# Hypothetical Protein PA5198 (Yestervin) indicated to be an LD-Carboxypeptidase within Pseudomonas Aeruginosa

**Mathieu Norcross, Co: William Conrad, Makenzie Kamm, Tate Rosenhagen, Allyson Richards, Kayenath Khan, Leslie Gonzales, Hannah Kobienia, Sergio Sosa, Anthony Sullivan, George Vladimirov. Bomby Ahuja-Kittiphanh**
Lake Forest College
Lake Forest, Illinois 60045

**Abstract**
The present experiment investigates the function of protein PA5198 found in *Pseudomonas Aeruginosa*, known as Yestervin, using a series of different bioinformatic methods. Preliminary research from the protein data bank provides an enzyme consortium number of 3.4.17.13. These findings led to the hypothesis that Yestervin acts as an LD-Carboxypeptidase in biological systems. Bioinformatic tools used to confirm the function of Yestervin are as followed: PyMol identification of active sites within Yestervin using motif comparison, BLAST/SMART BLAST sequence comparisons between Yestervin and proteins of known function, Dali global alignment comparing carbon backbone structures of known proteins and Yestervin, Pyrx autodocking identifying ligand binding to a Yestervin-like protein's active site amino acid residues, protein purification and kinetic evaluation through PNPA enzymatic assay. Results from these bioinformatic tools indicate that Yestervin is indeed a LD-Carboxypeptidase used to cleave amide bonds between L and D amino acids occurring naturally in bacterial peptidoglycan. These findings could play an important role in treatment development of *Pseudomonas Aeruginosa* infections, as well as provide mechanisms for fighting antibiotic-resistant bacteria in a hospital setting.

**Introduction**
The vast gap of knowledge surrounding discovered proteins and their function serves as a setback for advancing knowledge in many fields of Biology. To date, around 20% of protein in even the most well-studied organisms have no descriptions on their roles in biological systems (Wood et al. 2019). These proteins could serve to explain a multitude of mechanisms that have been previously misunderstood and could open doors to many new discoveries in the field of Biology.

The hypothetical protein PA5198, also known as Yestervin, found in *Pseudomonas Aeruginosa*, is a protein of unknown function requiring further inquiry. The unknown protein has an R-free value of 0.125 with 0.2% Ramachandran outliers, a total of six unique ligands (L(+)-Tartaric Acid, Di(Hydroxyethyl)ether, Glycerol, 1,2-Ethanediol, Potassium and Sodium ions), and PDB ID labeling it protein 1ZL0. Little is known about the protein of unknown function, however, it is suggested that the protein is possibly an LD-Carboxypeptidase. The enzyme consortium number (EC 3.4.17.13), which matches that of the unknown protein, suggests that the unknown protein is a LD-Carboxypeptidase known for the ability to cleave amide bonds between L and D amino acids occurring naturally in bacterial peptidoglycan. This severing of L and D amino acids takes tetrapeptides, truncates them into tripeptides, and then to peptidoglycan building blocks for the bacteria. LD-Carboxypeptidases have been the target of antibiotics due to their importance. The sensitivity to lactam antibiotics varies widely between antibiotics of the same class but correlates with the chirality of the amino acids of the antibiotic used.

The catalytic class of LD-Carboxypeptidases is unknown, being named the U61 family. In *Pseudomonas aeruginosa* this U61 enzyme was found to have LD-Carboxypeptidase activity. A study by Korza & Bochtler (2005) found that the *Pseudomonas aeruginosa* recombinant converted tetrapeptides to tripeptides and had a structure of a serine peptidase with a Ser-His-Glu catalytic triad. The Ser115 and His285 of *Pseudomonas aeruginosa* were ruled out as part of the enzyme as the activity was not homologous to LD-Carboxypeptidase, further narrowing down the structure of the unknown protein in question. This carboxypeptidase in question was further investigated by Xu et al. (2020) focusing on the Carbapenem Resistance in *Pseudomonas aeruginosa* infections. Carbapenem is used to treat *Pseudomonas aeruginosa* infections in a clinical setting, but recent resistance to Carbapenem has led to further investigation into the mechanism behind the resistance. It was found that the *ldcA* (PA5198) gene encoding for murein tetrapeptide carboxypeptidase negatively influences *ampC* expression in *Pseudomonas aeruginosa*. The *ampC* expression is an important mechanism in bacteria that helps them develop resistance to many anti-bacterial drugs. With murein tetrapeptide carboxypeptidase expression leading to decreased expression of *ampC*, a mechanism using carboxypeptidases to combat drug resistance in bacteria could arise (Xu et al. 2020). The resistance to Carbapenem is due to a multitude of overexpressed and under-expressed genes that produce varying amounts of proteins. One of these overexpressions that allows Carbapenem resistance is the overexpression on the chromosomal gene *ampC* encoding for intrinsic cephalosporinase in *Pseudomonas aeruginosa*. This overexpression links carboxypeptidase to Carbapenem resistance, allowing for further research into therapeutic treatments regarding the carboxypeptidase in *Pseudomonas aeruginosa*.

Understanding the unknown protein Yestervin in *Pseudomonas aeruginosa* is important because of the possible clinical implications it could have. *Pseudomonas aeruginosa* is the most common-gram negative bacterium found in nosocomial infections. The bacteria is responsible for 16% of nosocomial pneumonia cases, 12% hospital-acquired urinary tract infections, 8% of surgical wound infections, 10% of bloodstream infections, and 38% of ventilator-associated pneumonia deaths (Bodey et al. 1983). The bacterium is most problematic in immunocompromised patients, as it is an opportunistic infection because of the bacteria's presence throughout the natural environment. Once in the body, the infection is fast at causing problems throughout the body, from endocarditis in the heart, to eye problems, to urinary tract infections. The bacteria produces several extracellular products responsible for the extensive bodily damage and has been resistance or poorly treated with a number of different antibiotics throughout history (Van Delden & Iglewski 1998). The Yestervin protein could be the missing link in effective therapeutic treatments for *Pseudomonas aeruginosa*, which is why further investigation is a necessary process.

In this experiment, the Yestervin protein (1ZL0) was investigated using a series of different structural and sequential bioinformatic databases. These databases compared the hypothetical protein's structure and amino acid sequence to proteins of known function. From these results, it is hypothesized the Yestervin protein is acting as a LD-Carboxypeptidase within biological systems.

**Methods**
**PyMol/ProMol Methods:**
To identify proteins with similar structural active sites, PyMol/ProMol structural alignment was performed. The Pymol program was accessed through Microsoft Remote Desktop app. Once opened, the Promol application was accessed through the Plugin tab. Under the Promol interface, the Motif Finder tab was accessed, and the PDB name (1ZL0) was entered in the query box. Once a PDB name was entered, a set including all motifs is selected and RMSD values were calculated. The Motif Finder search parameters were unaltered, and a full protein structure was generated for Yestervin. The result pane under the motif finder lists all the active sites found on the protein entered. The results list Levenshtein distance, identity of the motif template, RMSD values, and an EC class for the motif selected. Using the RMSD values found under each motif, find the motifs with the lowest RMSD values. A motif with a low RMSD value is selected, and motif alignment is visualized through Promol. A visualization of active site alignment between the motif and the protein was generated and analyzed. The molecules generated are then identified for their residues. Clicking on each residue of the alignment shows the amino acid linked in the motif and how many atoms are in the respective alignment in the Pymol GUI interface. All values and visual representations of motifs were recorded, and values were analyzed to see if the unknown protein being studied best fit the motif of certain EC classes.

**BLAST/Pfam Methods:**
To identify proteins with similar sequences and domains compared to Yeservin, we performed a BLASTP and Pfam sequence search. The FATSA sequence for the Yestervin protein was obtained from https://www.rcsb.org/. This FASTA sequence was put into the BLAST query box and submit-

ted with no changes to the original BLASTP search criteria and the protein sequence was analyzed. Once the sequence was analyzed, a graphical summary is presented showing the top 100 sequences in the BLAST databank that overlap with the sequence of the query protein. The graphic summary was captured and analyzed. The top three alignments of each protein were looked at in detail and examined for their E-values and percent overlap. Once a general BLAST was run, a SMART BLAST was run Yestervin. The SMART BLAST provided a phylogenetic tree that showed how each protein relates to proteins found in other organisms, and where common ancestors arose between Yestervin and known proteins in specific organisms. Once all information is collected, the data is analyzed, and a conclusion regarding the possible function of the unknown protein was reached. The FASTA sequence for Yestervin was then put into the sequence query box (https://pfam.xfam.org/search#tabview=tab0) and run based on default search parameters. Once the protein's sequence was analyzed, all the protein superfamilies that match up with query protein sequence were displayed. Clicking on each protein superfamily gives a summary of the family with an example structure, and literature where the protein superfamily is discussed. Under the "domain organization" tab, common sequence architecture found within the protein superfamily is detailed. Under the "HMM logo" tab, the conserved amino acid residues are detailed on each sequence position. Once all Pfam data is collected, the data is analyzed. From this data, the possible function of the unknown protein is deduced and compared to data and conclusions from the BLAST and ProMol data collected earlier.

**Dali Global Alignment Methods:**

To identify proteins with similar 3D carbon backbones, we performed a Dali 3D Global alignment. The Dali search query was accessed at http://ekhidna2.biocenter.helsinki.fi/dali/. From there, the PDB search tab was selected and the PDB ID for each respective protein was added into the search bar. After Dali search was completed, the matches against full PDB were selected. These results were downloaded and put into raw data storage for further analysis following this lab. A summary of all matching proteins is presented with RMSD, Z-score, LALI, NRES, and %ID values. A description of each protein is provided at the end of each matching chain. From this summary page, three proteins of differing function were selected for further analysis. These proteins with different functions were analyzed using a 3D image with cartoon view and structural conservation. Each protein and its overlap with the query protein were captured and added to their respective tables for further data analysis. Following the analysis of 3D image superposition, the three proteins of differing functions were compared to the active sites of Yeservin found in the previous ProMol lab. Each protein of known function was selected, and the "structural alignment" tab was selected. Using the previously determined active sites for Yestervin, each protein of known function was analyzed to see if the active sites of our unknown proteins were conserved in the proteins of known functions. This was done by determining the first three amino acids before the conserved active site amino acid from the PDB ID of our unknown sequence. Using this four-letter amino acid sequence, a search was performed on the "structural alignment" tab. The four-letter amino acid sequence was found for the query protein and was compared with the protein of known function in the "structural alignment" tab. The four-letter amino acid sequence for the known protein that matched up with the query protein was recorded to see if the active site amino acid was conserved. Using all this data, the proteins of known function were determined to be good fits or not, and a hypothesis on both protein 2O14 and 1ZL0's function was created.

**Autodock Pyrx Methods:**

A hypothetical reaction for the Yestervin protein has been hypothesized using previous experiments. The AutoDock program was used to look at what ligands are used in the hypothesized reaction. The protein used in the AutoDock experiment is protein 5Z01 which is Murein Tetrapeptide Carboxypeptidase. This protein was used because Yestervin could not be loaded into the AutoDock program, so the protein 5Z01 was used because it has a similar carbon backbone and similar hypothetical function to the Yestervin protein. The PDB file for protein 5Z01 was downloaded from the protein data bank, and was uploaded to the AutoDock program. The E.C. Class 3.4.17.13 was put into the protein data bank search bar and the ligands that are commonly found for that EC class were observed. A total of five different ligands that were found in that EC class were selected (TLA, PEG, DMU, DEU, and ABU). Each ligand was saved as a .sdf file onto the computer and added into the AutoDock program. Once the protein and the ligands were loaded into the AutoDock program, Vina Wizard was run with all the ligands and protein 5Z01 selected. Once the program was run, the analysis for each ligand compared to protein 5Z01 was displayed. The list of all the ligands and the statistical analysis for each ligand were saved and stored in raw data storage. Once data was saved, the free energies of binding for each conformation of each ligand were observed. The ligand conformation with the most negative binding energy was visually observed in the AutoDock program and saved as a PDB file. Once the PDB file is saved, PyMol is opened and both the ligand and protein PDB files are added into the program. The PyMol visualizes the ligand inside the protein and the surface function shows the ligand fitting into binding sites on the protein. The amino acid residues that bind to the ligand at the binding site are visualized within PyMol and recorded in the results section. Pictures of the ligand within the protein binding site and the amino acid residues that help bind the ligand to the protein were recorded as figures within the results section.

**Nickel-His purification of Yestervin:**

The purpose of the Nickel-His purification is to separate the Yestervin protein from other non-specific proteins in order to further study the properties of Yestervin. A column with nickel ions present is used to separate Yestervin from other proteins. A binding buffer composed of 50mM Na2HPO4, 300mM NaCl pH 8.0 supplemented with 10mM imidazole is added to the Nickel-His column along with the protein solution. This buffer solution reduces the non-specific interactions between other proteins and the column, allowing our protein of interest to bind to the column specifically while also allowing Yestervin to remain in its native state. The pH of the buffer allows for Nickel-His binding to occur within the column, while also not denaturing the protein because of its neutral pH level. A wash buffer composed of 50mM Na2HPO4, 300mM NaCl pH 8.0 supplemented with 20mM imidazole is then added to the column in three successive additions. This wash buffer specifically elutes all proteins that are not of interest, leaving the POI attached to the nickel ions within the column. Finally, an elution buffer composed of 50mM Na2HPO4, 300mM NaCl, and 250mM imidazole at pH 8.0.A is added to the column. The excess amount of imidazole binds to the nickel ions, forcing the POI off the column, causing it to elute into fractions. A total of 3 elution additions are added to the column until there is no POI being eluted from the column anymore. The different fractions are run through a Gel Electrophoresis to see the contents of each of the fractions. The lysate shows all proteins present within the protein solution, including the Yestervin protein. The flow-through shows proteins that are eluted during the addition of binding buffer. The three different wash additions elute the proteins that remain in the column that are not the Yestervin protein. The elution fractions contain all of the purified Yestervin proteins separated from all other proteins. These separated Yestervin proteins are set aside for further research in order to determine their function in biological systems.

**Protein yfkn and 2O14 purification through 6X-His protein spin column:**

The purpose of the 6X-His protein column purification was to purify the protein of interest, yfkn, along with protein 2O14 for positive control. The procedure for this experiment was performed by Dr. William Conrad at Lake Forest College. Professor Conrad followed the HOOK protein purification spin protocol to purify protein yfkn and 2O14. The BL21 (DE3) E. Coli with six different plasmids were transformed and the bacteria was grown overnight in LB broth containing 100 ug/ml ampicillin at 37 C shaking at 240 rpm. Once grown, a total of 1 ml of bacteria was transferred into autoinduction media which contained lactose that induced bacterial expression. The HOOK bioscience protocol was then followed by Professor Conrad as directed, except no protease inhibitor was added into the samples. Once samples were created following the HOOK bioscience protocol, protein expression was observed in bacterial strains by placing 15 ul of bacteria in 5 ul 4X SDS PAGE loading buffer, which breaks the bacterial cells and linearizes the protein. The buffer also contains 1 ul TCEP which breaks any disulfide bonds. The samples were then denatured at 95 degrees C and 10 ul of both protein samples were added to a 15-lane 16% tricine gel. A 10 ul bio-rad precision protein was run through the gel to ensure identification of molecular weights of the proteins. The proteins were then lysed according to the HOOK purification kit and all fractions were saved and loaded into the gels. The bacterial cells were harvested from a 50 ml bacterial culture by centrifugation at 5,000xg for 10 minutes. The bacterial pellets were then resuspended in 2mL bacterial PE LB until the suspension was deemed homogenous. The sample was vortexed and 5-50 ul PEL-B™-Lysozyme was added before the sample was incubated at 37 degrees C for 30-60 minutes. The sample was vortexed again for 30 minutes after

incubation. This vortex occurred before soluble proteins were separated and was centrifuged at 25,000xg for 15 minutes. The now clarified lysate was transferred to a 15 ml centrifuge tube and was swirled until a homogenous slurry formed. A 0.8 ml 50% resin slurry solution was then transferred to the lysate before being incubated for 15 minutes at room temperature. The tube was then centrifuged at 1,500xg for 5 minutes to pellet the resin. The supernatant was discarded, and the resin was suspended in a 250 ul wash buffer. The 6X-His spin column was then added to a collection tube and then centrifuged for 2 minutes at 2,000xg. The flow-through was collected and the spin column was returned to the collection tube where 500 ul of wash buffer was added to wash away any unbound or nonspecific proteins. Once the wash buffer was added a second time, the column was eluted. The elution of the 6X His tagged proteins occurred through the addition of 0.5 ml elution buffer to the resin and the incubation of the column for 5 minutes at room temperature. The spin column was then centrifuged for 2 minutes at 2,000xg. The flow through was transferred to a 1.5 ml centrifuge tube and the spin column was then returned to the collection tube. This process was repeated three times to create three total elution trials. All of the proteins from the collection tubes were then added to a 15-lane 16% tricine gel where the results of the purification were analyzed to determine if protein yfkn and 2O14 were properly purified (methods were obtained from Professor Conrad's description on lab archives and through the HOOK 6X His Protein Spin Purification Protocol provided).

**Protein Kinetics analysis:**

The purpose of this experiment is to determine enzyme kinetics such as Vmax and Km values for both protein 2O14 and yfkn 3GVE. This was done using an assay measuring absorbance of the conversion of substrate to product of the hydrolysis of PNPA. This hydrolysis occurs in the presence of hydrolase enzymes, therefore high absorbance and Vmax values indicate an enzyme is a hydrolase. A 2O14 elution 1 and 3GVE (yfkn) elution 1 sample was created after being dialyzed into 50mM sodium phosphate buffer. The samples were spun in a Pierce 3 kDa MWCO centrifiugal filter concentrator three times using 20 mL of buffer and 1 mL of elution. The samples' protein abundance was measured using absorbance at 280nm with a path length of 1 cm, while the extinction coefficient for each protein was found through https://web.expasy.org/protparam/. Using a 96 well plate, each well was filled with 160 ul of 50mM sodium phosphate buffer. Then in each column, 10mM PNPA was added in the wells with a volume of 0-20 ul. A total of 20 ul of 2O14 and 3GVE were added to Rows A and C respectively and observed for changes in absorbance over time. The data was recorded into an Excel document and Michaelis-Menten plots, Lineweaver-Burke plots, and other protein kinetic values were calculated using the absorbance data. Substrate concentration and initial velocity values (calculated from absorbance) were calculated and used to create the Michaelis-Menten graph. The substrate concentration and initial velocity values were then used to find the 1/V0 and 1/[S] values used to create the Lineweaver-Burke plot. The lineweaver-burke plot was used to calculate the Km and Vmax values for each protein, which then was used to calculate other protein kinetic values that quantitively showed the protein's enzymatic activity.

## Results

**PyMol/ProMol active site structural analysis suggest EC range for Yestervin protein:**

The Yestervin protein is a protein of unknown function from *Pseudomonas aeruginosa*. The protein is a hypothetical protein with a sequence length of 311 and an EC= 3.4.17.13 (Source: https://www.rcsb.org/structure/1ZL0). Research into the EC class 3.4.17.13 suggests the protein is a LD-Carboxypeptidase known for the ability to cleave amide bonds between L and D amino acids occurring in bacterial peptidoglycan (Korza & Bochtler 2005). The EC class suggested in the PDB was tested through Promol analysis, which is a local structure-based program comparing active site residues between known proteins and Yestervin. The best fit criteria used in this experiment were motifs with a Levenshtein value of 0 and a RMSD value below 2. The Yestervin protein was expected to have active site alignment with motifs around EC = 3.4. The best fit active site for Yestervin protein was suggested to be motif 2ace. The 2ace motif has a Levenshtein value in the range of 0-4, RMSD All: 1.5116, RMSD alpha: 1.5092, and RMSD alpha and beta: 1.3696. The 2ace motif showed 3 alignments with similar residues in those alignments (His 285 Yestervin- His 440 2ace, Glu 217 Yestervin- Glu 327 2ace, and Ser 115 Yestervin- Ser 200 2ace) (*Figure 1*). The alignment visualization shown in *Figure 1* shows differences between the motif and Yestervin structures. The motif 2ace has an EC class of 3.1.1.7, which is not in the range of the expected EC of 3.4 for the protein Yestervin. The second-best active site was determined to be 1o2u. The 1o2u active site has a Levenshtein value of 0, RMSD All: 1.4783, RMSD alpha: 0.058, and RMSD alpha and beta: 0.6694. The overall motif alignment with the Yestervin protein is shown in *Figure 1*. The 1o2u motif was chosen as the second-best active site because it had less residue-matches compared to 2ace, but still had a Levenshtein value of 0 and a RMSD value below 1.5. The residue alignment between the 1o2u motif and the protein were found to be similar (His 51 Yestervin- His 57 1o2u, and Asp 69 Yestervin- Asp 102 1o2u) (*Figure 1*). The motif 1o2u has an EC class of 3.4.21.4, which falls in the predicted range of the Yestervin protein. The two best fit active sites found were from differing EC classes. The best fit active site belonged to the EC class of 3.1, while the second active site belonged to the predicted EC class of 3.4. Research into the difference between EC classes 3.1 and 3.4 shows the differing functions the two motifs that have significant active site overlap with Yestervin have. While both EC classes are hydrolases, EC class 3.1 are esterases acting on ester bonds, while EC class 3.4 are peptidases that act on peptide bonds. The two best fit active sites suggest that the Yestervin protein is either an esterase acting on ester bonds or a peptidase acting on peptide bonds. More analysis is needed to confirm Yestervin's EC class and its function in biological systems. In conclusion, the protein Yestervin needs further analysis in order to properly determine its function, as the active sites show conflicting EC values and functions in biological systems.

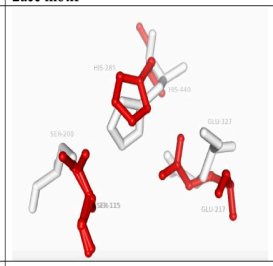
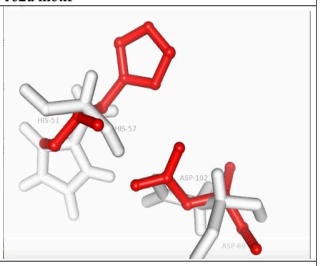
| Motif | 2ace motif | 1o2u motif |
|---|---|---|
| Image of motif from PyMol program |  |  |
| EC class | 3.1.1.7 | 3.4.21.4 |
| RMSD All | 1.5116 | 1.5855 |
| Levenshtein value range | 0-4 | 0 |
| # of residues | 3 | 2 |
| Yestervin residue | His 285, Glu 217, Ser 115 | His 51, Asp 69 |
| Matching motif residues | His 440, Glu 327, Ser 200 | His 57, Asp 102 |

**Figure 1:** ProMol active site alignment between Yestervin (red, 1ZL0) and the 2ace and 1o2u motifs (white). The 2ace-Yestervin amino acid residue overlap occurred between three residues (a His, Glu, and Ser). Yestervin and the 2ace motif showed significant overlap with an RMSD value of 1.5116, meaning the motif and protein are closely aligned and thus may share similar functions. The 1o2u-Yestervin amino acid overlap occurred between two residues (a His, and Asp). Yestervin and the 1o2u motif showed significant overlap with an RMSD value of 1.5855 between the two amino acid residues suggesting the motif and Yestervin may share similar functions.

**BLAST/SMART BLAST sequence alignment results indicate LD-Carboxypeptidase as a significant match to protein Yestervin.**

From PyMol/ProMol analysis, protein Yestervin could not be conclusively determined to belong to a specific EC class. To help narrow down the possible functions, sequence alignment through BLAST programs were run to determine possible proteins with similar sequences. The protein Yestervin had significant BLAST findings that helped narrow down the function. The expected results for both the BLAST experiment were that the Yestervin protein would have significant sequence alignment with LD-Carboxypeptidases and any close relatives to the LD-Carboxypeptidases. The BLAST sequence alignment tool found significant overlap with LD-Carboxypeptidases. Both the BLAST and SMART BLAST showed that protein Yestervin had significant sequence overlap with LD-Carboxypeptidases found in a multitude of different bacteria, which is consistent as LD-Carboxypeptidases belong to prokaryotic bacteria and not in eukaryotic or archaic or-

ganisms (*Figure 2*). The BLAST experiment also indicated the superfamily Peptidase_S66 as the potential protein family that the protein Yestervin may belong to. The BLAST superfamily analysis showed a conserved catalytic triad (Ser-Glu-His) between the Yestervin protein, and multiple LD-Carboxypeptidases were found within different organisms in the Peptidase_S66 Family (*Figure 3*). These findings suggest that the protein Yestervin matches sequentially with LD-Carboxypeptidases. These LD-Carboxypeptidases have a function of hydrolyzing the amide bond that links dibasic amino acids to C-terminal D-amino acids. These LD-Carboxypeptidases have an EC: 3.4.17.13, which is the same EC value as Yestervin. From this data, we can hypothesize that the protein Yestervin shares the same function as LD-Carboxypeptidase and indeed does belong to the EC class 3.4.17.13 .

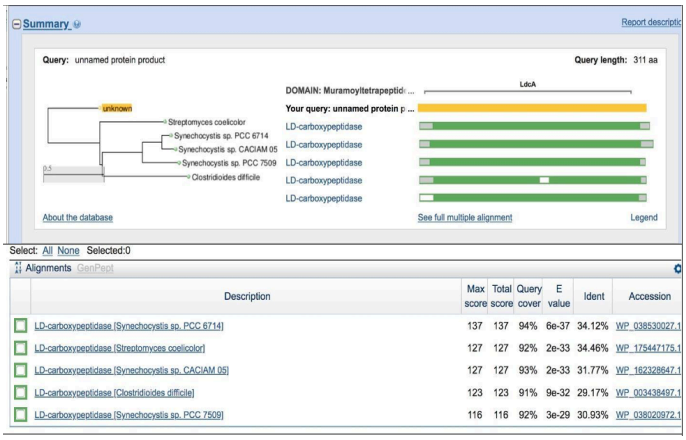



**Figure 2:** A SMART BLAST sequence search showing a phylogenetic tree of closely related evolutionary relatives between Yestervin and proteins from a multitude of proteins from different bacteria. This phylogenetic tree compares Yestervin to proteins of known function in a series of different model organisms. Yellow band shows the Yestervin sequence, while the green bands show sequence hits from different bacteria species. All significant hits for Yestervin show LD-Carboxypeptidases from different bacteria species.

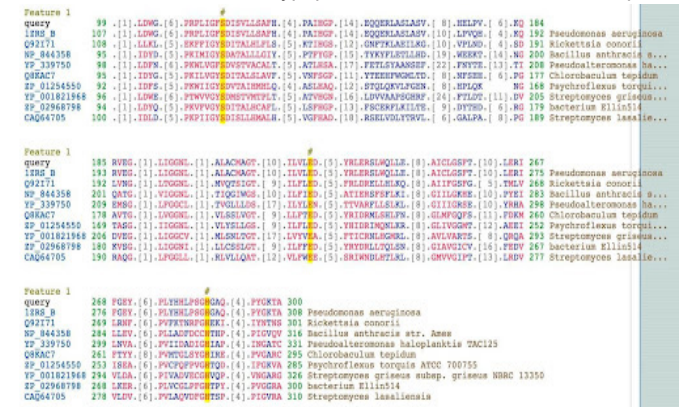


**Figure 3:** Conserved domain alignment from BLAST search between different bacteria species showing conserved catalytic triad between the Yestervin and LD-Carboxypeptidases from a multitude of different species. The Ser-His-Glu Triad was found in previous research by Korza & Bochler (2005). Highlighted amino acids indicate the conserved Ser-His-Glu triad. Red amino acids indicate conserved amino acids throughout all the different bacteria species, while the blue amino acids indicate amino acids that are not conserved throughout all observable bacteria species.

**Dali Global Structural alignment shows the protein Yestervin's significant overlap with the Murein Tetrapeptide Carboxypeptidase protein 5Z01.**

By BLAST sequence alignment, Yestervin was observed to have significant sequential overlap with known LD-Carboxypeptidases. These results were followed by searching for homologous structures rather than primary sequences using Global 3D alignment through the Dali search engine. This Dali search compared proteins of known function with similar alpha carbon backbone structure to Yestervin. These results allowed for both carbon backbone and active site structural site alignments to be examined between proteins of known function and Yestervin. The 3D superposition alignment between Yestervin and three proteins of known function provided more insight into Yestervin's function. The gap in knowledge surrounds Yestervin's function. The expected results are that Yestervin lines best with LD-Carboxypeptidases as indicated in previous BLAST experiments. The best fit carbon backbone alignment with Yestervin is protein 5Z01-A. The protein has a Z-score of 33.6, RMSD value of 2.1, and a 27% ID. The overlap between the two proteins is significant, with multiple coils and middle carbon backbone chains being conserved between 5Z01-A and Yestervin. However, there is no overlap with the Yestervin ligand groups and some of the outer carbon backbone chain of the 5Z01-A (*Figure 4*). The significant overlap is underscored by the high Z-score, low RMSD, and significant LALI value of 284. From these values and the visual, it can be concluded that Yestervin has significant overlap with protein 5ZO1-A which is found to be a Murein Tetrapeptide carboxypeptidase (*Figure 4*). Another significant alignment came from protein 3TYX-B. This protein had a Z-score of 33.3, an RMSD of 2.3, and percent ID of 20%. The 3D superposition image shows significant overlap between the protein's carbon backbone, but no overlap in the middle ligands of Yestervin. There also is lacking overlap with some of the outer carbon backbone of protein 3TYX-B. Besides the missing overlap on these structures, the alignment looks significant and is underscored by the high z-score, low RMSD, and high LALI of 332. The protein 3TYX-B is a Microcin immunity protein MCCF, which goes against the expected results (*Figure 4*). The third protein with significant overlap was protein 4INJ-A. This protein had a Z-score of 33.2, an RMSD of 2.1, and a percent ID of 26%. The carbon backbone shown on the 3D superposition image shows significant overlap between the middle of the carbon backbone for both proteins. However, the proteins don't match at the Yestervin middle ligands (green ball-like structures in *Figure 4* images), and at the top part of the Yestervin carbon backbone. The proteins also do not overlap at middle ligands of protein 4INJ-A and at the outer carbon backbone of protein 4INJ-A. Despite these discrepancies in overlap, the significant overlap between the middle backbones is underscored by the high z-score, high percent ID, low RMSD, and high LALI value of 326. Protein 4INJ-A is described as a LMO1638 Protein which is described as a hydrolase (*Figure 4*). This goes against the expected results of the matching protein being a carboxypeptidase. The active sites of the proposed catalytic triad within Yestervin were then compared to the Dali fit proteins to see if there was conserved active site alignment. The goal of this was to determine if the Dali proteins were good fits when it came to Yestervin's active site. To determine this, the active site residues for Yestervin were compared to each of the Dali fit protein sequences to see if the active site residues were conserved. In protein 5Z01-A, 2 of 3 active sites found in Yestervin (His 285, Ser 115) were conserved. The Glu residue in Yestervin was replaced with an Asn in protein 5Z01-A. These findings, along with 3D superposition data, suggest that Yestervin and 5Z01-A have similar structures and thus similar functions (*Figure 4*). The protein 5Z01-A is a Murein Tetrapeptide carboxypeptidase, which suggests that Yestervin acts as a carboxypeptidase in biological systems. Protein 3TYX-B had 2 of 3 conserved amino acids when compared to protein 1ZL0. The active site residue Glu 217 in Yestervin is replaced with a phenylalanine in protein 3TYX-B. Despite this discrepancy, all other data suggests that protein 3TYX-B is a good fit to Yestervin (*Figure 4*). However, the protein 3TYX-B does not match as well as protein 5Z01-A does, meaning the function of Yestervin is more likely to be a carboxypeptidase rather than a microcin immunity protein MCCF. The protein 4INJ-A showed little active site overlap when compared with Yestervin, having only the His 285 active site conserved. Using this data, protein 4INJ-A was determined to not be a good fit for Yestervin, and therefore does not share the same function as Yestervin. From all of this data, we can hypothesize that Yestervin is a carboxypeptidase. Yestervin fits best with protein 5Z01-A, which is a Murein Tetrapeptide carboxypeptidase. This data, along with data collected in BLAST and ProMol sequencing, suggest that Yestervin is a carboxypeptidase. Therefore, it is hypothesized that Yestervin acts as a carboxypeptidase within biological systems.

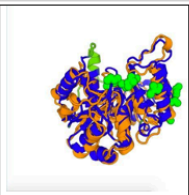
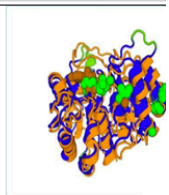
| Chain/Description | 5Z01-A<br>Murein Tetrapeptide carboxypeptidase | 3TYX-B<br>Microcin immunity protein MCCF | 4INJ-A<br>LMO1638 Protein |
|---|---|---|---|
| Image Captured (Cartoon with Structure Conservation view) |  |  |  |
| Active Site: His 285<br>PSGH | DFGH | DCCH | DFGH |
| Active Site: Glu 217<br>EDVG | EDIN | EDSS | EDDF |
| Active Site: Ser 115<br>IGFS | CGHS | IGYS | CGFA |
| # of active site matches | 2 | 2 | 1 |
| Good Fit? | Yes | Yes | No |
| Z-Score | 33.6 | 33.3 | 33.2 |
| RMSD | 2.1 | 2.3 | 2.1 |
| LALI (length of alignment) | 284 | 287 | 280 |
| NRES (Number of Residues) | 305 | 332 | 326 |
| %ID | 27 | 20 | 26 |

**Figure 4:** Dali Global Alignment results showing significant carbon backbone and active site matching between Yestervin and proteins 5Z01-A and 3TYX-B. All three proteins (5Z01-A, 3TYX-B, 4INJ-A) were compared to Yestervin through protein superposition. The blue ribbons represent overlap between the protein in question and Yestervin. The orange ribbons represent the protein in the database that doesn't overlap with Yestervin, while the green ribbons represent Yestervin structures that don't overlap with the proteins in the database. The green balls within image represent Yestervin ligands not found on query proteins. After sequential comparisons, Protein 5Z01-A and 3TYX-B were found to be good matches with 2 out of 3 active site amino acids being conserved. Protein 4INJ-A was determined not to be a good fit because of the lack of overlap between Yestervin and 4INJ-A active site amino acids. Red letters show amino acids conserved between Yestervin and query protein while black letters show amino acids that are not conserved. The highlighted amino acids are the active site residues in Yestervin.

**Pyrx Autodocking show ligands DEU and TLA binding to expected active site in protein 5Z01**

Following Dali Global alignment matches, the ligands involved in the potential Yestervin reaction were found. The goal of this Autodock experiment was to determine specific ligands/substrates that bind to Yestervin and possibly play a role in the enzymatic reaction. The ligand binding of the Murein Tetrapeptide Carboxypeptidase 5Z01 was used to provide conclusive results into what the ligands are involved in the hypothetical reaction of the protein of interest Yestervin. The protein Yestervin was unable to be used for this experiment, as the AutoDock program was unable to load the protein into the program. As a result, protein 5Z01 was used, as it was the best fit protein found in the Dali sequence which compared carbon backbones of the proteins. This switching of the protein being observed led to changes in the expected amino acid residues involved in the ligand binding. The Yestervin protein had an active site composed of His 285, Glu 217, and Ser 115. However, in protein 5Z01, the Glu 217 was not conserved and instead had an Asn at that site. Therefore, we would expect ligands to bind with or near a His, Asn, and Ser amino acid (*Figure 5*). Protein 5Z01 had the greatest affinity for ligands DMU and TLA, with binding affinity of -5.9 and -4.9, respectively (*Figure 5*). The ligands DMU, TLA shared the same binding site on protein 5Z01 and share similar amino acid residue interactions (*Figure 5*). This binding site shared by both ligands was also shared by the other two analyzed ligands ABU and PEG. The ABU and PEG ligands were not analyzed in detail because their binding affinity of above -4 suggested they were not the strongest fit for the active site. Another ligand DEU was analyzed as well but did not interact with the expected active sit and was thus not analyzed. The ligand TLA has a binding affinity of -4.9 with interactions with 2 expected amino acid residues. The TLA ligand has two amino acid residue interactions with Ser 106 and Asn 176, which is consistent with the proposed active sites. The visualization of these amnio acid residues interacting with TLA are found in *Figure 5*. Both the Ser and Asn were involved in the binding of the TLA ligand, suggesting that the ligand is involved in the hypothetical reaction for Yestervin. The DMU ligand has similar results to that of the TLA ligand. With a binding affinity of -5.9, the DMU ligand binds a little stronger to the active site compared to TLA. The DMU ligand also has active site amino acid residue interactions, having close interactions Asn 176, Ser 106, and His 270. DMU differs from TLA in the fact that it shares a close interaction with His 270, which is the final expected amino acid residue involved in the active site. The visualization of amino acid residues interacting with ligand DMU are found in *Figure 5*. These results indicate that the ligand DMU is involved in the hypothetical reaction of the Yestervin protein as it has interactions with the catalytic triad of amino acids typical of the Yestervin protein. The DMU ligand is only present in 4 currently known hydrolases, but the significant alignment with the expected amino acid residues at the active site suggests that the ligand is involved in this hypothetical hydrolase reaction. In conclusion, both TLA and DMU have significant binding affinities under -4 and have significant overlap with the expected amino acid residues, showing that they bind to the expected active site. TLA has a less significant binding affinity and only binds to 2 of 3 expected amino acid residues, but is a prevalent ligand in hydrolases, which is the expected reaction for the Yestervin protein as Yestervin is a hydrolase. The ligand DMU has a more significant binding affinity and binds to all 3 of the expected amino acid residues. While it is not often found in hydrolases, the binding affinity and amino acid residue interactions suggest that it is an important ligand for the Yestervin reaction.

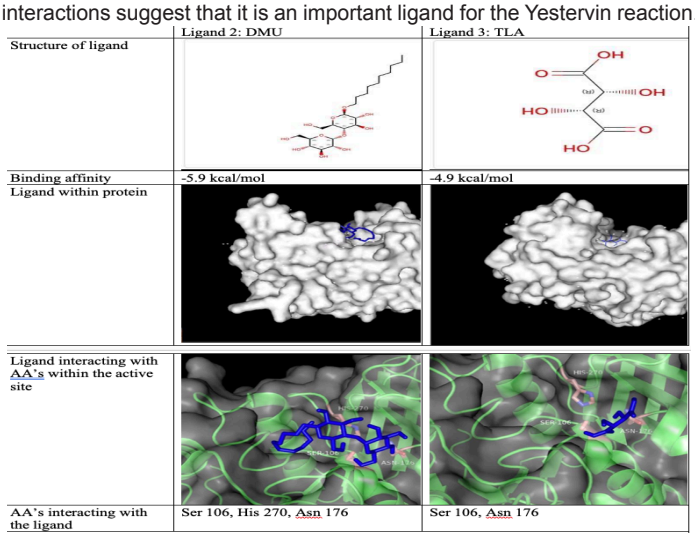



**Figure 5:** Pyrx Autodocking of two ligands that had significant binding with protein 5Z01. Ligands DMU, and TLA structures from the PDB were provided and their binding affinities indicated they bind significantly to 5Z01 as they have a binding affinity under -4. Both DMU and TLA share the same active site. Both DMU and TLA are found within the same active site and interact with expected amino acids within that active site. The ligand DMU (left side in blue) interacts with all three amino acid residues (Ser 106, His 270, Asn 176 in pink) attached to the green carbon backbone, while TLA (right side in blue) only interacts with two of the active site residues (Ser 106, Asn 176 in pink).

**Protein purification successfully purified target protein yfkn**

Following Autodocking experiments, where the protein active site and amino acid residues were identified and observed, the protein yfkn was purified in solution for further investigation into Yestervin's unknown function. The protein yfkn replaces Yestervin for this experiment as yfkn was the closest available protein to Yestervin that was purified by Dr. William Conrad. The goal of the protein purification experiment is to transform **E. Coli** with plasmid expressing yfkn, then purify the protein from the

*E. Coli* using the HOOK 6X His Protein Spin protocol which performs rapid purification of 6X His tagged proteins from bacterial cultures. To accomplish this goal, a His protein Spin protocol was implemented in which a figure showing the expected kDa for each protein was created and then the protein of interest was purified with nickel chelating resin to elute and purify the protein. Protein yfkn was purified in a 15 lane 16% gel and took up lanes 9-15. The expected results would show protein yfkn at approximately 80 kDa and would show pure purification in the last elution fraction. The protein yfkn shows a band in between 75-100 kDa, which is the expected result as protein yfkn has a molecular weight of 80 kDa. *Figure 6* shows protein yfkn solution running through a 15 lane 16% tricine gel. The protein lysate lane shows an abundant presence of yfkn. This abundance of yfkn causes bands to appear in the flow through and wash lanes (*Figure 6*). These results occur because there is such an abundance of protein that there is not enough nickel to bind to within the column, and thus some of the protein cannot bind to the column and is washed out. The elution trials show a strong band at 80 kDa, which shows the appearance of yfkn but the first two elutions show a protein at 10-15 kDa. Elution 3 shows the protein yfkn pure within its fraction, which is the expected result (*Figure 6*). In conclusion, the protein yfkn was purified in the final elution trial as the band at 80 kDa in the elution 3 fraction indicates that the only protein present within that fraction is the protein yfkn. This means that the protein was successfully purified and thus can be examined further in future experiments.

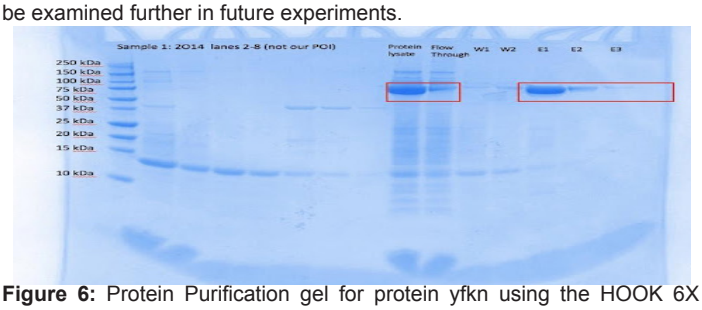



**Figure 6:** Protein Purification gel for protein yfkn using the HOOK 6X His Protein Spin protocol which performs rapid purification of 6X His tagged proteins from bacterial cultures. Lanes 2-8 show the successful purification of 2O14 that was used as the positive control for protein kinetics experiments. Lanes 9-15 show significant portions of purified protein yfkn with significant bands within the protein lysate lane each elution trial at the expected 80 kDa. So much yfkn was purified that there was not enough room for all purified protein to bind, explaining the band on the flow through. The red boxes show the lanes where yfkn was purified.

**Protein yfkn demonstrated no activity in enzymatic assay when compared to positive control 2O14**

After successful purification of both protein 2O14 and yfkn, both proteins were added to an enzyme activity assay in order to determine protein kinetic values such as Vmax, Km, and Kcat among other values. The goal of this experiment was to quantify both proteins' enzymatic activity with the values of Vmax, Km, Kcat. The technique involved included adding both purified proteins into a 10mM PNPA solution and watching for changes in color of the solution. The color changes were measured in absorbance, which indicated enzymatic activity and was used to calculate all figures and values for both proteins. The expected results for each protein were that a Vmax, Km, and Kcat value would be calculated, and that the proteins would have quantified enzymatic activity values. For this experiment, the discussion of protein 2O14 will be used as a positive control to indicate that enzymatic assay errors or other variables are not what caused the results for protein yfkn. For protein 2O14, enzymatic activity was determined through the assay's absorbance calculations and initial velocity, Michaelis-Menten, and Lineweaver-Burke plots were used to quantify this enzymatic activity. The initial velocity plot for protein 2O14 demonstrates that as substrate concentration increased, the product concentration also increased (*Figure 7*). This increase in product as substrate increases demonstrates that enzymatic activity is present in the reaction. This initial velocity plot was followed by a Michaelis-Menten plot for protein 2O14. The Michaelis-Menten plot was calculated and further demonstrated enzymatic activity. The plot showed that as substrate concentration increased, so did the initial velocity of the reaction. The reaction did not meet a Vmax as the peak did not level off within the given substrate concentration range (*Figure 7*). This inability to distinguish Vmax and Km from the Michaelis Menten led to the production of a Lineweaver-Burke plot. The Lineweaver-Burke plot demonstrated quantitative values that allowed for the calculations of Vmax and Km. The lineweaver-burke plot demonstrated a y-intercept of 0.0415, which was used to calculate a Vmax value of 24.1 (*Figure 7*). This high Vmax explains why the Vmax was not seen on the Michaelis-Menten graph as the graph did not go as high as 24 for the initial slope. The slope of the plot was found to be .1238, which was used to calculate a Km value of 2.98 since slope=Km/Vmax. The Vmax and Km values demonstrate that there is indeed enzymatic activity for protein 2O14. From these two values, the Kcat which is the turnover number, or the number of substrate molecules each enzyme site converts to product per unit of time, was found to be 4075361, demonstrating that the 2O14 protein is a catalytic enzyme within the assay. Enzyme efficiency was determined to be 1366024, demonstrating that 2O14 is an efficient enzyme. The units of microliter per enzyme and activity of enzyme were found to be 0.00241 and .120 respectively, further quantifying the enzymatic activity of 2O14. The assay was effective in determining protein kinetic values for protein 2O14, and thus showed that protein 2O14 is indeed able to catalyze the hydrolysis reaction of PNPA. The Vmax, Km, and Kcat values give a quantitative measure of the protein 2O14's enzymatic function, which were used as a positive control for the experimental protein yfkn.

For the protein yfkn, the results showed that the protein had no enzymatic activity within the assay. The protein yfkn was used instead of the Yestervin protein because of restraints in time and the availability of protein yfkn as a purified protein. Protein yfkn was the closest protein available to Yestervin and was thus used as a substitute for Yestervin. The protein yfkn showed no enzymatic activity within a PNPA assay as shown in *Figure 7*. The initial velocity plot shows that there is no enzymatic activity as there is no product formation no matter how much substrate is present in the system (*Figure 7*). This lack of increases in product as the substrate increases demonstrates that there is no enzymatic activity for protein 3GVE in the PNPA assay. The Michaelis Menten graph for protein 3GVE further confirms these findings (*Figure 7*). The Michaelis-Menten graph shows no consistent increase in initial velocity with substrate. This inconsistent trend further demonstrates that there is no enzymatic activity for 3GVE in the current assay. The lineweaver-burke plot was unable to find a distinguishable Vmax, Km, or any protein kinetic values (*Figure 7*). The lineweaver-burke had a negative slope and no distinguishable y-intercept meaning that protein kinetic values like Vmax, Km, and Kcat were not distinguishable. These inconclusive graphs lead to the conclusion that protein 3GVE is not involved in hydrolysis and may be a catalyst for other reactions. The current hydrolysis assay shows no function for protein 3GVE, and thus we can conclude that protein 3GVE is not a catalyst for the PNPA hydrolysis reaction and may be a catalyst for other types of reactions. Different assays with differing reactions should be performed in order to determine protein 3GVE's function and its protein kinetic quantitative measures such as Vmax, Km, and Kcat.

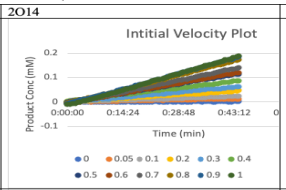
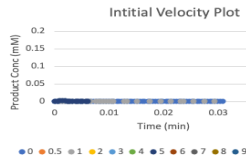
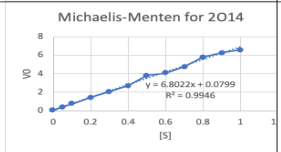
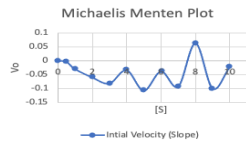
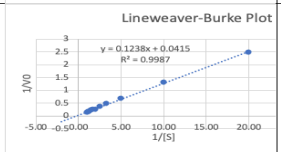
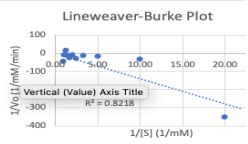
| Protein being analyzed | 2O14 | yfkn |
|---|---|---|
| Initial Velocity Plot |  |  |
| Michaelis Menten plot |  |  |
| Lineweaver-Burke Plot |  |  |
| Vmax | 24.09639 | N/A |
| Km | 2.983373 | N/A |
| Kcat | 4075361 | N/A |
| Enzyme efficiency | 1366024 | N/A |
| Units per microliter of enzyme | 0.00241 | N/A |
| Activity | 0.120482 | N/A |



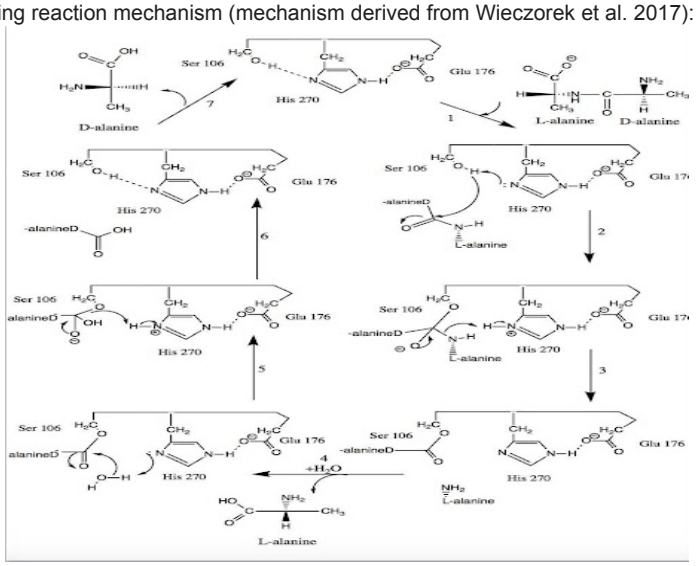
**Figure 7:**

Comparison of protein kinetic measures between positive control protein 2O14 and protein yfkn. The enzymatic assay where hydrolysis of PNPA

occurred was used for both proteins, and absorbance values were used to determine protein kinetic values. Protein 2O14 provides a positive control, showing the expected initial velocity, Michaelis-Menten, and Lineweaver-Burke plot from which kinetic values like Vmax and Km were derived. The initial velocity plot looked at product concentration in mM over a given time within the reaction system. This allowed for initial velocities measures to be calculated which were used to create the Michaelis-Menten and Lineweaver-Burke plots. Protein yfkn showed no activity as evident from its initial velocity plot and thus kinetic values were unable to be derived.

**Discussion**

Using all the data from the different bioinformatic tools, the hypothesis that Yestervin is an LD- Carboxypeptidase can be supported. The PyMol active site analysis implicated Yestervin as either belonging to EC class 3.1 or 3.4, while also suggesting the possibility of a Ser-His-Glu catalytic triad through the amino acid residue comparison between Yestervin and 2ace. The sequential comparison through BLAST/SMART BLAST indicated Yestervin was likely a LD Carboxypeptidase belonging to the Peptidase_S66 superfamily while further supported the presence of a Ser-His-Glu catalytic triad through catalytic triad conservation among different bacteria species with similar sequences as Yestervin. Running Yestervin through a Dali Global alignment found a significant carbon backbone match with the Murein Tetrapeptide Carboxypeptidase protein 5Z01, which further supports the hypothesis that Yestervin is a LD carboxypeptidase. The protein 5Z01 also had significant overlap with the active site amino acid residues in Yestervin. The protein 5Z01 was used in Pyrx autodocking and showed that ligands DMU and TLA bind within the expected active site and interact with the expected Ser-His-Glu catalytic triad (in the case of 5Z01, the Glu was replaced with an Asn meaning the catalytic triad observed was a Ser-His-Asn triad). The ligand TLA is a very common ligand among hydrolases, with involvement in 110 known hydrolase proteins. With the TLA ligand being so common in hydrolases along with the interaction with the expected active site, it can be concluded that TLA is an important ligand involved in the hypothetical reaction of the Yestervin protein. A possible function for the ligand TLA in Yestervin is the function of inhibitor. The TLA molecule shown in *Figure 5* shows that the TLA has two carboxy groups but no amine group. This presence of no amine group means the cleaving of L and D amino acids by Yestervin would not be possible as there is no amine group that connects to a carboxy group that would thus be cleaved by Yestervin. The Yestervin protein was then replaced by similar protein yfkn due to the unavailability of Yestervin, which was purified through the HOOK 6X His Protein Spin protocol as it showed up at the expected 80 kDa within the 15 lane 16% tricine gel. The purified protein yfkn was unable to be accurately assessed for protein kinetic measures due to inactivity within the assay used. The positive control of 2O14 did show activity, meaning the lack of activity of yfkn was not due to a systematic error, but rather that yfkn did not react with PNPA in the assay. All of these key findings support the hypothesis that Yesetervin is indeed an LD-Carboxypeptidase within biological systems. The Ser-His-Glu catalytic triad within Yestervin acts to cleave L and D amino acids through the following reaction mechanism (mechanism derived from Wieczorek et al. 2017):




The results showing the Ser-His-Glu catalytic triad within Yestervin support the hypothesis that Yestervin is an LD-Carboxypeptidase as well. A review of different serine proteases showed that a Ser-His-Glu triad has been found in the LD-Carboxypeptidase (PDB: 1ZRS) within the organism *Pseudomonas Aeruginosa*, which is the same organism Yestervin has been found in (Korza & Bochtler 2005) (Ekici et al. 2008). With the protein 1ZRS being an LD-Carboxypeptidase within the same organism as Yestervin and containing the same Ser-His-Glu catalytic triad as Yestervin, the hypothesis that Yestervin is a LD-Carboxypeptidase just like 1ZRS can be concluded from the research performed in this study along with the research from other studies previously cited. The Yestervin protein has the potential to be an alternative protein than the suggested LD-Carboxypeptidase. The PyMol active site alignment suggested the possibility that Yestervin belongs to EC class 3.1. The LD Carboxypeptidases belong to EC class 3.4, while EC class 3.1 are esterases that cleave ester bonds. The PyMol bioinformatic program suggests the possibility of Yestervin belonging to the class of esterases, as 2 of the 3 best fit active sits belong to proteins that are known esterases. These results can be explained by the similarities between esterases and peptidases. Both enzymes use hydrolysis to break bonds within proteins. Both of them belong to the same EC class, meaning both are hydrolases in nature. Therefore, it is possible that there is some overlap between the amino acid residues between esterases and peptidases as they both cleave bonds because they are both hydrolases. The Dali global alignment also suggests the possibility that Yestervin could be a Microcin immunity protein MCCF. This immunity protein that matched with Yestervin showed similar Z-score values with the Dali best fit Murein carboxypeptidase (33.3 vs 33.6 respectively) and higher RMSD values than the matched carboxypeptidase (2.3 vs 2.1 respectively). Both Dali protein matches also have two matching active site residues, with the Glu active site amino acid not being conserved in both cases. From these findings, it is plausible to consider an alternate function of Yestervin as an immunity protein in biological systems. The reason for these findings has to do with the similar functions between Microcin immunity proteins and Carboxypeptidases. Uniprot analysis of the Microcin immunity protein indicated it had molecular functions indicative of hydrolase activity, which is the same molecular function as a carboxypeptidase. The biological process of these Microcin immunity proteins is to support bacteriocin immunity. Bacteriocin immunity is the resistance to bacteriocins, which are polypeptide antibiotics that could harm bacteria. This bacteriocin immunity is similar to carboxypeptidases' role in the *ampC* mechanism, which is a mechanism for antibiotic resistance in bacteria. The role of carboxypeptidases in antibiotic resistance is further discussed in future studies. Both the Microcin immunity protein and carboxypeptidases have similar hydrolase molecular functions, and both play a role in antibiotic resistance in bacteria. These similarities between the two proteins can explain why Yestervin had results that suggested the protein may be a Microcin immunity protein. The data from these experiments could be refined on a multitude of different fronts. The data obtained from the Pyrx Autodocking lab used the protein 5Z01 which was a known protein with the function of a Murein Tetrapeptide carboxypeptidase. This caused the amino acid residues within the active site to change when looking at ligand binding through Autodock (Glu residue changed to Asn). A better refinement of the data would be using Yestervin within the Pyrx program to show how the ligands interact with the Ser-His-Glu catalytic triad instead of looking at how the ligands using Yestervin within the Pyrx program to show how the ligands interact with the Ser-His-Glu catalytic triad instead of looking at how the ligands interact with a similar Ser-His-Asn catalytic triad. Another refinement of the data would be using Yestervin within the purification and protein kinetic studies. A limitation of this study was the unavailability of Yestervin within solution that could be purified. Instead, the protein yfkn was used, which is a different EC class compared to Yestervin and has a completely different structure when compared to Yestervin. The protein yfkn has an EC class of 3.1, which means it is an esterase, while Yestervin has an EC class of 3.4, which is the class of peptidases. The use of yfkn was out of availability and thus the results were different than if Yestervin was purified. Yestervin contains a different kDa and pI value than yfkn and may have reacted differently than yfkn acted within the protein kinetic assay. The lacking availability of Yestervin and the use of the different proteins 5Z01 and yfkn in the Autodocking, protein purification, and protein kinetics experiments is a glaring limitation of this study. In future studies, the Pyrx program will be run using Yestervin and Yestervin will be purified, and the kinetics of the protein will be analyzed to provide further evidence that supports or refutes the hypothesis that Yestervin

is an LD-Carboxypeptidase. This future study would be performed using all the same methods as previously described within this report, but the protein Yestervin would be used in all experiments instead of substitute proteins like 5Z01 and yfkn. The expected results would be that the protein Yestervin is indeed a LD-Carboxypeptidase in all experiments. Another future study revolves around the ligand TLA, which was implicated to act as an inhibitor to Yestervin's active site. TLA will be further studied and its function within the active site will be determined. This study would be performed using an assay with TLA present in the assay. The TLA would bind to the Yestervin protein and the activity of Yestervin would be analyzed in a similar fashion seen in the protein kinetics experiment. A positive control would be used to ensure that the system was not the issue. The expected results would be that TLA is indeed an inhibitor because its structure with two carboxyl groups indicate it cannot partake in the LD-Carboxypeptidase proposed mechanism. If TLA was found to be an inhibitor from these experiments, more studies would need to be performed looking into the application of TLA during Pseudomonas aeruginosa infections. The LD-Carboxypeptidases within Pseudomonas aeruginosa act to recycle peptidoglycan. If TLA is implicated to be an inhibitor of Yestervin, future studies could be centered around TLA's role in medical intervention fighting Pseudomonas aeruginosa infections by stopping peptidoglycan recycling leaving the bacteria vulnerable and weakened. The ligand DMU should also be investigated in future studies to see if it acts as a substrate for the Yestervin active site. This would be performed in a similar fashion to the future study regarding the ligand TLA. There are no current expected results, as DMU is a rare ligand within hydrolases and its function within the active site is currently unknown. This unknown interaction between DMU and Yestervin active site amino acid residues is the reason that this future study should be performed, as there is a gap in knowledge that needs to be filled. Another future study should focus on the potential of pharmaceutical manipulation of Yestervin activity. The importance of understanding how to manipulate Yestervin is large as it has implications for fighting antibiotic resistant bacteria. Carboxypeptidases within bacteria play a role in antibiotic resistance, as their presence leads to decreased expression of ampC. The expression of ampC is an important mechanism involved in developing antibiotic resistance in bacteria. Therefore, the presence of a carboxypeptidase could lead to a mechanism of using carboxypeptidases to combat drug resistance in bacteria (Xu et al. 2020). The hypothesis that Yestervin is an LD-Carboxypeptidase has been supported by multiple bioinformatic tools and previous studies, meaning that Yestervin can be further studied and possibly targeted for medical interventions that could save thousands of lives from in hospital antibiotic resistant infections that often end fatally. Antibiotic resistance bacteria are an impending global health crisis that could cause major problems in the future. Studying Yestervin, its role within antibiotic resistance bacteria, and how it can be manipulated to weaken antibiotic resistant bacteria is necessary research that could be the beginning of solving the major issue of antibiotic resistant bacteria.