Spring 5-2018

Functional studies of the *E. coli* proC and a putative ortholog Mrub_1345

Maureen Azar
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College, Rock Island Illinois

Functional studies of the *E. coli proC* and a putative ortholog Mrub_1345

Introduction

Proline biosynthesis

Proline is an important amino acid of proteins. It is unique because it is the only amino acid with a side chain connected twice to the protein backbone, forming a five-membered ring (Betts and Russell, 2003). It is often found where the polypeptide chain must change direction, and it can introduce kinks into α-helices. In addition, there are many published examples in both prokaryotes and eukaryotes of proline's importance in physiological processes, such as being used as an energy source, as an osmolyte for stress-protection and in redox control (*i.e.,* scavenger of free radicals), among other functions. Proline accumulation was originally implicated in osmotic stress tolerance in bacteria (Christian, 1955a,b) and it has been suggested that proline functions as a compatible solute that enables the cytoplasm to remain hydrated under conditions of high external osmolarity. For a recent review of proline biosynthesis and uses see Fichman *et al.* (2015).

Proline biosynthesis can occur *via* two pathways. The first and most common pathway begins with glutamate, while the other begins with arginine (Deutch *et al.,* 1982; Smith *et al*., 1980). The glutamate pathway is the main mechanism of proline biosynthesis in bacteria, whereas eukaryotes use this route predominantly under stress. The glutamate to proline pathway, as seen in Figure 1 for *E. coli* K12, is a 4-step pathway, which was first proposed by Vogel and Davis (1952). The pathway has been confirmed in many bacteria, including *E. coli* (see reviews Adams and Frank, 1980; Csonka and Leisinger, 2007; and Fichman *et al*., 2015). Fichman *et al.* (2015) used the SEED database to analyze how different organisms synthesize proline, which involved comparing 56 archaebacteria, 821 eubacteria and 13 eukaryotes. Of the genomes analyzed, 8 archaeal, 681 eubacterial (including *E. coli*) and all complete eukaryotic genomes have at least one ortholog for each of the three enzymes of the most common glutamate pathway. The level of proline biosynthesis is invariant regardless of proline starvation or excess (studies reviewed in Csonka and Leisinger, 200. Thus, proline is not subject to transcriptional control, as are most other amino acids. Feedback inhibition of the first enzyme in the pathway appears to be the only control mechanism.

Pyrroline-5-carboxylate reductase (P5CR/ProC)

The last reaction in the proline biosynthesis pathway is 1-pyrroline-5-carboxylate and NAD(P)H to proline and NAD(P)+, which is catalyzed by NADPH-dependent pyrroline-5-carboxylate reductase (P5CR or ProC in *E. coli*; EC 1.5.1.2). *E. coli K12* ProC was partially purified and characterized by Rossi *et al. (*1977) and estimated to be 320,000 D in size. *E. coli* ProC was subsequently purified and the gene *proC* cloned by Deutch *et al. (*1982). Its monomeric molecular weight determined to be 28,112 D, corresponding to 269 amino acids in length. The revised multimeric molecular weight was 280,000 D. The *E. coli* K12 MG1655 *proC* gene is 810 base pairs in length. It is located at map position 404,835-405,644 (8.72 centisomes, 31°), and its locus tag is b0368 as retrieved from the database Ecocyc (Keseler *et al*. 2013). Originally

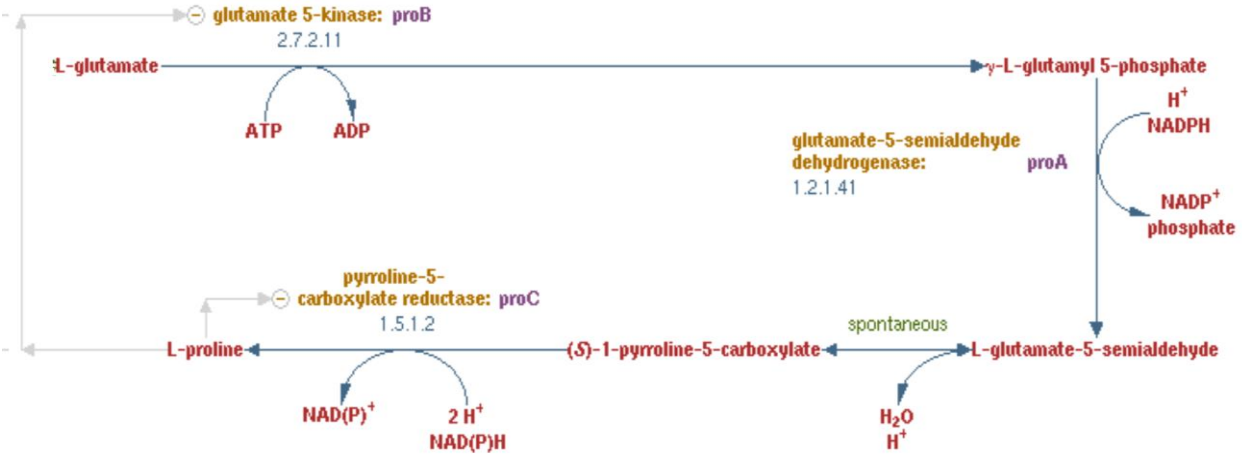


Figure 1. Proline biosynthesis pathway: The biosynthesis of L-proline in *E. coli* K12 MG1655 involves four steps beginning with L-glutamate. The first step reduces of L-glutamate to glutamate semialdehyde in two consecutive enzymatic steps, catalyzed by glutamate 5-kinase and glutamate-5-semialdehyde dehydrogenase. The next step is an exergonic spontaneous dehydration reaction, converting L-glutamate γ-semialdehyde to pyrroline-5-carboxylate. In the final step, the pyrroline derivative is reduced to L-proline, which is catalyzed by the NADPH-dependent pyrroline-5-carboxylate reductase. Enzymes in bold have been experimentally confirmed. Image taken with modification from Ecocyc, from

identified in *Neisseria meningitides* and *Streptococcus pyogenes (*Nocek et al, 2005*)*, P5CR consists of two domains, an N-terminal NADP-binding Rossmann fold and a C-terminal pyrroline-5-carboxylate binding site. As predicted by the Pfam database ((Finn *et al.*, 2014, Finn *et al.*, 2016); *E. coli* K12 ProC has the same functional domains: 1) F420_oxidored between amino acids 4-99 as the binding site of NADH/NADPH; and 2) the P5CR domain between amino acids 163-266 as the 1-pyrroline-5-carboxylate binding/catalytic site. Fichman *et al.* (2015) created a consensus sequence derived from P5CR of various organisms, including *E. coli*, using ConSurf server and the ConSeq program, and superimposed it onto the primary sequence of *Arabidopsis thaliana* P5CR and the 3D structure of human P5CR (*aka* PYCR1, Meng *et al.*, 2006). This allowed for the determination of highly conserved amino acids in ProC, as well as determined which amino acids are predicted to be functional or structural, and which are located on the surface or embedded in the protein. These predictions helped inform which amino acids were studied in our functional genomics experiments. The multiple sequence alignment tool T-Coffee (Notredame *et al.*, 2000). In conjunction with the program Weblogo Weblogo (Crooks et al., 2016) and the HMM logo of PFAM (Finn et al., 2014, Finn et al., 2016) were additional resources used in this investigation to identify highly conserved amino acids in *E. coli proC* and Mrub_1345.

We are using *E. coli* K12 MG1655 as a reference organism in the study of proline biosynthesis in *M. ruber* DSM1279 for these reasons: 1) the extensive research on proline

biosynthesis in *E. coli* (see reviews in Fichman *et al.,* 2015; Csonka and Leisinger, 2007*)*; 2) the availability *E. coli* K12 null strains called the Keio Collection (Baba et al., 2006); 3) the availability of *E. coli* K12-dedicated database Ecocyc (Keseler et al., 2013); 4) the availability of an *E. coli* complementation assay and applicable resources, which were provided through the NSF-funded online resource GENI-ACT (Dr. Kathryn Houmiel, GENI-ACT Program Coordinator, Seattle Pacific University, Personal Communication; Dr. Derek Wood, Seattle Pacific University, Personal Communication; GENI-ACT at http://geni-science.org/).

***Meiothermus ruber* DSM1279.**

Meiothermus ruber is a Gram-negative eubacteria belonging to the phylum Deinococcus-Thermus. The order Thermales, which is housed within the Thermus group and consists of 6 genera (Thermus, Marinithermus, Meiothermus, Oceanithermus, Rhabdothermus, and Vulcanithermus), is notable for containing thermophilic and hyperthermophilic genera possessing thermostable enzymes(Albuquerque and Costa, 2014]. *M. ruber* is one of eight currently known species in the genus *Meiothermus* (Euzeby, 1997). The genus name derives from the Greek words 'meion' and 'thermos' meaning 'lesser' and 'hot' to indicate an organism in a less hot place [Nobre et al., 1996; Euzeby, 1997]. It lives in natural hot springs and artificial thermal environments; it has an optimal growth condition of 60^{o}C and a growth range of 35-70^{o}C. The epithet "ruber" refers to the red pigment it produces.

The *Meiothermus ruber* DSM1279 genome was sequenced through a collaboration between the Joint Genome Institute and Leibniz-Institut DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH) called the Genomic Encyclopedia of Bacteria and Archaea (GEBA) project, the goal of which is to systematically fill in gaps in sequencing along the bacterial and archaeal branches of the tree of life (Wu *et al.,* 2009). Among the many projected benefits of the GEBA project are novel gene discovery, the identification of novel biochemical processes, and a better understanding of the processes underlying the evolutionary diversification of microbes (*e.g.*, lateral gene transfer and gene duplication). The complete genome consists of over 3 million base pairs, over 3000 protein-coding genes and 53 RNA genes; the genome has 63.4% GC content (Tindall *et al.,* 2010). A genome statistics summary is provided in Table 1, which was taken directly from Tindall *et al.* (2010, Table 3). Many features make *M. ruber* an interesting organism to study; for example, its tolerance to high growth temperatures. A few early bioinformatics studies have identified non-traditional biosynthetic pathways (*e.g.*, lysine and arginine biosynthesis). The study of proline biosynthesis and regulation is interesting because of proline's role as an osmolyte for stress-protection.

Bioinformatic analysis of Mrub_1345/Gene ID 646672805

The bioinformatics evidence predicting that the open reading frame Mrub_1345 (GenBank Gene ID 646672805) encodes pyrroline-5-carboxylate reductase (*aka* ProC) was originally performed via automated annotation by Tindall *et al.* (2010), which was subsequently confirmed by manual annotation (Wills and Scott, 2015). Mrub_1345 is 783 nucleotides and 260 amino acids in length; it is located at map position 1378760-1379542. Among other similarities to *E. coli* ProC, Mrub_1345 is assigned to the same TIGRfam family (Haft *et al.,* 2001), has the same two PFam domains and it belongs to the same COG (*e.g.*, tigrfam00112: pyrroline-5-carboxylate

Table 1. Genome statistics from the *Meiothermus ruber* genome (original table from Table 3 of Tindall *et al.*, 2010)

Attribute	Value	% of Total
Genome size (bp)	3,097,457	100.00%
DNA Coding region (bp)	2,807,535	90.64%
DNA G+C content (bp)	1,963,304	63.38%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	3,105	100.00%
RNA genes	53	1.71%
rRNA operons	2	
Protein-coding genes	3,052	98.29%
Pseudo genes	38	1.22%
Genes with function prediction	2,229	71.79%
Genes in paralog clusters	390	12.56%
Genes assigned to COGs	2,286	73.62%
Genes assigned Pfam domains	2,394	77.10%
Genes with signal peptides	1,079	34.75%
Genes with transmembrane helices	697	22.45%
CRISPR repeats	6	

reductase; pfam03807/NADP oxidoreductase coenzyme F420-dependent; pfam14748/P5CR_dimer; COG0345: Pyrroline-5-carboxylate reductase, respectively). Further evidence of an orthologous relationship is from a BLASTp (Altschul, *et al.,* 1990) comparison between *E. coli* ProC and *M. ruber* Mrub_1345, which resulted in an E-value of 1e-40, an Identities score of 33% and a Positive score of 53% (Wills and Scott, 2015). Both proteins are localized to the cytoplasm, as determined by the lack of transmembrane helices and signal peptides. The bioinformatics tools used for these analyses included TMHMM (Krogh *et al.,* 1998; Krogh *et al.,* 2001), PSORTb (Yu *et al.,* 2010), SignalP (Peterson *et al.,* 2011), LipoP (Juncker *et al.,* 2016) and Phobius (Kall *et al.,* 2004; Kall *et al.,* 2007).

Complementation as a functional genomics tool

While bioinformatics tools can predict the function of gene/protein using computational means, functional genomics uses experimentation to determine function. In our lab, *M. ruber* open reading frames (ORFs) are cloned into the pKt1 expression vector via a ligase independent cloning process developed by Dr. Kathryn Houmiel (Seattle Pacific University, personal

communication; protocol taken from GENI-ACT at http://geni-science.org/). The pKt1 vector (Figure 7) contains the *sacB* gene within the cloning site, which is integrated downstream of an arabinose inducible promoter. Components of the arabinose operon are the *araC* and the *araBAD* promoter, which allows for transcriptional activation of an inserted ORF in the presence of arabinose. A pKt1 containing a putative *M. ruber* gene was transformed into a corresponding null strain from the Keio Collection (Baba *et al.,* 2006). In these null strains, open-reading frame coding regions were replaced with a kanamycin cassette flanked by FLP recognition target sites to create in-frame deletions upon excision of the resistance cassette. These mutants provide a resource for the systematic analyses of unknown gene functions and gene regulatory networks and for genome-wide testing of mutational effects in a common strain background, *E. coli* K-12 BW25113. Clones are screened on medium with or without supplemental arabinose or proline to identify those that can restore the strain to prototrophy. One published example of using the Keio strains for a complementation assay of P5CR is the oomycete plant pathogen *Phytophthora nicotianae* (Ambikapathy *et al.,* 2002). Complementation of a Keio strain defective in the P5CR protein by the *P. nicotianae* P5CR cDNA confirmed that the gene encoded a functional P5CR. The *proC* genes of three species of *Agrobacterium* have been tested using the pKt1 system (Brad Goodner, Hiram College, Hiram, OH; personal communication; protocol taken from GENI-ACT at http://geni-science.org/).

Site-directed mutagenesis as a functional genomics tool

Site-directed mutagenesis (SDM) is a useful technique for exploring the structure and therefore biological activity of DNA, RNA, and protein molecules, since structure and function are interconnected and a change to structure can induce changes in function (for reviews see Smith, 1982; Plapp, 1995; Betts and Russell, 2003). It is also commonly called site-specific mutagenesis or oligonucleotide-directed mutagenesis and has been practiced since about 1980. In its simplest form, SDM involves the deletion of a segment from a circular virus, bacteriophage or recombinant DNA, followed by religation of the shortened DNA, and subsequent genetic complementation analysis in which the specifically shorted DNA fails to complement a genetically-mapped defective gene obtained by conventional mutagenesis (Smith, 1982). Through the years, many approaches have been developed. In this project, two PCR-based strategies, Q5 developed by New England Biolabs and QuikChange developed by Stratagene, were employed to make specific amino acid substitutions for the purpose of identifying essential or critical residues for catalysis and ligand binding. General acids, general bases and catalytic nucleophiles seem to be the most essential residues in an active site as they directly participate in the formation and breaking of covalent bonds (Peracchi, 2001). Alanine-scanning, the systematic substitution of potentially critical amino acids with alanine has proven especially powerful in determining the role of sidechain functional groups at specific positions (Morrison and Weiss, 2001; Betts and Russell, 2003). Alanine is a non-polar, uncharged amino acid with a small methyl R group. As described by Betts and Russell (2003), "[alanine] is probably the dullest amino acid" and it is commonly found in noncritical positions in the polypeptide sequence. It rarely seems to play a role in substrate recognition or specificity.

The usual approach of measuring the impact of mutations on protein function is via enzyme kinetics (*e.g.,* the determination of the K_{cat} and K_m for a reaction) (Peracchi, 2001). While no systematic mutational studies have been applied to *proC* in *E. coli* K12 or *M. ruber*,

Perez-Arellano *et al.* (2006) induced a series of mutations in glutamate-5-kinase (GK) of *E. coli* K12, the first step of proline biosynthesis. Among other outcomes, the results supported the functional roles of K10, K217, and T169 in the catalysis and ATP binding; the roles of D148 and D150 in glutamate binding at the active site; and the roles of D148 and N149 in proline binding.

Precautions in interpreting outcomes of SDM must be considered, however. Peracchi (2001) and Plapp (1995) describe examples of mutations in "essential" residues that generally did not annihilate activity and sometimes leave most of the catalytic power of the enzyme intact. Also, the criterion that an amino acid residue is essential if its substitution totally inactivates is obsolete (Plapp, 1995). While mutagenesis of residues involved in catalysis has highlighted the interplay between different catalytic devices and strategies, it has also softened the distinction between essential and nonessential groups and demonstrating the functional plasticity of enzyme active sites (Peracchi, 2001). The GK example described above illustrates this issue. Except for the D150A mutant, all the other 13 mutants were active to some degree as compared to the wild-type GK (Perez-Arellano *et al.*, 2006). However, activity was less than 1% of wild-type for the K10, N149, D150, D170 and K217 mutants.

Previous functional genomics work within the *Meiothermus ruber* genome analysis project

Early work in this lab demonstrated the feasibility of expressing *M. ruber* ORFs in an auxotrophic Keio *E. coli* host, as well as the ability to discern activity levels. For example, several *M. ruber* ORFs predicted to encode enzymes in arginine, lysine and histidine biosynthesis have been confirmed as orthologs of *E. coli* genes. Figure 2 shows the complementation assay for the Mrub_1080-1079/pKt1 transformed into a Keio *E. coli proB* null strain. Mrub_1080 and Mrub_1079 are adjoining ORFs predicted to be orthologs of *E. coli proB* and *proA*, respectively. ProB and ProA are the enzymes catalyzing the first two steps in proline biosynthesis, respectively (See Figure 1), and appear to form a functional complex critical to the activity of both enzymes (Hayzer and Moses, 1978; Leisinger, 1996). As previously observed in *E. coli* (Deutch *et al.*, 1984; Leisinger, 1996), *proB* and *proA* constitute an operon with a single promoter proximal to *proB*. In *M. ruber*, the putative orthologous genes Mrub_1079 and Mrub_1080 are predicted to be an operon due to their adjacent position and similar transcriptional orientation in the genome.

Functional complementation has been observed between *E. coli proB* and a *proB*74 mutant individually transformed into *S. cerevisiae proB*/pro1 mutant strain (Orser *et al.*, 1998). In our system, *E. coli proB* and *E. coli proA* readily complemented their respective null strains on both minimal and minimal supplemented with arabinose. Mrub_1080/pKt1 and Mrub_1079/pKt1 weakly complemented their respective null strains after an extended incubation period on minimal with arabinose. We hypothesized that the weak growth response is due to incompatibility in the interspecies complex formed between the *E. coli* and *M. ruber* proteins. The Mrub_1080-1079 construct also weakly complemented both the Keio *E. coli proB* and *proA* auxotrophic strains. The addition of an *E. coli* Shine Delgarno sequence upstream of *proA* in the Mrub_1080-1079 construct, did not enhance complementation. Thus far, complementation of all Mrub genes tested requires arabinose induction; extending the incubation time from the usual two days to four days has thus far resulted in identifying weaker complementation outcomes for *M. ruber* genes grown on minimal with arabinose.

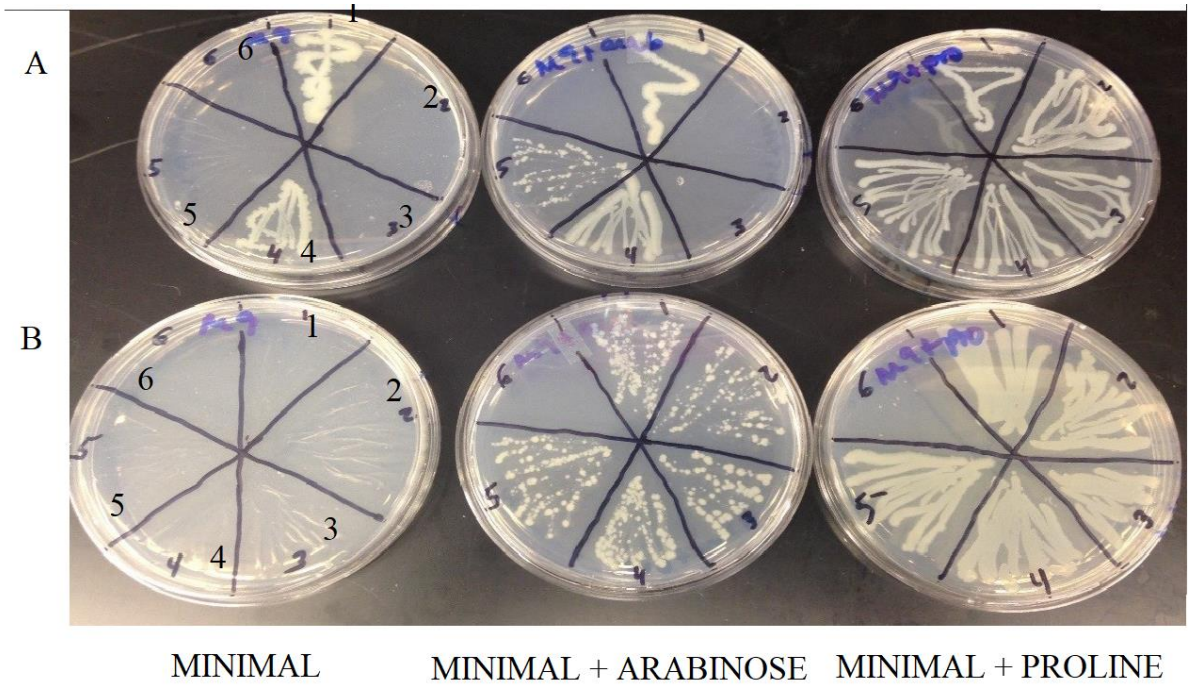


Figure 2. An Mrub_1080-1079 construct weakly but successfully complemented a Keio *E. coli proA* auxotroph. Unless otherwise noted, the *Keio E. coli proA* auxotroph is the host strain for all samples. Panel A: 1) Wild-type *E. coli* K12; 2) *Keio E. coli proA* auxotroph; 3) pKt1; 4) *E. coli proBA* in pKt1; 5) Mrub_1080-1079 in pKt1; 6) empty; Panel B: 1-5 are individual isolates of Mrub_1080-1079 in pKt1; sector 6 is empty.

In an initial mutagenesis study of *proC*, *E. coli proC* (positive control), Mrub_1345, *E. coli proC* Gly11Asp mutant and Mrub_1345 Gly9Asp mutant were cloned separately into the pKt1 expression vector and transformed into the Keio *E. coli proC* null strain JW0377-1. Glycine/Gly/G is a small amino acid with a hydrogen as its side chain. It is hydrophobic and often burired inside the protein core. According to Schulze-Gahmen *et al.* (1996), glycines play a distinct functional role, such as using its backbone to bind to phosphates, which means that substituting glycine for any other amino acid could have drastic impact on a protein's function. Gly17 and Gly19 in *A. thaliana proC* were identified by Fichman *et al* (2006) as highly conserved among prokaryotes and eukaryotes, and possessing a critical structural function. Therefore, these residues were predicted to be critical to NAD(P)H binding (Figure 3, Panel A). Panel B of Figure 3 shows a PFam HMM logo (Finn *et al.*, 2014, Finn *et al.*, 2016) for the NADPH binding domain of P5CR (PF03807), which supports the Fichman *et al.* (2015) findings. A Weblogo (Crooks *et al.*, 2004) constructed from species within the Deinococcus-Thermus phylum (Panel C, Figure 3), which includes *M. ruber*, shows the corresponding Gly7 and Gly9 as highly conserved. These positions correspond to Gly11 and Gly13 in *E. coli proC* (data not shown).

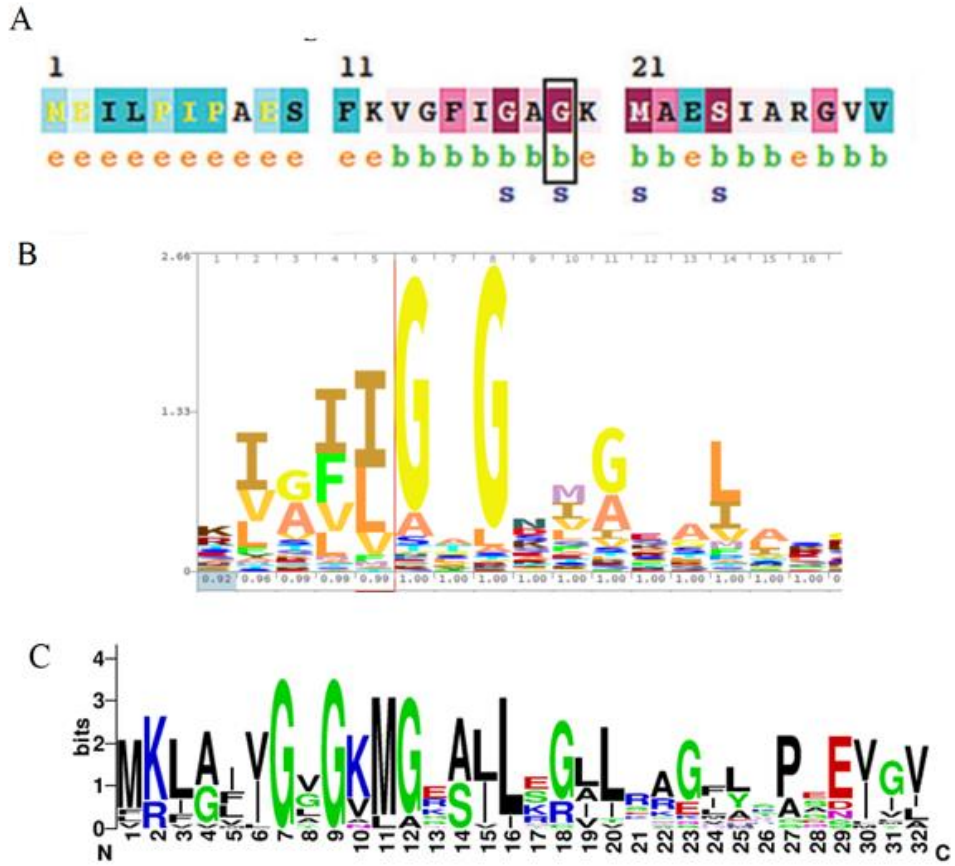


Figure 3. Consensus sequences of the NH_2-terminus of the NADPH binding domain of P5CR reveals the two highly conserved glycine (G) residues. Panel A - ConSurf sequence (Panel A) (Fichman *et al.*, 2015); Panel B - Pfam PF03807 HMM logo ((Finn *et al.*, 2014, Finn *et al.*, 2016); and Panel C - Weblogo (Crooks *et al.*, 2004). The Weblogo was generated from species in the Deinococcus -Thermus phylum only. In Panel A, black boxes denote conserved amino acid residues that are also indicated for human P5CR. Lower case letters mean: e = exposed amino acid; b = buried amino acid.

It was predicted that a Gly to Asp mutation would result in loss of function for *proC* due to their differing chemistries and function. Aspartate/Asp/D has a large charged side chain that is polar, and aspartates are typically located on the surface of proteins. According to Betts and Russell (2003) aspartate is frequently involved in protein active or binding sites due to its negative charge and it ability to interact with positively charged R groups and non-protein atoms. Surprisingly, both the wild-type and mutant G→D *E. coli* clones complemented on minimal and minimal with arabinose, while the wild-type and mutant G→D *M. ruber* clones complemented only on minimal supplemented with arabinose (Figure 4). We hypothesized that the difference in optional growth temperature between *E. coli* and *M. ruber* (37°C and 60°C, respectively) might negatively impact the function of *M. ruber* genes in the *E. coli* background, which could be

compensated by arabinose induction. Consequently, we confirmed that Mrub_1345 encodes pyrroline-5-carboxyate reductase via complementation, but we were unable to detect a loss of function for the mutants under these conditions.

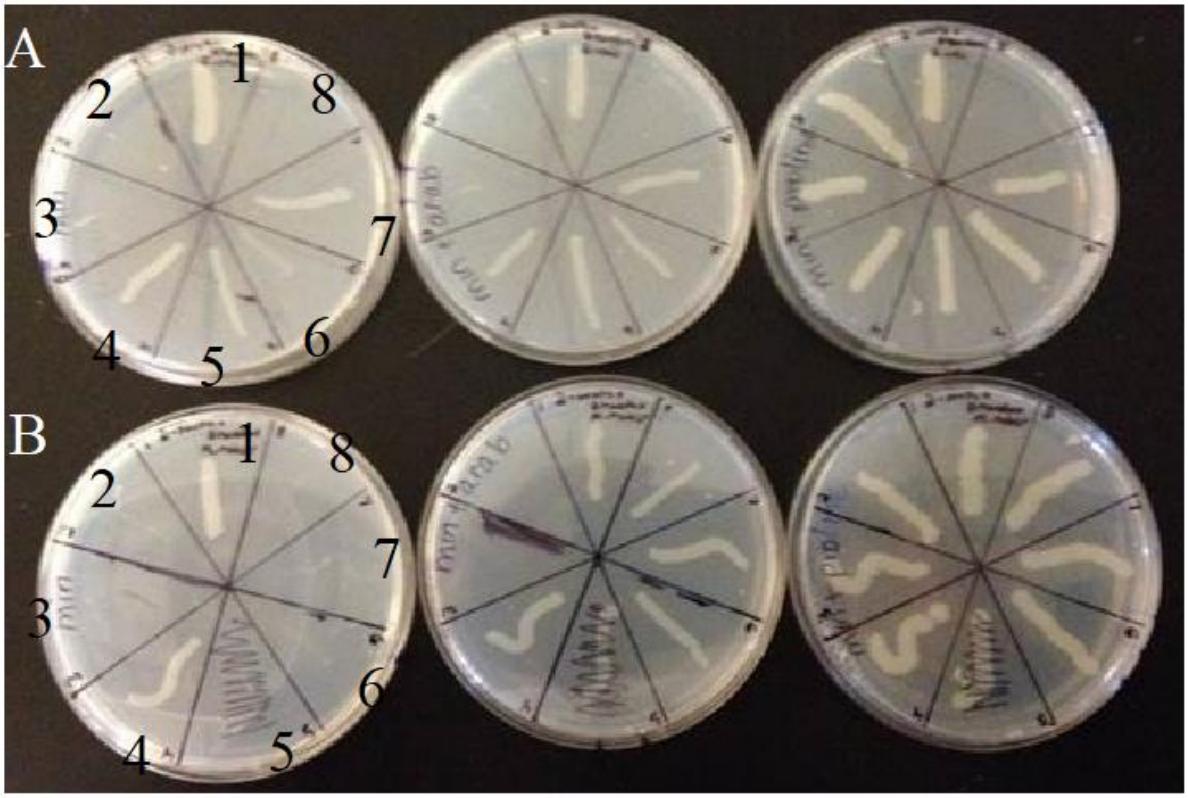


Figure 4. Mrub_1345 is orthologous to *E. coli proC*, and neither a Gly9Asp in Mrub_1345 or Gly11Asp in *E. coli proC* demonstrated a loss of function in the complementation assay. Unless otherwise noted, the *Keio E. coli proC* auxotroph is the host strain for all samples. Far left plates are minimal media; middle plates are minimal + arabinose; far right plates are minimal + proline. Panel A. Sector 1) wild-type *E. coli* K12 strain; 2) Keio *E. coli proC* null strain; 3) pKt1; 4) *E. coli proC*+/pKt1; 5-7) *E. coli proC* Gly11Asp mutant/pKt1; 8: empty sector. Panel B. 1) wild-type *E. coli* K12 strain; 2) Keio *E. coli proC* null strain; 3) pKt1; 4) *E. coli proC*+/pKt1; 5) Mrub_1345/pKt1; 6-8) Mrub_1345 Gly11Asp mutant/pKt1.

In this project, additional amino acids predicted to be functionally important to ProC were identified by utilizing at least two of the consensus sequences described above. Summarized below are descriptions of the roles of these amino acids to protein function, as described by Betts and Russell (2003).

- Arginine (R) is a large positively charged polar amino acid. It tends to be on the exterior of the protein, but its amphipathic nature allows it to be buried in the protein. Arginine form a

stabilizing hydrogen bond via a salt-bridge with a negatively charged aspartic acid or glutamic acid.

- Histidine (H) is a polar amino acid, but it substitutes poorly with any other amino acid. It can move protons on and off of the side chain, which makes it a good residue as the protein functional center. Histidines are the most common amino acids in protein active or binding sites. Histidine can rarely be exchanged for other amino acids.
- Threonine (T) can reside both within the interior of a protein or on the protein surface, and is commonly in protein functional centers. The hydroxyl group is fairly reactive, and can form hydrogen bonds with a variety of polar substrates.
- Tyrosine (Y) is usually buried in protein hydrophobic cores. A common role for tyrosines is phosphorylation by protein kinases. In this context, substitutions or loss of tyrosine is rarely tolerated due to the specificity of tyrosine kinases (Hanks *et al.*, 1988).
- Proline (P) is unique in that it is the only amino acid with a side chain is connected to the protein backbone twice, forming a five-membered ring. It is often found where the polypeptide chain must change direction. It can also introduce kinks into α-helices.

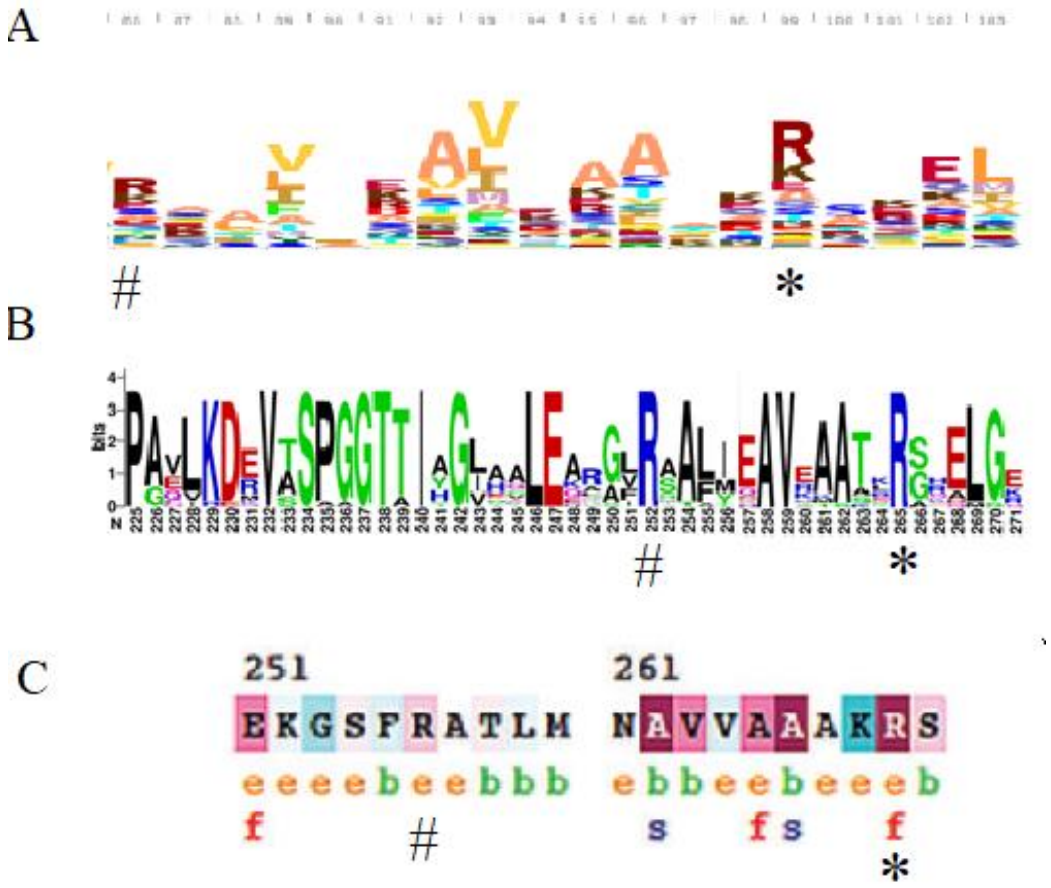
The Pfam HMM logo, Weblogo and the predictions from Fichman *et al.* (2015) did not always agree on the degree of conservation for a particular residue, however. For example, Fichman *et al.* (2015) and the Weblogo of 10 Deinococcus-Thermus species were used to predict that Arg239 (#) and Arg252 (*) in Mrub_1345 (Figure 5, Panels B and C) and Arg249 in *E. coli proC* (Weblogo not shown) are highly conserved, but the Pfam HMM logo suggested only marginal conservation of both residues in Mrub_1345 (Figure 5, Panel A). *E. coli proC* does not have an analogous residue to Mrub_1345 Arg252. Despite these issues, 7 amino acids in both Mrub_1345 and *E. coli proC* were tested in this project, using a combination of base-pair substitutions and deletion mutations (see Table 1). An attempt was made to mutate the analogous residues in both *E. coli proC* and Mrub_1345, but that was not always achieved.

The same three sequential amino acids were deleted in both *E. coli proC* and Mrub_1345. It was hypothesized that deletions of highly conserved amino acids might cause greater damage to the protein, especially if they were positioned sequentially. In Mrub_1345, the PAY residues beginning at position 169 and the same residues in *E. coli proC* beginning at position 176 were deleted, first the P, then PA and then PAY. All three consensus sequences suggested a high degree of conservation for this region (Figure 6). Serindipotously, a substantial deletion of the P5CR domain was obtained for both Mrub_1345 and *E. coli proC* as unintended products of PAY mutagenesis, as was a10-base pair duplication of Mrub_1345. All three of these mutants were tested for complementation.

Purpose of this study

This investigation highlights how the functional study of genes from the thermophile *M. ruber* can be performed using the Keio *E. coli* null strains (Baba *et al.*, 2006) and pKt1 expression vector (Dr. Kathryn Houmiel, Seattle Pacific University; houmik@spu.edu) in a complementation assay (Dr. Derek Wood, Seattle Pacific University; GENI-ACT at http://geni-science.org/). We applied specific bioinformatics tools (collected under the online

GENI-ACT platform) and the complementation assay to determine the function of Mrub_1345. In addition,



this

investig of
potentia ed
amino a me
kinetics uate
researc
pyrolli . We
also pro
substitu uld
negative *oli*
compl

Figure 5: The COOH-terminus of the P5CR domain contains several highly conserved amino acids, two corresponding R are identified by "#" across the three consensus sequences. Panel A – a portion of the Pfam HMM logo for PF14748 (Finn *et al.*, 2014, Finn *et al.*, 2016); Panel B - Weblogo (Crooks *et al*., 2004) for 10 species within the Deinococcus-Thermus phylum including Mrub_1345; Panel C – a portion of Figure 6 from Fichman *et al.*, 2015, where maroon means highly conserved and lower case letters mean: e = exposed amino acid; b = buried amino acid; f=predicted functional amino acid; and s = a predicted structural amino acid.

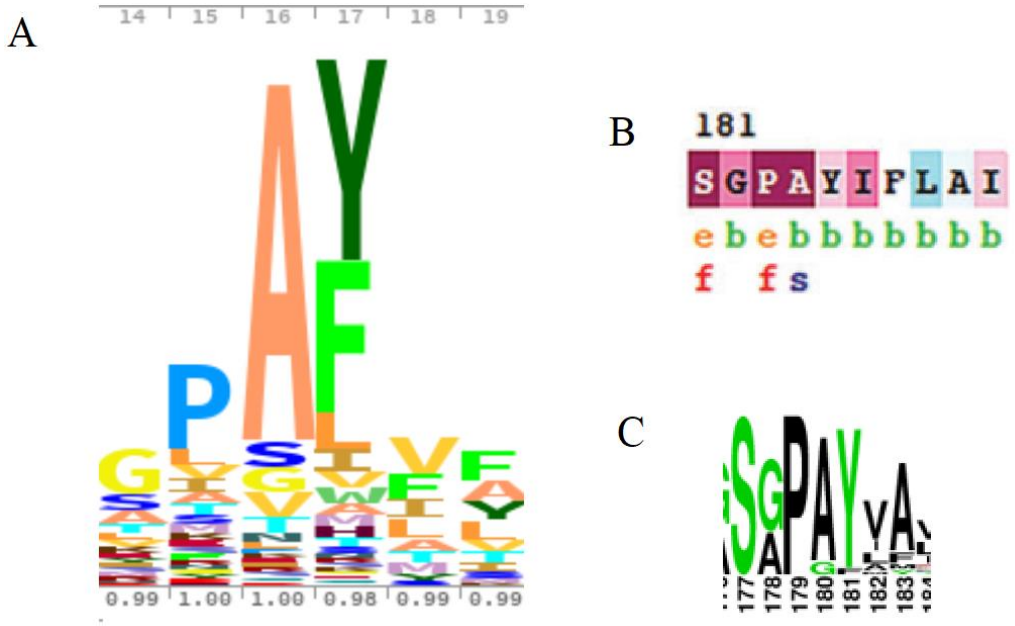


Figure 6. The PAY region of *proC* shows high conservation, based on the Pfam Hmm logo PF14748, Fichman *et al* (2015) and 10-species Weblogo. Panel A – a portion of the Pfam HMM logo for PF14748 ((Finn *et al.*, 2014, Finn *et al.*, 2016); Panel B – a portion of Figure 6 from Fichman *et al. (*2015) of *A. thaliana proC*, where lower case letters mean: e = exposed amino acid; b = buried amino acid; f=predicted functional amino acid; and s = a predicted structural amino acid. Panel C – Weblogo (Crooks *et al.*, 2004) for 10 species within the Deinococcus-Thermus phylum, including Mrub_1345.

Materials and Methods:

Materials.

Reagents and bacterial media were purchased from Fisher Scientific. Bacterial genomic DNA isolation kit, the High-Speed Plasmid Mini Kit and PCR/Gel extraction mini kit were purchased from IBT Scientific. Two site directed mutagenesis kits were used in this project: the Q5 site-directed mutagenesis kit from New England Biolabs and the Agilent QuikChange mutagenesis kit. PCR primers were purchased from Integrated DNA Technologies (Coralville, IA). PCR was performed using the ProMega: GoTaq® Green Master Mix kit. Sanger DNA sequencing was provided by the Iowa Institute of Human Genetics, Genome Division, University of Iowa (Iowa City, IA). The Keio strain JW0377-1 (Baba *et al*.,2006) and *E. coli* K12 1655 were purchased from the Yale *E. coli* Stock Center. *M. ruber* was provided by the ATCC (ATCC Number 35948). The pKt1 plasmid (Figure 7) was provided by Kathryn Houmiel, GENI-ACT Program Coordinator, Seattle Pacific University, houmik@spu.edu).

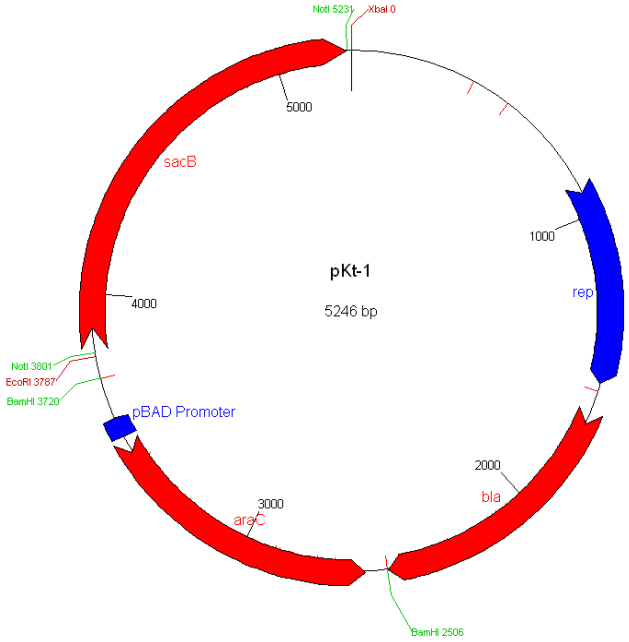


Figure 7. The pKt1 map. The expression vector pKt1 was developed from pBTB backbone modified to contain ligation independent cloning homology regions and the *sacB* gene between the *EcoRI* and *XbaI* sites. The incorporation of the *araC* and *pBAD* promoter regions from the arabinose operon allows for transcription induction in the presence of arabinose. PKt1 was created and provided by Kathryn Houmiel, GENI-ACT Program Coordinator, Seattle Pacific University.

Methods:

Cloning of *proC.* Mrub_1345 and *E. coli proC* (locus tag b0386) were cloned into pKt1 vector using a ligase independent cloning protocol described on the online resource GENI-ACT (http://geni-science.org/), which replaces the *sacB* with either an *E. coli proC* or Mrub_1345. Briefly, a 12-15 nucleotide region of DNA lacking thymine and containing a *not*1 restriction site was engineered into the multiple cloning site of the vector pKt1 (K. Houmiel, personal communication). The complementary sequence was engineered onto the PCR primers (see Table 2) used to amplify Mrub_1345 and E. coli proC from the respective genomic DNA. PKt1 was linearized with not1 restriction enzyme, which cleaved at sites on either side of the 1.4 Kbp sacB gene. This created a 3.8Kbp pKt1fragment that was subsequently purified via agarose gel electrophoresis and gel extraction. The 3.8Kpb pKt1 and purified proC PCR product were treated with T4 DNA polymerase mixture that contained a single nucleotide (dTTP for pKt1 or dATP for proC PCR). Under these conditions, T4 DNA polymerase digested the DNA in the 3'→5' direction until the first specified nucleotide (A for the PCR and T for the pKt1) was reached. The polymerase subsequently idled as the dTTP or dATP were incorporated and then excised

repeatedly. In this way, long single stranded complementary regions on the ends of the pKt1 fragment and the proC PCR products were produced. The proC PCR sample is combined with the 3.8 Kbp pKt1 fragment, allowing the two to hydridize, which were then transformed into an E. coli K12 MG1655 strain. Once inside the host cell, the host provided DNA ligase to seal the final phosphodiester bonds of the recombinant proC/pKt1 molecule. One deviation from the GENI-Science protocol was the transformation step. PKt1 recombinants were transformed into host cells as described by Chung *et al.* (1989). Prior to performing the complementation assay, the sequencing of each recombinant pKt1 strain was confirmed by Sanger sequencing.

Table 2. Primers used in this study

NAME	PRIMER SEQUENCE
Ligase-independent Cloning	
E. coli proC-F	5'-CGA CAA GAG CGG CCG CAT GGA AAA GAA AAT CGG-3'
E. coli proC-R	AAC ACC AAG CGG CCG CCA TCA GGA TTT GCT GAG
Mrub1345-F	CGA CAA GAG CGG CCG CAT GAA ATT GGC TAT CG
Mrub1345-R	AAC ACC AAG CGG CCG CGA AGG ACG AAT AGA TA
Primer 5015 pKT specific F	TCT GAG GCT CGT CCT GAA T
Primer 5016 pKT specific R	TGA CGC TTT TTA TCG CAA CTC
Agilent QuikChange SDM	
Mrub_proC_G9D-F	GCT CTT ACC CAT CTT GTC TAC ACC CAC GAT AGC
Mrub_proC_G9D-R	GCT ATC GTG GGT GTA GAC AAG ATG GGT AAG AGC
Ecoli_proC_G11D_F	TGG CTT TTC CCA TAT TGT CGC AGC CAA TAA AAC CG
Ecoli_proC_G11D_R	CGG TTT TAT TGG CTG CGA CAA TAT GGG AAA AGC CA
Mrub_thr226ala-A	GCC GTG GAT GGC GGT ACC CCC GG
Mrub_thr226ala-B	CCG GGG GTA CCG CCA TCC ACG GC
Ecoli_thr236ala-A	CCG CTT CAA TGG CGG TGC CTC CCG G
Ecoli_thr236ala-B	CCG GGA GGC ACC GCC ATT GAA GCG G
NEB Q5 SDM	
Mr1345_R189A_F	GGC CGA TGT TGC GAT TGC CAC CGG CG
Mr1345_R189A_R	GCA AGC CGC AGG GCC TGG
Ec0386_R197A_F	GGA TGC CAC GGC CCC AGG CGT ATA AAT TTG CCG CTC AGG
Ec0386_R197A_R	CGC CCA GCA CGG CGG CGT
Ec0386_ R249A_F	GAA AGG CTT CGC TGC TGC AGT GAT CGA AG
Ec0386_ R249A_R	TCT TCC AGT ACG CGT ACC
Mr1345_ G253R_F	CTA CCC TGC GGC GGC ACG AAC TGG G
Mr1345_ G253R_R	CGG CCT CCA CTG CCT CCA
EC-delP176-F	GCC TAC GTA TTT ATG TTT ATC GAA G
EC-delP176-R	CGA AGA ACC GCT CAC ACC

MR-delP168-F	GCG TAT GTG GCC GTG GTG G
MR-delP168-R	CGC CGA GGC CGA CAT GCC
EC-delPA176-F	TAC GTA TTT ATG TTT ATC GAA GCG
EC-delPA176-R	CGA AGA ACC GCT CAC ACC
MR-delPA168-F	TAT GTG GCC GTG GTG GCG
MR-delPA168-R	CGC CGA GGC CGA CAT GCC
EC-delPAY176-F	GTA TTT ATG TTT ATC GAA GCG ATG G
EC-delPAY176-R	CGA AGA ACC GCT CAC ACC
MR-delPAY168-F	GTG GCC GTG GTG GCG GAG
MR-delPAY168-R	CGC CGA GGC CGA CAT GCC

The site-directed mutagenesis (SDM). SDM protocols were provided by the kit suppliers, Aligent and New England Biolabs. Table 1 lists the primers used in this study to create a series of base-pair substitutions and deletions of the *E. coli* K12 MG1655 and *M. ruber proC/Mrub_1345* genes. The Aligent QuikChange protocol used an overlapping primer strategy, while the New England Biolab's Q5 protocol used back-to-back primer strategy. In both cases, an online resource was used to create suitable primers. Putative mutants were confirmed by Sanger sequencing through the DNA sequencing facility at the Genomics Division, Human Genetics Institute, University of Iowa, Iowa City.

DNA sequencing analysis. The Expasy Translate Tool (https://web.expasy.org/translate/) was used to convert the DNA sequence data into an amino acid sequence, starting with the ATG start codon. An NCBI BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to create an alignment between the putative mutant ProC sequence and the wild-type sequence in GenBank. The putative mutant sequence was BLAST'ed against the respective genome *(i.e., Escherchia coli* K12 MG1655 or *Meiothermus ruber* DSM1279).

Complementation. The complementation assay used in this study is described on the GENI-ACT site (http://geni-science.org/). A collection of strains was made by transforming the Keio *E. coli proC* null strain with pKt1 containing different inserts (*e.g*., *E. coli proC+*, Mrub_1345, *sacB*, and mutants of the *E. coli proC* and Mrub_1345 genes). Three types of complementation media were made, which were minimal plates, minimal supplemented with arabinose plates and minimal supplemented with proline plates. The various strains were streaked onto these plates and incubated 48 hours at 37ºC. After the complementation assay was complete, we confirmed that each strain carrying the pKt1 vector (with and without the different varieties of *proC*) were confirmed by Sanger sequencing through the DNA sequencing facility at the Genomics Division, Human Genetics Institute, University of Iowa, Iowa City.

Results

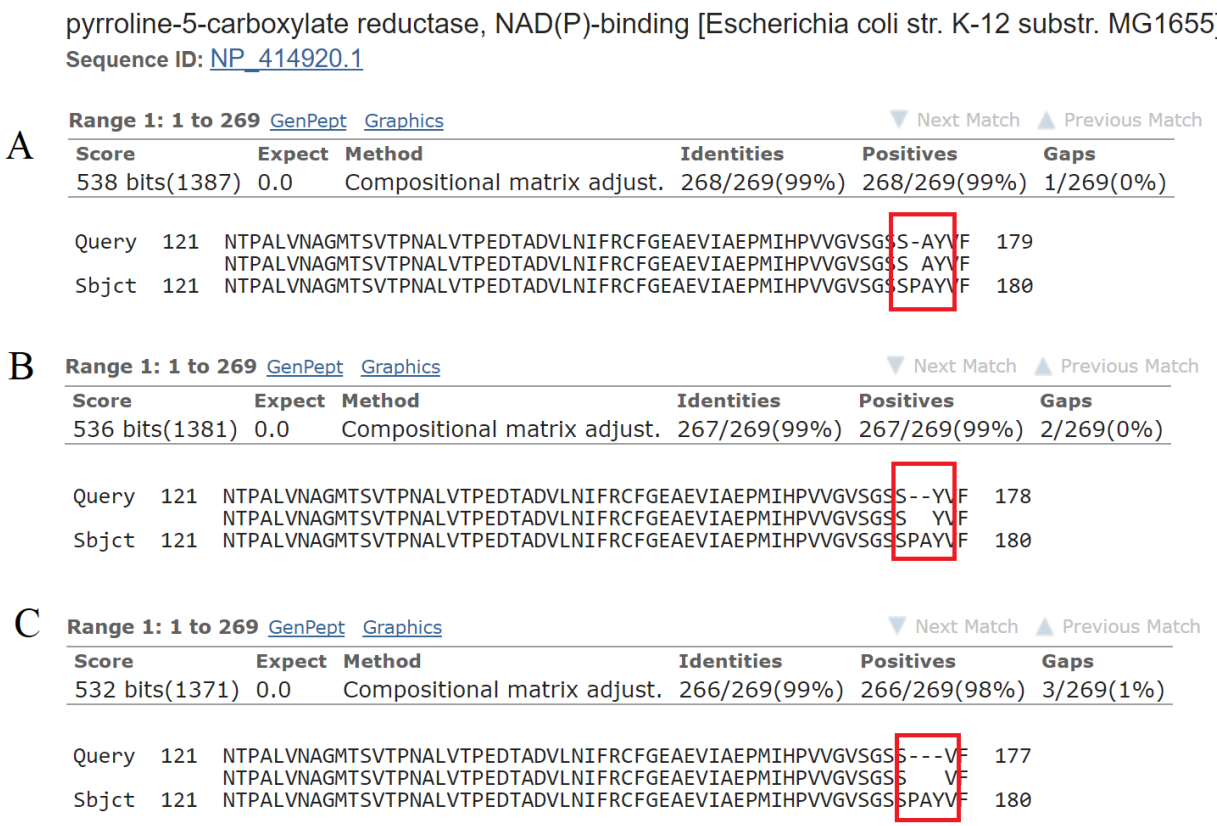


Figure 8. Partial BLASTp alignments between *E. coli* ProC deletion mutants ("query") and wild-type *E. coli* K12 MG1655 ProC ("sbjct"). Panel A is *E. coli* ProC ΔP176; Panel B is *E. coli* ProC ΔPA176-177 mutant, and Panel C is *E. coli* ProC ΔPAY176-178 mutant. The BLASTp tool from https://blast.ncbi.nlm.nih.gov/Blast.cgi

This is a partial protein BLAST between *E. coli* ProC deletion mutants and the wild type form of *E. coli*, *E. coli* K12 MG1655 ProC. A BLASTp against *E. coli* wild type was performed to confirm that the strains used for complementation had the intended strains. This confirmed that an amino acid duplication was isolated versus a premature stop codon. All degrees of mutation were aligned against the wild type; the single amino acid mutation, *E. coli* ProC ΔP176, is shown in panel A, the double amino acid duplication, *E. coli* ProC ΔPA176-177, is shown in panel B, and the triple amino acid duplication, *E. coli* ProC ΔPAY176-178, is shown in panel C.

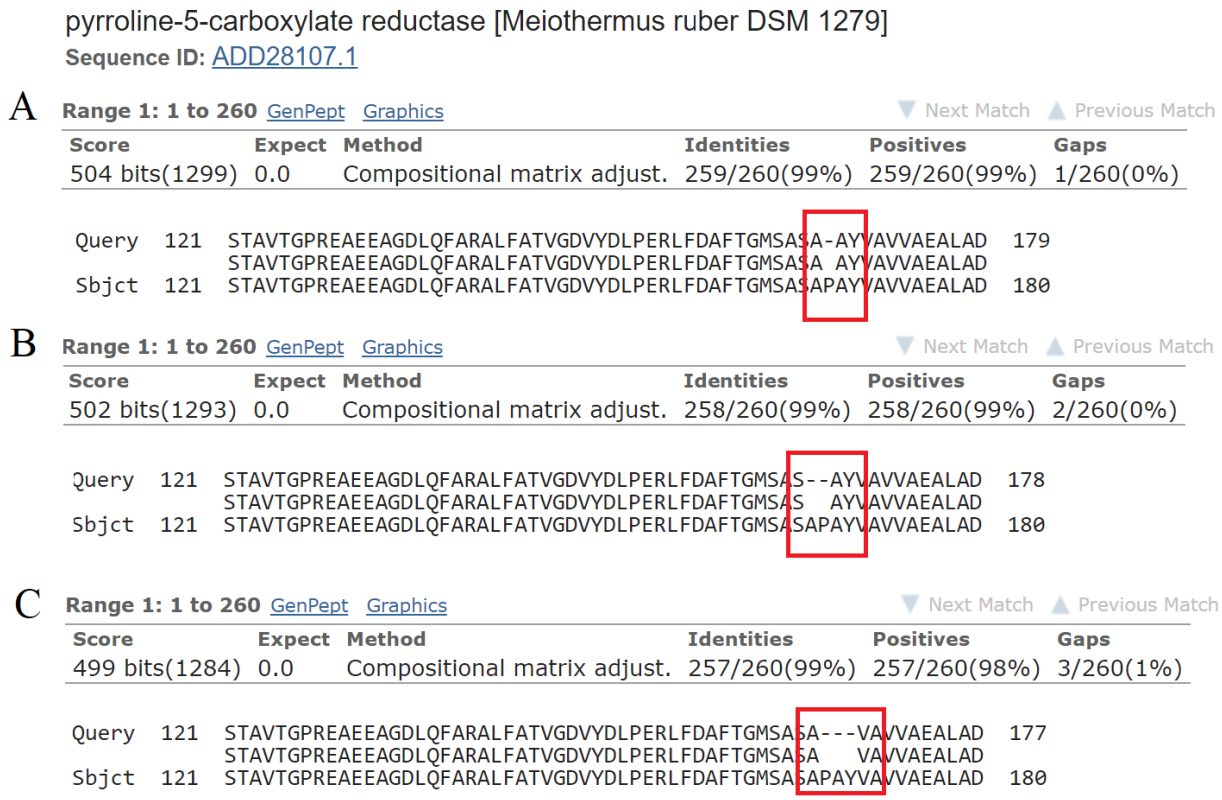


Figure 9. Partial BLASTp alignments between Mrub_1345 ProC deletion mutants ("query") and wild-type Mrub1345 ProC ("sbjct"). Panel A is Mrub_1345 ΔP168; Panel B is Mrub_1345 ΔPA168-169 mutant, and Panel C is Mrub_1345 ΔPAY168-170 mutant. The BLASTp tool from https://blast.ncbi.nlm.nih.gov/Blast.cgi

This is a partial protein BLAST between the wild-type Mrub1345 ProC and Mrub_1345 ProC deletion mutants that confirms the presence of the target strain. All amino acid deletion mutants of varying severity were aligned. The single amino acid deletion mutation (Mrub_1345 ΔP168) is depicted in panel A, the double amino acid deletion mutation (Mrub_1345 ΔPA168-169) is depicted in panel B, the triple amino acid deletion mutation (Mrub_1345 ΔPAY168-170) is depicted in panel C. The region impacted by the mutation is highlighted by the redbox.

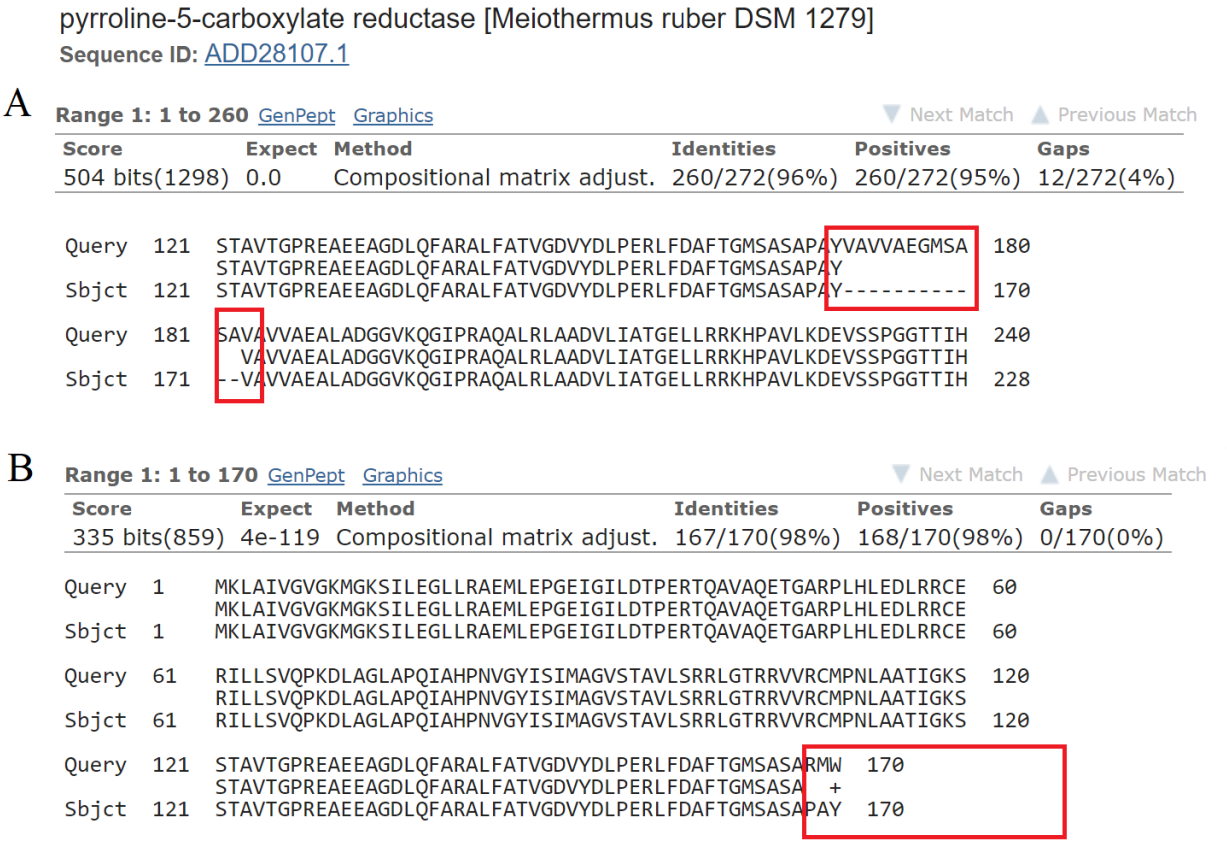


Figure 10. Partial BLASTp alignments between two putative Mrub_1345 mutants ("query") and the wild-type Mrub1345 ProC ("sbjct"). Panel A shows an Mrub_1345 ProC duplication after Y170; Panel B shows an Mrub_1345 frameshift/deletion mutant after A167. Red boxes highlight the region impacted by the mutation. BLASTp tool from https://blast.ncbi.nlm.nih.gov/Blast.cgi

Figure 10 shows a partial protein BLAST alignment between Mrub_1345 putative mutants and the wild-type version of *M. ruber*, Mrub1345 ProC. This confirmed the presence of the intended strain of duplication to be used for the complementation test. ProC is a protein that is 269 amino acids long and ends at 170. 170 is the end of the sequence. The alignment only shows the region impacted by the mutation, which is highlighted by the red box. Mrub_1345 ProC duplication after Y170 is depicted in panel A and Mrub_1345 frameshift/deletion mutant after A167 is in depicted in panel B.

pyrroline-5-carboxylate reductase, NAD(P)-binding [Escherichia coli str. K-12 substr. MG1655]
Sequence ID: NP_414920.1

Range 1: 1 to 217 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
439 bits(1129)	2e-159	Compositional matrix adjust.	217/217(100%)	217/217(100%)	0/217(0%)

```
Query  1    MEKKIGFIGCGNMGKAILGGLIASGQVLPGQIWVYTPSPDKVAALHDQFGINAAESAQEV  60
            MEKKIGFIGCGNMGKAILGGLIASGQVLPGQIWVYTPSPDKVAALHDQFGINAAESAQEV
Sbjct  1    MEKKIGFIGCGNMGKAILGGLIASGQVLPGQIWVYTPSPDKVAALHDQFGINAAESAQEV  60

Query  61   AQIADIIFAAVKPGIMIKVLSEITSSLNKDSLVVSIAAGVTLDQLARALGHDRKIIRAMP  120
            AQIADIIFAAVKPGIMIKVLSEITSSLNKDSLVVSIAAGVTLDQLARALGHDRKIIRAMP
Sbjct  61   AQIADIIFAAVKPGIMIKVLSEITSSLNKDSLVVSIAAGVTLDQLARALGHDRKIIRAMP  120

Query  121  NTPALVNAGMTSVTPNALVTPEDTADVLNIFRCFGEAEVIAEPMIHPVVGVSGSSPAYVF  180
            NTPALVNAGMTSVTPNALVTPEDTADVLNIFRCFGEAEVIAEPMIHPVVGVSGSSPAYVF
Sbjct  121  NTPALVNAGMTSVTPNALVTPEDTADVLNIFRCFGEAEVIAEPMIHPVVGVSGSSPAYVF  180

Query  181  MFIEAMADAAVLGGMPRAQAYKFAAQAVMGSAKMVLE  217
            MFIEAMADAAVLGGMPRAQAYKFAAQAVMGSAKMVLE
Sbjct  181  MFIEAMADAAVLGGMPRAQAYKFAAQAVMGSAKMVLE  217
```

Figure11. BLASTp alignment between an *E. coli* ProC deletion ("query") and wild-type *E. coli* ProC ("sbjct") showing a 100% alignment until E217; amino acids 218-269 are deleted from the query sequence. BLASTp tool from https://blast.ncbi.nlm.nih.gov/Blast.cgi

Figure 11 is a partial protein BLAST between *E. coli* ProC deletion mutants and the wild type version of *E. coli* proC to confirm the presence of the required strain. The result is complete alignment until E217, as seen in the figure. Because the BLASTp alignment only shows the impacted region of interest, the alignment ends at 217 as the amino acids in the region of 218-269 are deleted.

A

pyrroline-5-carboxylate reductase [Escherichia coli str. K-12 substr. MG1655]

Sequence ID: AUG15174.1

Range 1: 1 to 269 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
542 bits(1397)	0.0	Compositional matrix adjust.	268/269(99%)	268/269(99%)	0/269(0%)

```
Query  1    MEKKIGFIGCGNMGKAILGGLIASGQVLPGQIWVYTPSPDKVAALHDQFGINAAESAQEV  60
            MEKKIGFIGCGNMGKAILGGLIASGQVLPGQIWVYTPSPDKVAALHDQFGINAAESAQEV
Sbjct  1    MEKKIGFIGCGNMGKAILGGLIASGQVLPGQIWVYTPSPDKVAALHDQFGINAAESAQEV  60

Query  61   AQIADIIFAAVKPGIMIKVLSEITSSLNKDSLVVSIAAGVTLDQLARALGHDRKIIRAMP  120
            AQIADIIFAAVKPGIMIKVLSEITSSLNKDSLVVSIAAGVTLDQLARALGHDRKIIRAMP
Sbjct  61   AQIADIIFAAVKPGIMIKVLSEITSSLNKDSLVVSIAAGVTLDQLARALGHDRKIIRAMP  120

Query  121  NTPALVNAGMTSVTPNALVTPEDTADVLNIFRCFGEAEVIAEPMIHPVVGVSGSSPAYVF  180
            NTPALVNAGMTSVTPNALVTPEDTADVLNIFRCFGEAEVIAEPMIHPVVGVSGSSPAYVF
Sbjct  121  NTPALVNAGMTSVTPNALVTPEDTADVLNIFRCFGEAEVIAEPMIHPVVGVSGSSPAYVF  180

Query  181  MFIEAMADAAVLGGMPRAQAYKFAAQAVMGSAKMVLETGEHPGALKDMVCSPGGTTIEAV  240
            MFIEAMADAAVLGGMPRAQAYKFAAQAVMGSAKMVLETGEHPGALKDMVCSPGGTTIEAV
Sbjct  181  MFIEAMADAAVLGGMPRAQAYKFAAQAVMGSAKMVLETGEHPGALKDMVCSPGGTTIEAV  240

Query  241  RVLEEKGFAAAVIEAMTKCMEKSEKLSKS  269
            RVLEEKGF AAVIEAMTKCMEKSEKLSKS
Sbjct  241  RVLEEKGFRAAVIEAMTKCMEKSEKLSKS  269
```

B

pyrroline-5-carboxylate reductase [Meiothermus ruber DSM 1279]

Sequence ID: ADD28107.1

Range 1: 1 to 260 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
508 bits(1309)	0.0	Compositional matrix adjust.	259/260(99%)	259/260(99%)	0/260(0%)

```
Query  1    MKLAIVGVGKMGKSILEGLLRAEMLEPGEIGILDTPERTQAVAQETGARPLHLEDLRRCE  60
            MKLAIVGVGKMGKSILEGLLRAEMLEPGEIGILDTPERTQAVAQETGARPLHLEDLRRCE
Sbjct  1    MKLAIVGVGKMGKSILEGLLRAEMLEPGEIGILDTPERTQAVAQETGARPLHLEDLRRCE  60

Query  61   RILLSVQPKDLAGLAPQIAHPNVGYISIMAGVSTAVLSRRLGTRRVVRCMPNLAATIGKS  120
            RILLSVQPKDLAGLAPQIAHPNVGYISIMAGVSTAVLSRRLGTRRVVRCMPNLAATIGKS
Sbjct  61   RILLSVQPKDLAGLAPQIAHPNVGYISIMAGVSTAVLSRRLGTRRVVRCMPNLAATIGKS  120

Query  121  STAVTGPREAEEAGDLQFARALFATVGDVYDLPERLFDAFTGMSASAPAYVAVVAEALAD  180
            STAVTGPREAEEAGDLQFARALFATVGDVYDLPERLFDAFTGMSASAPAYVAVVAEALAD
Sbjct  121  STAVTGPREAEEAGDLQFARALFATVGDVYDLPERLFDAFTGMSASAPAYVAVVAEALAD  180

Query  181  GGVKQGIPRAQALRLAADVLIATGELLRRKHPAVLKDEVSSPGGTTIHGLAALEARGVRA  240
            GGVKQGIPRAQALRLAADVLIATGELLRRKHPAVLKDEVSSPGGTTIHGLAALEARGVRA
Sbjct  181  GGVKQGIPRAQALRLAADVLIATGELLRRKHPAVLKDEVSSPGGTTIHGLAALEARGVRA  240

Query  241  ALMEAVEAATLRRHELGKDE  260
            ALMEAVEAATLR HELGKDE
Sbjct  241  ALMEAVEAATLRGHELGKDE  260
```

Figure 12. Confirmation of *E. coli proC* (R249A) mutant (Panel A, query) and Mrub_1345 (G253R) mutant (Panel B, query) as determined by a BLASTp alignment to their respective wild-type protein sequence ("Sbjct"). Red boxes identify the mutated regions. BLASTp tool from https://blast.ncbi.nlm.nih.gov/Blast.cgi

Figure 12 is a BLASTp alignment of the wild-type protein sequence against its respective mutants, *E. coli proC* R249A in Panel A and Mrub_1345 mutant in Panel B. This confirms the mutant strains prior to performing a complementation test. The alignment only shows the region impacted by the mutation, which is depicted by the red box. In panel A, amino acid G in the "Sbjct" is substituted to R in the "query" and amino acid R in the "Sbjct" is substituted to A in the "query" in panel B.

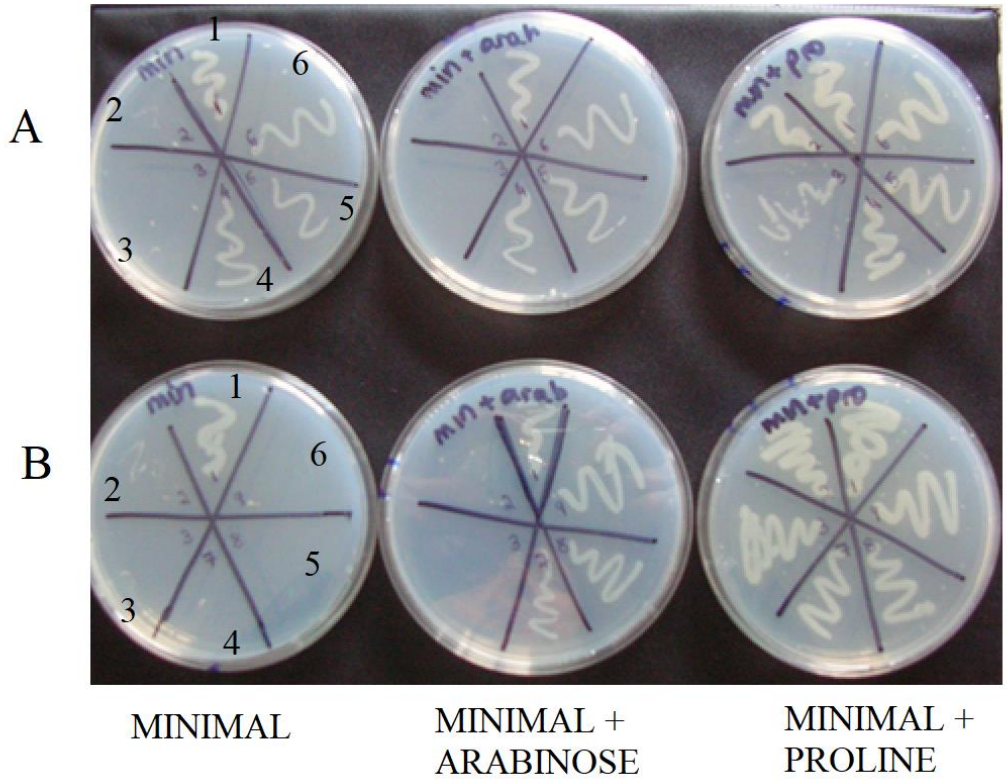


Figure 13. *E. coli proC*+ and *E. coli proC* (R249A) grew on minimal and min+arabinose, whereas the Mrub_1345 and Mrub_1345 (G253R) required arabinose activation to grow and the *E.coli* and Mrub_1345 deletions had to be supplemented with proline to grow. Unless otherwise noted, the *Keio E. coli proC* auxotroph is the host strain for all samples. Panel A. Sector 1) wild-type *E. coli* K12 strain; 2) Keio *E. coli proC* null strain; 3) pKt1; 4) *E. coli proC*+/pKt1; 5-6) *E. coli proC* (R249A)/pKt1. Panel B. Sector 1) wild-type *E. coli* K12 strain; 2) Keio *E. coli proC* null strain; 3) pKt1; 4) Mrub_1345/pKt1; 5-6) Mrub_1345 (G253R)/pKt1.

Sector 1 contains the wild-type *E. coli* K12 strain, which grew on all plates. Sector 2 contains the Keio *E. coli proC* null strain. This strain confirms no growth takes place without proline. It confirms the lack of growth because it is null for proline. Sector 3 contains pKt1, which does not contribute to proline production. Therefore, sectors 2 and 3 did not grow in on minimal and minimal+arabinose and only grew in minimal+proline. Sector 4 contains *E. coli proC*+/pKt1 in which the proC gene placed back in the null strain allows the production of proline. Panel A depicts *E. coli* mutants, *E. coli proC*+/pKt1 in sector 4 and *E. coli proC* (R249A)/pKt1 in sectors 5 and 6. Panel B depicts M. ruber mutants, Mrub_1345/pKt1 in sector 4 and Mrub_1345 (G253R)/pKt1 in sectors 5 and 6. E. coli proC+ and E. coli proC (R249A) grew on minimal and min+arabinose in the absence of proline. Because the mutants grew, this method is not the preferred method to discern mutation. Mrub_1345 and Mrub_1345 (G253R), however, required arabinose activation to grow. Both *E. coli* and *M. ruber* deletions only grew in the presence of proline.

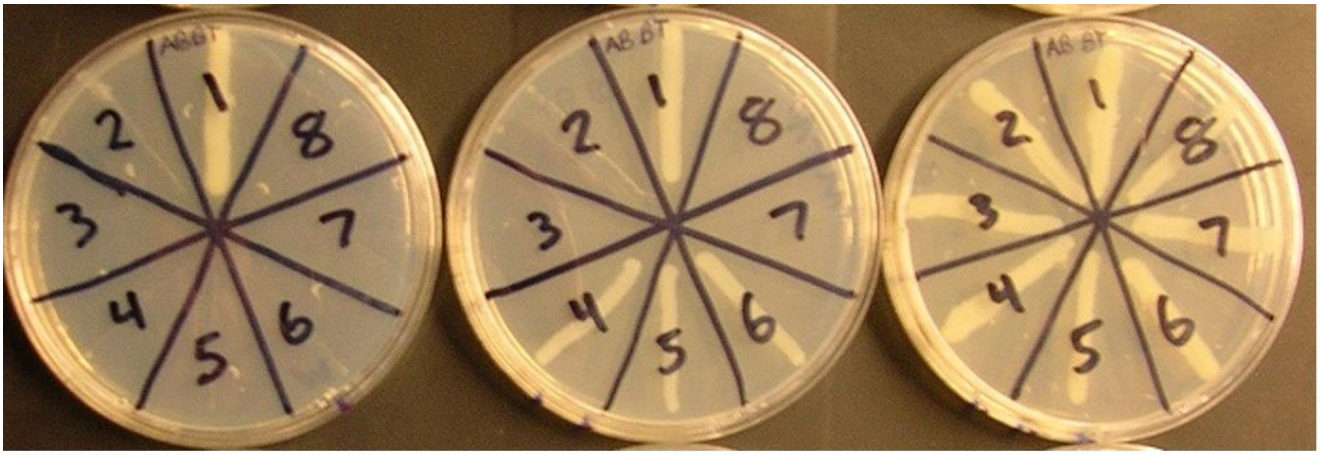
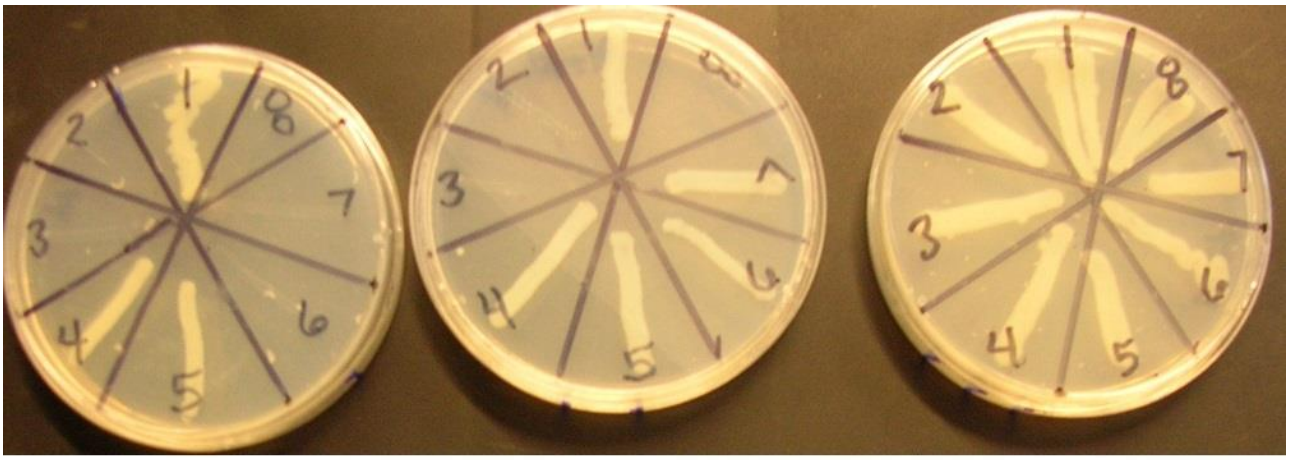


Figure 14. Complementation assay of Mrub_1345 wild-type and deletion mutants. Unless otherwise noted, the *Keio E. coli proC* auxotroph is the host strain for all samples. Sector 1) wild-type *E. coli* K12 strain; 2) Keio *E. coli proC* null strain; 3) pKt1; 4) Mrub_1345/; 5) Mrub_1345 (ΔP168)/pKt1; 6) Mrub_1345 (ΔPA168-169)/pKt1; 7) Mrub_1345 (ΔPAY168-170)/pKt1; 8) Mrub_1345 (premature stop)/pKt1 in Keio *E. coli proC* null strain.

Figure 14 shows a complementation assay of the wild type of *M. ruber*, Mrub_1345, and its respective deletion mutants of varying degrees. In addition to the wild-type *E. coli* K12 strain (sector 1), Mrub_1345/ (sector 4), Mrub_1345 (ΔP168)/pKt1 (sector 5), Mrub_1345 (ΔPA168-169)/pKt1 (sector 6) grew on minimal+arabinose. Keio *E. coli* proC null strain (sector 2), pKt1 (sector 3), Mrub_1345 (ΔPAY168-170)/pKt1 (sector 7), and Mrub_1345 (premature stop)/pKt1 in Keio *E. coli proC* null strain (sector 8) only grew on minimal+prokine. It was found that *M. ruber* deletion mutants show the same growth pattern as the wild type version in the case of one or two two base pair deletions. Mutants where three amino acids in a row were deleted showed no growth.



MINIMAL MINIMAL + ARABINOSE MINIMAL + PROLINE

Figure 15. Complementation assay of *E. coli proC*+ and deletion clones. Unless otherwise noted, the *Keio E. coli proC* auxotroph is the host strain for all samples. Sector 1) wild-type *E. coli* K12 strain; 2) Keio *E. coli proC* null strain; 3) pKt1; 4) *E. coli proC*+ /pKt1; 5) *E. coli proC* (ΔP176)/pKt1; 6) *E. coli proC* (ΔPA176-177)/pKt1; 7) *E. coli proC* (ΔPAY176-178)/pKt1; 8) *E. coli proC* (premature stop) pKt1.

Figure 15 is a complementation assay of the wild type of *E. coli proC*+ and clones of E. coli. These clones had varying degrees of mutation severity. The single amino acid deletion, *E. coli proC* (ΔP176)/pKt1 (sector 5) grew on minimal+arabinose along with the wild-type *E. coli* K12 strain (sector 1) and *E. coli proC*+ /pKt1 (sector 4). *E. coli proC* (ΔPA176-177)/pKt1 (sector 6) and *E. coli proC* (ΔPAY176-178)/pKt1 (sector 7) grew on minimal+arabinose. In E. coli, the last half of proC has the active site which binds the substrate. The *E. coli proC* (premature stop) pKt1 (sector 8) only grew on minimal+proline along with Keio *E. coli proC* null strain (sector 2) and pKt1 (sector 3). Therefore, it was found that *E. coli* can grow on minimal media whereas *M.ruber* needs arabinose. This is because it is growing in a new host and a new environment; it needs the added induction.

Table 3: Complementation assay of *E. coli proC and proC* mutants				
Host Cell	Plasmid	Minimal	Minimal + Arabinose	Minimal + Proline
E. coli K12 proC+	none	+	+	+
Keio *E. coli proC*	none	—	—	+
	E. coli proC+ in pKt1	+	+	+
	E. coli proC (Gly11Asp) in pKt1	+	+	+
	E. coli proC (Thr236Ala) in pKt1	+	+	+
	E. coli proC (Arg249Ala) in pKt1	+	+	+
	E. coli proC (ΔP176) in pKt1	+	+	+
	E. coli proC (ΔPA176-177) in pKt1	—	+	+
	E. coli proC (ΔPAY176-178) in pKt1	—	+	+
	E. coli proC premature stop in pKt1	—	—	+

Table 3 shows all the mutants generated and their growth pattern. It can be deduced from the data that loss of function from from base pair substitutions and small deletions cannot be discerned. *E. coli* was found to be more sensitive when grown on minimal media. The three substitution mutants and the single amino acid deletion grew under all growth conditions. More severe deletions, such as E. coli proC (ΔPAY176-178) in pKt1, grew on arabinose only. *E. coli* proC+ in pKt1, *E. coli proC* (Gly11Asp) in pKt1, *E. coli proC* (Thr236Ala) in pKt1, *E. coli proC* (Arg249Ala) in pKt1, and *E. coli proC* (ΔP176) in pKt1 grew under all growth conditions, even minimal. *E. coli proC* (ΔPA176-177) in pKt1 and E. coli proC (ΔPAY176-178) in pKt1 only grew with the supplement of arabinose. *E. coli proC* premature stop in pKt1 did not grow in minimal and minimal+arabinose and grew on minimal+proline only.

Table 4: Complementation assay of wild-type Mrub_1345 and Mrub_1345 mutants				
Host Cell	Plasmid	Minimal	Minimal + Arabinose	Minimal + Proline
E. coli K12 *proC+*	none	+	+	+
Keio *E. coli proC-*	none	—	—	+
	Mrub_1345 in pKt1	—	+	+
	Mrub_1345 (Gly11Asp) in pKt1	—	+	+
	Mrub_1345 (Thr226Ala) in pKt1	—	+	+
	Mrub_1345 (Gly253Arg) in pKt1	—	+	+
	Mrub_1345(ΔP168) in pKt1	—	+	+
	Mrub_1345 (ΔPA168-169) in pKt1	—	+	+
	Mrub_1345 (ΔPAY168-170) in pKt1	—	—	+
	Mrub_1345 duplication (region) in pKt1	—	+	+
	Mrub_1345 (premature stop) in pKt1	—	—	+

Table 4 shows all the mutants generated and their growth pattern. Similar to *E. coli*, loss of function and growth patterns cannot be discerned unless mutations are severe. *M. ruber* grew on arabinose, which is expected, as it needs this environment of increased activation due to the difference of temperature that *M. ruber* normally grows in. Mrub_1345 (Gly11Asp) in pKt1, Mrub_1345 (Thr226Ala) in pKt1, Mrub_1345 (Gly253Arg) in pKt1, Mrub_1345(ΔP168) in pKt1, Mrub_1345 (ΔPA168-169) in pKt1, and Mrub_1345 duplication (region) in pKt1 grew on minimal+arabinose and minimal+proline only. The three substitution mutations and the single amino acid deletion mutation showed a similar growth pattern. This is important to note because Mrub_1345 (Gly253Arg) in pKt1 depicts a relatively different case where a particular amino acid, glycine was altered. There is a smaller difference in this suctitution compared to alanine, which is the smallest amino acid that varies more significantly in its chemical nature from other amino acids. In the case of an increased severity of mutation, three amino acid deletion of Mrub_1345 (ΔPAY168-170) in pKt1, no growth was observed in minimal+arabinose. Similarly, Mrub_1345 (premature stop) in pKt1, only grew in minimal+proline.

Conclusion

The first goal of this project was to develop a set of mutants within *E. coli* and *M. ruber* ProC that reflected a hierarchy of severity to the functional state of the protein. We predicted more severe consequences to removing one or more conservative amino acids, as opposed to substituting chemically different amino acids for conserved amino acids. It was found that two or more amino acid deletions reduced or eliminated ProC function; amino acid substitutions did not impact ProC function under our experimental conditions. Uner our conditions, we could distinguish between a double and triple mutant. The *E. coli proC* double mutant ΔPA176-177/pKt1 grown on arabinose gave a very weak positive reaction, while the triple mutant ΔPAY176-178/pKt1 did not grow on minimal or minimal + arabinose. This observation suggests that our system might be able to distinguish the double and triple mutants in future investigations.

One observation we made in this experiment was the difference in the growth pattern of *M. ruber* ProC (nonmutated or mutated) as compared to the comparable nonmutant or mutant state in *E. coli*, respectively. For example, the *E. coli proC*-pKt1 in the null strain grew on minimal; however, the *M. ruber proC*-pKt1 in the null did not. If complementation was observed for the *M. ruber* strains, then it required transcriptional enhancement for growth, while *E. coli proC* did not. We hypothesize that under our experimental conditions (*e.g.*, 37^{o}C), *M. ruber* proteins function less due to their adaptation to higher temperatures (55-70^{o}C). Therefore, the prediction is that the lower temperature is likely inhibiting even the wild type version of *M. ruber* ProC. We predict that arabinose activation is overcoming this negative effect caused by the difference in temperature. Because *M.ruber* is adapted to growth under higher temperatures than *E. coli* and our experimental conditions, those *M. ruber* proteins are likely nonfunctional or not functioning as well as in normal conditions. Therefore, a transcriptional activation, more mRNA and therefore more protein product, is compensating for lower enzyme activity.

The second goal of this project was to evaluate the use of complementation test using the Keio *E. coli* null strains to assess the impact of mutants on ProC enzyme activity. While kinetic studies are a more accurate way for assessing enzyme function, it requires additional steps in enzyme purification and performing many enzyme assays, which can be prohibitively time consuming in a lab course or for independent student research projects. We had hoped to develop this complementation test as a less time-consuming strategy. While our approach hasn't allowed us to distinguish single amino acid substitutions or single amino acid deletions, we see potential in the study of two or more amino acid deletions. Nonetheless, this approach identified the level of severity needed in future research experiments, as well as possible future steps to be taken.

The next step to continue evaluating base pair substitutions would be to add a histidine tags to the ProC protein for easier purification, and then perform the enzyme kinetic assays. Adding histidine affinity tags has become the more functional route of purification in the study of biochemical proteins. The efficiency of histidine tags was investigated by Sinéad T. Loughran and Dermot Walls, where short affinity tags were used in purification of proteins (Loughran and Walls, 2011). In fact, this technique was employed on *E. coli* bacteria, which has lac operon sequences.

This is important because lac operon sequences are the most common form of expression used by different types of bacteria. Hence, this method shows potential for being successful with both *E.coli* and *M.ruber* bacteria in future implementations of histidine tag purifications. It was found that the six histidine aminoacids can be placed on either the N or C terminus. The difference in location is the added level of regulation through the C terminus, where purification is limited to full-length proteins (Walls and Loughran, 2011). Because histidines are uncommon amino acids in general protein structures, experimentation can be performed on these proteins at a pH of 8 as they maintain their uncharged condition. Histidines are also relatively smaller in size and, therefore, do not need to be isolated as they do not affect protein functioning (Walls and Loughran, 2011). The method of using his-tags was adapted from "Metal chelate affinity chromatography, a new approach to protein fractionation," the original developers of the process (Jerker Porath, Jan Carlsson, Ingmar Olsson, & Greta Belfrage, 1975).

References:

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410. PubMed.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori 2006. Construction of *Escherichia* coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:1-11

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E.. [Internet]. 2000. The Protein Data Bank. [2016 Dec 6]. Available from: http://www.rcsb.org/.

Betts, M. J., and Russell, R. B. 2003. Amino-Acid Properties and Consequences of Substitutions. Bioinformatics for Geneticists. 2003: 311–342.

Biolabs, N. E. Home - NEB | New England Biolabs. Home - NEB | New England Biolabs. https://www.neb.com/

Christian, J. H. B. 1955a. The influence of nutrition on the water relations of Salmonella oranienburg. Australian Journal of Biological Sciences 8:75–82.

Christian, J.H. B. 1955b. The water relations of growth and respiration of Salmonella oranienburg at 30 ◦C. Australian Journal of Biological Sciences 8, 490–497.

Chung CT, Niemela SL, Miller RH. 1989. One-step preparation of competent *Escherichia coli:* transformation and storage of bacterial cells in the same solution.
Proc Natl Acad Sci U S A. Apr;86(7):2172-5.

Csonka, L. and T. Leisinger. 2007. Biosynthesis of proline. EcoSalPlus 2013; 1-20

Crooks, GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator, Genome Research, 14:1188-1190, (2004).

Deutch, AH, Smith CJ, Rushlow KE, and Kretschmer PJ. 1982. *Escherichia coli* delta 1-pyrroline-5-carboxylate reductase: gene sequence, protein overproduction and purification. Nucleic Acids Research 10(23):7701-7714.

Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. Nucleic Acids Research. 2014;42(Database issue):D222-D230. doi:10.1093/nar/gkt1223.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A. 2016. The Pfam protein families database: towards a more sustainable future: Nucleic Acids Res., 44:D279-D285; [2016, Dec. 6]. Available from: http://pfam.xfam.org/

Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O. 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res 29(1):41-3.

Hayzer, D. J. & Moses, V.(1978). The enzymes of proline biosynthesis in *Escherichia coli*. Their molecular weights and the problem of enzyme aggregation. *The Biochemical Journal* **173**, 219–228.

Iowa Institute of Human Genetics.: University of Iowa Carver College of Medicine. https://www.medicine.uiowa.edu/humangenetics/genomics/

Juncker, A., H. Willenbrock, G. von Heijne, H. Nielsen, S. Brunak and A. Krogh. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 12(8):1652-62, 2003; [2016 Dec 6]. Available at: http://www.cbs.dtu.dk/services/LipoP/

Kall L, Krough A, Sonnhammer E. 2004. A combined transmembrane topology and signal Lukas Käll, Anders Krogh and Erik L. L. Sonnhammer. A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology*, 338(5):1027-1036, May 2004. (doi) (PubMed)

Käll,L., Anders Krogh and Erik L. L. Sonnhammer. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.*, 35:W429-32, July 2007 (doi) (PubMed)

Kanehisa M, Sato Y, Kawashima M, Furumichi M. and Tanabe M. (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res., 44, D457–D462; [2016 Dec 6]. Available from: http://www.genome.jp/kegg/

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Schroder, I., Shearer, A., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R.P., Paulsen, I., and Karp, P.D. 2013. EcoCyc: fusing model organism databases with systems biology Nucleic Acids Research 41:D605-612.

Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology, 305(3):567-580, January 2001. (PDF, 959503 bytes)

Krogh A, Rapacki K. TMHMM Server, v. 2.0. Cbs.dtu.dk. 2016 [accessed 2016 Dec 6]. http://www.cbs.dtu.dk/services/TMHMM/

Leisinger, T. 1996. Biosynthesis of proline. In *E. coli* and *Salmonella*: cellular and molecular biology (Neidhardt, F.C. ed), pp 434-441, ASM Press, Washington, DC.

Loughran S.T., Bree R.T., Walls D. (2017) Purification of Polyhistidine-Tagged Proteins. In: Walls D., Loughran S. (eds) Protein Chromatography. Methods in Molecular Biology, vol 1485. Humana Press, New York, NY

Madden T. The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from: **http://www.ncbi.nlm.nih.gov/books/NBK21097/** tool: BLASTp tool from https://blast.ncbi.nlm.nih.gov/Blast.cgi

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015 Jan 28;43(Database issue):D222-2. doi: 10.1093/nar/gku1221. Epub 2014 Nov 20. [PubMed PMID: 25414356]

Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, et al. 2012. IMG: The integrated microbial genomes database and comparative analysis system. Nucleic Acids Research 40(D1):D115-22. Available from: http://nar.oxfordjournals.org/content/40/D1/D115.full

Morrison KL and Weiss GA. 2001. Combinatorial alanine-scanning. Curr Opin Chem Biol 5(3):302-7.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology 302 (1):205-17 Available from: http://www.ebi.ac.uk/Tools/msa/tcoffee/

Orser, C., Goodner, B., Johnston, M., Gelvin, S. & Csonka, L. (1988). The Escherichia coli proB gene corrects the proline auxotrophy of Saccharomyces cerevisiae pro1 mutants. Molecular and General Genetics MGG212, 124–128.

Peracchi, A. Enzyme Catalysis: Removing Chemically 'Essential' Residues by Site-Directed Mutagenesis. Trends Biochem. Sci. 2001 26 (8), 497-503.

Pérez-Arellano I, Rubio V, Cervera J. 2006. Mapping active site residues in glutamate-5-kinase. The substrate glutamate and the feed-back inhibitor proline bind at overlapping sites. FEBS Lett. **2006** Nov 13;580(26):6247-53. Epub **2006** Oct 20. PMID:17069808

Petersen, Thomas, Søren Brunak, Gunnar von Heijne & Henrik Nielsen Discriminating signal peptides from transmembrane regions. Nature Methods, 8:785-786, 2011 Available from: http://www.cbs.dtu.dk/services/SignalP

Plapp, BV. 1995. Site-directed mutagenesis: a tool for studying enzyme catalysis. Methods Enzymol. 249:91-119. PMID: 7791629

Porath, Jerker, Carlsson, Jan, Olsson, Ingmar, Belfrage, Greta. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 1975; 258: 598–599.

Smith RJ, Downing SJ, Phang JM, Lodato RF, Aoki TT. Pyrroline-5-carboxylate synthase activity in mammalian cells. Proc. Natl Acad. Sci. USA. 1980;77:5221–5225.

Sonnhammer, ELL., G. von Heijne, and A. Krogh.A hidden Markov model for predicting transmembrane helices in protein sequences. In J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen, editors, Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, Menlo Park, CA, 1998. AAAI Press.

Tindall, B. J., Sikorski, J., Lucas, S., Goltsman, E., Copeland, A., Glavina Del Rio, T., … Lapidus, A. (2010). Complete genome sequence of Meiothermus ruber type strain (21T). Standards in Genomic Sciences, 3(1), 26–36. http://doi.org/10.4056/sigs.1032748

Vogel HJ, Davis BD (1952). "Glutamic γ-Semialdehyde and Δ1-Pyrroline-5-carboxylic Acid, Intermediates in the Biosynthesis of Proline." J Am Chem Soc 74:109-112.

Wills, B. and L. Scott. 2015. Cite the digital commons paper and his GENI-ACT website.

Yu, NY, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, L.J. Foster, F.S.L. Brinkman. 2010. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics 26(13):1608-1615