**Identification of Key Amino Acid Residues in L-Proline Catabolic Enzymes from Gram-Positive Bacteria**

**ABSTRACT**

Catabolism of the amino acid L-proline in most Gram-positive bacteria occurs through a common pathway that involves the FAD-dependent activity L-proline dehydrogenase (PRODH) and the NAD+-dependent activity L-Δ1-pyrroline-5-carboxylate dehydrogenase (P5CDH). The two activities are found in monofunctional proteins in the low-G/C *Firmicutes* and in monofunctional or bifunctional proteins in the high-G/C *Actinobacteria*. To assess the similarities of these proteins, the amino acid sequences from four representative species in the genera *Bacillus, Halobacillus, Staphylococcus, Streptomyces, Arthrobacter/Paenarthrobacter, Corynebacterium*, and *Mycobacterium* were retrieved from the UniProtKB database and aligned. The resulting phylogenetic trees showed a good clustering of the four sequences for each of the seven genera. However, the 20 monofunctional proteins showed an overall sequence identity of only about 15% while the eight bifunctional proteins showed an identity of about 60%. To identify the key amino acid residues in these proteins, the sequences from the Gram-positive bacteria were aligned with the monofunctional proteins from *Thermus thermophilus* and *Deinococcus radiodurans* and the bifunctional proteins from *Bradyrhizobium japonicum* and *Escherichia col*i. These proteins from Gram-negative bacteria have been crystallized and used as models of catalytic activity. While there was very little overall sequence identity, the same key residues were found at both the PRODH and P5CDH active sites with only occasional conservative substitutions. Although the evidence for channeling the L-Δ1-pyrroline-5-carboxylate (P5C) product of the first reaction (PRODH) through a structural tunnel to the active site of the second reaction (P5CDH) is limited among the Gram-positive bacteria, sequence analysis indicated that key residues affecting this process are also present in the Gram-positive bacteria. This suggests that the basic mechanism for proline catabolism has been retained during evolution and is the same in almost all microorganisms.

***Key Words:*** *Actinobacteria*, *Firmicutes*, Gram-positive bacteria, L-glutamate-γ-semialdehyde dehydrogenase,L-proline catabolism, L-proline dehydrogenase, L-Δ1-pyrroline-5-carboxylate dehydrogenase

**INTRODUCTION**

Many microorganisms can degrade amino acids as their sole sources of nitrogen and carbon (Bender, 2012; Gottschalk, 2012; Halvorson, 1972). L-proline is usually metabolized by a common catabolic pathway that involves an initial FAD-dependent oxidation by L-proline dehydrogenase (PRODH, EC 1.5.5.2, formerly called proline oxidase and designated EC 1.5.99.8) to form L-Δ1-pyrroline-5-carboxylate (Adams & Frank, 1980; Liu et al., 2017; Tanner, 2019). Electrons from FADH2 are then transferred to a quinone in a membrane-associated electron transport chain and eventually on to oxygen under aerobic conditions. L-Δ1-pyrroline-5-carboxylate (P5C) is in spontaneous equilibrium with L-glutamate-γ-semialdehyde (GSA), which then undergoes a NAD+-dependent oxidation by L-Δ1-pyrroline-5-carboxylate dehydrogenase (P5CDH, EC 1.2.1.88, formerly 1.5.1.12) to form L-glutamate. This enzyme is also called L-glutamate-γ-semialdehyde dehydrogenase (GSALDH). The resulting L-glutamate can then be converted to 2-oxoglutarate (α-ketoglutarate) by several different enzymes, including glutamate dehydrogenase, glutamine synthetase, and glutamate-pyruvate transaminase (Walker & van der Donk, 2016). This allows the distribution of the nitrogen in L-proline to other metabolites. The 2-oxoglutarate and its carbon atoms can be further metabolized through the citric acid cycle or other catabolic or anabolic pathways.

Bacteria are often divided into two major groups - the Gram-negative bacteria and the Gram-positive bacteria – based on the organization and staining of their cell walls. The Gram-positive microorganisms can be further separated into two major clades based on the G+C content of their DNA (Hogenhout & Loria, 2008). The *Firmicutes* or low G+C clade of bacteria includes microbes in the genera *Bacillus*, *Halobacillus*, *Lactobacillus, Clostridium*, *Staphylococcus*, and *Streptococcus*, as well as the cell-wall-deficient *Mycoplasmas* (Ludwig et al., 2009; Oren & Garrity, 2021; Zhang & Lu, 2015). The *Actinobacteria* or high G+C clade includes bacteria in the genera *Actinomycetes*, *Micrococcus*, *Arthrobacter/Paenarthrobacter*, *Corynebacterium*, *Streptomyces*, and *Mycobacterium* (Barka et al., 2016; Chater, 2016; Gao & Gupta, 2012). Both of these clades are well separated from the Gram-negative *Proteobacteria* and the *Archaebacteria* (Archaea).

The processes of L-proline catabolism in the Gram-positive bacteria were recently reviewed and compared to those in other bacteria (Christgen & Becker, 2019; Deutch, 2025; Tanner, 2008). For bacteria in the low G+C clade, such as *Bacillus* and *Staphylococcus* species, the two enzyme activities required for proline degradation always seem to occur as single monofunctional proteins with either PRODH or P5CDH activity. For bacteria in the high G+C clade, the two activities may occur either as monofunctional proteins in the case of *Streptomyces* and *Mycobacterium* species or as bifunctional proteins with both PRODH and P5CDH activities in the case of *Arthrobacter/Paenarthrobacter* and *Corynebacterium* species. The UniProKB database contains a large number of amino acid sequences coding for PRODH, also called PutA, PruB, and FadM, and for P5CDH (GSALDH), also called RocA and PruA. To assess the similarities of these proteins, the amino acid sequences from four commonly-studied or representative species in each of the well-characterized genera of Gram-positive bacteria were retrieved from the database, aligned, and analyzed. The results indicate that despite low overall sequence identity, the same key residues needed for enzyme activity in Gram-negative bacteria also occur in the Gram-positive microorganisms.

1. **MATERIALS AND METHODS**

The amino acid sequences of proteins involved in L-proline catabolism in Gram-positive bacteria were retrieved from the Uniport database at the Swiss Institute of Bioinformatics ([www.expasy.org](http://www.expasy.org)). Full descriptions of these proteins were also recovered from this site. The amino acid residues in these sequences are described and discussed here using the standard one-letter abbreviations. Sequence alignments were done with the Cluster Omega tool included in this database, printed out, and analyzed manually. Monofunctional and bifunctional PRODH and P5CDH sequences from the well-characterized Gram-negative bacteria *Thermus thermophilus, Deinococcus radiodurans*, *Bradyrhizobium japonicum,* and *Escherichia col*i were retrieved from the same database and used for comparisons.

1. **RESULTS AND DISCUSSION**

**3.1 SEQUENCE ANALYSIS OF L-PROLINE CATABOLIC ENZYMES**

The UniProKB database contains a large number of amino acid sequences encoding the proteins needed for L-proline catabolism. Table 1 lists the sequences from the four representative species in each of the genera described in detail by Deutch (2025), which were used in this study. Most refer to monofunctional proteins but the same sequence is shown for the *Arthrobacter/Paenarthrobacter* and *Corynebacterium* species because these bacteria make bifunctional proteins with both PRODH and P5CDH activities.

The phylogenetic trees created by the UniProtKB Clustal alignment program based on these sequences are shown in Fig. 1. Panel A shows the tree for the L-proline dehydrogenase (PRODH) sequences and Panel B shows the tree for the L-Δ1-pyrroline-5-carboxylate dehydrogenase (P5CDH) sequences. In these trees, there was a good clustering of the four sequences recovered for each of the seven genera used in the analysis. The trees would have been somewhat different if other species or strains in each genus had been used. The PRODH tree indicated a close relationship between the *Mycobacterium* and *Streptomyces* species, with decreasing relationships to the *Arthrobacter/Paenarthrobacter*, *Corynebacterium*, *Bacillus*, and *Staphylococcus* species. The P5CDH tree indicated a close relationship between the *Bacillus* and *Staphylococcus* species, with decreasing relationships to the *Arthrobacter/Paenarthrobacter*, *Corynebacterium*, *Mycobacterium*, and *Streptomyces* species. These trees should be compared to published phylogenetic results based on the16S rRNA sequences of these bacteria and other protein sequences (Gao & Gupta, 2012; Stackebrandt; 2009; Zhang & Lu 2015). Those phylogenies also show a close relationship between *Streptomyces* and *Mycobacterium* species as examples of high G+C Gram-positive bacteria (*Actinobacteria*) and between *Bacillus* and *Staphylococcus* species as examples of low G+C Gram-positive bacteria (*Firmicutes*).

**Table 1**. Sequences of PRODH and P5CDH used for alignment analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| clade | genus | species | PRODH sequence | P5CDH sequence |
| low G+C Gram+ | *Bacillus* | *subtilis* | P94390 | P94391 |
|  |  | *cereus* | A0A2B9D3D0 | Q81IP0 |
|  |  | *amyloliquefaciens* | A0AAP4DJE1 | I2CBH5 |
|  |  | *licheniformis* | A0A1Y0YAS3 | Q65NN2 |
|  | *Halobacillus* | *halophilus* | D9ZZC4 | D9ZZC6 |
|  |  | *karajensis* | A0A024P306 | A0A059NX79 |
|  |  | *trueperi* | A0A3D8VNF1 | A0A3E0JAW6 |
|  |  | *andaensis* | A0A917BCV3 | A0A917B635 |
|  | *Staphylococcus* | *saprophyticus* | Q49YJ5 | Q4A0E7 |
|  |  | *aureus* | A0A0H3KDE2 | A6QK44 |
|  |  | *epidermidis* | Q5HNE5 | Q8CN04 |
|  |  | *hemolyticus* | A0A2K0A772 | Q4L966 |
| high G+C Gram+ | *Streptomyces* | *coelicolor* | Q9Z560 | Q8CJR1 |
|  |  | *flaveus* | A0A917QH49 | A0A917QGY9 |
|  |  | *venezuelae* | A0A5P2CT34 | A0A5P2APZ4 |
|  |  | *niveus* | A0A1U9QQP5 | A0A1U9QQ38 |
|  | *Arthrobacter* | *aurescens* | A1R2L2 | A1R2L2 |
|  |  | *globiformis* | H0QP87 | H0QP87 |
|  |  | *oryzae* | A0A3N0C9F5 | A0A3N0C9F5 |
|  |  | *mobilis* | A0A7X6HD01 | A0A7X6HD01 |
|  | *Corynebacterium* | *glutamicum* | Q8NU48 | Q8NU48 |
|  |  | *freiburgense* | A0A1X8XLF1 | A0A1X8XLF1 |
|  |  | *variabile* | G0HBR5 | G0HBR5 |
|  |  | *jeikeium* | Q4JX46 | Q4JX46 |
|  | *Mycobacterium* | *tuberculosis* | A0A654TP71 | A5U1N5 |
|  |  | *kansasii* | X7ZLA3 | X7ZNA4 |
|  |  | *smegmatis* | A0R2H7 | A0R2H8 |
|  |  | *phlei* | A0A5N5UR16 | A0A5N5UR07 |

1. PRODH sequencesA close-up of a document

   AI-generated content may be incorrect.
2. P5CDH sequences

A close-up of a document

AI-generated content may be incorrect.

**Fig. 1**. Phylogenetic trees of L-proline dehydrogenase (PRODH, panel A) and L-Δ1-pyrroline-5-carboxylate dehydrogenase (P5CDH, panel B) based on the amino acid sequences of the proteins listed in Table 1.

* 1. **PROTEINS WITH L-PROLINE DEHYDROGENASE ACTIVITY**

The PRODH sequences for the four proteins within each genus showed a high degree of similarity. The *Bacillus* sequences were 303 or 305 amino acids long and showed 158/304 = 52.0% identity. The *Halo bacillus* sequences were 305 or 306 amino acids long and exhibited 223/305 = 73.1% identity. The *Staphylococcus* sequences were all 333 amino acids long and had 180/333 = 54.0% identity. The *Streptomyces* sequences were all 308 amino acids long and showed 209/308 = 67.8% identity. The bifunctional *Arthrobacter/Paenarthrobacter* sequences were 1185, 1157, 1174, and 1157 amino acids long and exhibited 742/1168 = 63.5% identity. The bifunctional *Corynebacterium* sequences were 1152, 1076, 1205, and 1158 amino acids long and had 401/1148 = 34.9% identify. The *Mycobacterium* sequences 282, 319, 317, and 317 amino acids long and showed 51.8% identity. The *M. tuberculosis* sequence was noticeably shorter than the others in this genus. These percentages of sequence identity might have been somewhat different if other species or examples of strains in each genus had been selected. However, it is interesting that the percentage of sequence identity was lowest for the long bifunctional *Corynebacterium* species and highest for the shorter monofunctional *Halobacillus* species.

The overall percentage of sequence identity for all 28 proteins with PRODH activity was much less, in part because the *Arthrobacter/Paenarthrobacter* and *Corynebacterium* sequences were much longer and contained both the PRODH and P5CDH activities. When the sequences of just the 20 monofunctional PRODH proteins were aligned, there was an overall sequence identity of 43/312 = 13.7%. There was very little sequence similarity in the amino-terminal portion of the proteins but more at the carboxyl-terminal end. When the sequences of just the eight bifunctional PRODH proteins were aligned, the overall sequence identity was 733/1158 = 63.3% identity. As shown in Fig. 1, the *Arthrobacter/Paenarthrobacter* and *Corynebacterium* species do share a common ancestor.

Two key features of the PRODH activity are its specificity for L-proline and its use of FAD as a cofactor. Studies from the laboratories of Tanner and Becker using crystallized proteins have provided a detailed analysis of the active site in both monofunctional and bi-or tri-functional proteins in Gram-negative bacteria (Liu et al., 2017; Tanner, 2019). The monofunctional proteins that have been used as models have included those from *Thermus thermophilus* (White et al., 2007) and *Deinococcus radiodurans* (Luo et al., 2012). Fig. 2 shows the primary contacts in the protein from *D. radiodurans*. The key amino acid residues for interaction with the substrate (shown here as tetrahydrofuroic acid or THFA) are R291, R292, and K98, which pair with the carboxylate group, and L257, Y278, and Y288, which surround the proline ring system. Other key residues located nearby are E64 and E295. The key residues in contact with the FAD cofactor are K98, D135, R187, and Y278.

Several images of a molecule

AI-generated content may be incorrect.

**Fig. 2**  Diagram of the active site of the monofunctional PRODH from *Deinococcus radiodurans* (from Luo et al., 2012)

To determine if the PRODH proteins from the Gram-positive bacteria have similar residues, the amino acid sequences of the *D. radiodurans* (Q9RW55) and *T. thermophilus* (Q72IB8) proteins were retrieved from the UniProtKB database and aligned with the 20 monofunctional Gram-positive PRODH sequences listed in Table 1. There was about 38/312 = 12.2% overall sequence identity for the 22 sequences, most of which occurred in the carboxyl-terminal portion of the proteins. The sequences from *D. radiodurans* and *T.* *thermophilus* were closely linked to one another in the phylogenetic tree created by the Cluster Omega program. These sequences were more closely related to those from the *Streptomyces* and *Mycobacterium* species than to those from the *Bacillus* and *Staphylococcus* species.

The following common sequences were observed in all 22 proteins. They are shown with the *D. radiodurans* amino acid sequence numbers, and the key residues shown in Fig. 2 are highlighted in bold. X indicates a conservative substitution or a poor match:

X59 D60 X61 X62 G63 **E64**

X95 S96 X97 **K98** X99

**D135** X136 E137 X138

**R187** X188 X189 K190 G191 A192 **Y193** X194 **E195**

X229 H230 D231

E253 X254 Q255 X256 **L257** X258 X259 X260 R261

X276 X277 **Y278**

**W285** X286 X287 **Y288** F289 X290 **R291 R292** X293 A294 **E295**

There are several places where a residue in the *D. radiodurans* and *T. thermophilus* sequences has been replaced by a conservative substitution in one or two groups. For example, L62 is replaced by valine (V) in the *Streptomyces* species and by isoleucine (I) in the *Bacillus* and *Halo bacillus* species. F254 is replaced by tyrosine (Y) in two of the *Mycobacterium* strains, and Y286 is replaced by phenylalanine (F) in the *Staphylococcus* species.

In a similar way, key residues at the active site of the PRODH domain have been identified for the bifunctional Put protein from *Bradyrhizobium japonicum* (Krishnan & Becker, 2005; Srivastava et al., 2010a) and the trifunctional PutA protein from *Escherichia col*i (Bogner & Tanner, 2022; Moxley et al., 2017, Srivastava et al., 2010b; Zhang et al., 2004). Fig. 3 shows the active site of the *E. coli* PutA PRODH domain complexed with L-lactate.

A collage of several images of a molecule

AI-generated content may be incorrect.

**Fig. 3** Diagram of the active site of the PRODH domain of the trifunctional PutA protein from *Escherichia coli* (from Bogner & Tanner, 2022).

The key residues are K329, D370, Y437, L513, Y540, Y552, R555, and R556. To determine if similar residues occur in the PRODH domain of the bifunctional proteins from *Arthrobacter/Paenarthrobacter* and *Corynebacterium* species, the amino acid sequences for the *B. japonicum* (Q59206) and *E. coli* (P09546) were retrieved from the UniProtKB database and aligned with the eight sequences listed in Table 1. The *B. japonicum* sequence with 1016 amino acids is shorter than those from the *Arthrobacter/Paenarthrobacter* and *Corynebacterium* species, while the *E. coli* sequence with 1320 amino acids is longer due to the DNA-binding domain at the amino-terminal end. The average length is 1159 amino acids. There was an overall sequence identity of 9.9% with short regions of identity in both the amino-terminal and carboxyl-terminal domains. The sequences from the Gram-negative bacteria *B. japonicum* and *E. coli* were closely linked to one another in a phylogenetic tree, and only distantly related to the Gram-positive *Arthrobacter/Paenarthrobacter* and *Corynebacterium* sequences.

The following common sequences were observed in all ten sequences, where they are given with the *E. coli* amino acid sequence numbers and the key residues shown in Fig. 3 are highlighted in bold. X indicates a conservative substitution or a poor match.

X286 L287 G288 E289

S327 X328 **K329** X330 S331

X369 **D370** X371 E372 E373

G400 X401 X402 X403 Q404 A405 Y406

X429 R430 X431 V432 K433 G434 A435

E509 F510 X511 X512 **L513** X514 G515 M516

**Y540** X541 P542 V543

A551 **Y552** L553 X554 **R555 R556** L557 X558 E559.

Again, almost all of the key residues identified before in the Gram-negative bacteria were found in the Gram-positive sequences, but there are a few variations. The key Y437 residue in the *E. col*i and *B. japonicum* sequences has been replaced by an asparagine (N) in the four *Arthrobacter/Paenarthrobacter* and four *Corynebacterium* sequences. Other conservative substitutions include the replacement of methionine (M) with leucine (L) at position 286, the replacement of alanine (A) with threonine (A) at position 541, and the replacement of valine (V) with isoleucine (I) at position 554. Alignment of the eight bifunctional Gram-positive sequences with the monofunctional *Thermus* and *Deinococcus* sequences confirmed that the same key residues are present in both types of proline dehydrogenase. This suggests that the basic mechanism for proline binding and oxidation and for FAD binding and reduction is the same in all of these microorganisms.

**3.3. PROTEINS WITH L-Δ1-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE ACTIVITY**

The P5CDH sequences for the Gram-positive bacteria listed in Table 1 also showed a high degree of amino acid similarity for the four proteins within each genus. The *Bacillus* sequences were 515 or 516 amino acids long and showed 304/515 = 59.0% identity. The *Halobacillus* sequences were 515 or 516 amino acids long and exhibited 323/516 = 62.6% identity. The *Staphylococcus* sequences were all 514 amino acids long and had 380/514 = 73.9% identity. The *Streptomyces* sequences were 543 or 546 amino acids long and showed 447/543 = 82.3% identity. The bifunctional *Arthrobacter/Paenarthrobacter* sequences again were 1185, 1157, 1174, and 1157 amino acids long and again displayed 742/1168 = 63.5% identity. The bifunctional *Corynebacterium* sequences again were 1152, 1076, 1205, and 1158 amino acids longand showed 401/1148 = 34.9% identify. The *Mycobacterium* sequences were 540, 534, 540, and 537 amino acids long and had 392/538 = 72.9% identity. In general, the P5CDH sequences showed a higher percentage of amino acid identity than those for PRODH except for the *Arthrobacter/Paenarthrobacter* and *Corynebacterium* sequences.

When the sequences of just the 20 monofunctional P5CDH proteins were aligned, the average length was 526 amino acids and there was overall sequence identity of 84/526 = 16.0%. When the sequences of just the eight bifunctional P5CDH proteins were aligned, the overall sequence identity again was 733/1158 = 63.3% identity because these sequences include the PRODH domain and the *Arthrobacter* and *Corynebacterium* species share a common ancestor.

A key feature of P5CDH is its use of NAD+ as a coenzyme for the oxidation of L-Δ1-pyrroline-5-carboxylate (L-glutamate-γ-semialdehyde) to form L-glutamate. The crystal structure of the monofunctional enzyme from *Thermus thermophilus* has been determined and a mechanism for its activity has been proposed (Inagaki et al., 2006). Fig. 4 shows the binding of NAD+ and P5C (GSA) to key residues at the active site.

A group of chemical structures

AI-generated content may be incorrect.

**Fig. 4.** Diagram of the active site of P5CDH from *T. thermophilus* (from Inagaki et al., 2006)

The key residues involved in binding the NAD+ coenzyme are E288, T289, and G290. The key residues involved in binding the substrate are N184, C322, S323, G477, and A478. Similar residues have been identified at the active site of PruA protein of *Mycobacterium tuberculosis* although the amino acid numbers are different (Lagautriere et al., 2014).

To determine if the sequences from the Gram-positive bacteria have similar residues, the amino acid sequence of the protein from *T. thermophilus* (Q72IB9) was aligned with the 20 monofunctional P5CDH sequences listed in Table 1. The average length was 528 amino acids, for which there was 79/528 = 15.0% sequence identity. The amino acid sequence of *T*. *thermophilus* sequence was more closely linked in the phylogenetic tree to the sequences from the *Bacillus* and *Staphylococcus* species than to those from the from the *Streptomyces* and *Mycobacterium* species.

The following common sequences were observed for the 21 proteins, where they are given with the *T. thermophilus* amino acid sequence numbers and the key residues in Fig. 4 are highlighted in bold. X indicates a conservative substitution or a poor match.

X39 X40 X41 I42 X43 G44

I180 X181 P182 X183 **N184**F185

G201 X202 X203 X204 X205 X 206 K207 P208

X256 X257 F258 T259 G260 S261

**E288** X289 **G290 G291** K292 D293

G319 Q320 K321 **C322** **S323** A324 X325 S326 R327

Y442 X443 L444 T445 G446

T476 X477 **A478** X479 V480 G481

P484 F485 G486 G487

There are some small differences between the reported *T. thermophilus* sequence and the others. The T289 residue in *T. Thermophilus* is replaced by methionine (M) in the sequences from *Bacillus, Halobacillus,* and *Staphylococcus*. The G477 residue in *T. thermophilus* is replaced by serine (S) or alanine (A) in the *Staphylococcus* species.

Crystallographic studies from the laboratories of Becker and Tanner have revealed the main interactions in the P5CDH domain of the bi- or trifunctional PutA proteins from *E. coli*, *B. japonicum*, and *Sinorhizobium meliloti* (Korasick et al., 2018; Campbell et al., 2021; Meeks et al., 2024). They used various analogues of L-proline which are inhibitors of the enzyme activity and can bind to the active site or structural fragments of the protein. Because there is no good diagram of the P5CDH domain which shows the amino acids that interact with both NAD+ and the substrate, the sequences of the four *Arthrobacter/Paenarthrobacter* species, the four *Corynebacterium* species, the bifunctional PutA protein from *B. japonicum* (Q59206), and the trifunctional PutA protein from *E. coli* (P09546) were aligned with the amino acid sequence of the P5CDH protein from *T. thermophilus* (Q72IB9). While there was only limited sequence identity, the following common sequences were observed for all 11 proteins, where they are given with the *E.* coli amino acid sequence numbers. X indicates a conservative substitution or a poor match. The key residues are again given in bold.

E730 X731 G732 K733 X734

E744 A745 X746 D747 F748 X479 X750 Y751 Y752 A753

X775 X776 P777 W778 **N779** F780 P781 X782 A783 I784

V799 X800 K802 P803 A804 X805

X882 **E883 T884** X885 **G886** X887 X888 A889 X890 I891

X905 X906 X907 S908 A909 X910

X913 G914 Q915 X916 **C917 S918** A919

E1018 X1019 F1020 G1021 P1022 X1023

X1046 G1047 L1048 T1049 X1050 G1051

G1070 N1071 X1072 Y1073 X1074 X1075 R1076

G1080 A1081 X1082 V1083 X1084

Q1086 X1087 F1088 G1089 G1090

S1094 G1095 X1096 X1097 K1098 X1099 G1100

The key residues in the *E. coli* sequence were correlated with the key residues in the T*. thermophilus* sequence. N779 in the *E. coli* sequence corresponds to N184 in the *T. thermophilus* sequence, E883 in the *E. coli* sequence corresponds to E288 in the *T. thermophilus* sequence, and T884 in the *E. coli* sequence corresponds to T289 in the *T. thermophilus* sequence. There are a few variations. There are two glycine residues (G885 and G886) in the *E. coli* sequence after T884. The first of these (G885) is replaced by serine (S) in the four *Arthrobacte*r*/Paenarthrobacter* and the four *Corynebacterium* sequences. However, G886, which corresponds to G291 in the *T. thermophilus* sequence, is conserved in all 11 sequences. C917 in the *E. coli* sequence corresponds to C322 in the *T. thermophilus* sequence, and S918 in the *E. coli* sequence corresponds to C323 in the *T. thermophilus sequence*.

The presence of the same key residues in the monofunctional and bifunctional Gram-positive P5CDH domains and in the model proteins from Gram-negative bacteria again suggests that the basic mechanism for P5C substrate binding and oxidation and NAD+ binding and reduction is the same in all of these microorganisms.

**3.4 SUBSTRATE CHANNELING DURING L-PROLINE CATABOLISM**

The proline catabolic pathway is short and only involves two enzymatic activities. Extensive studies in the laboratories of Becker and Tanner have shown that the efficiency of the degradative process can be greatly increased by channeling the L-Δ1-pyrroline-5-carboxylate (P5C) product of the first reaction (PRODH) through a structural tunnel to the active site of the second reaction (P5CDH, GSALDH). Along the way, it is spontaneously converted to L-glutamate-γ-semialdehyde which can then be oxidized to L-glutamate (Arentson et al., 2012; Liu et al., 2017; Tanner, 2019). This has been most clearly demonstrated for the bifunctional and trifunctional PutA proteins from the Gram-negative bacteria *Escherichia coli, Bradyrhizobium japonicum, Geobacter sulfurreducens, Rhodobacter capsulatus*, and *Sinorhizobium meliloti*. It has also been found to occur with the monofunctional enzymes from *Thermus thermophilus*, which are not fused together but are associated through their quaternary structure.

Evidence for substrate channeling among the Gram-positive bacteria is more limited. Korasick et al. (2017) crystallized the PutA protein from *Corynebacterium freiburgense* and found that it has both PRODH and P5CDH (GSALDH) activities. With L-proline as the variable substrate, the *Km* was 145 mmol L-1. With P5C as the variable substrate, the *Km* was 54 µmol L-1. In addition to these catalytic domains, the bifunctional protein has an aldehyde dehydrogenase superfamily domain at the carboxyl-terminal end, the removal of which reduced both enzymatic activities. Sedimentation velocity experiments indicated that in the presence of the proline analogue tetrahydro furoic acid, the PutA protein forms stable dimers. The authors used a trapping assay for P5C based on its binding to 2-aminobenzaldehyde to test for substrate channeling between the two active sites. The 2-aminobenzaldehyde/P5C complex has an intense yellow color and absorbs light at 443 nm. They found that the amount of this complex in solution was reduced by 52% in reactions where NAD+ was also present in the reaction mixtures. A kinetic model for NADH formation indicated that there was no substantial lag phase in the formation of the reduced coenzyme. These results were consistent with a substrate channeling mechanism.

Lagautriere et al. (2014) amplified the *pruA* and *pruB* genes from *Mycobacterium tuberculosis* and expressed them in *M. smegmatis*. The monofunctional PruA protein, containing a His6 tag and exhibiting P5CDH activity, was purified from M. smegmatis cells using a combination of metal affinity chromatography and size exclusion chromatography, based on an assay that measures the reduction of NAD+. The monofunctional PruB protein containing PRODH activity was also purified in this project using the same general methods. A coupled reaction leading to the proline-dependent formation of glutamate and an increase in A340 was observed over a period of 18 hr when the two proteins were combined in a ratio of 1 PruA:10 PruB. This was consistent with substrate channeling between the two subunits.

More recently, Kumar et al. (2023) demonstrated again that purified PruB and PruA proteins from *Mycobacterium tuberculosis* could be combined *in vitro* and were able to reduce menaquinol to menaquinone. The kinetic dependence of the PruA reaction on the P5C and NAD+ concentrations was defined. The reaction was found to be competitively inhibited by L-proline. The rate of the overall reaction was studied in the presence of varying concentrations of PruA and a fixed amount of PruB and a coupling of the two reactions was again demonstrated. To determine if the P5C formed as a result of the PRODH reaction can be channeled to the PruB protein, 2-aminobenzaldehyde was added to reactions with and without NAD+. As noted above, 2-aminobenzaldehyde forms a yellow complex with P5C that absorbs light at 443 nm. In reactions without NAD+, there was a rapid increase in A443 after the addition of the PruA protein to a mixture of PruB, proline, and ubiquinone. In reactions containing NAD+, the rate of this increase was greatly reduced. This supported the model of substrate channeling between the PruB and PruA proteins.

The nature of the tunnel connecting the PRODH active site and the P5CDH active site has been studied using mutants of the bifunctional proteins from *Bradyrhizobium japonicum* (Arentson et al., 2014) and *Geobacter sulfurreducens* (Korasick et al., 2021). In crystals of these proteins, the two sites are separated by 41 to 45 Å and a 75 Å tunnel is thought to connect them. During passage through this tunnel, P5C undergoes spontaneous hydrolysis to form L-glutamate-γ-semialdehyde. Using site-directed mutagenesis, Arentson et al. (2014) replaced four residues lining the central tunnel (T348, S607, D778, and D779) with bulky tyrosine (Y) or tryptophan (W) residues. The most dramatic reduction in NADH formation in a coupled reaction occurred with the mutants D779W and D779Y affecting a key aspartate (D) residue. In a similar way, Korasick et al. (2021) changed an alanine residue (A206) in a secondary tunnel in the *Geobacter sulfurreducens* sequence to a tryptophan (W). This mutant showed reduced PRODH activity and was less able to transfer P5C to the second active site, even though P5CDH (GSALDH) activity was not itself affected.

To determine if the same residues occur in the Gram-positive proteins, the amino acid sequences from *Bradyrhizobium japonicum*, *Geobacter sulfurreducens*, *Corynebacterium freiburgense*, and *Paenarthrobacter aurescens* were aligned. An aspartate residue corresponding to D779 in the *B. japonicum* sequence is also present in the *C. freiburgense*, and *P. aurescens* sequences. The corresponding residue in the *G. sulfurreducens* sequence is a histidine, although other aspartate residues are nearby. A comparison of the *B. japonicum* sequence to the eight bifunctional Gram-positive sequences listed in Table 1 indicated that the aspartate residue corresponding to D779 occurs in all of them. An alanine (A) residue corresponding to the A206 residue in the *G. sulfurreducens* appears as an alanine (A) in the *B. japonicum* sequence. However, it is replaced by a similarly small serine (S) in the *C. freiburgense* and *P. aurescens* sequences. A comparison of the *G. sulfurreducens* sequence to the eight bifunctional Gram-positive sequences listed in Table 1 indicated that the residue corresponding to A206 appears as a serine (S) in all of them. Transfer of P5C from one active site to the other, thus seems possible in the Gram-positive bacteria if it is not blocked by a bulky amino acid residue.

The examples of channeling in *C. freiburgense* and *M. tuberculosis* seem to be exceptions to the general pattern in Gram-positive bacteria, where proline catabolism more often involves a monofunctional membrane-associated PRODH activity and a separate cytoplasmic P5CDH activity. However, it would be worthwhile to explore this issue further with other monofunctional and bifunctional proline catabolic enzymes.

1. **CONCLUSIONS**

This analysis indicates that despite a long evolutionary separation of the low-G+C *Firmicutes* and the high G+C *Actinobacteria* clades of Gram-positive bacteria from the Gram-negative *Proteobacteria*, the same basic enzyme mechanisms have been retained in the L-proline catabolic pathway. The key amino residues needed for substrate and cofactor binding for both monofunctional and bifunctional PRODH and P5CDH (GSALDH) enzymes are almost always the same. There are occasional conservative substitutions of one polar residue for another, one nonpolar residue for another, or one small neutral residue for another. These changes may affect the *Km* for binding L-proline or P5C as well as the actual rates of catalysis. The large amount of amino acid sequence data now available provides a valuable tool for testing specific hypotheses about these reactions in Gram-positive microorganisms..

**DISCLAIMER (ARTIFICIAL INTELLIGENCE)**

The author hereby declares that no generative artificial intelligence (AI) tools such as Scalable Language Models (Chat GPT, COPILOT, etc.) or text-to-image generators were utilized in authoring or editing this paper.

**CONSENT AND ETHICAL APPROVAL**

Not applicable.

**REFERENCES**

Adams, E., & Frank, L. (1980). Metabolism of proline and the hydroxyprolines. *Annual Review of Biochemistry*, *49*(1), 1005-1061.

Arentson, B. W., Sanyal, N., & Becker, D. F. (2012). Substrate channeling in proline metabolism. *Frontiers in Bioscience*. 17, 375.

Arentson, B. W., Luo, M., Pemberton, T. A., Tanner, J. J., & Becker, D. F. (2014). Kinetic and structural characterization of tunnel-perturbing mutants in Bradyrhizobium japonicum proline utilization A. *Biochemistry*, *53*(31), 5150-5161.

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H. P., et al. (2016). Taxonomy, physiology, and natural products of *Actinobacteria*. *Microbiology and Molecular Biology Reviews*, 80(1), 1-43.

Bender, D. A. (2012). *Amino Acid Metabolism*. John Wiley & Sons.

Bogner, A. N., & Tanner, J. J. (2022). Structure-affinity relationships of reversible proline analog inhibitors targeting proline dehydrogenase. *Organic & Biomolecular Chemistry*, 20(4), 895-905.

Campbell, A. C., Bogner, A. N., Mao, Y., Becker, D. F., & Tanner, J. J. (2021). Structural analysis of prolines and hydroxyprolines binding to the L-glutamate-γ-semialdehyde dehydrogenase active site of bifunctional proline utilization A. *Archives of Biochemistry and Biophysics*, 698, 108727.

Chater, K. F. (2016). Recent advances in understanding *Streptomyces*. *F1000Research*, 5, 2795.

Christgen, S. L., & Becker, D. F. (2019). Role of proline in pathogen and host interactions. *Antioxidants and Redox Signaling*, 30(4), 683-709.

Deutch, C. E. (2025). L-Proline catabolism in Gram-positive bacteria: themes and variations. *Journal of Advances in Microbiology*, 25(8), 8-24.

Gao, B., & Gupta, R. S. (2012). Phylogenetic framework and molecular signatures for the main clades of the phylum *Actinobacteria*. *Microbiology and Molecular Biology Reviews*, *76*(1), 66-112.

Gottschalk G. (2012) *Bacterial Metabolism*, Springer

Halvorson, H. (1972) Utilization of single L-amino acids as sole source of carbon and nitrogen by bacteria. *Canadian Journal of Microbiology*, 18(11), 1647-1650.

Hogenhout, S. A., & Loria, R. (2008) Virulence mechanisms of Gram-positive pathogenic bacteria. *Current Opinion in Plant Biology*, 11(4), 449-456.

Inagaki, E., Ohshima, N., Takahashi, H., Kuroishi, C., Yokoyama, S., & Tahirov, T. (2006). Crystal structure of *Thermus thermophilus* Δ1-pyrroline-5-carboxylate dehydrogenase. *Journal of Molecular Biology*, 362(3), 490-501.

Korasick, D. A., Christgen, S. L., Qureshi, I. A., Becker, D. F., & Tanner, J. J. (2021). Probing the function of a ligand-modulated dynamic tunnel in bifunctional proline utilization A (PutA). *Archives of Biochemistry and Biophysics*, *712*, 109025.

Korasick, D. A., Gamage, T. T., Christgen, S., Stiers, K. M., Beamer, L. J., Henzl, M. T., et al. (2017). Structure and characterization of a class 3B proline utilization A: ligand-induced dimerization and importance of the C-terminal domain for catalysis. *Journal of Biological Chemistry*, 292(23), 9652-9665.

Korasick, D. A., Pemberton, T. A., Arentson, B. W., Becker, D. F., & Tanner, J. J. (2018) Structural basis for the substrate inhibition of proline utilization A by proline. *Molecules*, 23, 32.

Krishnan, N., & Becker, D. F. (2005). Characterization of a bifunctional PutA homologue from *Bradyrhizobium japonicum* and identification of an active site residue that modulates proline reduction of the flavin adenine dinucleotide cofactor. *Biochemistry*, 44(25), 9130-9139.

Kumar, S. Sega, S., Lynn-Barbe, J. K. Harris, D. L., Koehn, J. T., Crans, D. C., et al. (2023) Proline dehydrogenase and pyrroline-5-carboxylate dehydrogenase from *Mycobacterium tuberculosis*: evidence for substrate channeling. *Pathogens*, 12, 1171.

Lagautriere, T., Bashiri, G., Paterson, N. G., Berney, M., Cook, G. M., & Baker, E. N. (2014). Characterization of the proline-utilization pathway in *Mycobacterium tuberculosis* through structural and functional studies. *Biological Crystallography*, 70(4), 968-980.

Liu, L. K., Becker, D. F., & Tanner, J. J. (2017). Structure, function, and mechanism of proline utilization A (PutA). *Archives of Biochemistry and Biophysics*, 632, 142-157.

Ludwig, W., Schleifer, K.-H., & Whitman, W. B. (2009) Revised road map to the phylum *Firmicutes*. *Bergey’s Manual of Systematic Bacteriology*, pp. 1-13. Springer, New York, NY.

Luo, M., Arentson, B. W., Srivastava, D., Becker, D. F., & Tanner, J. J. (2012). Crystal structures and kinetics of monofunctional proline dehydrogenase provide insight into substrate recognition and conformational changes associated with flavin reduction and product release. *Biochemistry*, 51(50), 10099-10108.

Meeks, K. R., Bogner, A. N., Nix, J. C., & Tanner, J. J. (2024). Crystallographic fragment screening of a bifunctional proline catabolic enzyme reveals new inhibitor templates for proline dehydrogenase and L-glutamate-γ-semialdehyde dehydrogenase. *Molecules*, 29(22), 5408.

Moxley, M. A., Zhang, L., Christgen, S., Tanner, J. J., & Becker, D. F. (2017). Identification of a conserved histidine as being critical for the catalytic mechanism and functional switching of the multifunctional proline utilization A protein. *Biochemistry*, 56(24), 3078-3088.

Oren, A., & Garrity, G. M. (2021) Valid publication of the names of forty-two phyla of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 71(10), 005056.

Stackebrandt, E. (2009). Phylogeny based on 16S rRNA/DNA. *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons.

Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L., et al. (2010a). Crystal structure of the bifunctional proline utilization A flavoenzyme from *Bradyrhizobium japonicum*. *Proceedings of the National Academy of Sciences USA*, 107(7), 2878-2883.

Srivastava, D., Zhu, W., Johnson Jr, W. H., Whitman, C. P., Becker, D. F., & Tanner, J. J. (2010b). The structure of the proline utilization A proline dehydrogenase domain inactivated by N-propargylglycine provides insight into conformational changes induced by substrate binding and flavin reduction. *Biochemistry*, 49(3), 560-569.

Tanner, J. J. (2008) Structural biology of proline catabolism. *Amino Acids*, 35(4), 719-730.

Tanner, J. J. (2019). Structural biology of proline catabolic enzymes. *Antioxidants & Redox Signaling*, 30(4), 650-673.

Walker, M. C., & van der Donk, W. A. (2016). The many roles of glutamate in metabolism. Journal of Industrial Microbiology and Biotechnology, 43(2-3), 419-430.

White, T. A., Krishnan, N., Becker, D. F., & Tanner, J. J. (2007). Structure and kinetics of monofunctional proline dehydrogenase from *Thermus thermophilus*. *Journal of Biological Chemistry*, 282(19), 14316-14327.

Zhang, M., White, T. A., Schuermann, J. P., Baban, B. A., Becker, D. F., & Tanner, J. J. (2004). Structures of the *Escherichia coli* PutA proline dehydrogenase domain in complex with competitive inhibitors. *Biochemistry*, 43(39), 12539-12548.

Zhang, W., & Lu, Z. (2015) Phylogenetic evaluation of members above the species level within the phylum *Firmicutes* based on conserved proteins. *Environmental Microbiology Reports*, 7(2), 273-281.