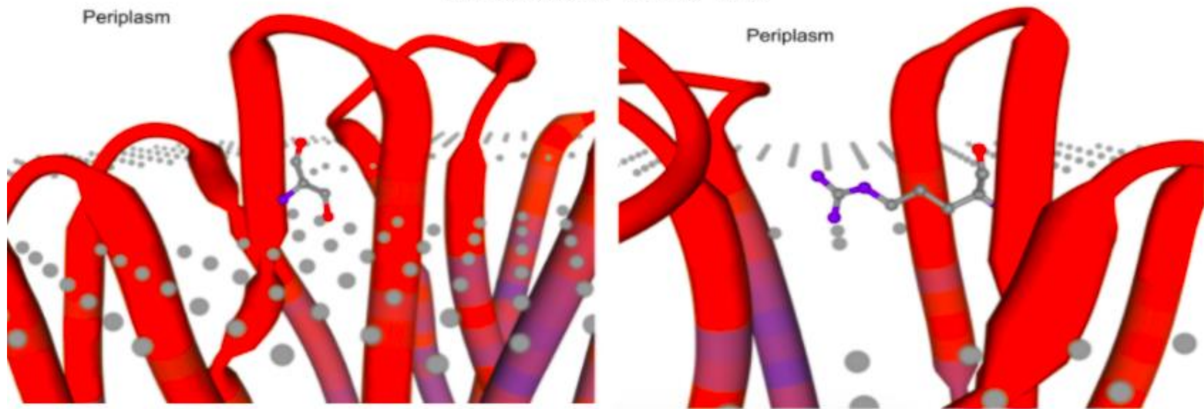


Figure 3: The 3d models of the 6 mutations and the wildtype ProP structure. The wild-type structure is shown on the left and the mutated structure is shown on the right. While the software was not able to show how the entire confirmation of ProP changed, it was a great visualization of how the amino acids changed.

Mutation	Feature change	Significance on Structure and Function
M65R	**Hydrophobic->Ionized**	**Significant**
S73F	**Polar->Hydrophobic**	**Significant**
L89R	**Hydrophobic->Ionized**	**Significant**
V183G	**Hydrophobic->Hydrophobic**	**Inconsequential**
S284R	**Polar->Ionized**	**Significant**
I300M	**Hydrophobic->Hydrophobic**	**Inconsequential**

Table 1: The mutations, the nature of the mutation on the side chain, and whether or not the change would be significant on the structure and function.

Discussion

Based on **figure 2** it appears that the regions where the mutations occurred are conserved among the 22 orthologs. Moreover, this implies that these amino acids are important for function. Additionally, from **figure 2**, it is clear that the specific amino acids, where the mutations occur, are just as conserved as the regional values, indicating they are as important to the function of ProP. This conclusion is backed up by the conservative principle in evoluntionary biochemistry, which states that systems designed to allow organisms cope with their enviornemnt are unlikely to change, especially when essential functions depend on them. So, it is not surprising to see that these ProP orthologs share so much similarity (Eck & Dayhoff, 1966).

The models that were generated do not show a significant difference between the wild type and mutated ProPs. Moreover, the models are still great visualizations of how drastic the mutations were. While the models were thought-provoking, it should be stated that they were generated by altering the A. thaliana Sugar Transport Protein 10 to *Salmonella typhimurium's* ProP sequence. Moreover, the software works through finding the most similar protein in their database and it just so happened A. thaliana Sugar Transport Protein 10 had the closest sequence of 15.25% alignment. Further, since protein sequence dictates structure, and structure dictates function, such a drastically different amino acid alignment means that the proposed model of ProP should only be taken with a grain of salt (Berg, 1970).

Table 1 is a coherent summary of the mutations, and as shown, four of the six mutations were drastic. These four mutations involved the difference between neutrally charged amino acids and charged amino acids, which would have a significant effect on function. Moreover, since all of the mutations occurred on sites involved in transport pore formation, that could certainly be an explanation as to why the mutants experienced increased proline transport at low osmolarity. Further, these mutants likely have confirmations that mimic wild-type ProP proteins that have undergone osmotic stress. This conclusion is supported by the fact that phosphorylation as a post-translational modification is the most common modification and that the practice of mimicking phosphoruylated proteins through mutation of amino acids to aspartate or glutamate is a well studied practice. So, it is

likely that one of the sites where a mutation occurred was where a phosphorylation event occurs during periods of osmotic stress to allow ProP to take up more proline (Yang, Cho, & Park, 2018).

Conclusion

The regions where the mutations occurred on Dr. Gasper's 6 strains were all located in the transport pore of the protein, and all mutants experienced greater proline uptake in the face of glycine betaine antagonism. These mutants likely have mutations that mimic post-translational modifications, which lend them similar confirmations. This statement is further supported by the nature of these mutations as seen in **table 1**. Mutations that involved uncharged amino acids to charged amino acids are similar to phosphorylation events. Further, each of the mutated regions had amino acids that were highly conserved among 22 ProP orthologs, which indicates they are important to function. The next steps should be to either crystallize ProP or measure the growth rates of the mutants and wild type with proline analogs to further characterize the mutations.

Reference list

Berg, J. (1970, January 01). Protein Structure and Function. Retrieved May 02, 2021, from https://www.ncbi.nlm.nih.gov/books/NBK21177

Brocker, C., Thompson, D., & Vasiliou, V. (2012, August). The role of hyperosmotic stress in inflammation and disease. Retrieved April 28, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3438915/

Eck, R., & Dayhoff, M. (1966, April 15). Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences. Retrieved May 02, 2021, from https://pubmed.ncbi.nlm.nih.gov/17775169/

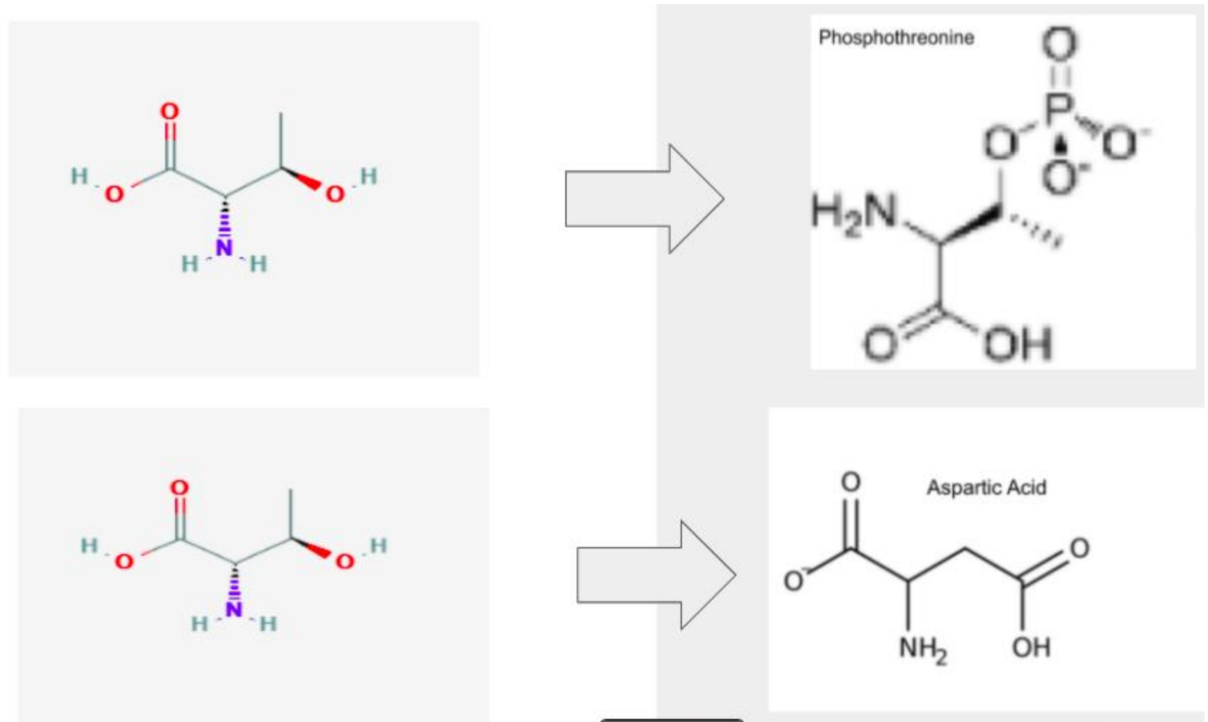
Gasper, B. J. (2012, June 31). Isolation and preliminary characterization of amino acid substitution mutations that increase the activity of the osmoregulated prop protein of salmonella enterica serovar typhimurium. Retrieved April 28, 2021, from https://pubmed.ncbi.nlm.nih.gov/22360681/

Lefebvre, V., Poormohammad Kiani, S., & Durand-Tardif, M. (2009, August 13). A focus on natural variation For abiotic constraints response in the model species Arabidopsis thaliana. Retrieved April 29, 2021, from https://pubmed.ncbi.nlm.nih.gov/20111677/

Poolman, B., Spitzer, J., & Wood, J. (2004, July 23). Bacterial osmosensing: Roles of membrane structure and electrostatics in lipid–protein and protein–protein interactions. Retrieved May 03, 2021, from https://www.sciencedirect.com/science/article/pii/S0005273604001609

Yang, A., Cho, K., & Park, H. (2018). Chemical biology approaches for studying posttranslational modifications. Retrieved May 02, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103722/

Supplementary Data



Supplementary 1: How a post-translational modification such as a phosphorylation and a mutation from an uncharged to a charged amino acid like aspartic acid could lead to similar structural and functional changes.

```
Salmonella_typhimurium          FV-AYALGKVFFPG- DPSVQMIAALA FSVPFLIRPLGGLFFGMLGDKYGRQKILAITI  103
Citrobacter_freundii            FV-AYALGKVFFPG- DPSVQMVAALA FSVPFLIRPLGGLFFGMLGDKYGRQKILAITI  103
Escherichia_coli                FV-AYALGKVFFPG- DPSVQMVAALA FSVPFLIRPLGGLFFGMLGDKYGRQKILAITI  103
Enterobacter_aerogenes          FV-AYALGKVFFPG- DPSLQMIAALG FSVPFLIRPLGGLFFGMLGDKYGRQKILAITI  103
Cronobacter_sakazakii           FV-AYALGKVFFPD- NPSVQMIAALA FSVPFLIRPLGGLFFGMLGDKYGRQKILAITI  104
Pantoea_agglomerans             FV-AYALGKVFFPD- SPGVQMIAALA FSVPFLIRPLGGLFFGMLGDKYGRQKILSITI  104
Pseudomonas_aeruginosa          FV-AYALGKVFFPD- NPSVQMIAALG FSVPFLIRPLGGLFFGMLGDKYGRQKILAITI  103
Dickeya_dadantii                FV-AYALGQVFFPG- DPGVQMIAALA FSVPFLIRPLGGVFFGALGDKYGRQKILAITI  104
Yersinia_pestis                 FF-AAVIGDLFFPAD PEWLRQVQTFG FAAGYLARPLGGIIMAHFGDLVGRKKMFTLSI  117
Providencia_stuartii            FL-AYVLGQVFFPG- SPSVQMIAALA FSVPFLVRPLGGVVFGILGDKYGRQKVLAVTI  104
Serratia_marcescens             FV-AYALGQVFFPG- SPGVQMIAALA FSVPFLVRPLGGLFFGAMGDKFGRQKVLSVTI  104
Proteus_mirabilis               FL-AYVLGQVFFPG- SPGVQMIAALA FSVPFLVRPLGGVVFGMLGDKFGRQKVLSVTI  104
Acinetobacter_baumannii         YV-AYVLGKVFFPD- SPSVQMIAALA FSVPFIFRPLGGLFFGHLGDKYGRQKVLAITV  104
Agrobacterium_radiobacter       TASALVFNKVFFPS- DALVGTLLAFG FASAYLARILGAALFGHFGDRLGRKSMLLFSL  82
Bacillus_anthracis              YL-AVILSQLFFSGV NSGLQLVLTFG FAAAFLVRPIGGVFFGRIGDKYGRKIVLSTTI  103
Stenotrophomonas_maltophilia    YL-AVTIGQVFFPS- NPTAQVIAAFA FTVAFLVRPLGGLVFGPLGDRYGRQKVLAFTM  113
Rickettsia_rickettsii           VF-SLIIGQVFFPG- SEFIRILLSLG FAVGFLTRPVGGILFGYIGDRYGRRIALIISM  85
Rhodococcus_rhodochrous         ILAATVLGPLFFPN- NAVASLLMALA QGLGFIARPLGGIVFGHLGDKFGRKPILVTTF  106
Streptomyces_fulvissimus        YL-AGTLGKVFFPS- SPGAQVVSTFA FAAAFLVRPLGGLVFGPLGDRVGRQKVLALTM  106
Burkholderia_cepacia            YI-AVTLGKVFFPS- SPSAQLLATFG FAAAFLVRPLGGMVFGPLGDRIGRQRVLAATM  118
Erwinia_amylovora               FV-AFALGQVFFPG- DSGTQMIAALA FSVPFLIRPLGGLFFGALGDKYGRQKILSITI  104
Staphylococcus_aureus           YT-TAYIGANFFSPV NADIRQMLTFA LAIAFLLRPIGGVVFGIIGDKYGRKVVLTSTI  97
Campylobacter_jejuni            FF-AEYIANVFFPKD SEFWALLNTYG FAAGYLARPLGGIVMAHFGDKFGRKNMFMLSI  91
```

Organism	Sequence	Position
Salmonella_typhimurium	FV-AYALGKVFFPG-ADPSVQMIAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Citrobacter_freundii	FV-AYALGKVFFPG-ADPSVQMVAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Escherichia_coli	FV-AYALGKVFFPG-ADPSVQMVAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Enterobacter_aerogenes	FV-AYALGKVFFPG-ADPSLQMIAALGTFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Cronobacter_sakazakii	FV-AYALGKVFFPD-ANPSVQMIAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	104
Pantoea_agglomerans	FV-AYALGKVFFPD-VSPGVQMIAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILSITI	104
Pseudomonas_aeruginosa	FV-AYALGKVFFPD-ANPSVQMIAALGTFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Dickeya_dadantii	FV-AYALGQVFFPG-ADPGVQMIAALATFSVPFLIRPLGGVFFGALGDKYGRQKILAITI	104
Yersinia_pestis	FF-AAVIGDLFFPADMPEWLRQVQTFGIFAAGYLARPLGGIIMAHFGDLVGRKKMFTLSI	117
Providencia_stuartii	FL-AYVLGQVFFPG-ASPSVQMIAALATFSVPFLVRPLGGVVFGILGDKYGRQKVLAVTI	104
Serratia_marcescens	FV-AYALGQVFFPG-ASPGVQMIAALATFSVPFLVRPLGGLFFGAMGDKFGRQKVLSVTI	104
Proteus_mirabilis	FL-AYVLGQVFFPG-ASPGVQMIAALATFSVPFLVRPLGGVVFGMLGDKFGRQKVLSVTI	104
Acinetobacter_baumannii	YV-AYVLGKVFFPD-ASPSVQMIAALATFSVPFIFRPLGGLFFGHLGDKYGRQKVLAITV	104
Agrobacterium_radiobacter	TASALVFNKVFFPS-FDALVGTLLAFGTFASAYLARILGAALFGHFGDRLGRKSMLLFSL	82
Bacillus_anthracis	YL-AVILSQLFFSGVDNSGLQLVLTFGTFAAAFLVRPIGGVFFGRIGDKYGRKIVLSTTI	103
Stenotrophomonas_maltophilia	YL-AVTIGQVFFPS-SNPTAQVIAAFATFTVAFLVRPLGGLVFGPLGDRYGRQKVLAFTM	113
Rickettsia_rickettsii	VF-SLIIGQVFFPG-ESEFIRILLSLGVFAVGFLTRPVGGILFGYIGDRYGRRIALIISM	85
Rhodococcus_rhodochrous	ILAATVLGPLFFPN-GNAVASLLMALATQGLGFIARPLGGIVFGHLGDKFGRKPILVTTF	106
Streptomyces_fulvissimus	YL-AGTLGKVFFPS-SSPGAQVVSTFATFAAAFLVRPLGGLVFGPLGDRVGRQKVLALTM	106
Burkholderia_cepacia	YI-AVTLGKVFFPS-SSPSAQLLATFGTFAAAFLVRPLGGMVFGPLGDRIGRQRVLAATM	118
Erwinia_amylovora	FV-AFALGQVFFPG-ADSGTQMIAALATFSVPFLIRPLGGLFFGALGDKYGRQKILSITI	104
Staphylococcus_aureus	YT-TAYIGANFFSPVENADIRQMLTFAALAIAFLLRPIGGVVFGIIGDKYGRKVVLTSTI	97
Campylobacter_jejuni	FF-AEYIANVFFPKDMSEFWALLNTYGAFAAGYLARPLGGIVMAHFGDKFGRKNMFMLSI	91

Organism	Sequence	Position
Salmonella_typhimurium	FV-AYALGKVFFPG-ADPSVQMIAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Citrobacter_freundii	FV-AYALGKVFFPG-ADPSVQMVAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Escherichia_coli	FV-AYALGKVFFPG-ADPSVQMVAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Enterobacter_aerogenes	FV-AYALGKVFFPG-ADPSLQMIAALGTFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Cronobacter_sakazakii	FV-AYALGKVFFPD-ANPSVQMIAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	104
Pantoea_agglomerans	FV-AYALGKVFFPD-VSPGVQMIAALATFSVPFLIRPLGGLFFGMLGDKYGRQKILSITI	104
Pseudomonas_aeruginosa	FV-AYALGKVFFPD-ANPSVQMIAALGTFSVPFLIRPLGGLFFGMLGDKYGRQKILAITI	103
Dickeya_dadantii	FV-AYALGQVFFPG-ADPGVQMIAALATFSVPFLIRPLGGVFFGALGDKYGRQKILAITI	104
Yersinia_pestis	FF-AAVIGDLFFPADMPEWLRQVQTFGIFAAGYLARPLGGIIMAHFGDLVGRKKMFTLSI	117
Providencia_stuartii	FL-AYVLGQVFFPG-ASPSVQMIAALATFSVPFLVRPLGGVVFGILGDKYGRQKVLAVTI	104
Serratia_marcescens	FV-AYALGQVFFPG-ASPGVQMIAALATFSVPFLVRPLGGLFFGAMGDKFGRQKVLSVTI	104
Proteus_mirabilis	FL-AYVLGQVFFPG-ASPGVQMIAALATFSVPFLVRPLGGVVFGMLGDKFGRQKVLSVTI	104
Acinetobacter_baumannii	YV-AYVLGKVFFPD-ASPSVQMIAALATFSVPFIFRPLGGLFFGHLGDKYGRQKVLAITV	104
Agrobacterium_radiobacter	TASALVFNKVFFPS-FDALVGTLLAFGTFASAYLARILGAALFGHFGDRLGRKSMLLFSL	82
Bacillus_anthracis	YL-AVILSQLFFSGVDNSGLQLVLTFGTFAAAFLVRPIGGVFFGRIGDKYGRKIVLSTTI	103
Stenotrophomonas_maltophilia	YL-AVTIGQVFFPS-SNPTAQVIAAFATFTVAFLVRPLGGLVFGPLGDRYGRQKVLAFTM	113
Rickettsia_rickettsii	VF-SLIIGQVFFPG-ESEFIRILLSLGVFAVGFLTRPVGGILFGYIGDRYGRRIALIISM	85
Rhodococcus_rhodochrous	ILAATVLGPLFFPN-GNAVASLLMALATQGLGFIARPLGGIVFGHLGDKFGRKPILVTTF	106
Streptomyces_fulvissimus	YL-AGTLGKVFFPS-SSPGAQVVSTFATFAAAFLVRPLGGLVFGPLGDRVGRQKVLALTM	106
Burkholderia_cepacia	YI-AVTLGKVFFPS-SSPSAQLLATFGTFAAAFLVRPLGGMVFGPLGDRIGRQRVLAATM	118
Erwinia_amylovora	FV-AFALGQVFFPG-ADSGTQMIAALATFSVPFLIRPLGGLFFGALGDKYGRQKILSITI	104
Staphylococcus_aureus	YT-TAYIGANFFSPVENADIRQMLTFAALAIAFLLRPIGGVVFGIIGDKYGRKVVLTSTI	97
Campylobacter_jejuni	FF-AEYIANVFFPKDMSEFWALLNTYGAFAAGYLARPLGGIVMAHFGDKFGRKNMFMLSI	91

```
Salmonella_typhimurium        MGSWLDFGSIAGF LGAGVVVLIST VGEENFLEWGWRIPFFIALPLGIIGLYLRHALEE  223
Citrobacter_freundii          MGSWLDFGSIAGF LGAGVVVLIST VGEENFLEWGWRIPFFLALPLGLIGLYLRHALEE  223
Escherichia_coli              MGSWLDFGSIAGF LGAGVVVLIST VGEENFLDWGWRIPFFIALPLGIIGLYLRHALEE  223
Enterobacter_aerogenes        MGSWLDFGSIAGF MGAGVVVLIST VGEENFLDWGWRIPFFLALPLGIIGLYLRHALEE  223
Cronobacter_sakazakii         MGSWLDFGSIAGF LGAGVVVLISA LGEENFLSWGWRLPFFLALPLGLIGLYLRHALEE  224
Pantoea_agglomerans           MGSWLDFGSIAGF LGAGLVVLISS IGEESFLEWGWRLPFFVALPLGIIGLYLRHALEE  224
Pseudomonas_aeruginosa        MGSWLDFGSIAGF MGAGVVVLISS VGEQNFLDWGWRIPFFLALPLGIIGLYLRHALEE  223
Dickeya_dadantii              MGSWLDFGSIAGF LGAGVVVLIST IGEQAFLEWGWRLPFFLALPLGLIGLYLRHALEE  224
Yersinia_pestis               ACGTLTAGLTAGI LGSLVATVMNT LGHQAILEGGWRIPFFLGGIFGLFAMYLRRWLQE  237
Providencia_stuartii          MGSWLDFGSIAGF LGAGVVVLISS VGEENFHEWGWRIPFFLALPLGIIGLYLRHALEE  224
Serratia_marcescens           MGSWLDFGSIAGF MGAGVVVLISS VGEANFLDWGWRIPFFIAAPLGLIGLYLRHALEE  224
Proteus_mirabilis             MGSWLDFGSIAGF MGAGVVVLIST MGEAAFHEWGWRIPFFLALPLGLIGLYLRHALEE  224
Acinetobacter_baumannii       MGSWLDFGSIAGF LGAATVALITH VGEARFAEWGWRIPFFLALPLGIIGLYLRNRLEE  224
Agrobacterium_radiobacter     YGSWVQIGVPAGT IANLVFLAIAS MSSEDLLAWGWRVPFLASILLVAVGAYVRLNTAE  202
Bacillus_anthracis            LGSGLEIGTLSGY AASVIVTILTL LTDEQMLSWGWRIPFLIAAPIGLVGLYLRRHLDE  223
Stenotrophomonas_maltophilia  MGSWLEFGTLGGY AGAGTVTALHM LSSEQMLDWGWRIPFLVAGPLGLLGLYMRMKLEE  233
Rickettsia_rickettsii         TAGLVHGSNIAGT IATLIGIIIER F---SHIDFAWRFAFLLGGFMGLAGFYLRLRVSE  202
Rhodococcus_rhodochrous       WAAWPQSGAPAGT LATVTVGILAL FPGDAFDNWGWRVAFLLAVPLLIIGFLIRRGVEE  226
Streptomyces_fulvissimus      LGSWLDFGTFVGY LGSGLVTVLTA LGTDGMTDWGWRIPFFVAGPMGIIGLYMRLKLEE  226
Burkholderia_cepacia          MGSFLEFGTLIGY MGAGVVALLTA LSQEALLSWGWRVPFLIAGPLGLIGLYIRMKLEE  238
Erwinia_amylovora             MGSWLDFGSIAGF LGAGLVVLISA IGEASFLDWGWRIPFFVALPLGIIGLYLRHALEE  224
Staphylococcus_aureus         LGSGLEIGTLSGY AASIMIAVLTF LTDEQMASFGWRIPFLLGLFLGLFGLYLRRKLEE  217
Campylobacter_jejuni          FLSCLNSAMALGI LGSIVFLIINA FSIEEIAAYAWRIAFFVGGIFGIISIYLRRFLQE  211
```

```
Salmonella_typhimurium        M LTYMPSYLSHN HYSEDHGVLIIIAIMIGMLFVQPVMGLLSDRFGRRPFVIMGSIALF  336
Citrobacter_freundii          M LTYMPSYLSHN HYSEDHGVLIIIAIMIGMLFVQPIMGLLSDRFGRRPFVIMGSIALF  336
Escherichia_coli              M LTYMPSYLSHN HYSEDHGVLIIIAIMIGMLFVQPVMGLLSDRFGRRPFVLLGSVALF  336
Enterobacter_aerogenes        M LTYMPSYLSHN HYSEDHGVLIIIAIMVGMLFVQPIMGLLSDRFGRKPFIILGSVALF  336
Cronobacter_sakazakii         M LTYMPSYLSHN HYSEEHGVLIIIAIMIGMLFVQPVMGLLSDRFGRRPFVIFGSVALM  337
Pantoea_agglomerans           M LTYMPSYLSHN HYSEDHGVMIIIAIMIGMLFVQPMIGMMSDRFGRRPFVIIGSIALM  337
Pseudomonas_aeruginosa        M LTYMPSYLSHN HYSEDHGVLIIIAIMVGMLFVQPVIGMLSDRFGRRPFILIGSVALF  336
Dickeya_dadantii              M LTYMPSYLSHS HYSENHGVLIIIAIMIGMLFVQPVMGLLSDRFGRKPFVVIGSVAMF  337
Yersinia_pestis               V ILMTPTYLQKQ NVPPELALQANSLAIIALVIGCVVAGLAIDRFGASKTFIVGSLMLA  342
Providencia_stuartii          M LTYMPSYLSHN NYSADHGVLIIIAIMIGMLFVQPIIGLTSDKIGRRPFVIAGSLGLI  337
Serratia_marcescens           M LTYMPSYLSHN HYSEDHGVLIIIAIMIGMLFVQPVIGLTSDRIGRKPFIIGGSIGLL  337
Proteus_mirabilis             M LTYMPSYLSHN NYSADHGVLIIIAIMIGMLFVQPVIGLLSDKIGRKPFVIGGSVGLF  337
Acinetobacter_baumannii       M LTYLPSYFSHN GYSEAHGALIIIAVMVGMLFVQPVIGYLSDKFGRRPFIFIGSFSLI  332
Agrobacterium_radiobacter     L VAFGLTYGTQA KISRNEMLVIVLIACAVCIVLLPLFGWLSDRIGRRPVILGGIIAEA  306
Bacillus_anthracis            M LSYIPSYLTQV KVKETTGLLIISITMALMIPLALYFGKLSDKIGNKRVVQIGLLGLT  331
Stenotrophomonas_maltophilia  M LTYMPSYLSVT GYAESKGLLLIIIVMLVMMPLNIVGGLFSDRLGRRPMIIGACIALL  340
Rickettsia_rickettsii         L KTYINVFYYNV HLSNTIALSYLAYSSFIAMIAMPLAGGTADIIGKFKMAMLVGVAIL  306
Rhodococcus_rhodochrous       V TIFVIAYATTY DYTRGAIVTTVAFASVCQFLGMIGGGWWSDRVGRKIAMLVPAVSLV  337
Streptomyces_fulvissimus      M TSYLPTYMSQT GEPETTSQLLVLGTMLLVVLTITTVGRSSDRWGRRPVFMAGSVALI  344
Burkholderia_cepacia          M LSYLPSFMSST HFDESHSLVLVLIVMVLMMPLTLAAGRLSDRIGRKPVMLAGCVGLL  347
Erwinia_amylovora             M LTYMPSYLSHN HYSEDHGVMIIIAIMLGMLFVQPVMGLMSDKFGRRPFVIIGSIALL  337
Staphylococcus_aureus         M TAYLPTYLEQV KLDATTTSVLITCVMAIMIPLALMFGKLADKIGEKKVFLIGTGGLT  322
Campylobacter_jejuni          V VLLMPKFMPSI NLSGVEGSYLQILGILGIALGGAFMGYLVDKFGLFKICIFFSLTFV  317
```

Salmonella_typhimurium	MLLTYMPSYLSHNLHYS	DHGVLIIIAIM	GMLFVQPVMGLLSDRFGRRPFVIMGSIALF	336
Citrobacter_freundii	MLLTYMPSYLSHNLHYS	DHGVLIIIAIM	GMLFVQPIMGLLSDRFGRRPFVIMGSIALF	336
Escherichia_coli	MLLTYMPSYLSHNLHYS	DHGVLIIIAIM	GMLFVQPVMGLLSDRFGRRPFVLLGSVALF	336
Enterobacter_aerogenes	MLLTYMPSYLSHNLHYS	DHGVLIIIAIM	GMLFVQPIMGLLSDRFGRKPFIILGSVALF	336
Cronobacter_sakazakii	MLLTYMPSYLSHNLHYS	EHGVLIIIAIM	GMLFVQPVMGLLSDRFGRRPFVIFGSVALM	337
Pantoea_agglomerans	MLLTYMPSYLSHNLHYS	DHGVMIIIAIM	GMLFVQPMIGMMSDRFGRRPFVIIGSIALM	337
Pseudomonas_aeruginosa	MLLTYMPSYLSHNLHYS	DHGVLIIIAIM	GMLFVQPVIGMLSDRFGRRPFILIGSVALF	336
Dickeya_dadantii	MLLTYMPSYLSHSLHYS	NHGVLIIIAIM	GMLFVQPVMGLLSDRFGRKPFVVIGSVAMF	337
Yersinia_pestis	VVILMTPTYLQKQFNVP	ELALQANSLAI	ALVIGCVVAGLAIDRFGASKTFIVGSLMLA	342
Providencia_stuartii	MLLTYMPSYLSHNLNYS	DHGVLIIIAIM	GMLFVQPIIGLTSDKIGRRPFVIAGSLGLI	337
Serratia_marcescens	MLLTYMPSYLSHNLHYS	DHGVLIIIAIM	GMLFVQPVIGLTSDRIGRKPFIIGGSIGLL	337
Proteus_mirabilis	MLLTYMPSYLSHNLNYS	DHGVLIIIAIM	GMLFVQPVIGLLSDKIGRKPFVIGGSVGLF	337
Acinetobacter_baumannii	MLLTYLPSYFSHNLGYS	AHGALIIIAVM	GMLFVQPVIGYLSDKFGRRPFIFIGSFSLI	332
Agrobacterium_radiobacter	LIVAFGLTYGTQALKIS	NEMLVIVLIAC	VCIVLLPLFGWLSDRIGRRPVILGGIIAEA	306
Bacillus_anthracis	MILSYIPSYLTQVLKVK	TTGLLIISITM	LMIPLALYFGKLSDKIGNKRVVQIGLLGLT	331
Stenotrophomonas_maltophilia	MLLTYMPSYLSVTMGYA	SKGLLLIIIVM	VMMPLNIVGGLFSDRLGRRPMIIGACIALL	340
Rickettsia_rickettsii	LVKTYINVFYYNVMHLS	TIALSYLAYSS	IAMIAMPLAGGTADIIGKFKMAMLVGVAIL	306
Rhodococcus_rhodochrous	VYTIFVIAYATTYFDYT	GAIVTTVAFAS	CQFLGMIGGGWWSDRVGRKIAMLVPAVSLV	337
Streptomyces_fulvissimus	MITSYLPTYMSQTLGEP	TTSQLLVLGTM	LVVLTITTVGRSSDRWGRRPVFMAGSVALI	344
Burkholderia_cepacia	MVLSYLPSFMSSTLHFD	SHSLVLVLIVM	LMMPLTLAAGRLSDRIGRKPVMLAGCVGLL	347
Erwinia_amylovora	MLLTYMPSYLSHNLHYS	DHGVMIIIAIM	GMLFVQPVMGLMSDKFGRRPFVIIGSIALL	337
Staphylococcus_aureus	MVTAYLPTYLEQVIKLD	TTTSVLITCVM	IMIPLALMFGKLADKIGEKKVFLIGTGGLT	322
Campylobacter_jejuni	VLVLLMPKFMPSILNLS	VEGSYLQILGI	GIALGGAFMGYLVDKFGLFKICIFFSLTFV	317

Supplementary 2: The 6 regions used to create figure 2, and the ProP orthologs.