

Wilfrid Laurier University

Scholars Commons @ Laurier

Theses and Dissertations (Comprehensive)

2024

Structural and Functional Analysis of the Putatively Virulent Gene Products Tde1414, Tde1511, and Tde2714 from the Oral Pathogen *Treponema denticola*

Brock W. Corbett

Wilfrid Laurier University, corb6680@mylaurier.ca

Follow this and additional works at: <https://scholars.wlu.ca/etd>



Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

Recommended Citation

Corbett, Brock W., "Structural and Functional Analysis of the Putatively Virulent Gene Products Tde1414, Tde1511, and Tde2714 from the Oral Pathogen *Treponema denticola*" (2024). *Theses and Dissertations (Comprehensive)*. 2605.

<https://scholars.wlu.ca/etd/2605>

This Thesis is brought to you for free and open access by Scholars Commons @ Laurier. It has been accepted for inclusion in Theses and Dissertations (Comprehensive) by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact scholarscommons@wlu.ca.

**Structural and Functional Analysis of the Putatively
Virulent Gene Products Tde1414, Tde1511, and Tde2714
from the Oral Pathogen *Treponema denticola***

By

Brock William Corbett

B.Sc. Biochemistry, University of Guelph, 2019

THESIS

Submitted to the Department of Chemistry and Biochemistry

in partial fulfilment of the requirements for

Master of Science in Chemistry

Wilfrid Laurier University

© 2023 Brock William Corbett

Abstract

Periodontitis is a destructive polymicrobial disease of the teeth and gums characterized by dental plaque biofilms on the tooth surface, severe inflammation of surrounding gum tissues, and the resorption of alveolar bone. Discussed as one of the most prevalent conditions plaguing humankind, increased clinical parameters of periodontitis is associated with the increased presence of three bacteria known as the “red complex”: *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*. Various transcriptomic and proteomic studies of *T. denticola* when incorporated into mature biofilms have illuminated key gene targets the bacterium may use as virulence factors to aid in pathogenesis. Of the identified gene targets, *Tde1414*, *Tde1511*, and *Tde2714* were further investigated on their roles to promote infection. *Tde1414* was annotated as a putative phosphonopyruvate decarboxylase. The first reported structural model of a phosphonopyruvate decarboxylase has been successfully solved to a resolution of 1.95 Å, bound to its cofactors thiamine diphosphate and a magnesium cation. The protein target *Tde1511* was annotated as a putative pathogen-specific surface antigen, belonging to a family of periplasmic iron-binding proteins. The structures of apo- (PDB 87TK), copper-reconstituted (PDB 87TL), and copper/iron-reconstituted (PDB 87TM) structures of recombinant *Tde1511* have been solved to 1.78 Å, 1.89 Å, and 1.82 Å, respectively. Using inductively coupled plasma optical emission spectrometry and differential scanning fluorimetry analyses, copper and iron have been shown to be the preferred metals this protein interacts with when given the choice between several biological divalent transition metals. A novel and significant iron import operon in this bacterium has also been identified and proposed. The protein target *Tde2714* was annotated as a putative sulfatase-modifying factor enzyme domain-containing protein. Current peptide TLC results point to the function of *Tde2714* as a putative clostripain-like protease. Structurally, a solution has

been found excluding the putative catalytic domain. Understanding the biological processes in which these proteins function will unveil a deeper mechanistic understanding to infection. Our characterization has provided insight into what we have hypothesized to be important biological contributions that may contribute to keystone biological processes for *T. denticola* and as such, may provide details to exploit for therapeutic design.

Acknowledgements

Firstly, I would like to thank my supervisor, Dr. Michael Suits. Thank you for taking a chance on me to enter your lab, for patiently coaching me throughout all the big milestones, and for being there as a supervisor, as a financial mentor, and as a friend. I looked up to you immensely and still do to this day. I believe we get along so well, which made the late-night crystal collections go by even better. I will never forget the accomplishments we were able to achieve together and wish nothing but the best for the Suits' lab going forward. I am also thankful for my TAC committee members Dr. Lillian DeBruin and Dr. Joel Weadge. Thank you for challenging me during my presentations to make me the best I can be. You always pushed me a little bit further every time, but I wouldn't have found the success I have today without the help of both of you. To couple this, I extend a warm appreciation to the funding agencies supporting the lab which allowed me to excel and discover major findings along the way. I also thank the family of William Nikolaus Martin for the science scholarship I was selected for. That allowed me to dedicate even more time in the lab to pursue what I am passionate about.

Secondly, Brittany Nagy, I appreciate you to no ends. You and I had so many long days in that lab together, long nights of crystal collecting, and shared so many stories to make the purification times fly by. We aligned so much in so many different aspects, I wouldn't wish for anyone else to be my partner in crime while completing this degree. I wish you all the best my heart can offer as you tackle that PhD. You got this girl.

Thirdly, Joe Meissner. As a boss and as a friend, you've been one of my favourite people to meet along my journey at Laurier. You've allowed me to grow as a leader and have taught me a handful of tips along the way on interpersonal skills and communicating with others that you

may not have realized. I looked up to you in your ability to deliver information in a fun way but could also switch to supervisor mode when needed which I find is a delicate balance to execute that you do so effortlessly.

Lastly, my fiancé, Ally. Thank you for pushing me to be my very best through the ups and downs of this process. You deserve nothing but the world. I appreciate your patience with me through all the long nights of marking, typing, replying to students, and late nights at the lab. Your sacrifices to allow me to live out my dream and passion will not be forgotten.

Declaration

I declare that this thesis and the data discussed within are the direct result of my own work. I hereby declare that I, Brock William Corbett, am the sole author and composer of this paper. The results contained in this thesis have not been submitted in whole or in part to another institution for any degree requirements.

Table of Contents

Abstract.....	i
Acknowledgements	iii
Declaration.....	v
Table of Contents	vi
List of Tables.....	ix
List of Figures.....	x
List of Abbreviations.....	xiii
Chapter 1. Introduction.....	1
1.1 The Oral Cavity and Dental Plaque Biofilms	1
1.2 Periodontal disease and Periodontitis	3
<i>1.2.1 Disease Symptoms and Prognosis</i>	<i>3</i>
<i>1.2.2 Initial Detection, Prevalence, and Societal/Economic Burdens.....</i>	<i>5</i>
<i>1.2.3 Current Treatment Options for Periodontal Disease</i>	<i>7</i>
<i>1.2.4 “Cluster Theory” and the Red Complex Bacteria</i>	<i>8</i>
1.3 Treponema denticola.....	10
<i>1.3.1 Characteristics, Species Evolution, and Genetic Analysis.....</i>	<i>10</i>
<i>1.3.2 Use of Virulence Factors for Periodontal Infection.....</i>	<i>12</i>
1.4 Phosphonates	14
<i>1.4.1 Characterization and Biosynthetic Pathway.....</i>	<i>14</i>
<i>1.4.2 Mechanism of Thiamine Diphosphate Catalysis in Decarboxylation Reactions</i>	<i>16</i>
1.5 Iron Acquisition.....	18
<i>1.5.1 Significance of Iron and Bacterial Metal Homeostasis.....</i>	<i>18</i>
<i>1.5.2 Using Periplasmic Surface Antigens for Iron Import</i>	<i>18</i>
<i>1.5.3 Tde1511 – High affinity ferrous transport lipoprotein</i>	<i>19</i>
1.6 Proteases	20
<i>1.6.1 Function, Classifications, and Biofilm Remodelling Enzymes.....</i>	<i>20</i>
<i>1.6.2 Cysteine Protease Mechanism</i>	<i>21</i>
<i>1.6.3 Tde2714 – a putative Clostripain-like enzyme.....</i>	<i>22</i>
1.7 Purpose of Study	24
Chapter 2. Materials and Methodology	26

2.1	Materials	26
2.1.1	<i>Bacterial Transformations and DNA Miniprep</i>	26
2.1.2	<i>Bacterial Growth and Recombinant Expression</i>	27
2.1.3	<i>Cell Lysis</i>	27
2.1.4	<i>Immobilized Metal Affinity Chromatography.....</i>	28
2.1.5	<i>Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)</i>	28
2.1.6	<i>Dialysis</i>	28
2.1.7	<i>Anion Exchange Chromatography, Dynamic Light Scattering, and Protein Concentration</i> 29	
2.1.8	<i>Protein Crystallization, Crystal Optimization, and X-Ray Diffraction Analysis</i>	29
2.1.9	<i>Inductively Coupled Plasma Optical Emissions Spectroscopy (ICP-OES)</i>	30
2.1.10	<i>Differential Scanning Fluorimetry (DSF)</i>	31
2.1.11	<i>Peptide Thin Layer Chromatography</i>	31
2.1.12	<i>Proteolytic Degradation Assay</i>	32
2.2	Methodology	32
2.2.1	<i>DNA miniprep, Bacterial Transformations, and Recombinant Protein Expression</i>	32
2.2.2	<i>Cell Lysis</i>	34
2.2.3	<i>Immobilized Metal Affinity Chromatography.....</i>	35
2.2.4	<i>SDS-PAGE and Dialysis.....</i>	36
2.2.5	<i>Anion Exchange Chromatography and Dynamic Light Scattering.....</i>	37
2.2.6	<i>Protein Concentration and Initial Protein Crystal Screening.....</i>	38
2.2.7	<i>Crystal Optimization, X-Ray Diffraction Analysis, and Data Manipulation</i>	39
2.2.8	<i>Diffraction Data Manipulation and Protein Data Bank Submissions</i>	41
2.2.9	<i>Inductively coupled plasma optical emission spectrometry</i>	41
2.2.10	<i>Differential scanning fluorimetry.....</i>	43
2.2.11	<i>Peptide Thin Layer Chromatography</i>	44
2.2.12	<i>Proteolytic Degradation Assay</i>	46
Chapter 3.	Results	46
3.1	Bioinformatics	46
3.1.1	<i>Predicted Biophysical Properties.....</i>	46
3.1.2	<i>Predicted Protein Structures</i>	50
3.1.3	<i>Predicted Function of Tde2714.....</i>	51
3.1.4	<i>Multiple Sequence Alignments</i>	55

3.1.5	<i>Identification of a putative Iron Import System</i>	59
3.2	Protein Expression and Purification	61
3.2.1	<i>Transformations and Induction Optimization</i>	61
3.2.2	<i>Primary Purification through Ni-NTA IMAC</i>	62
3.2.3	<i>Anion Exchange and DLS</i>	63
3.3	Crystallization	65
3.3.1	<i>Crystal Screens</i>	65
3.3.2	<i>Crystal Optimization</i>	67
3.4	X-Ray Diffraction Analysis and Structural Solutions	71
3.4.1	<i>Initial Data Quality</i>	71
3.4.2	<i>Molecular Replacement and Structural Solutions</i>	73
3.5	Functional Assays	81
3.5.1	<i>ICP-OES with Tde1511</i>	81
3.5.2	<i>DSF of Tde1511</i>	82
3.5.3	<i>Peptide TLC Analysis of Tde2714</i>	83
3.5.4	<i>Proteolytic Degradation Assay with Tde2714</i>	85
Chapter 4	Discussion	88
4.1	Analysis of Tde1414	88
4.1.1	<i>Tde1414 and the Family of Thiamine Diphosphate Requiring Enzymes</i>	88
4.1.2	<i>Bioinformatics</i>	90
4.1.3	<i>Structural Characterization of Tde1414</i>	91
4.2	Analysis of Tde1511	94
4.2.1	<i>Bioinformatics</i>	94
4.2.2	<i>Structural Characterization</i>	98
4.2.3	<i>Functional Properties in Metal Binding</i>	101
4.3	Analysis of Tde2714	103
4.3.1	<i>Bioinformatics</i>	103
4.3.2	<i>Structural Characterization</i>	105
4.3.3	<i>Functional Properties in Proteolysis</i>	106
Chapter 5	Conclusion	108
References	112

List of Tables

Table 1. Predicted structural, functional, and biochemical properties of the gene products Tde1414, Tde1511, and Tde2714 from <i>T. denticola</i>	47
Table 2. Bioinformatic analysis of the gene products predicted to be in an iron import operon with Tde1511 from the pathogen <i>T. denticola</i>	60
Table 3. Scaled crystallographic data statistics for the structural analysis of Tde1414 from the pathogen <i>T. denticola</i>	72
Table 4. Scaled crystallographic data statistics for the structural analysis of apo, copper reconstituted, and copper/iron reconstituted structures of Tde1511 from the pathogen <i>T. denticola</i>	73
Table 5. Refinement statistics for the structures of Tde1511 submitted to the Protein Data Bank.	79
Table 6. DSF analysis of the shift in melting temperature after incubating EDTA-treated and copper loaded Tde1511 with various biological divalent metal cations..	83

List of Figures

Figure 1. Biofilm containing a co-culture of the red complex bacteria.	2
Figure 2. The progression of periodontal disease from healthy teeth and gums to advanced periodontitis.	4
Figure 3. Cluster theory for the progression of periodontal disease.	8
Figure 4. Illustration of the morphology of the oral spirochete <i>T. denticola</i>	11
Figure 5. Phosphonate biosynthetic pathway present in <i>T. denticola</i>	15
Figure 6. Chemical structure of thiamine diphosphate.	16
Figure 7. Proposed reaction mechanism of the ThDP-dependent decarboxylation of phosphonopyruvate.	17
Figure 8. Proteolytic mechanism of typical cysteine proteases.	22
Figure 9. Predicted protein structures of Tde1414, Tde1511, and Tde2714 from the pathogen <i>T. denticola</i> using AlphaFold2.	51
Figure 10. Multiple sequence alignment of known FGEs with Tde2714.	52
Figure 11. Deep learning model DeepFRI for predictive functional annotations on Tde2714 from <i>T. denticola</i>	53
Figure 12. I-TASSER suite of programs for predictive functional annotations for Tde2714 from the pathogen <i>T. denticola</i>	54
Figure 13. Multiple sequence alignment between other phosphonopyruvate decarboxylases and Tde1414 from the pathogen <i>T. denticola</i>	55
Figure 14. Multiple sequence alignment between other thiamine diphosphate-dependent enzymes and Tde1414 from the pathogen <i>T. denticola</i>	56
Figure 15. Multiple sequence alignment between characterized proteins within the COG3470 family and Tde1511 from the pathogen <i>T. denticola</i>	57

Figure 16. Multiple sequence alignment between Tde2714 and clostripain shows a conserved His/Cys dyad.....	58
Figure 17. Multiple sequence alignment between putative clostripain-like enzymes and Tde2714 from the pathogen <i>T. denticola</i>	58
Figure 18. Induction parameter optimization for the gene target Tde1511 from the pathogen <i>T. denticola</i>	62
Figure 19. SDS-PAGE analysis of the primary IMAC purifications for Tde1414, Tde1511, and Tde2714 from the pathogen <i>T. denticola</i>	63
Figure 20. SDS-PAGE analysis of the secondary anion exchange chromatography purifications for Tde1414, Tde1511, and Tde2714 from the pathogen <i>T. denticola</i>	65
Figure 21. Initial sparse matrix crystallization of Tde1414 from the pathogen <i>T. denticola</i>	66
Figure 22. Optimized protein crystallization of Tde1414 from the pathogen <i>T. denticola</i>	68
Figure 23. Protein crystal optimization of the protein target Tde1511 from the pathogen <i>T. denticola</i>	69
Figure 24. Using crystal seeding techniques to promote effective crystallization of Tde2714 from the pathogen <i>T. denticola</i>	70
Figure 25. Novel structural determination of Tde1414, a phosphonopyruvate decarboxylase, at a resolution of 1.95 Å from the pathogen <i>T. denticola</i>	75
Figure 26. Thiamine diphosphate binding pocket in Tde1414.....	75
Figure 27. Structural determination of Tde1511, a high-affinity ferrous iron binding/transport protein from the pathogen <i>T. denticola</i>	77
Figure 28. Topological map depicting the 2D structure of recombinant Tde1511 from the pathogen <i>T. denticola</i>	77
Figure 29. Coordination of copper and iron in the structure of Tde1511 from the pathogen <i>T. denticola</i>	79
Figure 30. ICP-OES analysis of the ability for Tde1511 to bind various biological divalent metal ions. .	82

Figure 31. Peptide modification reactions using Tde2714 with peptides containing FGE recognition sequences and a negative control analyzed through TLC.	84
Figure 32. TLC analysis of the reaction of Tde2714 with a peptide containing the SxAxR FGE recognition sequence in the presence of increasing concentrations of a protease inhibitor cocktail tablet.	85
Figure 33. Day 0: SDS-PAGE analysis of the proteolytic degradation using Tde2714 with various protein substrates in the presence and absence of PIC.	87
Figure 34. Day 7: SDS-PAGE analysis of the proteolytic degradation using Tde2714 with various protein substrates in the presence and absence of PIC.	86
Figure 35. Day 14: SDS-PAGE analysis of the proteolytic degradation using Tde2714 with various protein substrates in the presence and absence of PIC.	88
Figure 36. Phylogenetic tree comparing the sequence divergence between thiamine diphosphate dependent enzymes.	89
Figure 37. Proposed iron import operonic system, HftABCD, identified in <i>T. denticola</i>	97

List of Abbreviations

ABC transporter	ATP Binding Cassette transporter
AEP	2-aminoethylphosphonate
AEPT	2-aminoethylphosphonate-specific aminotransferase
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BAEE	N-benzoyl-L-arginine ethyl ester
βME	beta-mercaptoethanol
Bis-Tris	2-hydroxyethylamino-trisaminomethane
BLASTp	Basic Local Alignment Search Tool for proteins
CHESS	Cornell High Energy Synchrotron Source
CLS	Canadian Light Source
COG	Clusters of Orthologous Groups
CPI	Community Periodontal Index
C-type	Calcium dependent-type
Da	Dalton
DGR	Diversity-Generating Retroelement
DLS	Dynamic Light Scattering

DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic acid
EPS	Extra Polymeric Substance
FeS	Iron-sulphur cluster
FGE	Formylglycine-generating enzyme
FMN	Flavin mono nucleotide
FPLC	Fast Protein Liquid Chromatography
GBD	Global Burden of Disease
GPNA	N-glutamyl-L-phenylalanine p-nitroaniline
GRAVY	Grand average of hydrophilicity
Hft	High affinity ferrous transport
HSAB theory	Hard soft acid base theory
ICP-OES	Inductively coupled plasma optical emission spectrometry
Ig-like	Immunoglobulin-like
IMAC	Immobilized metal affinity chromatography

IPTG	β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
kDa	kiloDalton
LB	lysogeny broth
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LOD	Limit of Detection
LOQ	Limit of Quantification
MSA	Multiple Sequence Alignment
MWCO	Molecular weight cut-off
Ni-NTA	Nickel-nitrilotriacetic acid
OD ₆₀₀	Optical density at 600 nanometers
OFeT	oxidase-dependent ferrous transport
OMV	Outer membrane vesicle
P-Cho	phosphocholine
PEA	phosphoethanolamine
<i>P. gingivalis</i>	<i>Porphyromonas gingivalis</i>
PCR	polymerase chain reaction
PDB	Protein Data Bank

PEG	polyethylene glycol
PEP	phosphoenolpyruvate
pH	power of Hydrogen
pI	isoelectric point
PIC	protease inhibitor cocktail
PnAA	phosphonoacetaldehyde
PntC	phosphonyl tailoring cytidyltransferase
Pnpy	phosphonopyruvate
PP	pyrophosphate
Ppd	phosphonopyruvate decarboxylase
Ppm	phosphoenolpyruvate mutase
PSI	pounds per square inch
Pyr	pyrimidine
SAM	S-adenosyl methionine
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SeMet	selenomethionine
<i>T. denticola</i>	<i>Treponema denticola</i>

<i>T. forsythia</i>	<i>Tannerella forsythia</i>
TEMED	tetramethylethylenediamine
ThDP	thiamine diphosphate (thiamine pyrophosphate)
TLC	thin layer chromatography
TM	transmembrane
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
TvpA	Treponema variable protein A
USD	United States Dollar
WHO	World Health Organization
XRF	X-Ray Fluorescence

Chapter 1. Introduction

1.1 The Oral Cavity and Dental Plaque Biofilms

The human oral cavity is home to a microflora of unique bacteria that work interdependently or frequently symbiotically to prevail within their host ¹. The sequencing of 16S rDNA found in dental plaque samples have illuminated the presence of more than 600 species in the human oral cavity ^{1,2}. With adequate oral hygiene in a healthy host, non-threatening bacterial species are able to thrive and function cohesively, while the concentration of pathogenic bacteria remain at benign levels. To flourish efficiently in the oral cavity, bacteria undergo the process of biofilm formation, which results in the generation and adherence of dental plaques to the teeth and gum surface ^{2,3}.

Biofilms are the result of bacteria secreting a self-produced matrix known as extracellular polymeric substances (EPS) ^{3,4}. The EPS is typically comprised of various carbohydrates, lipids, proteins, and DNA, and the ratio of each is influenced by the species of bacteria producing the biofilm ^{3,5}. These components are what give the biofilm its mechanical stability, while also mediating adhesion to the target surface ⁵. As seen in Figure 1, biofilm production allows for intimate contacts to be made between the bacteria participating in the biofilm, as well as with surrounding materials comprising the matrix. Biofilm secretion and maturation typically occurs in four summarized steps: it initiates with the reversible attachment to a surface and aggregation of bacteria, then microcolonies are able to form with the secretion of EPS, the biofilm is then allowed to grow and mature with the introduction of various bacteria who also contribute to the EPS production, and finalizes in the dispersion and release of bacteria to colonize new sites distal to the initial biofilm ³⁻⁵.

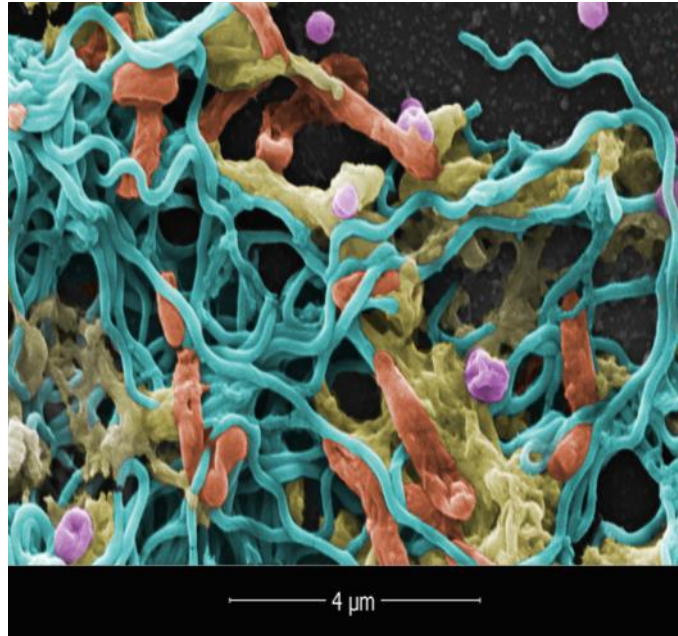


Figure 1. Biofilm containing a co-culture of the red complex bacteria. Showcased above is a microscopy image of the red complex bacteria embedded within a biofilm matrix. *T. denticola* (cyan), *P. gingivalis* (purple), and *Tannerella forsythia* (orange) can be seen in intimate contact with each other, as well as with EPS (yellow) comprising the biofilm. Microscopy image was taken as 24, 000x magnification and falsely coloured. Photo was recreated with permission and authored by Chris Bartlett, Thomas Brenner, Joel Weadge, and Michael Suits.

Due to the range of environmental conditions the oral cavity can experience through everyday living (varying levels of available oxygen, water, salt, and exposure to acids), the production of biofilms allow bacteria to thrive in various surroundings ⁴. This energetically expensive process helps to retain water and nutrients for bacterial growth and can provide protection against toxic metals and antimicrobial compounds in the environment ⁴. Extracellular enzymes embedded within the matrix may also serve as an external digestion system to extract nutrients out of the environment ⁵. Polysaccharides within the biofilm may form mesh-like structures which help to impede or inhibit the entry of antimicrobial compounds or host immune cells, protecting the bacteria encapsulated in the matrix ³. An additional benefit to biofilm formation is the increase in intimate contacts and interactions experienced by the bacteria; this

facilitates the sharing of nutrients and metabolites, quorum sensing, and horizontal gene transfer⁵.

Mechanical brushing, flossing, and professional dental cleanings are excellent tools to promote healthy teeth and gums due to the removal of these dental plaque biofilms^{1,6}. When the dental plaques are not removed and the biofilm is allowed to mature, damaging consequences begin to arise due to the proliferation of pathogenic bacterial species. Inflammation and bleeding of the gums may be seen, dental carries can form and damage the structural integrity of the teeth, and most importantly, periodontal disease is allowed to initiate and progress where irreversible damage takes place^{1,7}. The significance, prevalence, disease progression, and current treatment options for periodontal disease were explored. The goal of this research was to provide foundational knowledge with respect to key proteins putatively at play in the progression of periodontal disease.

1.2 Periodontal disease and Periodontitis

1.2.1 Disease Symptoms and Prognosis

Periodontal disease is characterized by its inflammatory and degradative properties that encompasses the disorders of the tissue surrounding and supporting the tooth^{1,6,8}. Classified based on the severity level of the observed symptoms, periodontal disease ranges from mild gingivitis to severe advanced periodontitis as shown in Figure 2^{1,9}. When dental plaque on the tooth begins to build resulting from poor oral hygiene, gingivitis occurs, and the surrounding gum tissue becomes inflamed and bleeds easier when probed¹. Proper teeth cleaning at this time

can reverse and/or inhibit the progression of the disease with minimal to no irreversible damage incurred.

With no intervention, the biofilm matures, progressing the disease into periodontitis. At this stage there is increased inflammation, and crevices between the tooth root and gingiva begin to form ⁶. This is due to the destruction and resorption of alveolar bone that structurally supports the tooth and forms the tooth socket ^{1,6,8}. This process causes irreversible damage to the oral cavity and will continue if treatment does not occur swiftly. If corrective actions are not sought, the disease evolves into advanced periodontitis. In this phase, more inflammation occurs, increased alveolar bone destruction forms even deeper crevices between the tooth root and gingiva, and periodontal ligaments are destroyed ^{1,6-8}. Due to the significant and irreversible damage done to the tissues supporting the tooth, worst case scenarios result in the loss of teeth all together.

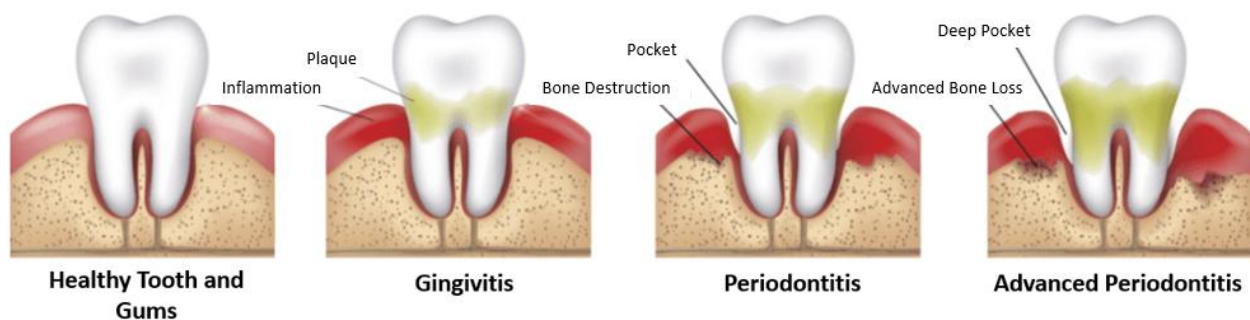


Figure 2. The progression of periodontal disease from healthy teeth and gums to advanced periodontitis. The image above highlights the hallmark signs of the progression of periodontal disease. Starting with a healthy tooth and gums, the tooth is clear of any debris and the supporting structures remain intact. When gingivitis occurs, dental plaque biofilms begin to deposit along the gum line and surrounding tissues become inflamed. Periodontitis is established when greater amounts of dental plaque biofilm lie on the tooth surface, surrounding tissues experience much more inflammation, and alveolar bone begins to be resorbed. During advanced periodontitis, significant rates of dental plaque biofilm deposition are seen, surrounding tissues are extremely inflamed, and supporting bone and ligament structures are deteriorated. This image was adapted from <https://www.istockphoto.com/illustrations/periodontal-disease>.

1.2.2 Initial Detection, Prevalence, and Societal/Economic Burdens

The Global Burden of Disease (GBD) stated that periodontitis was the sixth most prevalent condition across humankind from 1990-2010 ^{7,10}. As recently as 2022, the World Health Organization (WHO) estimated that 1.086 billion people across the world suffer from severe periodontitis ¹¹. Current knowledge of the extent and severity in which the global population is suffering from this disease is attributed to the initial research and findings by A.L. Russell in 1956 ¹². Due to the lack of data on periodontal disease in the 1950's, the disease was initially evaluated using the same criteria as dental caries ¹². With a diagnosis given only when blatant damage to the oral cavity was observed, Russell created the first quantitative scoring index for periodontal disease ^{8,12}. Although this attracted more awareness to the disease progression and negative consequences concerning periodontal disease, the index was basic in design; a score would increase from 0 to 8 (a higher score equating to more severe disease progression) based on the observed conditions a patient is experiencing ^{12,13}. Useful in its initial release, a more comprehensive index system was needed to more accurately diagnose a patient's condition to provide more effective treatment options.

In the late 1990s, the WHO began to support and endorse the use to the Community Periodontal Index (CPI) when evaluating the status of a patient's periodontal disease progression ^{8,14}. This allowed the limitations of Russell's scoring index to be surpassed by the direct measurement of gingiva pocket depth and tooth attachment loss by dental professionals ⁸. A CPI scoring of code 0 represents healthy gums with no observed issues. Code 1 is assigned when gingival bleeding is observed upon probing of the tissues. Code 2 constitutes an observation of calcified dental plaque and a greater degree of bleeding of the tissues surrounding the tooth. Code 3 is assigned when a gingival pocket depth of 4-6mm is measured while Code 4 is given

when a pocket depth greater than 6mm is recorded ^{8,15}. These updated guidelines to assess the global presence and severity of periodontal disease has highlighted the need for immediate action towards combatting and inhibiting the destructive development of this disease.

The GBD surveyed 17 countries in various continents for the presence of periodontal disease and found a 62% prevalence rate throughout the population ⁷. The trend in the number of positive diagnoses is on the rise, thought to be due to the growing total global population, as well as the number of elderly individuals ⁷. People aged 65 and older have shown a 79.3% prevalence of periodontitis ^{7,11}. In regions such as Germany and Quebec, adults between the ages of 35-44 have shown a prevalence of CPI code 4 periodontal disease in about 21% of the population ⁸. As a result, it has been characterized that periodontal disease is experienced by a large population of individuals across many age brackets and geographical locations.

An issue with treating dental diseases, such a periodontal disease, is the large cost burden placed on the individual who is seeking professional treatment. A trip to the dentist is not cheap to those without dental insurance, and this burden is additive between the direct and indirect costs involved. Direct costs are defined as the dollar value associated with professional dental treatment. Indirect costs are summarized as monetary consequences incurred by the individual due to situations such as time away from work. Using data from 2018 on the economic burden of periodontal disease in the United States and Europe, the requirement for immediate action to impede the progression of this disease can be acknowledged for more than just the physical ramifications that individuals experience. Direct costs of periodontitis in the United States have been estimated to total about \$3.49 billion USD while in Europe the total falls around \$2.72 billion USD ¹⁰. Astonishingly, the indirect costs of periodontitis in the United States and Europe were estimated to reach about \$150 billion USD and \$197 billion USD, respectively ¹⁰. Due to

the extensive costs that come with uninsured dental work and treatment, it is not surprising that low-income countries and families experience a higher risk for oral diseases ^{10,11}. Therefore, a cost-effective method for deterring the initiation of periodontal disease is required to support the large population of individuals at risk across the globe.

1.2.3 Current Treatment Options for Periodontal Disease

The number one line of defence to mitigate the onset of periodontal disease is consistent mechanical removal of dental plaque that forms on the tooth surface ⁶. The personal use of a toothbrush to brush all surfaces of the teeth for 2 minutes twice a day, and the use of interdental aids, such as dental floss, is sufficient to promote healthy tooth, gum, and oral health ^{1,6,16}. When consistent oral hygiene is unable to be maintained, the situation becomes more complex due to the polymicrobial infection that takes place to promote periodontal disease ^{1,17}. With no single causative agent responsible for the irreversible damage done to the oral cavity, aggressive and invasive therapies are required to remove the various pathogenic bacterial species at the site of infection and to remodel the damaged tissues and bone ^{6,18}.

With an early diagnosis, antibiotics can be administered to resolve the issue ⁶. To couple this, mouth washes containing the antimicrobial compound chlorhexidine are typically prescribed to decrease the levels of bacteria in the mouth and impede the rate of biofilm formation and maturation ¹⁶. When symptoms become too severe, professional intervention is required. Surgical excision of the infected tissue is typically required to prevent pathogenic bacteria from colonizing neighbouring tissues at the site of infection ^{16,18}. Once this objective is met, bone replacement grafts can be performed to rebuild lost connective tissues and support structures needed for tooth placement and stabilization ⁶.

As demonstrated, periodontal disease is a significant issue that negatively impacts the physical and financial lives of individuals worldwide. Consequences are felt no matter the age, race, or geographical location of the patient, with an emphasis on the elderly and low-income populations. Cost effective solutions are required with a high level of accessibility to all who seek them to combat the deleterious effects that accompany periodontal disease.

1.2.4 “Cluster Theory” and the Red Complex Bacteria

In 1998, an initiative led by Sigmund Socransky aimed to further understand the pathology of periodontal disease on a microbial level ¹⁹. Using dental plaque samples from patients at various stages of the disease, a cluster analysis revealed bacteria that coordinate into

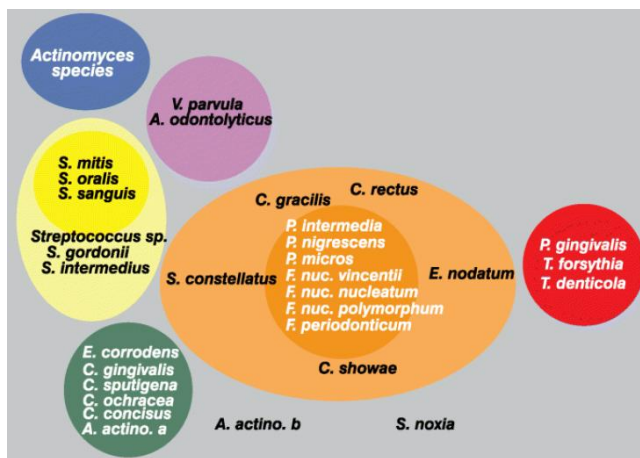


Figure 3. Cluster theory for the progression of periodontal disease. Highlighted above are the different microbial complexes that sequentially colonize the dental plaque biofilm to promote and progress periodontal disease. The red complex is composed of *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*. Photo adapted from Socransky et al., 1998 (ref. 19).

discrete consortia; the results of which have intriguing implications for the progression of periodontal disease ¹⁹. This technique established five conserved groups of bacterial species, shown in Figure 3, that can be used as biomarkers to access at what stage of the disease a patient may be experiencing ¹⁹. Using a colour-coded system, Socransky’s “cluster theory” identified that bacteria within the yellow, green, and purple complexes are responsible

for the initial colonization of the tooth surface and biofilm matrix production ^{17,19}. He then showed that bacteria within the orange complex colonize the dental plaque next to further the

maturation of the biofilm and promote more aggressive destruction of the surrounding tissues^{17,19}. Finally, it was found that bacteria within the red complex were late-stage colonizers of the mature biofilm and were rarely isolated without the presence of orange complex bacteria¹⁹. A striking association was noted between an increased presence of red complex bacteria to increased clinical parameters of periodontal disease¹⁹. Thus, identifying the mechanisms of infection used by the red complex may allow for professional intervention to mitigate further irreversible damage to the oral cavity.

Representing 68% of the bacterial species at sites of periodontitis, the red complex is comprised by *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*^{19,20}. Typically, isolated adjacent to the epithelial lining of deep periodontal pockets, the presence of these bacteria is associated with increasing gingiva pocket depths and an increase in bleeding of the gums and surrounding tissues¹⁷. This may be attributed to the significant number of proteolytic enzymes this complex is known to produce that deteriorate surrounding tissues and ligaments¹⁹. *T. forsythia* had been found to colonize the biofilm first out of the three and isolates of *P. gingivalis* almost always included the presence of *T. forsythia*¹⁹. It has also been established that *P. gingivalis* and *T. denticola* have synergistic properties when working together in both nutrient acquisition and disease progression^{17,21}. Thus, the ways in which these three bacteria thrive independently and how they complement one another during infectious states requires further exploration to gain a greater mechanistic understanding of disease progression. With this knowledge, *T. denticola* and its role in periodontal disease is further characterized. While independent characterization of this organism is important, in the context of polymicrobial coexistence in the oral microbiome, focus is placed on upregulated factors during dense co-cultures in periodontal-conferring timeframes within the oral cavity.

1.3 *Treponema denticola*

1.3.1 Characteristics, Species Evolution, and Genetic Analysis

Treponema denticola (*T. denticola*) is a Gram negative, anaerobic, asaccharolytic, highly motile spirochete^{17,22}. The phylum “spirochaetes” also includes members responsible for Lyme disease, syphilis, and a few borreliosis relapsing fevers²³. A unique feature of spirochetes is their long, corkscrew-like morphology as shown in Figure 4, coupled with the presence of periplasmic flagella¹⁷. These features allow these bacteria to be highly motile in their environments, aiding in nutrient acquisition and promoting infection in tissues distal to the initial site of infection¹⁷. *T. denticola* typically adheres to host tissues, such as tooth surfaces, and preferentially interacts with actively dividing cells in areas of inflammation and repair^{20,23}. It also interacts with biofilm matrix proteins and microorganisms within the biofilm, such as cell surface receptors on *P. gingivalis*^{20,23}.

Throughout the evolution of the treponeme species, genetic analysis has revealed a significant number of horizontal gene transfer events have taken place with other bacterial, eukaryotic, and archaeal species²⁴. Due to the extent of genetic diversity that treponemes possess, *T. denticola* has about 26% of its genome encoding proteins with no major similarities to previously sequenced proteins²². Treponemes exhibit a strong conservation of DNA

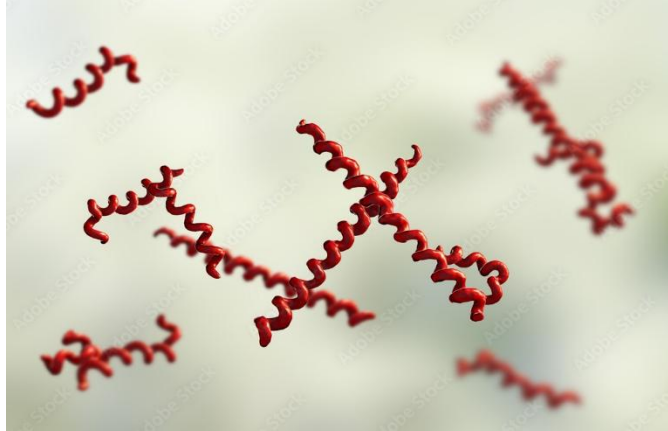


Figure 4. Illustration of the morphology of the oral spirochete *T. denticola*. The figure above showcases the morphology of the oral pathogen *T. denticola* (in red). Hailing from the phylum Spirochaetota, they are characterized as Gram negative, anaerobic bacteria with a long, helical, corkscrew-like morphology. Illustration was retrieved from <https://stock.adobe.com/ca/images/treponema-denticola-bacteria-3d-illustration-gram-negative-anaerobic-bacteria-that-live-in-oral-cavity-and-are-involved-in-periodontal-diseases/186445212>.

replication/repair genes and transcription factors to one another, but genetic variation is observed in genes used for functional activities that help an individual species thrive in their respective environments ²⁴. Specific to *T. denticola*, genome annotations have revealed the presence of eight ATP binding cassette (ABC)-type importers for peptides and amino acids and another eight ABC-type importers for iron ²⁵. The significant occurrence of these two import systems may be attributed to *T. denticola* requiring protein as a main source of energy (due to being asacchrolytic), while iron has been shown to increase growth promotion and *T. denticola* infection severity ^{20,25}. Furthermore, 83 proteins have been identified for drug efflux functions comprising 47 unique systems to accomplish this ²⁵. Finally, a significant number of proteolytic enzymes have been identified within *T. denticola*'s genome ^{17,25}. These enzymes have proposed functions that aid in biofilm remodelling and host tissue destruction, both of which may aid in host tissue penetration and/or enabling dispersion from the biofilm to colonize other host tissues ^{17,20,25}. It is evident this bacterium is equipped with an arsenal of tools to thrive and promote

infection in the oral cavity, but how these pieces work together in the pathology of periodontal disease requires more investigation.

1.3.2 Use of Virulence Factors for Periodontal Infection

Virulence factors are biochemical features used by pathogens to colonize, invade, and persist within their host^{26,27}. They can be used to undergo quick adaptive changes in response to their environment, promote adhesion, or used to evade host immune systems²⁶⁻²⁸. With the number of antibiotic resistant bacteria on the rise, orienting therapeutic design to inhibit critical virulence factors for infection may disarm the bacteria to give time for host immune systems to eliminate the pathogen^{28,29}. Two benefits to targeting virulence factors in bacterial diseases are that the number of potential drug targets becomes quite large, providing multiple avenues where a successful solution can be found³⁰. Secondly, virulence factors are typically strain-specific, resulting in the protection of host microflora during drug administration²⁸. With *T. denticola*'s genome sequenced to completion, identifying possible virulence factors and how they relate to infection becomes the next step to uncover thoughtful remedies to impede the effects of periodontal disease.

To elucidate potential protein targets for further analysis, a few studies have been evaluated to determine appropriate genes to examine. A comparative genome analysis using various spirochetes highlighted the presence of a nucleotide sequence for a putative phosphonopyruvate decarboxylase in *T. denticola* (Tde1414)²⁵. The protein target of this gene sequence is hypothesized to aid in the biosynthesis of phosphonates to alter the chemistry of lipids and glycans incorporated into the cell membrane or presented at the cell surface^{25,31}. It was also identified that Tde1413 (a phosphoenolpyruvate mutase) and Tde1415 (2-

aminoethylphosphonate-specific transaminase and a phosphonyl tailoring cytidylyltransferase) were found to be major proteins in *T. denticola* following a proteomic characterization ³². Due to the protein targets Tde1413 and Tde1415 hypothesized to work in concert with Tde1414, understanding how this bacterium uses phosphonates during infectious time frames is a relatively unexplored area. With little information in the literature on the structure and function of this enzyme, Tde1414 is a topic waiting to be explored. A study quantifying the enrichment and depletion of proteins in outer membrane vesicles (OMVs) from *T. denticola* was also evaluated ³³. OMVs are composed of a single membrane with the same lipid composition as the outer membrane from the bacterium in which they originated ^{33,34}. They contain outer membrane proteins embedded in their surface, while the lumen of the vesicle is enriched with cytoplasmic proteins ^{33,34}. The secretion of OMVs has been identified as a method of bacterial virulence and have been known to aid in cell adhesion and for the delivery of toxins to neighbouring cells and tissues ^{33,34}. Of the annotated gene targets, Tde1511, a putative pathogen-specific surface antigen, was identified to be significantly depleted from these vesicles ³³. Primary functions for this protein have been annotated as being involved in high affinity ferrous iron transport. Iron is required for *T. denticola* growth promotion and used as a cofactor in enzymes for structural stability or catalytic activity ^{35,36}. The significance of this protein and why this bacterium prefers to exclude it from OMVs will be investigated. Conversely, the same study identified lipoproteins containing formylglycine-generating enzyme (FGE)-sulphatase domains are significantly enriched in these OMVs ³³. Of the nine recorded FGE-sulphatase domain containing proteins in the *T. denticola* genome, Tde2714 was found to have the second highest enrichment score out of the seven found in OMVs ³³. To couple this, a genome wide transcriptomic analysis of *T. denticola* had been done while in a planktonic state versus when incorporated into a mature

biofilm²². This identified 126 genes in which expression increased by a fold change of 1.5 or greater, where Tde2714 was found to be one of these genes²². Why *T. denticola* chooses to highly enrich this protein onto the surface of OMVs and its function related to disease progression will be probed.

1.4 Phosphonates

1.4.1 Characterization and Biosynthetic Pathway

Phosphonates comprise a diverse class of compounds with a shared commonality of a C-P bond within the molecule³⁷. Considered secondary metabolites synthesized by some organisms (non-essential to life, but hold critical functions), characterized phosphonates found in nature typically exhibit herbicidal or antimicrobial properties³⁷. Structurally, the carbon-phosphorus bond has been shown to be resistant to biochemical, thermal, and photochemical hydrolysis³⁸. Hence, the inherent anti-degradative properties of these molecules highlight why about half of all bacterial phosphonate biosynthesis is thought to produce phosphonoglycans and phosphonolipids as cell wall constituents³⁹. With few exceptions, a majority of biological phosphonates arise from the isomerization of phosphoenolpyruvate (PEP) to phosphonopyruvate (Pnpy), although this reaction heavily favours the reverse direction as shown in Figure 5^{37,39}. To overcome this barrier, a phosphonopyruvate decarboxylase (Ppd) is employed that produces a second irreversible reaction to create phosphonoacetaldehyde (PnAA)³⁹.

Due to the revolution in the acquisition and analysis of sequence data available in the genomic era of biology, a significant number of new and unique phosphonate products have been identified with new enzymes that perform unprecedented reactions⁴⁰. Genomics data has also

allowed for the analysis of biosynthetic gene neighbourhoods to be conducted in which *T. denticola* was identified as containing the enzymatic machinery for Pep mutase (Ppm, *Tde1413*) and Ppd (*Tde1414*)³⁹. The gene *Tde1415* was also found, and its protein target has been characterized as an AEP transaminase that is fused to a phosphonyl tailoring cytidylyltransferase (PntC) domain, used to convert PnAA into an activated form of 2-aminoethylphosphonate (AEP)^{31,39}. Activated AEP may then be captured by nucleophilic carbohydrates or lipids to be integrated into the cell wall thereafter³¹. *T. denticola* was also found to have these three genes in a locus previously implied for the synthesis of capsular polysaccharide B in *Bacteroides fragilis*, which contains AEP in an ester linkage to the hydroxyl group of a carbohydrate³⁷.

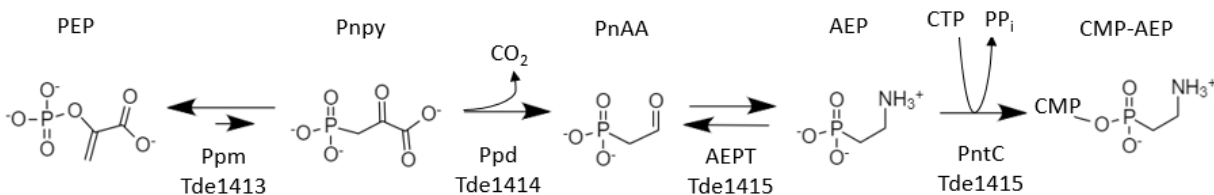


Figure 5. Phosphonate biosynthetic pathway present in *T. denticola*. This figure depicts the predicted chemical transformations that occur within *T. denticola* where the initial substrate phosphoenolpyruvate (PEP) is converted into the target product 2-aminoethylphosphonate (AEP). An isomerization of PEP into phosphonopyruvate (Pnpy) is facilitated through the action of phosphoenolpyruvate mutase (Ppm). Phosphonopyruvate decarboxylase (Ppd) then performs a decarboxylation on Pnpy to produce phosphonoacetaldehyde (PnAA). PnAA is acted on by an AEP-specific aminotransferase (AEPT) to produce 2-aminoethylphosphonate (AEP). Finally, a PntC activates AEP with the addition of CMP, forming CMP-AEP. This molecule is now primed for a nucleophilic attack by lipids or glycans^{31,39}. Image was generated using ChemDraw®

AEP is the phosphonate counterpart to phosphoethanolamine (PEA), a common lipid headgroup^{31,41}. It shows resistance to chemical, enzymatic, acid/base, and phosphotransferase-catalyzed hydrolysis^{41,42}. This molecule performs a majority of its functions at the cell surface, incorporated into surface exposed proteins, lipids, and/or carbohydrates⁴¹. Although the complete role of phosphonolipids/phosphonoglycans is currently unknown, current hypotheses in the literature speculate a role in cell-to-cell signalling, increasing cell membrane integrity in extreme environments, or as a phosphorus reserve when nutrients are limited for uptake³⁸. Thus,

a reasonable hypothesis can be created here that *T. denticola* may use AEP to modify its cell surface to become more resistant to various forms of hydrolysis during infectious timeframes. Experiments with *T. denticola* have showcased the presence of phosphonate products using solid-state NMR, while PntC has shown the preferential activation of AEP over phosphocholine (P-Cho) due to steric discrimination in the active site³¹. This raises questions on the evolutionary advantage this decision confers for the bacterium.

1.4.2 Mechanism of Thiamine Diphosphate Catalysis in Decarboxylation Reactions

ThDP, a vitamin B1 derivative, binds at the dimeric interface of enzymes anchored in place via coordinating interactions of the diphosphate moiety with a magnesium ion, as well as hydrogen bonds from neighbouring amino acids in the enzyme structure^{43,44}. The PP domain of enzymes interacts with the diphosphate portion of ThDP while the Pyr domain binds and interacts with the aminopyrimidine moiety of ThDP⁴⁵. Binding in its strained (“V”) conformation as shown in Figure 6, this orients the molecule to act as an electrophilic catalyst by promoting an intramolecular proton transfer within itself^{45,46}

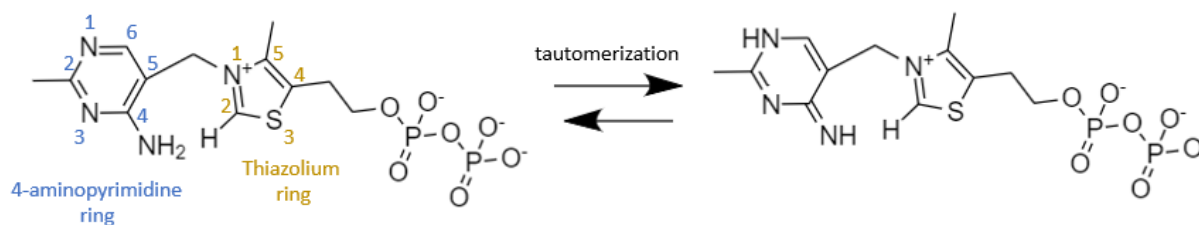


Figure 6. Chemical structure of thiamine diphosphate. The structure above shows the initial state of ThDP. The N1 nitrogen of the 4-aminopyrimidine ring can accept a hydrogen and the structure can tautomerize, forming an imino group at the C4 carbon. This allows for an intramolecular proton transfer from C2 on the thiazolium ring where the reactive ylide is formed⁴⁶. Structure was created using ChemDraw®.

Shown in Figure 7 is the proposed mechanism of ThDP-dependent Pnpy decarboxylation which follows similar reaction steps of already established ThDP-dependent decarboxylation mechanisms^{46–48}. The strained conformation allows for less than a 3.5 Å distance between the C4' imino group of the pyrimidine ring from the C2 hydrogen of the thiazolium ring as shown in the figure above^{45,48}. At this distance, the C2 hydrogen is transferred to the C4 imino group, forming a reactive ylide⁴⁷. The ylide is able to attack a carbonyl on the substrate to form a tetrahedral intermediate^{47,48}. An electrophilic substitution reaction occurs where the decarboxylation of the substrate results in the formation of an enamine product^{47,48}. Protonation of the enamine and elimination of the substrate reforms ThDP for another cycle of catalysis⁴⁸.

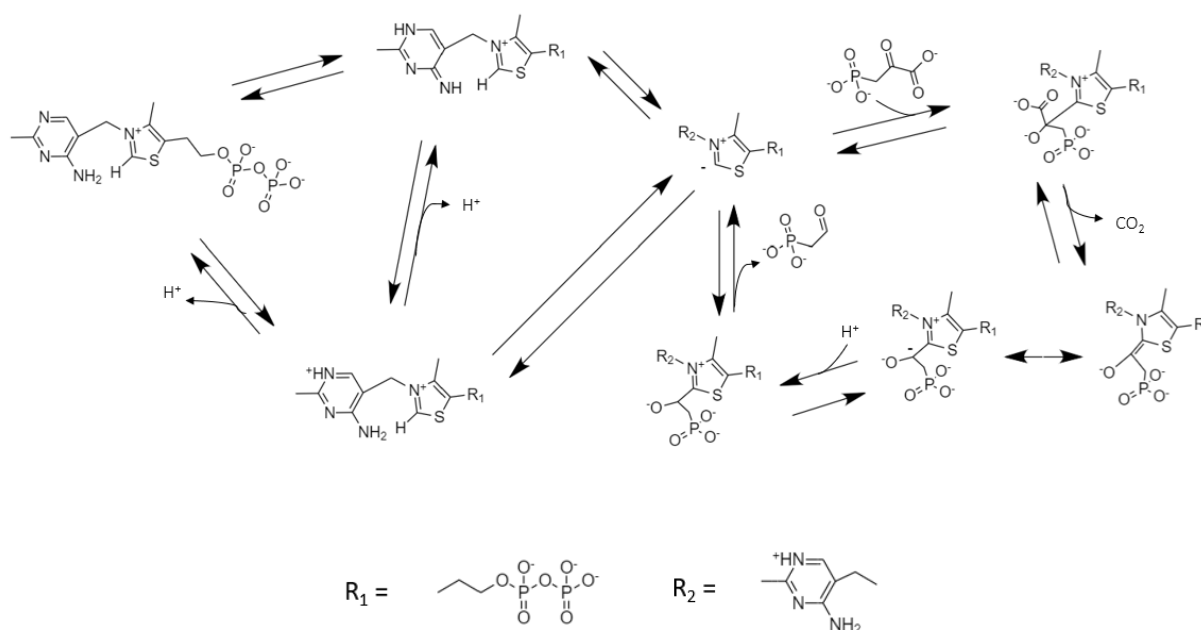


Figure 7. Proposed reaction mechanism of the ThDP-dependent decarboxylation of phosphonopyruvate. When bound and stabilized in its protein binding site, ThDP can tautomerize, allowing the 4'-imino group to be within proximity to abstract the C2 hydrogen of the thiazolium ring. This process creates an ylide, which attacks the carbonyl of phosphonopyruvate, forming a tetrahedral intermediate. Once decarboxylation occurs, the protonation of an enamine product results in the release of phosphonoacetaldehyde and regeneration of the ylide structure of ThDP. From here, ThDP is capable of regenerating to its initial structure, or initiate another cycle of catalysis^{46–48}.

1.5 Iron Acquisition

1.5.1 Significance of Iron and Bacterial Metal Homeostasis

Iron is one of the most highly abundant elements on Earth and has been harnessed by biology to exploit its two stable valence species. Before the great oxygenation event, reduced ferrous iron would be the predominant species found ⁴⁹. Introduction of oxygen into the environment caused ferrous iron to be oxidized and found primarily in its ferric form ^{49,50}. Transition metals found in nature are almost exclusively a part of proteins as cofactors, components of biomolecules, used as signalling molecules, or used in structural rearrangements in electron transfer chemistry ⁵¹. What makes ferrous iron significant outside of its productive applications in biology is its ability to initiate and propagate a reaction system that generates hydroxyl radicals through the Fenton reaction and the Haber-Weiss reaction ^{52,53}. These hydroxyl radicals promote damaging and deleterious effects to DNA, proteins, and lipids in the cell wall, resulting in strategies employed to protect cells from free iron in the environment ⁵⁴. In mammals, proteins, such as lactoferrin and ferritin, bind free ferric cations to mitigate the chances of the Fenton reaction from occurring ^{50,55}. In bacteria, different iron translocation systems have evolved to maintain intracellular metal concentrations to avoid levels of toxic accumulation.

1.5.2 Using Periplasmic Surface Antigens for Iron Import

In 1975 it was identified that surface antigens may not only be used to discriminate between differing cell types but may also employ critical functions to promote pathogenesis ⁵⁶. An example of this is the surface antigen PsaA from *Streptococcus pneumoniae*. On top of eliciting an immune response when detected within a host, PsaA had also been identified to have

metal binding properties, while also facilitating adhesion to type II pneumocyte cells ⁵⁷. *S. pneumoniae* infections using strains deficient in PsaA showed avirulence in mice models, highlighting the dual functionality these surface antigens may hold ⁵⁷.

With this knowledge, vaccine development can be targeted towards surface antigens that hold critical functions, while also being a common cell marker on all serotypes of a bacterium ⁵⁸. Luckily, treponeme species share a surface antigen with high sequence similarity that may also play a key role in pathogenesis. Classified in the Clusters of Orthologous Groups (COG) family number 3470, this group of proteins are annotated as uncharacterized proteins probably involved in high affinity ferrous iron transport ³⁶. Recently, a small handful of proteins within this group have been structurally and functionally characterized ^{35,36,59,60}. Identified as periplasmic pathogen-specific surface antigens, all characterized members have shown the requirement of copper binding as a cofactor to allow the binding and translocation of iron to occur ⁶⁰⁻⁶².

1.5.3 *Tde1511 – High affinity ferrous transport lipoprotein*

The protein product of the gene *Tde1511* identified in *T. denticola* has been bioinformatically annotated as a putative pathogen-specific surface antigen. Gene annotations show a periplasmic-targeting signal sequence from residues 1-20, and N-terminal acylation sites are predicted to tether this surface antigen to the cell wall within the periplasm ^{62,63}. As stated above, proteins in COG3470 show putative functions in iron transport and facilitate metal homeostasis within the cell. Characterized members of this family have shown functions in iron translocation as part of EfeUOB, FetMP, FeO, and FtrABCD iron import systems ^{50,54,59}. *Tde1511* has been suggested to be in an operon with other iron transport proteins and has been directly linked to critical iron sequestering when key proteases that interact with iron binding

proteins are knocked out ⁶⁴. Determining this protein's biochemical role and function will be attempted.

1.6 Proteases

1.6.1 *Function, Classifications, and Biofilm Remodelling Enzymes*

Proteases are enzymes that can hydrolyze peptide bonds within protein secondary or tertiary structures ⁶⁵. Their function most likely arose for protein catabolism to gain amino acids needed for the cell to grow ⁶⁵. Today, proteases have known functions that aid in apoptosis, inflammation, alter protein-protein interactions, and create bioactive molecules ^{65,66}. They are broadly divided into exopeptidases and endopeptidases. Exopeptidases (aminopeptidases and carboxypeptidases) cleave at the ends of a peptide, while endopeptidases cleave internal peptide bonds ^{65,67}. These categories can be further subdivided based on key mechanistic attributes: aspartic, glutamic, metallo-, cysteine, serine, and threonine proteases ^{65,67}. Some proteases have strict specificity for certain substrates, but many show non-specific cleavage outside of amino acids they cleave beside ^{4,65}

Extracellular proteases found within biofilm matrices may be used to remodel the local environment or secreted as a form of bacterial virulence ⁴. Proteases within biofilms have been known to cleave certain ECM proteins, such as E cadherin, collagen type IV, fibrinogen, and laminin ^{4,68}. They can disrupt epithelial barriers to help bacteria penetrate and invade host tissues, or cleave host cell surface proteins, exposing hidden receptors (cryptitopes) to bind to ^{4,68}. Therefore, the function of a putative clostripain-like protease in *T. denticola* will be explored and how it may contribute to the virulence of this bacterium.

1.6.2 Cysteine Protease Mechanism

The mechanism of cysteine proteases uses a cysteine thiolate ion as a nucleophile for the cleavage of peptide bonds as shown in Figure 8⁶⁶. A catalytic dyad of cysteine and histidine are often found in the active site of cysteine proteases⁴. The imidazolium of histidine abstracts the proton from the cysteine sulfhydryl, forming the nucleophilic thiolate ion^{4,66}. The thiolate ion attacks the carbonyl group of the peptide backbone to form a tetrahedral intermediate^{4,66}. The peptide accepts a hydrogen from the imidazolium, releasing the carboxyl terminus and leaving an acylated enzyme intermediate^{4,66}. A water molecule is then introduced across the peptide-enzyme intermediate. A hydrogen ion reincorporates with the thiolate ion while the hydroxyl ion is added to the peptide, forming a carboxylic acid, and prompting the release of the peptide from the enzyme^{4,66}. The result of this catalytic cycle regenerates the initial cysteine and histidine side chains, while cleaving a peptide bond in the process.

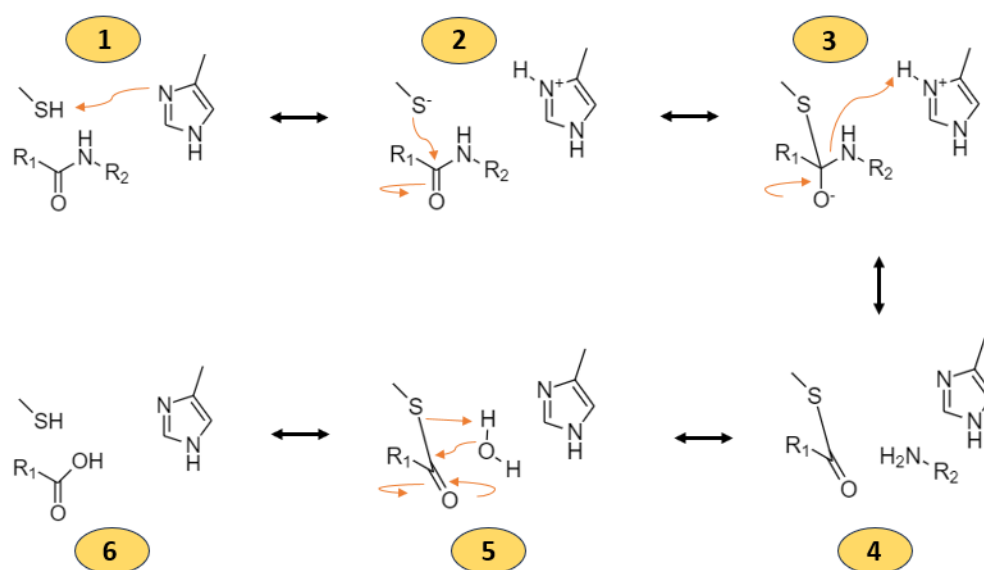


Figure 8. Proteolytic mechanism of typical cysteine proteases. The catalytic dyad of cysteine and histidine provide the necessary requirements for hydrolytic cleavage of peptide bonds to be achieved. Histidine abstracts a proton from the sulfhydryl group, forming the nucleophilic thiolate ion. The thiolate ion then attacks the carbonyl of a peptide backbone to form the tetrahedral intermediate. The peptide backbone accepts a hydrogen from histidine, cleaving the peptide bond. A water molecule is activated by the thiolate ion to allow a proton to associate with the thiolate ion while the hydroxyl group is incorporated with the other piece of the peptide backbone. This allows for all catalytic residues to be regenerated while the peptide dissociates from the enzyme ⁶⁶. Diagram was created using ChemDraw. ®

1.6.3 *Tde2714 – a putative Clostripain-like enzyme*

The protein target of the gene *Tde2714* identified in *T. denticola* has been bioinformatically annotated as a putative sulfatase-modifying factor enzyme domain-containing protein. Performing initial Basic Local Alignment Search Tool for proteins (BLASTp) searches provides overwhelming data on the function of this protein most likely being annotated as a formylglycine-generating enzyme (FGE). Another group of researchers then identified *Tde2714* along with six other “FGE domain containing proteins” to be highly enriched on the outer leaflet of OMVs secreted by *T. denticola* ³³. One of these identified “FGE domain containing proteins”, Tde2269, had been characterized eight years prior and found to be a diversity-generating retroelement (DGR) (Treponema variable protein A, TvpA) that could generate an extraordinary amount of sequence variation through a calcium-dependent (C)-type lectin protein scaffold ⁶⁹. The core components of a DGR are a reverse transcriptase (RT), a variable region of DNA (VR), and a template region of DNA (TR) ^{70–72}. The TR is composed of an invariant DNA sequence, which the RT transcribes into cDNA and places this in the VR locus ^{70,72}. The RT has been characterized to have poor accuracy when transcribing adenine nucleotides, resulting in adenine-directed mutagenesis, and altering adenine-containing codons ^{69,70,72}. The C-type lectin fold has shown to aid in the surface exposure of VR regions while maintaining a constant core scaffold,

allowing a significant number of DGRs to take advantage of this to promote the generation of novel ligand binding motifs at these surface exposed sites ^{69–72}.

The FGE fold is a subclass of the C-type lectin fold, which is present in Tde2714, highlighting why BLASTp searches mainly identified FGE-related functions ⁷³. Another group of researchers looked into the presence of DGR proteins in *T. denticola* as a method of creating sequence variation in surface exposed proteins to promote the binding of novel ligands, allowing the bacterium's incorporation into various biofilms ⁷¹. Of the seven previously identified FGE domain containing proteins, four had signatures of DGR activity, this did not include Tde2714 ⁷¹. Proteins containing C-type lectin folds typically exhibit calcium-dependent glycan binding using conserved glycan binding motifs, such as “EPN” and/or “WND” ⁷⁴. Both of these hallmark motifs for glycan binding are absent in Tde2714.

After much thought, the accumulated data on the function of Tde2714 points to it being a protease, more specifically, a clostripain-like protease. Clostripain (EC 3.4.22.8) is a cysteine endopeptidase with specificity for cleavage on the carboxyl side of arginine residues ^{75–77}. With a catalytic dyad of His/Cys, these proteases also contain a conserved serine and a conserved location for an aromatic residue to aid in stabilizing the peptide substrates in the binding pocket ⁷⁵. Secreted as an inactive prepro-enzyme, calcium is required to activate the protease, and the autocleavage of a linker chain results in an active proteolytic enzyme ^{75,77–79}. With evidence of this enzyme class being able to degrade human proteins as well as various ECM components, enriching this protease on the surface of OMVs may allow for local degradation of the ECM to occur, allowing the cargo within the OMV to be delivered to its destination with ease ^{80–82}

1.7 Purpose of Study

Despite the increase in oral hygiene knowledge and teachings, periodontitis is still one of the most prevalent conditions experienced across all humankind^{7,11}. The revelation of “Cluster Theory” allowed for connections to be made for the first time on specific groups of bacteria to direct symptoms of the disease as it increases in severity¹⁹. Identification of the red complex became significant due to the association of an increased presence of red complex members to increased clinical parameters of periodontitis¹⁹. On top of the overwhelming direct and indirect costs attributed to living with this disease, red complex members (mainly *T. denticola*) have been found to disperse from initial colonization sites in the mouth and have been found to be implicated in diseases regarding the brain, heart, and spleen^{10,83}. Nevertheless, periodontitis is still a polymicrobial infection that causes irreversible damage to periodontal structures in the oral cavity that require invasive professional procedures to rectify^{1,6}. With no single causative agent for this disease, creating meaningful therapeutic treatments becomes difficult to pinpoint the best perspective to approach solving this issue.

To better understand the mechanisms that promote periodontal disease, three virulence factors from *T. denticola* were structurally and functionally characterized. Tde1414 is a putative phosphonopyruvate decarboxylase that is hypothesized to help form phosphonolipids and -glycans that are inserted into the cell membrane to increase membrane integrity. Tde1511 is a putative pathogen-specific surface antigen that is hypothesized to be used for the uptake of iron for growth promotion and infection. Tde2714 is a putative clostripain-like protease located on the outside of OMVs and hypothesized to locally degrade surrounding ECM proteins to allow the vesicle to reach its destination more efficiently.

To achieve this, three experimental aims were followed:

- Experimental Aim I: to express and purify the targeted gene products in pET vectors using affinity and ion exchange chromatography, while optimizing solubility and induction parameters that promote the greatest protein yield.
- Experimental Aim II: protein crystallization in sparse matrix screens and crystal growth optimization using a la carte chemicals to use in X-ray diffraction analysis for structural determination.
- Experimental Aim III: to assess protein activity using biochemical and biophysical methods, such as:
 - IIIa. Analyzing potential metal-binding properties of Tde1511 using inductively coupled plasma optical emission spectroscopy. This will be coupled using differential scanning fluorimetry to assess structural stability in the presence of various biological divalent metal cations.
 - IIIb. Examining peptide modifications that Tde2714 may produce using thin layer chromatography to identify putative novel products.

The long-term goal of this research is to biochemically characterize proteins that *T. denticola* may use to progress periodontal disease and to create theories as to the mechanisms at play which facilitate this process. Illuminating different processes which increase the clinical severity of periodontitis may uncover potential targets for novel therapeutic design. These compounds or small molecules may be added to toothpastes and mouth washes to mitigate irreversible damage caused in patients experiencing periodontal disease.

Chapter 2. Materials and Methodology

2.1 Materials

2.1.1 *Bacterial Transformations and DNA Miniprep*

Custom plasmids containing the target gene sequences codon optimized for *E. coli* in pET-21b (+) (Tde1511 and Tde2714) and pET-28b (+) (Tde1414) vectors were ordered from Bio Basic Inc. (Markham, Ont.). Calcium competent NEB 5- α *E. coli* cells were ordered from New England Biolabs (Ipswich, MA, USA). Ultra pure water was purchased from Bio Basic Inc. Bio-tryptone, yeast extract, and reagent grade sodium chloride were ordered from Bio Shop (Burlington, Ont.). Lab companion SI-300R shaking incubator was purchased from Artisan Technology Group (Champaign, IL, USA). Microtubes (1.5mL), 92x16mm petri dishes, and 13mL 100x16mm culture tubes were ordered from SARSTEDT (Des Grandes-Prairies, QC). Analog heat block and the Clinical 100 laboratory centrifuge were purchased from VWR (Mississauga, Ont.). Agar A and kanamycin sulfate were purchased from Bio Basic Inc. (Markham, Ont.). QIAprep spin Miniprep Kit was ordered from Qiagen Inc. (Toronto, Ont.). Agarose, ethylenediaminetetraacetic acid (EDTA) disodium salt dihydrate, ampicillin sodium salt, and glacial acetic acid were purchased from Bio Shop (Burlington, Ont.). Ethidium bromide was received from Thermo Fisher Scientific (Waltham, MA, USA). Molecular Imager Versa Doc and a 1kB DNA ladder was ordered from Bio Rad (Hercules, CA, USA). Cary 60 UV-Vis spectrophotometer was purchased from Agilent Technologies Canada (Mississauga, Ont.). A quartz cuvette (45mm x 12.5mm x 12.5mm, 3.5mL) was purchased from Sigma Aldrich Canada (Oakville, Ont.).

2.1.2 Bacterial Growth and Recombinant Expression

Calcium competent BL21 (DE3) *E. coli* cells were ordered from New England Bio labs (Ipswich, MA, USA). Cultures grown in 2L Pyrex® shaker flasks were purchased from Sigma Aldrich Canada (Oakville, Ont.). Large Avanti Jxn-26 Centrifuge, J-lite® series JLA 8.1 rotor, J series JA-20 rotor, and 1L centrifuge bottles were purchased from Beckman Coulter Canada Inc. (Mississauga, Ont.). Bio Ultrapure grade isopropyl β -D-1-thiogalactopyranoside (IPTG) was purchased from Bio Shop (Burlington, Ont.). GENESYS™ 105 UV-Vis spectrophotometer was obtained from Thermo Fisher Scientific (Waltham, MA, USA).

2.1.3 Cell Lysis

Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), 2-hydroxyethylamino-trisaminomethane (Bis-Tris), glycerol, DL- dithiothreitol (DTT), Triton X-100, hydrochloric acid, magnesium chloride hexahydrate, hen egg white lysozyme, and sucrose were ordered from Bio Shop (Burlington, Ont.). Deoxycholic acid sodium salt and DNase I were purchased from Bio Basic (Markham, Ont.). Sodium hydroxide pellets were ordered from Anachemia Science (Mississauga, Ont.). Nalgene centrifuge tubes (50mL) were ordered from Sigma Aldrich Canada (Oakville, Ont.). The TS Series 0.75kW cell disruptor was obtained from Constant Systems (Daventry, Northamptonshire, England). Conical tubes (114x28mm, 50mL) were ordered from SARSTEDT Canada (Des Grandes-Prairies, QC).

2.1.4 *Immobilized Metal Affinity Chromatography*

The Econo-Column® Chromatography columns and Econo-Gradient pump were purchased from Bio-Rad Laboratories Canada Ltd. (Mississauga, Ont.). HisPur™ nickel-nitrilotriacetic acid (Ni-NTA) resin was purchased from Thermo Fisher Scientific (Waltham, MA, USA). Guanidine hydrochloride and a rocking platform were ordered from VWR (Mississauga, Ont.). Reagent grade imidazole was purchased from Bio Shop (Burlington, Ont.). Nickel (II) chloride hexahydrate was ordered from Bio Basic Inc. (Markham, Ont.). Reagent alcohol (95% (v/v)) was purchased from Cancer Diagnostics (Durham, NC, USA). Plastic 15mL falcon tubes were ordered from SARSTEDT Canada (Des Grandes-Prairies, QC).

2.1.5 *Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)*

Sodium dodecyl sulphate (SDS), tetramethylethylenediamine (TEMED), 40% (w/v) acrylamide/ bis-acrylamide (19:1), Coomassie Brilliant Blue G-250, and reagent grade methanol were all purchased from Bio Shop (Burlington, Ont.). Ammonium persulfate was ordered from Fisher Scientific (Hampton, NH, USA). Mini-Protean tetra system, Power Pac basic, glass plates, and precision plus protein™ dual colour ladder was purchased from Bio-Rad Laboratories (Hercules, CA, USA). Bromophenol blue was ordered from Sigma Aldrich (Oakville, Ont.). Glycine was purchased from Bio Basic (Markham, Ont.).

2.1.6 *Dialysis*

45mm molecular porous membrane dialysis tubing was ordered from VWR (Mississauga, Ont.). SpinBar magnetic stirring bar was obtained from Fisher Scientific (Hampton, NC, USA).

A 4L plastic graduated container was ordered from Owens Corning Canada (Guelph, Ont.). Hot plate/stirrer used was purchased from VWR. Plastic transfer pipettes (3mL) were ordered from STARSED Canada (Des Grandes-Prairies, QC).

2.1.7 Anion Exchange Chromatography, Dynamic Light Scattering, and Protein Concentration

Fast protein liquid chromatography (FPLC) was conducted using the ÄKTA Pure FPLC, a 5mL HiTrap™ Q FF anion exchange chromatography column and operated using the UNICORN 6.4 Workstation program all purchased from GE Life Sciences Canada (Mississauga, Ont.). Dynamic light scattering (DLS) was done using the DynaPro PlateReader-II and visualized using the program Dynamic 7.1.9 from Wyatt Technology Corporation (Santa Barbra, CA, USA). Black polystyrene 96-well microplates were purchased from Owens Corning Canada (Guelph, Ont.). Protein concentration was done using Amicon Ultra-15 centrifugal filter units (MWCO 10kDa and 30kDa) from Sigma Aldrich (Oakville, Ont.). Centrifugation took place using the Allegra X-14R with the SX4750 swinging bucket rotor all purchased from Beckman Coulter Canada Inc. (Mississauga, Ont.).

2.1.8 Protein Crystallization, Crystal Optimization, and X-Ray Diffraction Analysis

Sparse matrix crystal screens MSCG 1-4 were purchased from Anatrace Inc. (Maumee, OH, USA). Crystal Gryphon, intelli-plate 96-3 3-well sitting drop plates, and 0.2mL microtubes were purchased from Art Robins Instruments (Sunnyvale, CA, USA). HD clear packaging tape was ordered from ShurTech Brands, LLC (Avon, OH, USA). A 55L benchtop incubator was purchased from Molecular Dimensions Ltd. (Maumee, OH, USA). SteREO Discovery V8

stereomicroscope was purchased from Carl Zeiss Canada Ltd. (North York, Ont.). TC-Plate 24 well, cell +, F plates were purchased from STARSED Canada (Des Grandes-Prairies, QC). Circular siliconized glass cove slips were purchased from Hampton Research (Aliso Viego, CA, USA). High vacuum grease was ordered from Owens Corning Canada (Guelph, Ont.).

Sodium citrate, sodium acetate, and ammonium sulphate were all purchased from Bio Shop (Burlington, Ont.). Polyethylene glycol 3350 was ordered from Hampton Research (Aliso Viejo, CA, USA). Sodium nitrate was ordered from Bio Basic Canada (Markham, Ont.). Thiamine pyrophosphate was ordered from Sigma Aldrich (Oakville, Ont.).

Goniometer Base B1A and 100, 300, and 600 μ m dual thickness micro loops LD were purchased from MiTeGen (Ithaca, NY, USA). Liquid nitrogen was ordered from Linde Canada (Cambridge, Ont.). XT20 cryogenic storage tank was purchased from Taylor-Wharton (Bollvar, OH, USA). Shipping canes, ALS pucks, and cassette loading and shipping kit was purchased from Crystal Plate Positioning Systems (Jamestown, NY, USA). X-ray diffraction data was obtained using the Canadian Light Source Synchrotron (Saskatoon, SK, CA) and the Cornell High Energy Synchrotron Source (CHESS) (Ithaca, NY, USA). Data processing was completed using the CCP4 suite of programs (version 8.0). Structure refinement and modelling was done using Phenix version 1.19.2, XQuartz version 2.8. and *Coot* version 0.9.6.

2.1.9 Inductively Coupled Plasma Optical Emissions Spectroscopy (ICP-OES)

Calcium chloride dihydrate, magnesium chloride hexahydrate, manganese chloride tetrahydrate, nickel (II) chloride hexahydrate, and zinc chloride were all purchased from Bio Shop (Burlington, Ont.). Cobalt (II) chloride was ordered from Sigma Aldrich (Oakville, Ont.).

Cupric chloride was purchased from Matheson, Coleman, & Bell (Los Angeles, CA, USA).

Ferric chloride hexahydrate was purchased from Bio Basic Canada (Markham, Ont.). ICP-OES was conducted using the Optima 8000 from Perkin Elmer (Waltham, MA, USA).

2.1.10 Differential Scanning Fluorimetry (DSF)

Includes all divalent metal salts listed above and DL-DTT. Ferrous iron chloride was purchased from Sigma Aldrich (Oakville, Ont.). 0.2mL polymerase chain reaction (PCR) tube strips were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Differential scanning fluorimetry (DSF) was conducted on the QuantStudio™ 6 Pro Real Time PCR instrument and visualized using the program Design and Analysis software version 2.6, both obtained from Thermo Fisher Scientific (Waltham, MA, USA). A 5000x SYPRO® orange protein gel stain was purchased from Sigma Aldrich (Oakville, Ont.).

2.1.11 Peptide Thin Layer Chromatography

High purity (>95%) peptide substrates were synthesized and purchased from Pepmic Co. Ltd. (Suzhou, Jiangsu, China). Silica 60 aluminum thin-layer chromatography (TLC) sheets, ACS reagent ninhydrin, ACS grade sodium phosphate, and 99.8% pyridine (anhydrous) were purchased from Sigma Aldrich (Oakville, Ont.). 1-butanol was from Bio Shop (Burlington, Ont.). Centrивap concentrator was obtained from Labconco (Kansas City, MO, USA). Precision Model 18EG Gravity Convection Lab Oven was purchased from GCA (Chicago, IL, USA). Protease inhibitor cocktail tablets cOmplete, EDTA free was purchased from Sigma Aldrich (Oakville, Ont.).

2.1.12 Proteolytic Degradation Assay

Albumin from chicken egg white, N-glutamyl-L-phenylalanine p-nitroanilide (GPNA), and conalbumin from chicken egg white were all obtained from Sigma Aldrich (Oakville, Ont.). Hen egg white lysozyme was purchased from Bio Shop (Burlington, Ont.).

2.2 Methodology

2.2.1 DNA miniprep, Bacterial Transformations, and Recombinant Protein Expression

DNA minipreps of the target genes started by removing three 50µL aliquots of calcium-competent *E. coli* NEB 5-α cells from the -80°C freezer and allowed to thaw on ice. The three gene targets were removed from the -20°C freezer and allowed to thaw on ice. Next, 1µL of each plasmid was placed into its own aliquot of competent cells and the mixtures were left on ice for 30min to equilibrate. The three solutions were placed into the heat block at 42°C for 50sec and placed on ice for 3min thereafter. Next, 500µL of lysogeny broth (LB) (8g tryptone, 5g yeast extract, 2.5g NaCl) was added to each sample and they were left to incubate with shaking at 37°C, 240rpm, for 1h. Then, 200µL of each sample was aseptically placed onto individual LB agar plates supplemented with the required antibiotic (100 µg/mL ampicillin for Tde1511 and Tde2714, 50 µg/mL kanamycin for Tde1414), evenly spread across each surface, and placed upside down to incubate at 37°C for 16h. The next day, a single bacterial colony was aseptically isolated from each plate and individually placed into three tubes containing 10mL of LB and supplemented with the required antibiotics. These solutions were left to incubate for 16h with shaking at 37°C and 240rpm. The following day, DNA minipreps were carried out following the

protocol provided in the Qiagen DNA miniprep kit. The resulting 50µL aliquots of the three target plasmids were stored in the -20°C freezer for future use.

Bacterial transformations for protein expression initiated by removing a 50µL aliquot of calcium-competent *E. coli* BL21 (DE3) from the -80°C freezer to thaw on ice. Miniprep DNA of the target gene was removed from the -20°C freezer and allowed to thaw on ice. Next, 1µL of plasmid was added to the competent cells and the mixture was left on ice for 30min to equilibrate. Solutions were then placed in a heating block set to 42°C for 50sec, followed by 3min on ice. Then, 500µL of LB was added to the mixture which was then placed to incubate with shaking at 37°C, 240rpm, for 1h. Then, 200µL of this solution was evenly spread over a LB agar plate supplemented with the required antibiotic and allowed to sit for 5min. This plate was turned upside down and incubated at 37°C for 16h. The next day, one bacterial colony was aseptically isolated and placed into a sterile tube containing 10mL LB supplemented with the required antibiotic. This solution was then incubated with shaking at 37°C and 240rpm for 16h. The next day, 9 sterile microfuge tubes were filled with 1mL of the overnight bacterial culture and glycerol was added to a final concentration of 25% (v/v). These bacterial glycerol stocks were flash frozen in liquid nitrogen and placed in the -80°C freezer for future use. The remaining 1mL of the bacterial culture was added to 1L of autoclaved 2YT media (16g tryptone, 10g yeast extract, 5g NaCl) and allowed to shake at 240rpm, 37°C until an optical density at 600nm (OD₆₀₀) of 0.6 was achieved. When the appropriate OD was reached, IPTG was added to a final concentration of 1mM to induce protein production, and the culture was left for 16h at 16°C and 160rpm.

2.2.2 *Cell Lysis*

Following the 16h gene induction for protein expression, cell cultures were pelleted through centrifugation at 5000rpm, 16°C, for 15. For chemical lysis of Tde2714, cell pellets were resuspended in 25mL of lysis buffer (750mM sucrose, 50mM Tris HCl, 300 mM NaCl, pH 8.0) and left to mix for 20min on ice. Next, 10mg of hen egg white lysozyme was added to the mixture and left to stir for another 10min. Following this, 50mL of a solution containing 24mM deoxycholic acid, 1% (v/v) Triton X-100, 50mM Tris HCl pH 8.0, and 300mM NaCl was added and left to stir for an additional 10min. Next, 2.6 µg/mL DNase I and 5mM magnesium chloride was added and the solution was left to stir on ice for an additional 10min.

Using mechanical cell lysis for Tde1414 and Tde1511, two separate lysis buffers were created. The cell pellet of Tde1414 was resuspended in a solution containing 50mM Tris HCl pH 7.6, 200mM NaCl, and 5mM magnesium chloride. The cell pellet of Tde1511 was resuspended in a solution containing 100mM Bis-Tris pH 6.5 and 300mL NaCl. During cell disruption, a pressure of 15-20 pounds per square inch (PSI) was used throughout the procedure. The instrument was cleaned and primed by flowing deionized water, 70% (v/v) ethanol, and desired buffer through the machine in that order. Then the resuspended sample was poured into the cell disruptor and collected over ice on the other end. The instrument was then cleaned following proper protocols and turned off. At this stage, both chemical and mechanical cell lysate solutions were centrifuged at 15,000 xg, 4°C, for 35min. Supernatants were collected for further purification while solid pellet waste was discarded.

2.2.3 *Immobilized Metal Affinity Chromatography*

The 20% (v/v) ethanol storage solution was drained out of the IMAC column, and two column volumes of water was passed through the column. One column volume of a IMAC buffer A (appropriate lysis buffer for each species (outlined in section 2.2.2), supplemented with 5mM imidazole) was then passed through the column while stirring the beads. Finally, two column volumes of lysis buffer were passed through the column while stirring the beads. Resin beads were then evenly distributed throughout the conical tubes containing the cell lysis supernatants from the prior step, and these solutions were left to incubate on a rotating platform for 40min at 4°C. The lysate/resin solutions were then poured into the chromatography column, and the flowthrough was collected and stored in the fridge. Lysis buffer was gently poured on top of the resin beads in the column as to not disturb them. Following the Econo-Gradient pump protocol, the system was first primed with IMAC buffer B (appropriate lysis buffer for each species (outlined in section 2.2.2), supplemented with 500mM imidazole), followed by IMAC buffer A. The Econo-Gradient wand was secured in the chromatography column, and following the Econo-Gradient pump protocol, the machine was turned on and set to the purification program. Actively collecting fractions during the step gradient of 5-500mM imidazole, the program finished in about 70min. Collected fractions were placed in the fridge. The Econo-Gradient pump enclosure and tubing was stored in 20% (v/v) ethanol for future use. The resin beads within the column were cleaned using one column volume of 500mM sodium hydroxide, one column volume of 10% (v/v) acetic acid, one column volume of deionized water, and stored in 20% (v/v) ethanol for future purifications.

2.2.4 SDS-PAGE and Dialysis

To analyze the location and purity of protein samples throughout purification, SDS-PAGE analysis was employed. A 12 % (w/v) separating gel was first prepared by adding in a beaker 3.1mL of the 40% (w/v) acrylamide/ bis-acrylamide (19:1), 4.24mL of milliQ water, 2.5mL of a 1.5M Tris HCl pH 8.8 buffer, 100 μ L of 10% (w/v) SDS solution, 50 μ L of 10% (w/v) APS solution, and 15 μ L TEMED. This was poured into a glass casting stand with a plastic transfer pipette until solidified. Following this, a 4% (w/v) stacking gel was prepared by adding into a beaker 1.01mL of the 40% (w/v) acrylamide/ bis-acrylamide (19:1), 6.33mL of milliQ water, 2.5mL of a 0.5M Tris HCl pH 6.8 buffer, 100 μ L of 10% (w/v) SDS solution, 50 μ L of 10% (w/v) APS solution, and 10 μ L TEMED. This was poured over top of the separating gel, a comb was added, and the solution was left to solidify. Once the gel was prepared, 20 μ L of each sample to be analyzed was added into a 0.6mL microfuge tube. Next, 5 μ L of a SDS sample dye (1mL 2M Tris HCl pH 6.8, 5mL glycerol, 1g SDS, 1mL 0.2% (w/v) bromophenol blue, 3mL milliQ water) was added to each sample to be analyzed. Microtubes containing samples were placed in the heat block at 92°C for 5min to denature proteins. The gel was placed into a chamber containing 1x SDS running buffer (25mM Tris HCl, 200mM glycine, 0.1% (w/v) SDS) and the comb was removed. Following this, 16 μ L of each sample was added to an individual well in the gel. A 5 μ L sample of the precision plus dual stained protein ladder was also placed into a well. The lid was placed on the chamber and the battery pack was turned on. The samples were left to migrate through the gel at 180 volts for 55min.

Once completed, the gel was removed from the glass casting case and placed into a shallow plastic container. Protein visualization occurred by staining with Coomassie brilliant blue G-250 solution (supplemented with methanol and glacial acetic acid) and microwaved for

two sets of 30sec. Stain solution was recollected, and the gel was rinsed with deionized water. A destaining solution containing 10% (v/v) methanol and 10% (v/v) glacial acetic acid was added to the gel and microwaved for 30sec. The gel was left in the destain solution on a rocking platform for 30min for protein banding patterns to begin to resolve.

Protein containing fractions identified by SDS-PAGE were pooled together for dialysis. For Tde1414, the 4L dialysis buffer was composed of 50mM Tris HCl pH 7.6, 20mM sodium chloride, and 5mM magnesium chloride. For Tde1511, the 4L dialysis buffer was composed of 100mM Bis-Tris pH 6.5. For Tde2714, the 4L dialysis buffer was composed of 50mM Tris HCl pH 8.0. Porous cellulose dialysis tubing was used with a molecular weight cut off (MWCO) of 3000-5000 Daltons (Da). Tubing was soaked in buffer for 30sec until it was able to be opened with ease. The bottom was clipped shut, and a plastic transfer pipette was used to load the pooled protein sample into the dialysis tubing. When complete, the top was clipped, and the dialysis tubing set up was secured to the side of the container using thread and tape. A stir bar was added to the container, and the dialysis set up was placed in the cold room on a stir plate and left to dialyze for 16h.

2.2.5 Anion Exchange Chromatography and Dynamic Light Scattering

When dialysis was completed, protein samples were emptied from the dialysis tubing and subjected to a brief centrifugation at 17, 000 xg, 4°C, for 15min to remove any precipitation. At this time, the ÄKTA Pure FPLC was cleaned and primed for protein purification. FPLC buffers used were similar conditions to the protein target's dialysis buffer. For FPLC buffer A, only the buffer species was added to solution and vacuum filtered through a 0.2 µm filter to filter and

degas the buffer. For FPLC buffer B, the buffer species was added to the solution supplemented with 1M sodium chloride before being vacuum filtered through a 0.2 μm filter to filter and degas the buffer. Following the line priming of the FPLC and protein sample clarification, the apparatus was assembled following proper protocol and the anion exchange procedure was initiated on the UNICORN 6.4 Workstation. Fractions were automatically analyzed at 280nm to produce a chromatogram and samples were collected in 3mL aliquots. Using the chromatogram produced, fractions over the peak protein absorption were analyzed via SDS-PAGE to determine purity and to select fractions to pool together for protein concentration.

Samples used in SDS-PAGE analysis were also subjected to dynamic light scattering to assess protein purity and homogeneity. Using black styrene 96-well plates, 30 μL of each sample was loaded into the tray in triplicate. The DynaPro Platereader-II was turned on and allowed to reach optimal temperature for 30min. Using the Dynamics program, the collect acquisition image was set to 20, the temperature was set to 25°C, and the plate reader was set to the appropriate well to begin the analysis on. The 96-well plate was inserted into the machine, the door was closed, and the collection of dynamic light scattering from samples was performed at room temperature. Upon completion, the 96-well plate was removed, the door was closed, the machine was turned off, and the well plate was cleared of any samples in the wells.

2.2.6 Protein Concentration and Initial Protein Crystal Screening

Samples analyzed from anion exchange chromatography via SDS-PAGE and DLS that were deemed suitable to move forward with in protein crystallization trials were pooled together and placed on ice. Amicon Ultra-15 centrifugal filters with a MWCO of 10 kDa were used for Tde1414 and Tde1511 while the MWCO of 30 kDa was used for Tde2714. Filter units were

initially primed using two rounds of milliQ water and two rounds of the respective buffer for each protein target. Solutions were added to the filter unit and centrifuged at 4300 rpm, 4°C, for 5min. Once the filter units were primed and ready for use, protein sample was loaded into the top of the unit and centrifuged at 4300 rpm, 4°C, for 15min. Concentrating continued until the total protein sample was reduced to about 1mL or when a protein concentration of 35 mg/mL was achieved. To assess protein concentration, 5µL of protein sample was added to 1mL of milliQ water in a quartz cuvette. An absorbance value at 280nm was taken and placed into a dilution calculator using Beer's Law which resulted in a concentration in mg/mL to be produced.

Concentrated protein samples were centrifuged at 17, 000 xg, 4°C, for 10min to remove any precipitation or aggregation. Using a Gryphon Crytal Robot, an initial cleaning protocol was used to clear the lines and needles of any debris. Then, 180µL of the concentrated protein sample was added to the instrument, and a fresh intelli-plate 96-3 3-well sitting drop plate was positioned on the platform. MCSG crystal screens 1-4 were individually added to the platform and the program was initiated. Protein to reservoir ratios of 1:1, 2:1, and 1:2 were used. Once the crystal screen was complete, HD clear packaging tape was used to cover the top of the plates and they were incubated at 18°C. Periodically, protein crystallization was monitored using the ZEISS SteREO microscope at 1000x magnification.

2.2.7 Crystal Optimization, X-Ray Diffraction Analysis, and Data Manipulation

Successful protein crystallization conditions were replicated and optimized using TC-Plate 24 well plates and the hanging drop vapour diffusion method. Salt concentrations and precipitating agents had their concentrations altered, while introducing varying precipitation

agents and divalent metal cations (such as glycerol, PEG3350, Cu^{2+}). The pH and buffering agents were also manipulated to identify optimal crystal growth conditions. Protein to reservoir ratios were also optimized, using ratios of 1:1, 2:1, and 1.5:1. Samples were placed on cleaned, siliconized glass cover slips, and suspended over a reservoir of 1000 μL using vacuum seal grease to create an air-tight enclosure for vapour diffusion to occur. These prepared plates were left to incubate at 18°C until crystal growth was observed.

When diffraction-grade crystals were observed, nylon micro loops from MiTeGen were used to remove the crystal and placed them into freshly prepared cryoprotectant. Cryoprotectants were made to mirror the crystallization condition the crystals initially grew in. For Tde1414 and Tde2714, solutions were supplemented with 35% (v/v) and 30% (v/v) glycerol, respectively, to mitigate the formation of ice crystals. For Tde1511, a solution of saturated ammonium sulphate (3.7M) was used as the cryoprotectant. Crystals were then looped from their cryoprotectant and submerged in liquid nitrogen. The crystals were then packaged appropriately in containers filled with liquid nitrogen, and samples were sent to the CLS or CHESS facilities for X-ray diffraction analysis. At the synchrotron, testing screens were taken of the crystals at 0°, 45°, and 90° to determine crystal statistics, such as the potential quality of the data, early resolution estimates, and crystal space groups. When an appropriate crystal was screened, a full native data set was taken covering all 360° of the crystal using a beam energy of 13.0 keV, an exposure time of 0.01sec, an attenuation of 0.0%, and an aperture of 50 μm .

2.2.8 *Diffraction Data Manipulation and Protein Data Bank Submissions*

Data files were analyzed for the appropriate range of frames to include, such as using the graph of R_d (R-factor) as a function of frame number difference. As R_d increases with frame development, this typically indicates radiation damage of the protein crystal where data obtained may not be as accurate. The unmerged.mtz file was scaled using Scala in the CCP4 suite of programs. Matthew's Coefficient was also determined for each crystal. Using Phenix Phaser-MR, molecular replacement was completed using known homologous structures with greater than 35% sequence identity for Tde1511 and using AlphaFold 2.0 structure prediction files for Tde1414. When a solution was converged on, 3 rounds of phenix.refine commenced to improve the structure. *Coot* was used to visualize the structure and electron density maps, and manual improvements of the structure occurred. After iterative rounds of refinement and model building, the Protein Data Bank (PDB) validation tool was used to identify potential issues in the structures. Once any errors were corrected, PDB deposition commenced. The experimental scaled unmerged.mtz file, as well as the most updated structure model was provided, and the online application was completed. Once completed, submitted protein structures were accepted into the PDB pending a publication release.

2.2.9 *Inductively coupled plasma optical emission spectrometry*

To determine metal-binding properties of Tde1511, ICP-OES was employed. After anion exchange, Tde1511 protein samples were concentrated to 30mL total volume and subjected to exhaustive EDTA dialysis using 50mM Bis-Tris pH 6.5 and 10mM EDTA. The sample was then dialyzed in just 50mM Bis-Tris pH 6.5 to remove excess EDTA. Ultracentrifugation was done and included 3 wash steps to remove excess EDTA even further. Protein was concentrated to 40

mg/mL and a sample was taken for analysis as the “EDTA-treated” protein sample. After treatment with EDTA, half of the protein sample was then dialyzed against 50mM Bis-Tris pH 6.5 and 10mM CuCl₂ for 3h. This copper reconstituted sample was then dialyzed against 50mM Bis-Tris pH 6.5 and concentrated and washed via ultracentrifugation to remove excess unbound copper ions in solution. At this stage, all beakers, equipment, and utensils were soaked in an acid bath for 24h to remove any excess contaminating metals and washed seven times with milliQ water. Seven beakers were used, filled with 50mL of 50mM Bis-Tris pH 6.5. For the analysis, the divalent metal salts used were CaCl₂, CoCl₂, ZnCl₂, MgCl₂, NiCl₂, MnCl₂, CuCl₂, and FeCl₃. The seven beakers of buffer were labelled 1μM, 5μM, 10μM, 25μM, 50μM, 100μM, and 200μM. The total concentration of each divalent metal salt was mirrored to the concentration labelled on the beaker. For example, the 5μM beaker contained 5μM CaCl₂, 5μM CoCl₂, 5μM ZnCl₂, and included all other metal salts. In dialysis buttons, a 1.25mL aliquot of 22.4 mg/mL EDTA-treated and copper reconstituted Tde1511 was added. Using foam as holders to maintain the correct orientation of the dialysis buttons, samples of Tde1511 were lowered into each beaker and allowed to dialyze in the cold room over night on a rocking platform.

The following morning, samples within the dialysis buttons were concentrated and washed four times via ultracentrifugation at 4300 rpm, 4°C, for 10min to remove excess unbound metals within each aliquot. Each sample was diluted up to 7mL in 50mM Bis-Tris pH 6.5 (final protein concentration of 5 mg/mL) and acidified with 2% (v/v) nitric acid. Prepared samples were brought to the ICP-OES machine at the Centre for Cold Regions and Water Science where Gena Braun, MSc. conducted the analysis. Metal standards were created and used to form a standard curve in the concentration range of the expected metal detection concentrations, and the

analysis commenced. Collecting data points in triplicates and averaging the values, final data values were exported as an excel file for future data analysis.

2.2.10 Differential scanning fluorimetry

To contextualize the metal binding results gained from ICP-OES, DSF was employed to determine if stable protein-metal complexes formed. Following a similar protocol as ICP-OES, protein samples of Tde1511 were exhaustively dialyzed against EDTA. Samples were then dialyzed against 50mM Bis-Tris pH 6.5 and concentrated and washed via ultracentrifugation to further remove excess EDTA. Half the sample was set aside in the fridge, while the other half was further dialyzed in 50mM Bis-Tris pH 6.5 and 10mM CuCl₂ for 3h. This copper reconstituted sample was then dialyzed against 50mM Bis-Tris pH 6.5 and concentrated and washed via ultracentrifugation to remove excess unbound copper ions in solution. Starting DSF, the first variable resolved was the correct ratio of protein to 20x SYPRO orange dye that gives an appropriate signal to quantify. When that was complete, the appropriate protein concentration was incubated with the various metals used in ICP-OES (CaCl₂, CoCl₂, ZnCl₂, MgCl₂, NiCl₂, MnCl₂, CuCl₂, FeCl₃ and FeCl₂ with 5mM DTT) in ratios of 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, and 1:10 to determine which ratio of protein to metal provided the greatest resolution in melting temperature shifts. After the preparation to ensure meaningful data can be collected, the optimal protein concentration was mixed with the appropriate metal concentration in triplicate and left to incubate on ice for 10min. Following this, 2μL of a 200x SYPRO orange stock was added to each PCR microtube to a final concentration of 20x, and the contents were mixed. PCR microtube strips were placed in the QuantStudio™ 6 Pro Real Time PCR and the temperature was set to increase from 18°C to 98.5°C, while fluorescence was continuously measured. Data from

the melting curve analysis was exported onto a USB and the data was saved for future data analysis.

2.2.11 Peptide Thin Layer Chromatography

5mg of peptide substrates were initially designed and purchased (>95% purity) following the four FGE recognition motifs and a negative control. The peptide sequences (N-term to C-term) with bolded FGE recognition sequences and short form names in brackets following were: DKEVMGIT**CSP**TRSMLLTGTDEHGDE (CxPxR peptide), DKEVMGITS**APAR**SMLLTGNDEHGDE (SxPxR peptide), DKEVMGIL**CTALR**SMLLTGRDEHGDE (CxAxR peptide), DKEVMGIL**STALR**SMLLTGRDEHGDE (SxAxR peptide), and DKEVMGIT**ATP**TRSMLLTGTDEHGDE (AxPxR peptide). Extinction coefficients for each peptide were determined experimentally at 205nm (absorption of a peptide backbone plus absorption of each residue's side chain). After reconstituting the lyophilized peptide powders with ultra pure, nuclease-free water, concentrations of the peptides were measured to be about 6.33mM. Peptides using FGE recognition sequences were incubated with Tde2714 with an attempt to observe FGE activity. Reaction tubes consisted of 10μM Tde2714, 100μM peptide substrate, 1mM DTT, and were diluted to a final volume of 500μL with 50mM sodium phosphate pH 8.0. Tde2714 was added last to initiate the reactions. Controls were also created where only enzyme was added with buffer, and only peptide was added with buffer. Reactions tubes were placed in the heat block at 37°C for 8h. While this occurred, the mobile phase for the TLC reaction was created and composed of 1-butanol, glacial acetic acid, and water in a 3:1:1 ratio. Multiple pieces of silica 60 aluminum TLC paper were cut to be one inch wide by twelve

centimeters long. Individual TLC plates were marked 1cm above the bottom of the strip in pencil and placed in an oven at 120°C for 10min to activate the silica.

Once reactions ran to completion, samples were concentrated in the Centrivap at 60°C until about 10µL of sample was left over. Three 3µL spots were spotted on each prepared TLC paper in the order of enzyme, peptide reaction, and just peptide and allowed to fully dry. TLC strips were placed in beakers containing the mobile phase and covered using parafilm. The mobile phase was allowed to travel up the strips until about 1cm from the top where the strips were removed from the TLC vessels and left in the fume hood to dry. To stain the TLC strips, a 1% (w/v) ninhydrin stain was used dissolved in a solution of 5:1 pyridine to glacial acetic acid. TLC strips were dunked entirely into the stain solution for 2sec, removed, and placed in an oven at 80°C for 10min to activate the stain and resolve spots along the TLC paper.

Following identical methods as listed above, the hypothesis of Tde2714 functioning as a protease was then tested. A protease inhibitor cocktail (PIC) tablet (EDTA free) was dissolved in 1.5mL ultra pure water to create a 25x stock. Using the SxAxR peptide as the reaction substrate, 4 reaction tubes were set up using 10µM Tde2714, 100µM SxAxR peptide, and 1mM DTT. To 3 reaction tubes, 1x, 2x, and 4x PIC was added. Controls were also created including just the peptide substrate, only enzyme, only 1x PIC, only 2x PIC, and only 4x PIC. Control mixtures of peptide plus 1x PIC, peptide plus 2x PIC, peptide plus 4x PIC, enzyme plus 1x PIC, enzyme plus 2x PIC, and enzyme plus 4x PIC were also created. Every micro tube was supplemented with 1mM DTT. The same reaction protocol, and TLC spotting and staining procedures were followed as stated above.

2.2.12 Proteolytic Degradation Assay

To observe the proteolytic potential of Tde2714, various proteins and substrates on hand were used in the assay along with the 25x stock of PIC. Stock solutions of 10mM were created for albumin from chicken egg whites, GPNA, conalbumin from chicken egg whites, and hen egg white lysozyme. All reactions and solutions were buffered in a solution of 50mM sodium phosphate pH 8.0. Four reaction mixtures of 500μL were created that contained 10μM Tde2714, 100μM substrates, and 1mM DTT. Another four reaction mixtures of 500μL were created that contained 10μM Tde2714, 100μM substrates, 1mM DTT, and 5x PIC. For controls, four 500μL solutions were created for each individual substrate at a concentration of 100μM. In addition, two 500μL solutions of 10μM Tde2714 were created, and one of them was supplemented with 5x PIC. All reaction tubes were kept at 4°C and agitated daily. Finally, 20μL samples were removed from each micro tube at time points of 0 days, 7 days, and 14 days and analyzed via SDS-PAGE.

Chapter 3. Results

3.1 Bioinformatics

3.1.1 Predicted Biophysical Properties

To initiate characterizing the three gene targets, bioinformatic research was completed to outline hypothetical biochemical and functional parameters. Identifying these variables early on allows for direct protocol modifications to be made to increase the chance of successful isolation, purification, crystallization, and functional assay results. The programs used to accumulate this data includes Expasy ProtParam, EggNOG, Uniprot, BLASTp, STRINGS, InterPro, and PSort

⁸⁴⁻⁹⁰. Uniprot was used to obtain FASTA sequences for each of the targets. ProtParam used the primary sequences of each target and predicts various biophysical and biochemical properties. EggNOG identifies clusters of orthologous groups (COG) characterization for a given gene target. BLASTp aided in finding homologous proteins based on similarities in primary sequence to outline potential functional annotations for the targets. STRINGS database is a helpful tool, which highlights potential protein-protein interactions that may be made with other proteins within the organism. InterPro identified key domains and motifs through homology with other known sequences. Lastly, PSort was useful in predicting the most likely cellular localization for each target. Key findings have all been summarized in Table 1.

Table 1. Predicted structural, functional, and biochemical properties of the gene products Tde1414, Tde1511, and Tde2714 from *T. denticola*.

Gene Target	Tde1414	Tde1511	Tde2714
# of Amino Acids	378	208	663
Molecular Weight (kDa)	41.0	22.6	73.9
Theoretical pI	6.47	4.84	5.31
Aliphatic Index	91.38	69.95	65.88
GRAVY Score	-0.101	-0.316	-0.509
Predicted Extinction Coefficient (M ⁻¹ cm ⁻¹)	22 390	32 555	170 225
Predicted Cellular Localization	Cytoplasm	Bacterial Inner Membrane	Microbodies
Predicted Function	ThDP-dependent Phosphonopyruvate decarboxylase	Metal/Iron Transporter	Formylglycine-generating enzyme

Uniprot has annotated the gene product of Tde1414 to have 378 amino acids in its primary sequence. Using this sequence in ProtParam, a molecular weight of 41.0 kDa and an isoelectric point (pI) of 6.47 were predicted. Furthermore, an extinction coefficient at 280nm in water was calculated assuming cysteines are not in disulphide bonds at $22\,390\text{ M}^{-1}\text{cm}^{-1}$. An aliphatic index of 91.38 was estimated, signifying the protein should have good thermostability (60-80 is considered good). A grand average of hydropathicity (GRAVY) was also predicted to be -0.101, where negative values indicate hydrophilicity and good water solubility. EggNOG analysis indicated this protein to be classified in COG0028, proteins requiring thiamine diphosphate. BLASTp results showed very high homology to other phosphonopyruvate decarboxylases from other treponeme species, as well as other bacterial species. STRINGS had identified potential protein-protein interactions with enzymes, such as Tde1413 (a phosphoenolpyruvate phosphomutase) and Tde1415 (aminotransferase-PntC). InterPro showcased Tde1414 containing domains relating to ThDP-binding folds and PSort predicted a cellular localization within the cytoplasm. All the predicted information is consistent and suggests Tde1414 to function as a ThDP-dependent phosphonopyruvate decarboxylase.

When analyzing Tde1511, Uniprot showed a total of 208 amino acids in its primary sequence. Using this sequence with ProtParam, a molecular weight of 22.3 kDa and a theoretical pI of 4.84 was calculated. An extinction coefficient at 280nm in water assuming all cysteines are not in disulphide bonds was predicted to be $32\,555\text{ M}^{-1}\text{cm}^{-1}$. An aliphatic index of 69.95 suggests this protein target has a fair thermostability, and a GRAVY score of -0.316 indicates good water solubility. EggNOG analysis indicated this protein to be apart of COG3470, proteins probably involved in high-affinity ferrous iron transport. The BLASTp results converge on Tde1511 having high sequence similarity to multiple iron transport proteins from various bacterial species.

STRINGS predicted Tde1511 is involved in a handful of protein-protein interactions, such as with a putative high-affinity iron transport permease (Tde1510), and putative ABC transport systems for macrolide export (Tde1514 and Tde1515). InterPro shows an iron transport domain within the primary sequence, as well as a metal-binding domain to couple this. Lastly, PSort has predicted the cellular localization of this protein target to the bacterial inner membrane. The proposed information above is all in agreeance and orients the function of Tde1511 as a pathogen-specific surface antigen with a function in metal binding and iron translocation.

Searching for Tde2714 in Uniprot shows this gene target to contain 663 amino acids in its primary sequence. Taking this sequence into ProtParam, a molecular weight of 74.0 kDa and a theoretical pI of 5.31 were predicted. An extinction coefficient at 280nm in water assuming all cysteines are not in disulphide bonds was estimated to be $170\,225\text{ M}^{-1}\text{cm}^{-1}$. An aliphatic index of 65.88 suggests fair thermostability, and a GRAVY score of -0.509 indicates good water solubility. EggNOG analysis indicates this protein to be apart of COG1262, formylglycine-generating enzymes required for sulphatase activity. Analysing BLASTp results has identified Tde2714 sharing the greatest sequence similarity to other FGE proteins from various treponeme and other bacterial species. It also shows similarities to a family of oxidizing enzymes, such as SUMF1 (FGE), EgtB (sulphoxide synthase), and PvdO (oxidation to form pyoverdine fluorophore). STRINGS outlines protein-protein interactions with three uncharacterized proteins. It also shows a potential interaction with Tde2713, an RNA methyltransferase belonging to the class IV-like S-adenosyl methionine (SAM)-binding methyltransferase superfamily. InterPro highlights the presence of a domain corresponding to FGE-sulphatase, as well as a SUMF FGE domain. Finally, PSort has predicted this protein to be localized within microbodies. This, the

above information all agrees with each other and points the function of Tde2714 to be related to FGE catalysis.

3.1.2 Predicted Protein Structures

With the success of artificial intelligence being able to confidently predict protein structures with a relatively high degree of accuracy, AlphaFold2 was employed to determine the three-dimensional structures of each protein target ^{91,92}. As shown in Figure 9, the structures with the highest confidence created by AlphaFold2 are displayed. The significance of these structure predictions is two-fold. Firstly, the accuracy of these models is so high it allowed for a successful molecular replacement solution to be found for Tde1414 and a partial solution to be successfully found for residues 126-375 of Tde2714. Secondly, the generated model of Tde2714 allowed for structure-guided functional annotation programs to be used. This allowed for one hypothesis of this protein's function to be dismissed and opened a new hypothesis with initial results that coincide with the predicted function. The model of Tde1511 generated was also used in molecular docking simulations with the predicted structure of Tde1510 to verify the possibility of iron translocation between the two proteins.

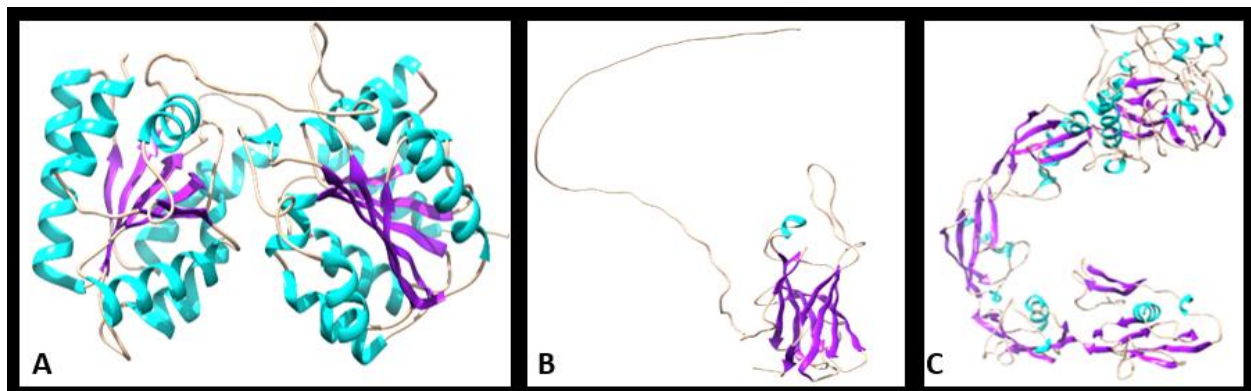


Figure 9. Predicted protein structures of Tde1414, Tde1511, and Tde2714 from the pathogen *T. denticola* using AlphaFold2. Shown above are the AlphaFold2 predicted three-dimensional structures of Tde1414 (A), Tde1511 (B), and Tde2714 (C). The structures were rendered in Chimera version 1.11.2 with alpha helices coloured cyan, beta strands coloured purple, and free coils coloured peach. AlphaFold2 predicted the entire structure of Tde1414 and Tde2714 with high accuracy. AlphaFold2 predicted the majority of Tde1511 (residues 56-208) with high confidence and had low confidence for the region (residues 1-55) previously annotated to be the signal sequence and a disordered region likely to be used in membrane anchoring of the protein.

3.1.3 Predicted Function of Tde2714

While the functions of Tde1414 and Tde1511 were either experimentally confirmed or having consistent convergence from bioinformatic analyses, Tde2714 required a broader evaluation of functional hypotheses. When evaluating FGE activity for Tde2714, a prior student in the Suits lab used a peptide with a FGE recognition sequence and tandem mass spectrometry to identify potential modifications to the peptide's target cysteine side chain to a formylglycine residue. This functional result was not detected. Expanding the recognition sequence using the four possible sequences FGEs discriminate for, as well as a negative control peptide, yielded similar reaction products for all peptides tested when only one recognition sequence was expected to have a unique spotting pattern. As seen in Figure 10, a multiple sequence alignment with Tde2714 and other known FGEs was conducted. For FGE catalytic activity to occur, all FGE enzymes contain a highly conserved serine and two highly conserved cysteines residues. Depicted is that Tde2714 does not share the conserved serine and one of the conserved/catalytic cysteine residues. This brought the idea that it was not the recognition sequence that was

important, but that this enzyme could interact with all peptide substrates similarly, despite minor differences in primary sequence. Thus, the hypothesis of a proteolytic enzyme was ignited.

tr Q73J62 Q73J62_TREDE	NPFGFLDMSGNVSEWCWDWIIQTT-----PEGGTDPV----GPLSGTNRTTCGGYSFFD	633
sp I6Y8I5 FGE_MYCTU	NGFGLDMIGNVWEWTTTEFYPHHRIDPPSTACCAPVKLATAADPTISQTLKGSHLCAP	265
sp Q8NBK3 SUMF1_HUMAN	NGYGLYNIVGNAWEWTSWWTVHHS----VEETLNPK----GPPSGKDRVKKGGSYNCHR	338
sp D1A7C3 FGE_THECD	NGYGLYNVAGNVWEWCADWWSADWHATESPATRIDPR----GPETGTARVTKGGSFLCHE	271
	* : ** : : ** : ** : : * : : : ** :	
tr Q73J62 Q73J62_TREDE	SACC---CAYRGLEEDKNKDDTTGFRIVCRYEF---	663
sp I6Y8I5 FGE_MYCTU	EYCHRYRPAARSPQSQDTATTHIGFRCVADPVSG--	299
sp Q8NBK3 SUMF1_HUMAN	SYCYRYRCAARSQNTPDSSASNLGFRCAADRLPTMD	374
sp D1A7C3 FGE_THECD	SYCNRYRVAARTCNTPDSSAAHTGFRCAADPL----	303
	. * * * : * . * * * .	

Figure 10. Multiple sequence alignment of known FGEs with Tde2714. The image above uses both prokaryotic organisms and a eukaryotic organism to highlight the strict conservation of the required residues that permit FGE catalysis to occur. The dark green box represents the position of the conserved serine, in Tde2714 that position aligns with a glycine residue. The red boxes highlight the positions of the conserved catalytic cysteine residues. The first occurrence shows Tde2714 containing a phenylalanine at this position. It is unlikely Tde2714 can participate in FGE catalysis due to the lack of conserved catalytic residues that promote FGE functionality. Multiple sequence alignments were generated using Clustal Omega. Q73J62_TREDE: Tde2714 from *T. denticola* ATCC 35405, FGE_MYCTU: FGE enzyme from *Mycobacterium tuberculosis* ATCC 25618, SUMF1_HUMAN: FGE enzyme from *Homo sapiens*, FGE_THECD: FGE enzyme from *Thermomonospora curvata* ATCC 19995.

To corroborate this initial thought, the deep learning function prediction tool DeepFRI was used⁹³. This model has been trained on over 40,000 known protein structures with experimentally verified functional annotations. With the improvements to the model algorithm through extensive training data sets, DeepFRI is robust enough to identify new functions on previously characterized proteins, as well as given functions to proteins with no experimental evidence on hand⁹³. As shown in Figure 11, DeepFRI was given the AlphaFold2 predicted structure of the proposed catalytic domain of Tde2714 (residues 378-663). The result of this exercise was a functional prediction of a cysteine-type endopeptidase.

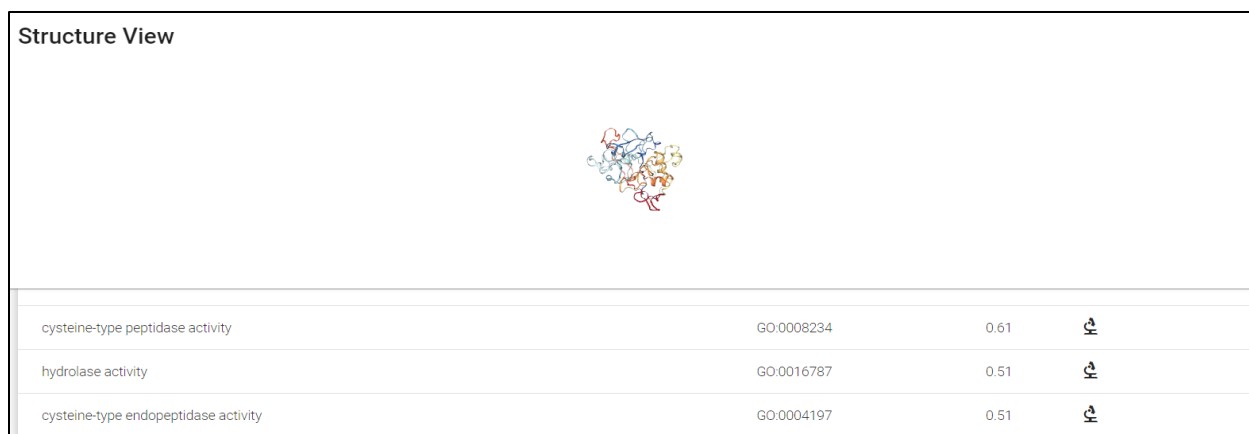


Figure 11. Deep learning model DeepFRI for predictive functional annotations on Tde2714 from *T. denticola*. Above is a screen view of the results produced by DeepFRI. This artificial intelligence learning model starts with identifying a large scope of potential cellular localization and biological processes the protein may be involved in and finishes with a small scope of direct functional annotations. AlphaFold2 was used to generate a predictive three-dimensional structure of the proposed catalytic domain of Tde2714 (residues 378–663). The most narrowed down functional annotation given for Tde2714 is that of a cysteine-type endopeptidase.

To further support the hypothesis of Tde2714 being a proteolytic enzyme, the I-TASSER suite of programs was used for their functional annotation prediction servers⁹⁴. The first program used was COACH, which generates protein-ligand binding interaction predictions and identifies potential ligand binding sites. The second program used was COFACTOR, which provides biological function annotations through comparisons to other homologous protein structures. As shown in Figure 12, the results from I-TASSER are comparable to proposed functions from DeepFRI. COACH had identified a predicted ligand binding site for a peptide substrate (shown as yellow spheres). COFACTOR had teased out a potential molecule that Tde2714 may interact with, N,N-dimethyl-2-pyrrolidin-1-ylquinazolin-4-amine (see Figure 11, right panel), that functions as a beta secretase (protease) inhibitor. Therefore, the predictive results of Tde2714 interacting with peptide substrates and protease inhibitors coincides with the predictive functional annotation provided by DeepFRI as a cysteine-type endopeptidase.

Lastly, initial SDS-PAGE gels showed banding patterns of multiple degradation products of Tde2714 during IMAC purification. Upon further research, the most prominent bands out of the many aligned with the molecular weights of the proposed catalytic domain of Tde2714 alone, and then with each linking repeat domain from the predicted structure. The primary sequence of Tde2714 was analyzed for protease recognition sequences to determine cut sites throughout the protein structure. The free coil regions connecting each linking domain were prioritized as these proteolytic degradation products would align with the banding pattern observed through SDS-PAGE. This produced three likely proteases that could produce the SDS-PAGE banding pattern observed: trypsin, proteinase K, and clostripain. The first two are serine proteases. Combining this knowledge with the results from DeepFRI, the suggested identity of Tde2714 as a clostripain-like protease (arginine-specific cysteine endopeptidase) became the most promising functional hypothesis.

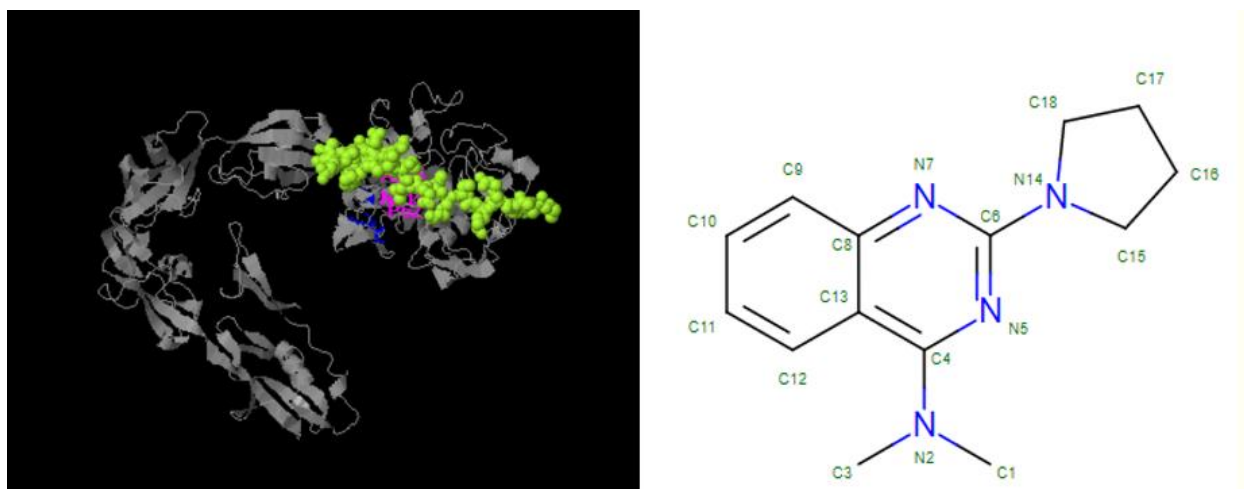


Figure 12. I-TASSER suite of programs for predictive functional annotations for Tde2714 from the pathogen *T. denticola*. Left: ligand identification using I-TASSER program COACH with Tde2714. A peptide substrate (yellow spheres) is predicted to be able to bind to the proposed catalytic domain. Right: ligand identification using I-TASSER program COFACTOR. N,N-dimethyl-2-pyrrolidin-1-ylquinazolin-4-amine has been proposed to interact with Tde2714. This molecule is a beta secretase (protease) inhibitor. Both images were taken directly from the results page generated by I-TASSER.

Using multiple sequence alignments (MSAs), Tde1414 had its primary sequence analyzed and compared against other Ppd enzymes, as well as other ThDP-dependent enzymes. Figure 13 highlights the comparison to other Ppds, while Figure 14 showcases the similarities shared between Tde1414 and other ThDP-dependent enzymes. Although a greater number of similarities had been identified between Tde1414 and other Ppd, the most significant result was the conservation of the GDG-X₂₅-NN motif proposed to facilitate ThDP cofactor binding in Tde1414, other Ppd, as well as other ThDP-dependent enzymes⁹⁵.



55

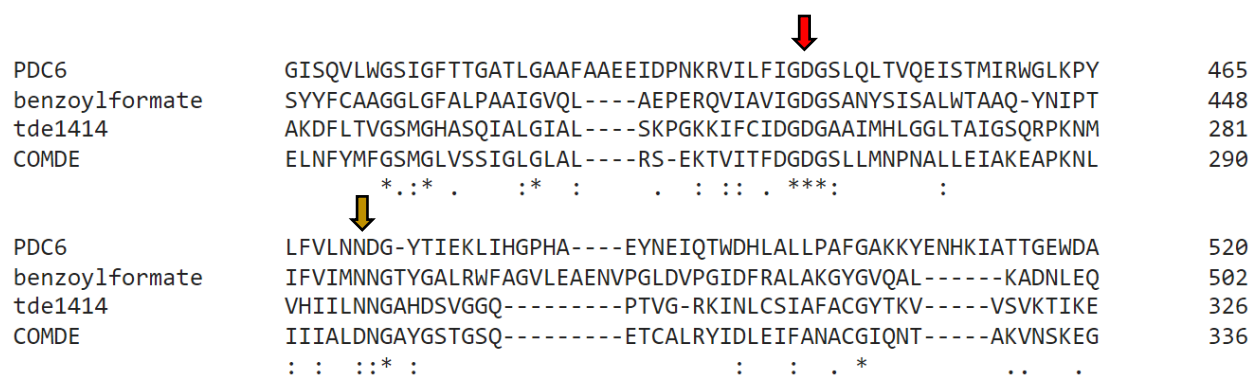


Figure 14. Multiple sequence alignment between other thiamine diphosphate-dependent enzymes and Tde1414 from the pathogen *T. denticola*. An MSA was created to compare the primary sequences of ThDP-dependent enzymes against Tde1414. Despite minimal sequence similarities, the single conserved motif expected to be found within all analyzed members was identified. The red arrow highlights the position of the conserved GDG motif. The gold arrow indicates the location of the conserved NN motif. The presence of these motifs is proposed to facilitate ThDP binding. Multiple sequence alignments generated using Clustal Omega. PDC6: pyruvate decarboxylase isozyme 3 from *Saccharomyces cerevisiae* ATCC 204508, benzoylformate: benzoylformate decarboxylase from *Pseudomonas putida*, tde1414: phosphonopyruvate decarboxylase from *T. denticola* ATCC 35405, and COMDE: sulfopyruvate decarboxylase from *Methanosarcina acetivorans* ATCC 35395.

To help support the hypothesis that Tde1511 is an iron transport protein within the COG3470 family, an MSA was generated using characterized members from this family of iron transporters with Tde1511. As shown in Figure 15, the results indicate that Tde1511 is likely related to these proteins and may hold a similar function as well. Within the compared sequences, the main sequence conservation identified was residues used for copper and iron coordination. Residues H94, M141, H148, and H178 for copper binding and residues E96, D98, and D145 that facilitate iron binding have all been characterized in homologous proteins to promote these interactions. The results indicate strong conservation for these residues between all members.



To further the hypothesis that Tde2714 is a clostripain-like protease, Figure 16 depicts an MSA generated with Tde2714 against the only recognized member and namesake of the clostripain family of proteases. When comparing Tde2714 with clostripain, a few similarities can be seen, with one cysteine and two histidine residues being conserved. Unfortunately, the identified catalytic cysteine of clostripain is not the cysteine that is conserved with Tde2714. Due to clostripain having a large sequence (526 residues) that may have skewed MSA attempts with the smaller proposed catalytic domain of Tde2714 (285 residues), a separate MSA was generated

```

tde2714      AHNPS-----YFQGSSHPTAS
CLOS_HATHI   EHANKANRLDGKNEFPEISTTSKYEANMGDPEVLKKFIDYCKSNYEADKYVLIMANHGGGA
              **      :      : :.*  :

tde2714      GESQEKRPVEQVSWFDAIAFCNELTRCCYS-----LGEAQCVCY
CLOS_HATHI   REKSNPRLNRAICWDDSNLDKNGEADCLYMGEISDHLTEKQSVDLLAFDACLMTAEVAY
              *..: *  . :.* *:      * : * *

```

Figure 16. Multiple sequence alignment between Tde2714 and clostripain shows a conserved His/Cys dyad. The primary sequences of Tde2714 and clostripain were compared to identify conserved residues that have potential to be involved in catalysis. When completed, two histidine and a cysteine residue were shown to be conserved. When viewed over a predicted structure of Tde2714, the highlighted conserved residues in red boxes show the highest likelihood of being close enough to perform proteolytic catalysis together. Multiple sequence alignment was generated using Clustal Omega. tde2714: sulfatase-modifying factor enzyme domain-containing protein from *T. denticola* ATCC 35405, and CLOS_HATHI: clostripain from *Hathewayia histolytica* (*Clostridium histolyticum*).

in Figure 17 with Tde2714 and putative clostripain-like protease sequences found on Uniprot.

This figure highlights a conserved histidine that matches the same position of the histidine conserved when Tde2714 was compared to clostripain. Similarly, a conserved cysteine residue is observed 6 residues away from the initial cysteine conserved from Figure 16. Therefore, these two images attempt to highlight potential conservation of a His/Cys dyad within Tde2714. More work is required to determine if these residues hold any functional properties.

```

Q65DM2_BACLD      VGKKNMDSPPGSVTDFITWGVKTYPAKKYVLIFWCHGLGSVDGYGGDENFGNKKMKISELQ
A0A1N6MVN2_9GAMM  LPNENMDSSDTLSSFIKWGLKSFSPDSYYIFFWCHGLGAAGGYGGDENYGNKKMSIGNMA
A0A6N1NL89_9VIRU  LPKQNMSPMTLSNFIEKTLAKATYRRTCLILACHGSGWFLKTEKHTVMGIPDMMN----
A0A1I0MGA6_9FIRM  TELRDMGETDTLKDFLKYGVEMYPAEHTAVILWNHGGGAMGGCCYDELFDLSHLTLPEMR
tde2714           -----DPHKVSLTAYRIGTTTEVTQELYELVMAHNPSYFQGSSHPTASGESQEKR----
                  ..  :.      :.  *  .

Q65DM2_BACLD      SGIKTAYEHTKQKFDLIGFDNCKMAGIETAYALRDYGYMLASVDYTNQNGWDYKRALQS
A0A1N6MVN2_9GAMM  LAFKNS-SMKQSKFEIIGFDNCKMANLETAYVLKSYGKYLVGSVDYTHKNGWNYKDIVES
A0A6N1NL89_9VIRU  -----ALKKVGIFIDLLIFDTCMMSVLECMNELVGTVEYVYVANQDYAENNGFVTKNLLSL
A0A1I0MGA6_9FIRM  DAFDEACAGYDKKLDFIAYDCCMACLELANILTPYADYFIASQETMPGTGFDYVVGFGSY
tde2714           -----PVEQVSWFDAIAFCNELTRCCYSLGEAQCVCYTYNGHTYTVEDAQAHNVPVMD
                  :      *      *      .

```

Figure 17. Multiple sequence alignment between putative clostripain-like enzymes and Tde2714 from the pathogen *T. denticola*. A multiple sequence alignment was completed between the primary sequence

of the proposed catalytic domain of Tde2714 with other putative clostripain-like proteases. Minimal sequence conservation was identified outside of proposed residues determined to be conserved within the clostripain family of proteases. A conserved serine and aromatic residue were conserved, implicated in peptide backbone stabilization in the binding pocket. A diad of His/Cys was also conserved shown in red boxes. Multiple sequence alignment was generated using Clustal Omega. Q65DM2_BACLD: clostripain from *Bacillus licheniformis* ATCC 14580, A0A1N6MVN2_9GAMM: clostripain from *Xenorhabdus innexi*, A0A6N1NL89_9VIRU: clostripain from *Tupanvirus soda lake*, A0A1I0MGA6_9FIRM: clostripain family protein from *Ruminococcaceae bacterium* KH2T8, and tde2714: sulfatase-modifying factor enzyme domain-containing protein from *T. denticola* ATCC 35405

3.1.5 Identification of a putative Iron Import System

Using a program offered through <https://microbesonline.org/>, genes are predicted for their probability to be in an operonic system based on variables, such as direction of translation and relative distance between the stop codon of one gene and the start codon of the next. It was identified that Tde1508-1509 and Tde1510-1518 were predicted to form an operon. The involved genes were bioinformatically characterized using Uniprot, Expasy ProtParam, STRINGS, BLASTp, and InterPro in an attempt to assign putative functions to each member and create a plausible method for iron import into the cytoplasm. As shown in Table 2, main functions identified were for iron import, facilitating electron transfer chemistry, and macrolide export to get lipoproteins across cell membranes. Tde1508, 1509, 1514, 1515, 1516, and 1518 are all involved in ABC transport systems with an emphasis on MacB and FtsX-like domains. These two domains are involved in macrolide export and positioning lipoproteins into the cell membrane. Tde1510 and Tde1511 show direct functions in iron binding and transport. Tde1510 shows high homology to other Ftr1-like permeases that function in iron translocation. Further analysis unveiled this protein contains the iron binding motifs “REGAE” and “REGLE” on its two surface exposed helices 1 and 4, coinciding with other iron permeases in this family. Evidence of a Fth1 domain alludes to this permease holding ferroxidase activity as well.

Tde1512 and Tde1517 show functions in electron transfer chemistry. Mirroring other iron import systems, such as the FtrABCD system, these proteins may have functions in resetting redox potentials of proteins within this operon to allow consecutive rounds of iron import to occur.

Table 2. Bioinformatic analysis of the gene products predicted to be in an iron import operon with Tde1511 from the pathogen *T. denticola*.

Gene	# Amino Acids	MW (kDa)	STRINGS	BLASTp (smallest E-value)	InterPro Domains
<i>Tde1508</i>	131	14.4	No annotation	DUF4418 (7e-08)	DUF4418, possibly MacB domain
<i>Tde1509</i>	221	24.5	ATP binding/ lipoprotein release	ABC transporter ATP binding protein (3e10-71)	ATPase, MacB, P-loop, Lipoprotein transport
<i>Tde1510</i>	422	46.9	High affinity iron transporter	Ftr1 family iron permease (4e-93)	Iron permease, Ftr1, Fip1, EfeU, Fth1
<i>Tde1511</i>	208	22.6	Pathogen-specific surface antigen	Iron transporter (2e-57)	Periplasmic metal-binding protein, Fe ²⁺ transport protein, membrane lipoprotein lipid attachment site
<i>Tde1512</i>	422	48	Membrane protein	FeS containing protein (1e-111)	FeS containing protein, FtrD-like
<i>Tde1514</i>	425	46.3	ABC permease protein	ABC transporter permease/ FtsX-like permease (2e-160)	ABC3 transporter permease, FtsX-like permease
<i>Tde1515</i>	373	41	ABC permease protein	ABC transporter permease/ FtsX-like permease (5e-121)	ABC3 transporter permease, FtsX-like
<i>Tde1516</i>	230	25.6	ATP binding protein	ABC transporter ATP binding protein (4e-117)	ATPase, MacB, P-loop, Lipoprotein transport

<i>Tde1517</i>	149	16.5	Conserved hypothetical protein	FMN binding protein (1e-13)	FMN-binding protein
<i>Tde1518</i>	403	45	ABC permease protein	ABC transporter permease/ FtsX-like permease (1e-114)	ABC3 transporter permease, FtsX-like permease

3.2 Protein Expression and Purification

3.2.1 Transformations and Induction Optimization

Transformation protocols using *Tde1414*-pET-28b, *Tde1511*-pET-21b, and *Tde2714*-pET-21b vectors with both NEB 5- α and BL21 (DE3) *E. coli* cells were successful. This was evident through the selective growth of colonies on agar plates supplemented with antibiotics. Using the Qiagen DNA miniprep kits, stock concentrations of each vector was determined through the spectroscopic analysis of the samples at 260nm. *Tde1414*-pET-28b was isolated at a concentration of 47.2 ng/ μ L, *Tde1511*-pET-21b was measured at a concentration of 36.3 ng/ μ L, and the concentration of *Tde2714*-pET-21b was observed to be 58.8 ng/ μ L.

Due to the induction parameters for *Tde1414* and *Tde2714* being previously identified, induction parameters for *Tde1511* required determination. An induction trial for this gene product was conducted which looked at variables of induction, such as the optical density of the bacterial culture (0.6, 0.8, 1.0), the temperature of induction and for how long it occurred for (37°C for 3h, 30°C for 5h, 23°C for 10h, or 16°C for 16h) , as well at the concentration of IPTG that induced the greatest production of protein (0.05mM, 0.1mM, 0.5mM, and 1mM). As shown in Figure 18, it was determined that the optimal protein induction parameters for *Tde1511* that

produced the largest yield of protein was an OD₆₀₀ of 0.6, an IPTG concentration of 0.5mM, and inducing at 30°C for 5h.

TDE 1511

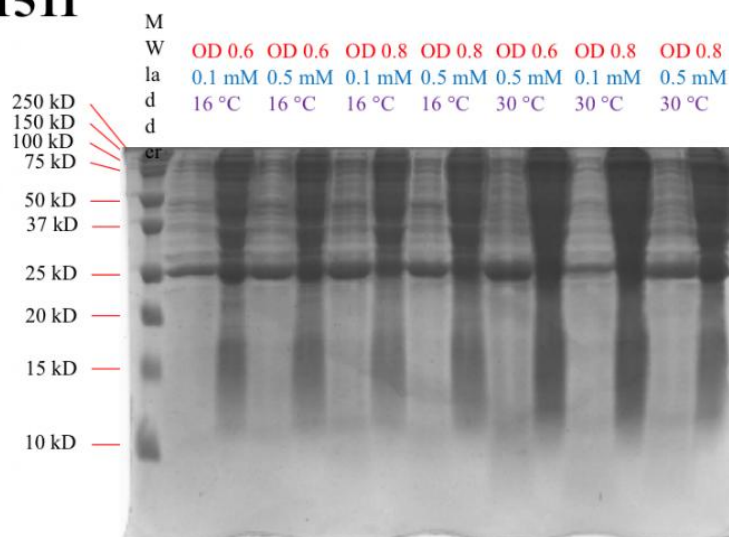


Figure 18. Induction parameter optimization for the gene target Tde1511 from the pathogen *T. denticola*. Variables of induction to produce Tde1511 were optimized to identify the condition which produced the largest yield of soluble protein. Optical densities at 600nm of 0.6, 0.8, and 1.0 were used at IPTG concentrations of 0.05mM, 0.1mM, 0.5mM, and 1mM. These variables were also tested against the temperature and time for induction using 37°C for 3h, 30°C for 5h, 23°C for 10h, or 16°C for 16h. Induced cells were lysed through four cycles of freeze-thaw into a buffer of 100mM Bis-Tris pH 6.5 and 300mM sodium chloride and centrifuged to separate the soluble and insoluble fractions. Samples of each soluble and insoluble fractions were analyzed via SDS-PAGE at 180 volts for 55min. Gels were stained using Coomassie brilliant blue protein dye and visualized using a VersaDoc molecular imager. The expected molecular weight of Tde1511 is 22.6 kDa.

3.2.2 Primary Purification through Ni-NTA IMAC

Following the induction of each protein target, cell lysis via chemical or mechanical methods was successful in isolating appropriate concentrations of protein to be used in downstream structural or functional assays. Crude cell lysates were incubated with Ni-NTA resin

for 40min to allow time for the nickel cations of the resin to coordinate with the electron dense hexa-histidine affinity tags recombinantly expressed on each protein target. The resin coordinated with each protein target was placed back into the column and a wash buffer containing 5mM imidazole was flowed through each column to allow the disassociation of weakly bound proteins from the resin. After a gradient of imidazole was flowed over the column from 5mM to 500mM, collected fractions were analyzed via SDS-PAGE as shown in Figure 19. Fractions containing the purest protein were pooled together and subjected to dialysis for 16h to remove excess imidazole and sodium chloride.

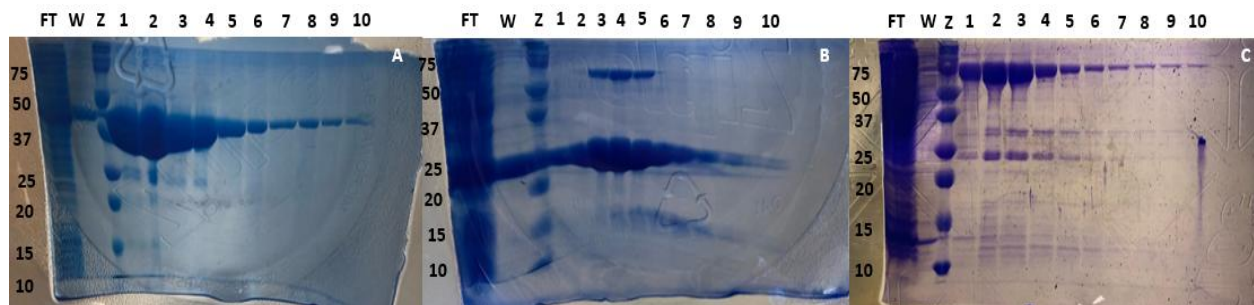


Figure 19. SDS-PAGE analysis of the primary IMAC purifications for Tde1414, Tde1511, and Tde2714 from the pathogen *T. denticola*. Crude cell lysates were incubated with Ni-NTA resin and subjected to IMAC purification using an imidazole gradient from 5-500mM. Results for Tde1414 (A), Tde1511 (B), and Tde2714 (C) are shown. All fractions collected from the IMAC purification had a 2 μ L aliquot removed and analyzed via SDS-PAGE. Coomassie brilliant blue was used as a protein stain. Numbers lining the top of each gel represent the fraction number sampled in that lane. Values up the left of each gel represent the molecular weight (in kDa) of the molecular weight standard (Z). The flow through (FT) and wash (W) fractions were also analyzed. Tde1414 is predicted to be about 44 kDa, Tde1511 is predicted to be about 23 kDa, and Tde2714 is predicted to be about 74 kDa.

3.2.3 Anion Exchange and DLS

Once dialysis was complete and the samples were briefly centrifuged to remove any denatured protein, FPLC was used to perform anion exchange chromatography. Due to each of the pH of each protein's respective buffer being more than 1.0 pH units above their theoretical

pI, a net negative surface charge was imposed over their surfaces, allowing anion exchange to be an effective secondary purification method. Once the protein targets were bound to the chromatography column, an increasing sodium chloride gradient from 0-1M was applied over the column. The free chloride anions competed for coordination with the positively charged quaternary amine of the Q Sepharose resin, eluting the target protein. Fractions spanning the peak of the FPLC-generated chromatogram were taken and analyzed via SDS-PAGE for protein concentration and purity as shown in Figure 20. These same fractions were also used in DLS analysis to ensure target protein purity and homogeneity prior to crystallization trials. In data not shown, each target had a radius between 3-5nm and a polydispersity below 25%. With molecular weights identified that coincided with expected monomers (Tde1511 and Tde2714) and dimers (Tde1414) in solution, these samples were deemed appropriate to move forward with in crystallization trials. Fractions of Tde1511 and Tde2714 that showed high purity and protein concentration were pooled together for protein concentration. Anion exchange for Tde1414 produced an abnormal chromatogram with a significant signal produced for protein in the flow through. Upon SDS-PAGE analysis, a reverse purification was observed. Pure protein was detected in the flow through, which is typically where contaminants would be present. Thus, the flow through was used in protein concentration.

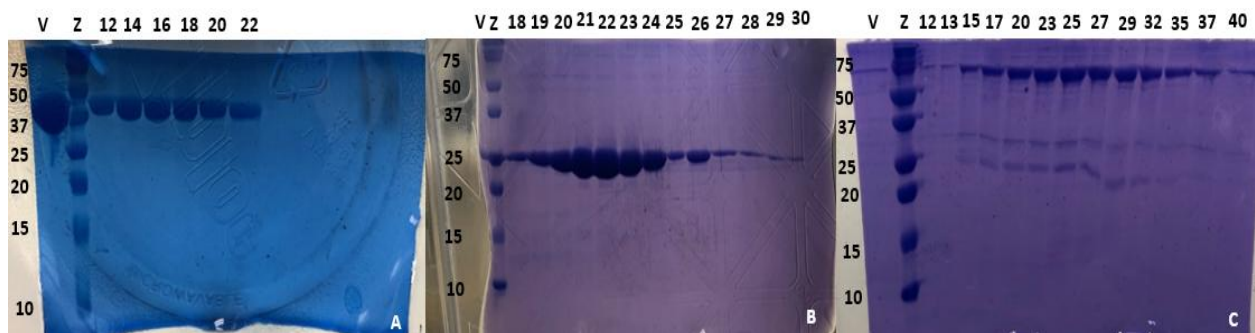


Figure 20. SDS-PAGE analysis of the secondary anion exchange chromatography purifications for Tde1414, Tde1511, and Tde2714 from the pathogen *T. denticola*. After dialysis, protein samples were further purified through anion exchange chromatography to achieve ultra pure, homogeneous samples for crystallization. Results for Tde1414 (A), Tde1511 (B), and Tde2714 (C) are shown. All fractions were sampled and had their contents separated via SDS-PAGE. Coomassie brilliant blue was used as a protein stain. Numbers lining the top of each gel represent the fraction number sampled in that lane. Values up the left of each gel represent the molecular weight (in kDa) of the molecular weight standard (Z). The void (V) fraction was analyzed to determine if protein failed to purify through the column. Tde1414 is predicted to be about 44 kDa, Tde1511 is predicted to be about 23 kDa, and Tde2714 is predicted to be about 74 kDa.

3.3 Crystallization

3.3.1 Crystal Screens

Protein concentration occurred following anion exchange chromatography and the protein concentration of the sample was determined using Beer's Law and the predicted extinction coefficients for the respective proteins. Tde1414 was concentrated to a final value of 45 mg/mL in a solution containing 20mM Tris pH 7.5, 20mM NaCl, 0.2mM MgCl₂, 0.2mM ThDP, and 1mM DTT. Tde1511 was concentrated to a final concentration of 51 mg/mL and diluted to a working concentration of 10 mg/mL in a solution containing 20mM Bis-Tris pH 6.5. Tde2714 was concentrated to a final concentration of 10 mg/mL in a solution containing 50mM Tris-HCl pH 8.0. Due to the initial crystallization parameters for Tde2714 being established by a previous student from the Suits laboratory, the MCSG 1-4 suite of sparse matrix screens were used to determine crystallization conditions for Tde1414 and Tde1511.

Prior to crystallization, Tde1414 was diluted to a working concentration of 30 mg/mL. A Gryphon Crystallization Robot was used to prepare sitting drop vapour diffusion crystallization plates using protein to reservoir ratios of 1:1, 1:2, and 2:1 that were left to incubate at 18°C. After a 6-week incubation time, Figure 21 depicts rod-like crystals observed in condition D7 from MCSG 1 containing a reservoir solution of 0.1M sodium citrate pH 5.5 and 20% (w/v) PEG

3000. No other condition produced any evidence of protein crystallization, so this solution was further optimized to promote the growth of larger crystals to be used in X-ray diffraction analysis.

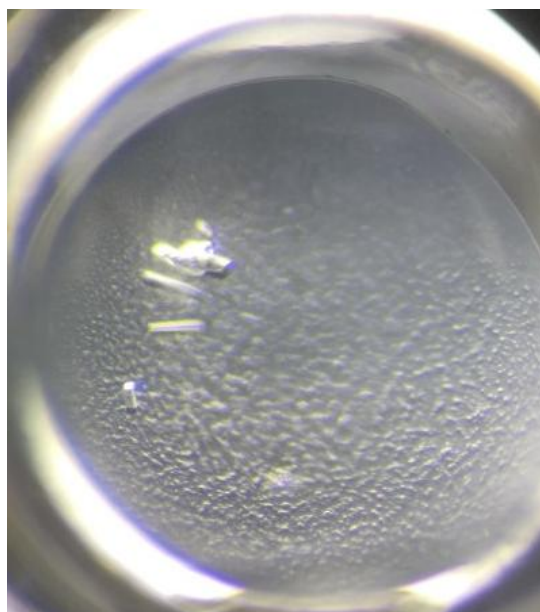


Figure 21. Initial sparse matrix crystallization of Tde1414 from the pathogen *T. denticola*. Tde1414 bound to cofactors thiamine diphosphate and Mg^{2+} was concentrated to 30 mg/mL for crystallization condition determination through sparse matrix screens. After 6 weeks of incubation at 18°C, rod-like crystals were found in the condition D7 from MCSG 1 containing 0.1M sodium citrate pH 5.5 and 20% (w/v) PEG 3000 at a protein to reservoir ratio of 1:1.

Using the working concentration of 10 mg/mL of Tde1511, a Gryphon Crystallization Robot was used to set up sitting drop vapour diffusion crystal plates using the MCSG 1-4 suite of sparse matrix crystal screens. Using protein to reservoir ratios of 1:1, 1:2, and 2:1, crystal trays were left to incubate at 18°C. After a 2-week period, tetragonal crystals were observed in condition H9 from the MCSG 2 crystal screen containing 0.1M sodium acetate:HCl pH 4.6 and

2.5M ammonium sulfate. This condition was further optimized to produce larger, diffraction grade crystals for structural determination.

3.3.2 *Crystal Optimization*

When optimizing the crystallization condition of Tde1414, initial optimizations using different buffer systems, pH's, salt concentrations, and organic solvents led to the production of two or three single crystals across 48 unique conditions with the use of PEG 3000. Finally, a switch to PEG 3350 was made. With this, two optimized plates were made, one that varied the PEG concentration (5% (w/v), 10% (w/v), 15% (w/v), 20% (w/v)) and pH (4.6, 4.9, 5.2, 5.5, 5.8, 6.1), while the other varied PEG concentration and protein concentration (25-50 mg/mL). Using the hanging drop vapor diffusion method in a 24-well plate, conditions containing 20% (w/v) PEG 3350, 0.1M sodium citrate pH 5.5 and 5.8, and protein concentrations ranging from 25 mg/mL to 50 mg/mL all produced a number of large, diffraction grade crystals. Figure 22 highlights one of the wells that showcased improved crystal growth after the final round of optimizations.

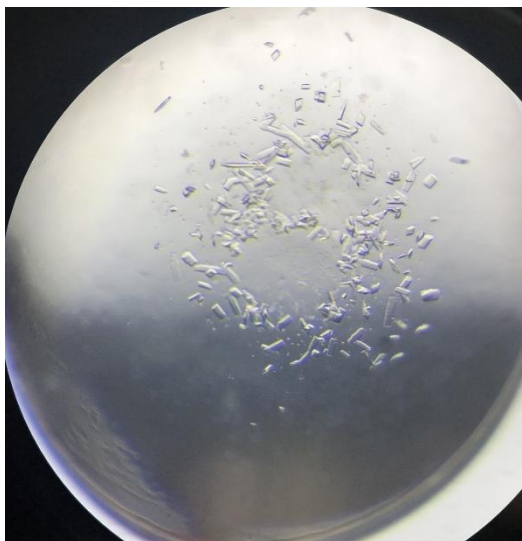


Figure 22. Optimized protein crystallization of Tde1414 from the pathogen *T. denticola*. The protein target Tde1414 was purified to homogeneity and concentrated to a working solution of 30 mg/mL for crystallization. The optimized crystallization plate was incubated at 18°C for 6 weeks. At this time, protein crystallization was observed in a condition containing 0.1M sodium citrate pH 5.5 and 20% (w/v) PEG 3350 at a protein to reservoir ratio of 1:1. These crystals were left to incubate for an additional 3 weeks to allow larger crystals to grow.

Optimization of Tde1511 protein crystallization followed similar methods as Tde1414. Using the initial condition of 2.5M ammonium sulfate and 0.1M sodium acetate pH 4.6, ammonium sulfate concentrations ranging from 1.8M to 2.8M were tested. The addition of organic compounds was included, such as glycerol and ethylene glycol, and pH ranges from 3.6 to 4.6 were used. The buffering species were also tuned using sodium citrate and succinic acid buffers due to their buffering regions being similar to that of sodium acetate. Although protein crystallization was still achieved using sodium acetate buffers, significant optimization was achieved using reservoir solutions containing sodium citrate pH 4.6, the addition of 5% (v/v) glycerol, and ammonium sulfate concentrations ranging from 2.0 M to 2.8 M. This crystallization condition was also used for copper reconstituted Tde1511 with success. Figure 23 exemplifies the success of this crystallization optimization through the production of large diffraction grade protein crystals.

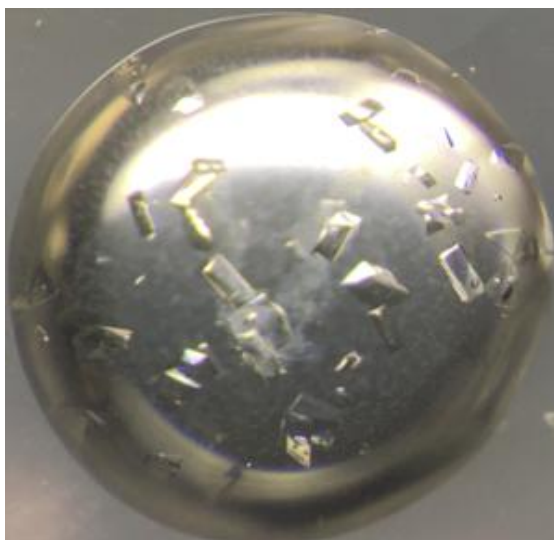


Figure 23. Protein crystal optimization of the protein target Tde1511 from the pathogen *T. denticola*. Tde1511 was purified to homogeneity and concentrated to 12 mg/mL for crystallization trials. Optimized plates were created and incubated at 18°C for two weeks. At this time, large tetragonal crystals were observed in a condition containing 50mM sodium citrate pH 4.6, 2.4 M ammonium sulfate, and 5% (v/v) glycerol. These crystals were harvested at a later date for X-ray diffraction analysis.

Although the crystallization condition for Tde2714 had been previously established, this target seemed refractive to crystallization at first. With sparse matrix screens reattempted to no prevail, a different avenue to optimization was taken. A small handful of Tde2714 protein crystals were harvested that had been grown over multiple crystallization plates in conditions containing 50mM Tris-HCl pH 7.5, 0.2 M sodium nitrate, 20% (w/v) PEG 3350, and 2.5% (v/v) glycerol; these crystals were placed into 100µL of this mother liquor and extensively vortexed to create a crystal seed stock. Using this crystal seed stock, two optimized crystal plates were designed. The first plate varied PEG 3350 concentrations over a range of 19.6% (w/v) to 20.6% (w/v) while using glycerol concentrations of 2% (v/v), 2.5% (v/v), and 3% (v/v). A row containing ethylene glycol at 5% (v/v) was also tested. Drops containing 1µL reservoir solution, 1.5µL purified Tde2714 at 10 mg/mL, and 0.5µL of the crystal seed stock were created to

promote larger crystal growth by introducing crystal nuclei right away. A secondary plate was created with all conditions containing 50mM Tris-HCl pH 7.5, 0.2 M sodium nitrate, 20% (w/v) PEG 3350, and 2% (v/v) glycerol. The crystal seed stock was then used to create 5 serial dilutions at 1:10 dilution factors to create seed stock samples at 1:10, 1:100, 1:1000, 1:10000, and 1:100000 dilutions. 0.5 μ L of these crystal seed samples were then introduced to drops containing 1 μ L reservoir solution and 1.5 μ L purified Tde2714 at 10 mg/mL at time points of 0 days, 7 days, 14 days, and 21 days. After one month of incubation at 18°C, it was observed that only the crystal seed dilutions of 1:10000, and 1:100000 added at day 0 were effective at generating large, diffraction grade crystals of Tde2714 as shown in Figure 24.

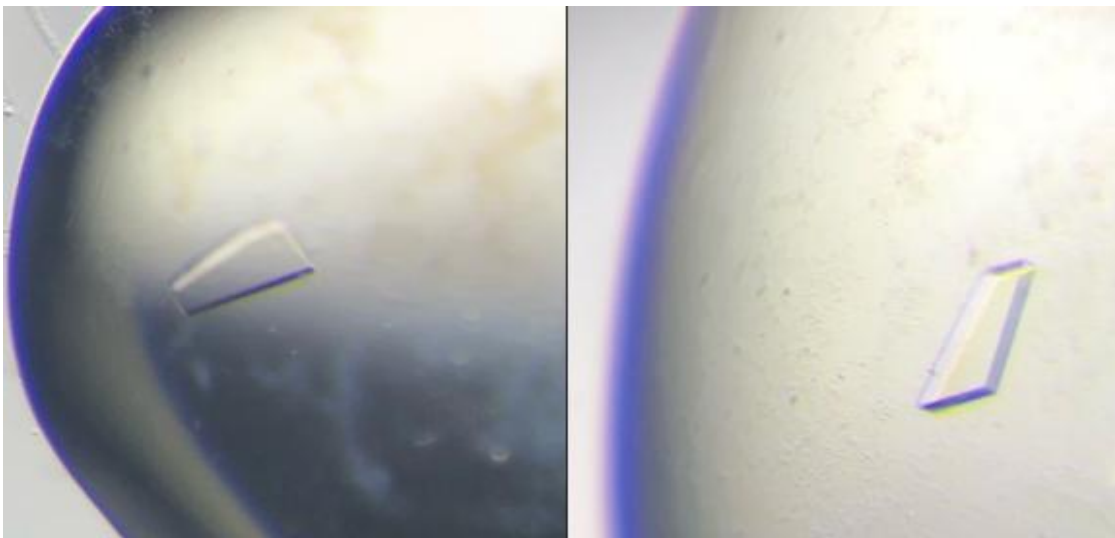


Figure 24. Using crystal seeding techniques to promote effective crystallization of Tde2714 from the pathogen *T. denticola*. Using a crystal seed stock of Tde2714, optimization of protein crystallization was carried out in conditions containing 50mM Tris-HCl pH 7.5, 0.2 M sodium nitrate, 20% (w/v) PEG 3350, and 2% (v/v) glycerol. The crystal seed stock was then diluted five times at a 1:10 dilution factor to create 1:10, 1:100, 1:1000, 1:10000, and 1:100000 seed stock dilutions. 0.5 μ L of these samples were added to drops containing 1 μ L reservoir solution and 1.5 μ L purified Tde2714 at 10 mg/mL at time points of 0 days, 7 days, 14 days, and 21 days. After one month of incubation at 18°C, large rhomboid crystals were observed in conditions using the 1:10000 seed dilution (left) and 1:100000 seed dilution (right) when introduced at day 0.

Following optimized protein crystal growth, crystals were harvested for X-ray diffraction analysis. Crystals of Tde1414 were placed into a cryoprotectant containing 0.1 M sodium citrate pH 5.5, 20% (w/v) PEG 3350, and 35% (v/v) glycerol. Apo and copper reconstituted Tde1511 protein crystals were cryoprotected in a solution containing 0.1 M sodium acetate pH 4.4 and 3.7 M ammonium sulfate. Other copper reconstituted Tde1511 crystals were soaked in a similar solution with the addition of 3mM FeCl₃ for 2h before being placed in the previously mentioned cryoprotectant for data analysis. Tde2714 protein crystals were cryoprotected in a solution containing 50mM Tris-HCl pH 8.0, 0.2 M sodium nitrate, 20% (w/v) PEG 3350, and 30% (v/v) glycerol. All cryoprotectants were successful in the mitigation of ice crystal formation upon flash freezing in liquid nitrogen.

3.4 X-Ray Diffraction Analysis and Structural Solutions

3.4.1 Initial Data Quality

After completing X-ray diffraction analysis at the synchrotron, data files were processed and scaled. Shown in Table 3 is the scaled experimental data statistics produced by the software Scala from the ccp4 suite of programs for the analysis of Tde1414. At an overall resolution of 1.95 Å, a space group of P2₂1₂1, and 2 molecules found in the asymmetric unit, a completeness of 100% was achieved. A CC(1/2) value of 0.999 was observed, meaning reflection intensities measured from one half of the data correlate very well with the other half, indicating reliable reflection data. I/sigma(I) at this resolution limit is 2.3 in the inner shell, which means an appropriate signal was detected over the background noise (cut off is a value typically of 2), resulting in an appropriate data resolution cut off to obtain meaningful structural information.

Table 3. Scaled crystallographic data statistics for the structural analysis of Tde1414 from the pathogen *T. denticola*.

Protein Target	Tde1414
Wavelength	0.95371 Å
Resolution Range	45.96 – 1.95 Å (2.06 – 1.95 Å)
Space Group	P22 ₁ 2 ₁
Unit Cell [a, b, c (α, β, γ)]	67.89 Å, 99.42 Å, 120.53 Å (90.0°, 90.0°, 90.0°)
Multiplicity	13.4 (14.0)
Completeness	100% (100%)
I/sigma(I)	13.2 (2.3)
CC(1/2)	0.999 (0.715)
R _{merge}	0.133 (1.437)
Total Number of Unique Observations	60 241 (8680)

With these variables in mind, Table 4 outlines the scaled experimental reflection data used in the structural determination of the apo, copper reconstituted, and copper/iron reconstituted structures of Tde1511. Initial statistics for the apo monomeric structure reveal a space group of C222₁ with one molecule in the asymmetric unit. A completeness of 100% was observed throughout the data. An I/sigma (I) of 3.5 in the highest resolution shell indicates an appropriate data resolution range was used. A CC(1/2) of 0.998 also supports the quality of the data used in structural determination. For the dimeric copper bound structure, a space group of P2₁2₁2₁ was identified with two molecules in the asymmetric unit. At a resolution of 1.89 Å, a completeness of 99.9% was achieved for the overall structure. An I/sigma(I) of 3.5 within the highest resolution shell indicates that this resolution cut off is appropriate for this parameter where a meaningful signal was detected over the background noise. A CC(1/2) of 0.997 also indicates a strong correlation between the reflections measured between each half set of data, indicating that a quality data set has been measured. Analyzing the copper and iron reconstituted data of Tde1511, a space group of P6₃ was identified with two molecules in the asymmetric unit. At a resolution cut off of 1.82 Å, a completeness of 99.9% was achieved for this dataset. An

I/sigma(I) of 2.1 within the highest resolution shell was found, and a CC(1/2) of 0.999 indicates a high quality data set was collected.

Table 4. Scaled crystallographic data statistics for the structural analysis of apo, copper reconstituted, and copper/iron reconstituted structures of Tde1511 from the pathogen *T. denticola*.

Protein Target	Apo Tde1511	Cu-Tde1511	Cu/Fe-Tde1511
Wavelength	0.9686	0.95374	0.95374
Resolution Range (Å)	29.62 – 1.78 (1.84 – 1.78)	42.17 – 1.89 (1.96 – 1.89)	45.57 – 1.82 (1.89 – 1.82)
Space Group	C 2 2 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 6 ₃
Unit Cell (a, b, c)	54.74 74.87 79.80	31.51 78.16 150.27	105.25 105.25 79.59
Unit Cell (α , β , γ)	90 90 90	90 90 90	90 90 120
Multiplicity	6.6 (6.9)	12.6 (12.2)	19.3 (14.7)
Completeness	100.0 (100.0)	99.9 (100.0)	99.9 (99.8)
I/sigma(I)	15.2 (3.5)	11.9 (3.5)	18.1 (2.1)
CC(1/2)	0.998 (0.950)	0.997 (0.853)	0.999 (0.863)
R _{merge}	0.061	0.139	0.084
Total Number of Unique Observations	16097 (2296)	30767 (4423)	44957 (6531)

3.4.2 Molecular Replacement and Structural Solutions

Due to absence of any structural information present in the literature for phosphonopyruvate decarboxylases, a successful molecular replacement solution was found using an AlphaFold2 predicted structure of Tde1414. This result was found very recently, so further structure refinement is necessary. The structure of Tde1414 shown in Figure 25 highlights the most refined structure obtained at the time of writing with an R-work of 0.182 and an R-free of 0.212. The dimeric complex includes residues 1-376 from chain A and residues 1-375 from chain B. The N-terminal domain of the first monomer interacts intimately with the C-terminal domain of the second monomer within the complex. Each domain comprises of an $\alpha/\beta/\alpha/\beta$ structure that forms a central beta pleated sheet surrounded by alpha helices. The N-terminal

domain and C-terminal domain are connected via a random coil that directly bridges the two subunits. Figure 26 takes a closer look at the cofactor binding interactions visible within the structure. The magnesium ion is coordinated with conserved GDG-X₂₅-NN residues D260 and N288, as well as the carboxylic acid group from A290, a water molecule, and two oxygens from the pyrophosphate in an octahedral coordination geometry. The pyrophosphate is also stabilized through interactions with the amino groups from A210 and A262. For the thiazolium and pyrimidine rings, interactions are noted with E49, Q72, S74, A210, M235, H291, S293, and V294. These polar residues provide hydrogen bonding stability to the sulphur and nitrogen atoms throughout ThDP, and the hydrophobic residues supply hydrogen bonding interactions to stabilize the molecule's carbon backbone within the binding site.

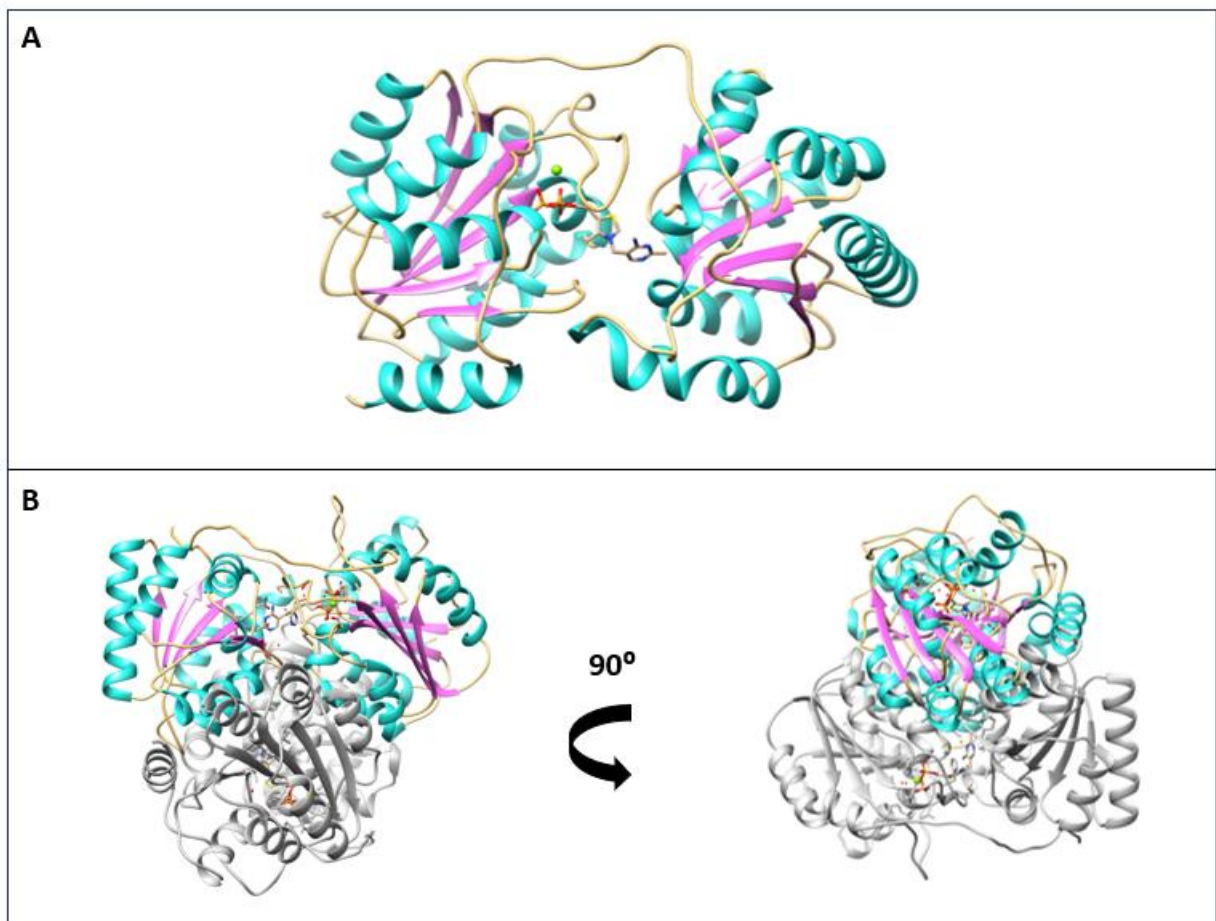


Figure 25. Novel structural determination of Tde1414, a phosphonopyruvate decarboxylase, at a resolution of 1.95 Å from the pathogen *T. denticola*. The first reported structure of a phosphonopyruvate decarboxylase is depicted in its monomeric state (chain A of the dimeric complex) (A) as well as its homodimeric complex (B) which was identified in the asymmetric unit of the protein crystal. Bound within the structure is the presence of two ThDP and two Mg^{2+} ions that associate within each monomer. The magnesium ion is used to anchor the pyrophosphate moiety within the C-terminal domain. Various other amino acids provide hydrogen bonding interactions or heteroatom stabilization to secure ThDP within its binding site. This results in the C-terminal domain being define as the pyrophosphate (PP) binding domain while the N-terminal domain facilitates the pyrimidine (Pyr) binding interactions. Structures were rendered in Chimera version 1.11.2 with alpha helices coloured cyan, beta sheets coloured purple, and random coils coloured peach. In the dimeric structure, one monomer is coloured grey to allow better differentiation between the two interacting monomers.

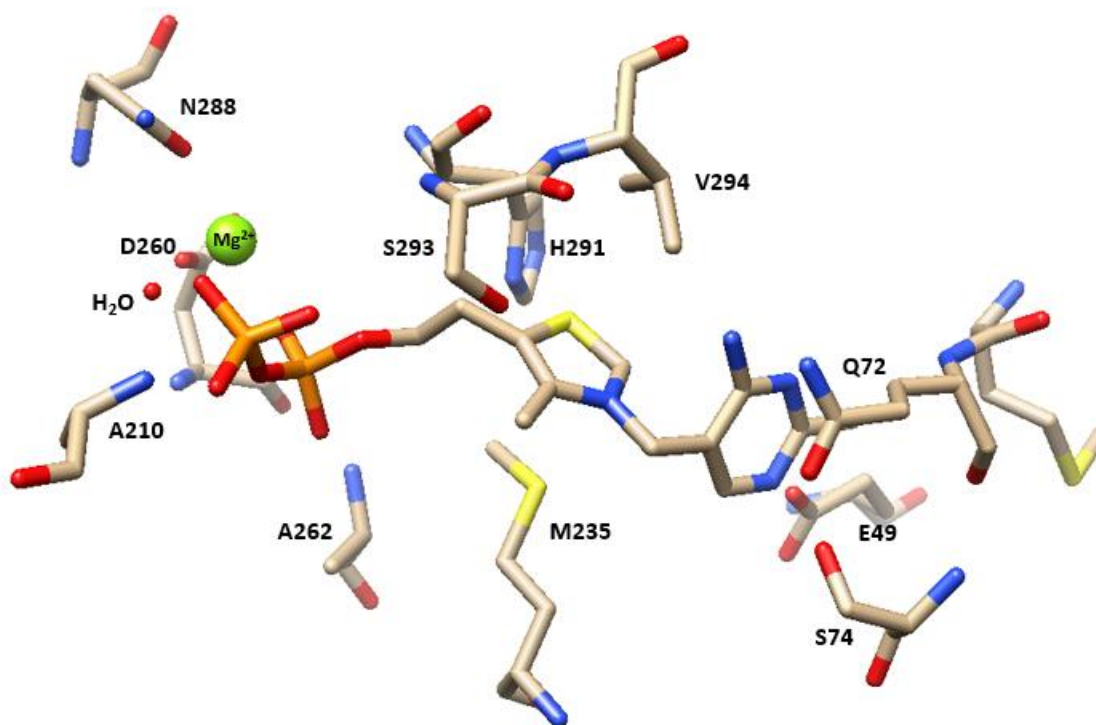


Figure 26. Thiamine diphosphate binding pocket in Tde1414. Removal of the protein back bone reveals the clear interactions which anchor the cofactors ThDP and Mg^{2+} within their binding sites. The magnesium cation is coordinated in an octahedral geometry through interactions with two oxygens from the pyrophosphate, OD1 from D260, OD1 from N288, a water molecule, and a carboxylic acid oxygen from the backbone of A290 (not pictured). The pyrophosphate group is anchored in place through its coordination with Mg^{2+} , nitrogen of the amino groups from the backbone from A210 and A262, and various water molecules (not pictured). The thiazolium ring stabilizes its sulphur with NE1 of H291, and its nitrogen with the δ -sulphur of M235. Both S293 and V294 are observed to provide hydrogen bonding interactions with this region as well. The pyrimidine ring is hydrogen bonded through interactions with its nitrogen atoms to residues E49, Q72, and S74. Structures were rendered in Chimera version 1.11.2. Red atoms represent oxygen, blue atoms represent nitrogen, yellow atoms represent sulphur, orange atoms represent phosphorus, and beige atoms represent carbon.

With a handful of homologous proteins to Tde1511 already having their crystal structure solved, a successful molecular replacement solution was found using the structure of Tp34 (PDB 3PJJ) from *Treponema pallidum* which shares a sequence similarity of 69%. All three structures presented of this protein have been refined to completion and accepted into the Protein Data Bank (Apo: PDB 8T7K, Cu: PDB 8T7L, Cu/Fe: PDB 8T7M). All structures of Tde1511 showcased in Figure 27 contain the expected amino acid residues from 57-208 with the addition of a few histidine's at the N-terminus from the recombinant affinity tag. Apo Tde1511 was resolved to a resolution of 1.78 Å in the space group C222₁ with one molecule in the asymmetric unit. Copper reconstituted Tde1511 was solved to a resolution of 1.89 Å in the space group P2₁2₁2₁ with two molecules in the asymmetric unit. Copper and iron reconstituted Tde1511 was solved to a resolution of 1.82 Å in the space group P6₃ with two molecules in the asymmetric unit.

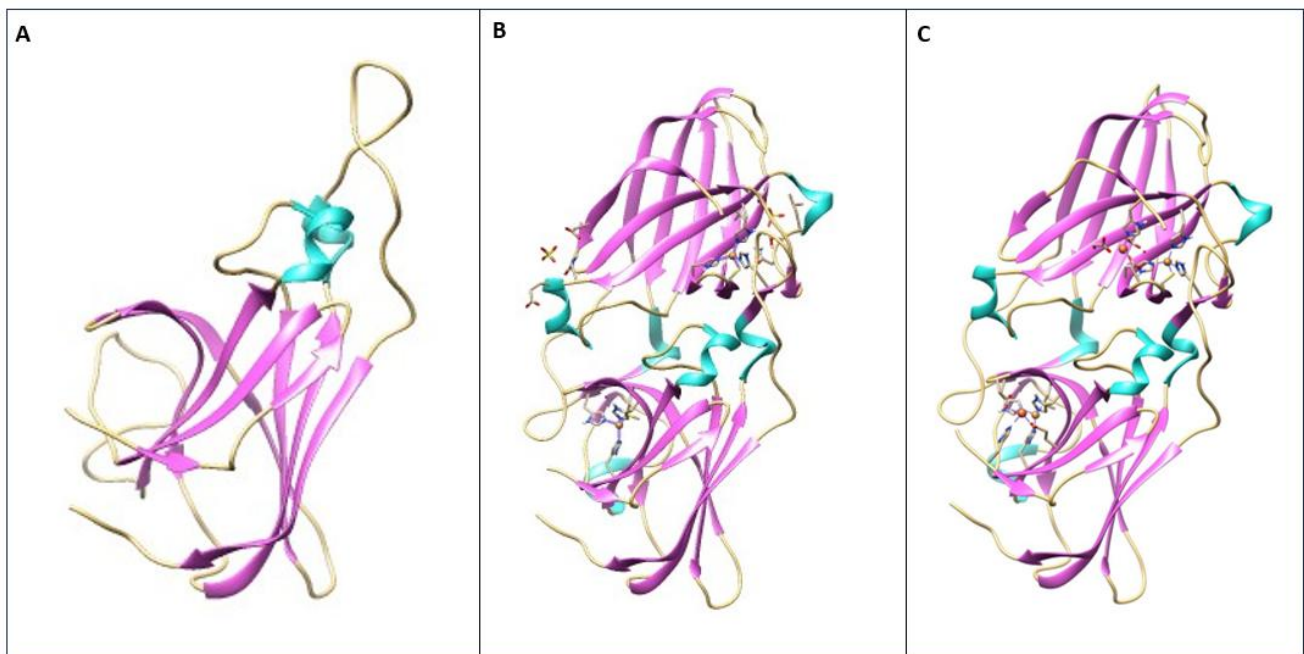


Figure 27. Structural determination of Tde1511, a high-affinity ferrous iron binding/transport protein from the pathogen *T. denticola*. The crystal structure of Tde1511 has been solved in its monomeric apo form at 1.78 Å (A), in its dimeric form when bound to its copper cofactor at 1.89 Å (B), and in its dimeric form when bound to its copper cofactor and coordinating its iron substrate at 1.82 Å (C). Copper coordination using residues H94, M141, H148, and H178* induces dimerization (* denotes residue from the other monomer). Once copper is bound, the iron binding site becomes available with coordinating residues E96, D98, and D145. Structure images were rendered in Chimera version 1.11.2 with alpha helices coloured cyan, beta sheets coloured purple, and random coils coloured peach.

Tde1511 has a beta sandwich structure which adopts an immunoglobulin (Ig)-like fold that can be further categorized as a hybrid (H)-type Ig-like fold. This is due to the presence of 8 total beta strands that comprise the two beta-sheets, while having its C' strand directly connected to strand D with this connection spanning from one sheet to the next as shown in Figure 28.

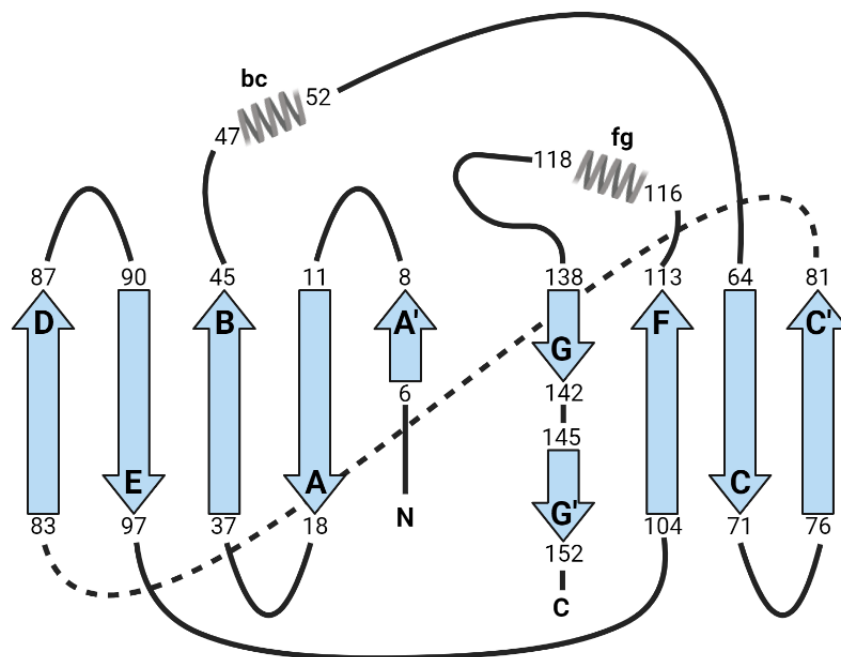


Figure 28. Topological map depicting the 2D structure of recombinant Tde1511 from the pathogen *T. denticola*. This 2D depiction of the recombinant structure of Tde1511 aids in the visualization of how neighbouring beta strands interact throughout the structure. The H-type characterization is due to the 8 total beta strands comprising the core beta sandwich. The C' strand links to strand D where this connection also spans the two beta sheets. The dashed line represents the direct transition between beta sheets as this distance consists of one amino acid.

Tde1511 coordinates its copper cofactor using residues H94, M141, H148, and H178* (* indicates residue originating from another monomer). The three imidazole nitrogens and the δ -sulfur of methionine coordinate the Cu^{2+} ion in a tetrahedral geometry. In homologous structures to Tde1511, iron is found to be coordinated with octahedral geometry in an acidic pocket containing the residues E96, D98, and D145 and one water molecule. In addition to these, this crystal structure has substituted the use of a second water molecule for coordination for an imidazole nitrogen originating from the affinity tag used in prior purification steps. As this is seen in both Fe^{3+} binding sites, it is assumed that this coordination geometry allowed for ideal crystal packing interactions to occur that resulted in the successful incorporation of iron into the structure as shown in Figure 29.

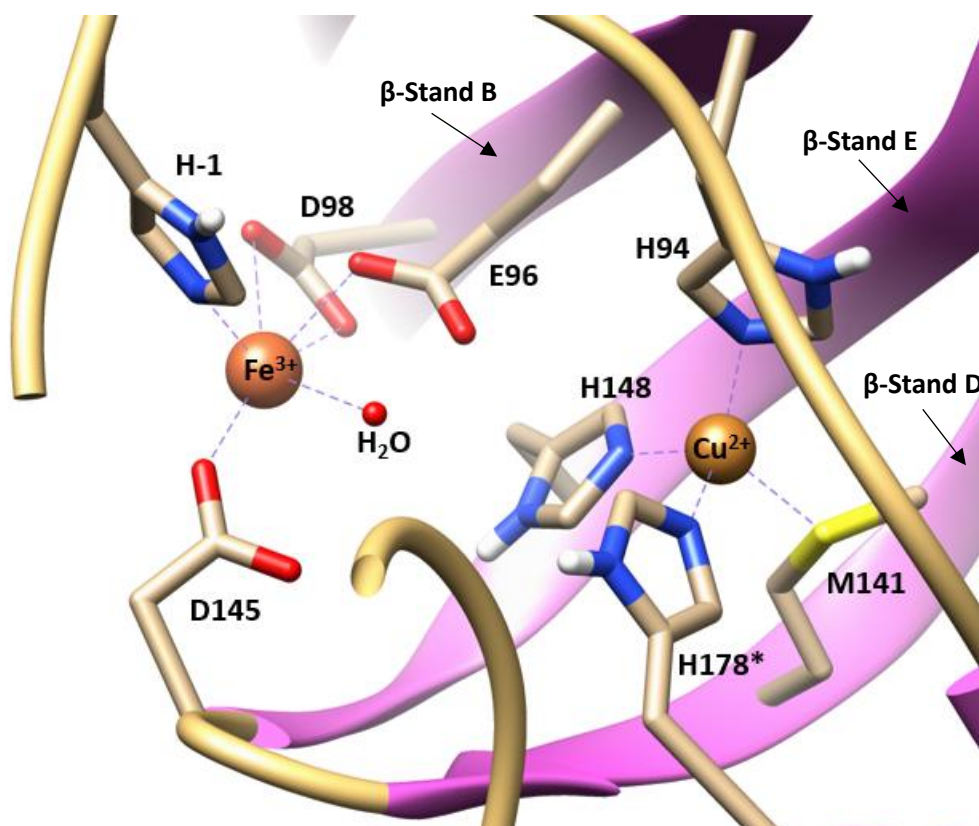


Figure 29. Coordination of copper and iron in the structure of Tde1511 from the pathogen *T. denticola*. A closer inspection of the metal binding sites of Tde1511 highlights residue involvement from beta strands B, D, E, and two random coils to facilitate metal binding. A cupric cation is coordinated in a tetrahedral geometry using the δ -sulfur from M141, and three N ϵ 2 atoms from H94, H148, and H178* (* denotes residue originating from another monomer). Ferric iron is coordinated in a trigonal prism geometry through a bidentate interaction of OD1 and OD2 from Asp98, monodentate interactions through OE1 of E96 and OD1 of D145, and the oxygen of one water molecule. Image was rendered in Chimera version 1.11.2 with alpha helices coloured cyan, beta sheets coloured purple, and random coils coloured peach. Red represents oxygen, blue represents nitrogen, and yellow represents sulfur.

After the refinement for each structure was complete, the final crystallographic data statistics are shown in Table 5. With this data, these three structures were submitted and accepted into the PDB by June 20th, 2023. They will be on hold until the release of a publication or until June 20th, 2024.

Table 5. Refinement statistics for the structures of Tde1511 submitted to the Protein Data Bank.

Structures	Apo Tde1511	Cu-Tde1511	Cu/Fe-Tde1511
Wavelength (Å)	0.9686	0.95374	0.95374
Resolution Range (Å)	29.62 – 1.78 (1.84 – 1.78)	42.17 – 1.89 (1.96 – 1.89)	39.79 – 1.82 (1.89 – 1.82)
Space Group	C 2 2 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 6 ₃
Unit Cell (Å, °)	54.7 75.0 79.8, 90 90 90	31.5 78.2 150.3, 90 90 90	105.3 105.3 79.6, 90 90 90
Total Reflections	32059 (3143)	61524 (6086)	89908 (8924)
Unique Reflections	16077 (1572)	30767 (3045)	44955 (4462)
Multiplicity	6.6 (6.9)	12.6 (11.3)	19.3 (19.0)
Completeness (%)	99.81 (99.94)	99.97 (100.00)	99.77 (99.73)
Mean I/sigma(I)	15.19 (3.26)	11.92 (3.16)	18.15 (2.10)
Wilson B-factor	27.34	33.67	37.12
R-merge	0.02418 (0.2089)	0.02964 (0.2104)	0.0197 (0.3945)
R-meas	0.03419 (0.2955)	0.04192 (0.2975)	0.02785 (0.558)
R-pim	0.02418 (0.2089)	0.02964 (0.2104)	0.0197 (0.3945)
CC1/2	0.999 (0.942)	0.998 (0.87)	0.999 (0.848)

CC*	1 (0.985)	1 (0.965)	1 (0.958)
Reflections used in refinement	16053 (1571)	30762 (3045)	44887 (4454)
Reflections used for R-free	747 (63)	1467 (131)	2288 (231)
R-work	0.2056 (0.3262)	0.2097 (0.2824)	0.2065 (0.3482)
R-free	0.2322 (0.3262)	0.2355 (0.3037)	0.2471 (0.3592)
CC(work)	0.959 (0.925)	0.949 (0.868)	0.957 (0.875)
CC(free)	0.934 (0.887)	0.926 (0.735)	0.949 (0.842)
Number of non-hydrogen atoms	1297	2533	2531
macromolecules	1247	2432	2447
ligands	0	12	4
solvent	50	89	80
Protein Residues	156	311	312
RMS (bonds)	0.011	0.012	0.012
RMS (angles)	1.44	1.34	1.36
Ramachandran favoured (%)	98.05	98.70	97.40
Ramachandran allowed (%)	1.95	1.30	2.60
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer Outliers (%)	3.10	1.20	0.40
Clashscore	2.90	0.85	1.48
Average B-factor	37.42	38.10	48.03
macromolecules	37.28	37.78	48.02
ligand	N/A	87.01	45.73
solvent	41.03	40.19	48.44

3.5 Functional Assays

3.5.1 ICP-OES with Tde1511

ICP-OES was conducted using EDTA-treated Tde1511 and solutions containing various concentrations of common biological divalent metal cations. A relevant metal interaction was identified when the measured concentration surpassed the respective LOQ, as this value indicates the concentration of analyte that can be consistently measured with reproducibility. This analysis has allowed conclusions to be drawn that Tde1511 has a high affinity for Cu^{2+} due to the presence of metal binding at low concentrations. All concentrations tested indicated a quantifiable concentration of copper being present in the sample. Other period 4 metal elements (Zn, Ni, and Co) also showed identifiable signals but at higher concentrations. At concentrations of $50\mu\text{M}$ and greater, Tde1511 bound nickel and zinc with similar affinities to copper. Calcium and magnesium both showed negligible interactions with Tde1511. Metals anticipated occupying the secondary metal binding site include iron and manganese. Quantifying iron at pH 6.5 is difficult due to the optimal solubility of ferric iron being around pH 3.5. Evidently, significant iron signals were only detected at the largest concentrations tested ($100\mu\text{M}$ and $200\mu\text{M}$). Manganese shares a similar mass-to-charge ratio as iron and acts as a good proxy ion to use in iron binding experiments. Significant manganese concentrations were detected as low as $10\mu\text{M}$ and can be quantified as low as $50\mu\text{M}$. These results are summarized in Figure 30 with each metal's respective limit of detection (LOD) and limit of quantification (LOQ) during the analysis.

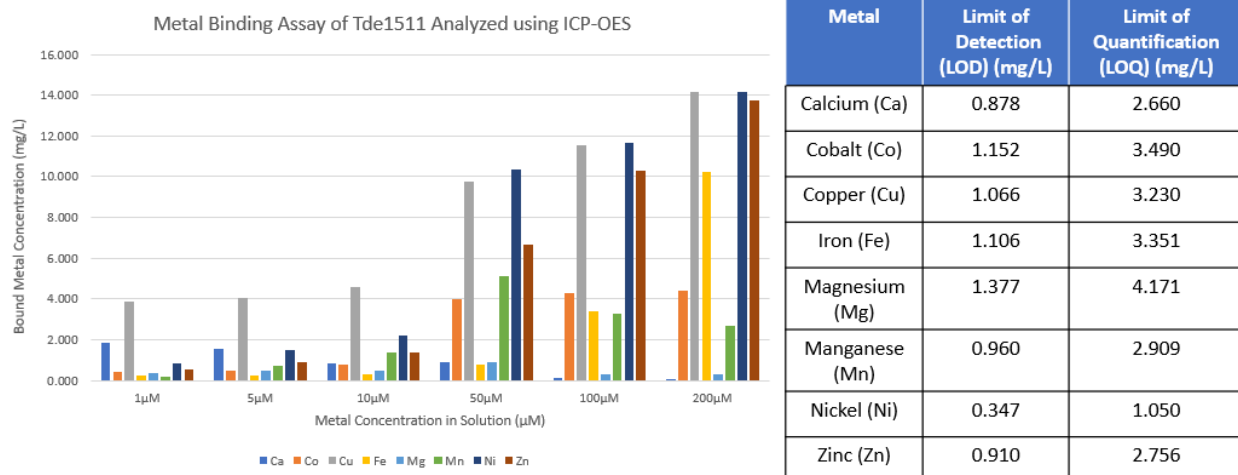


Figure 30. ICP-OES analysis of the ability for Tde1511 to bind various biological divalent metal ions. Tde1511 was dialyzed exhaustively against EDTA to ensure the absence of metals in solution. Various samples of 22.4 mg/mL Tde1511 were dialyzed against various concentrations of divalent metals ions in a buffer of 50mM Bis-Tris pH 6.5 for 24h. These samples had unbound metals removed from solution and diluted to a final protein concentration of 5 mg/mL and acidified with 2% (v/v) nitric acid. Following ICP-OES analysis, corrected intensities of each metal were converted to a concentration in mg/L and plotted in a bar graph. The table on the right outlines the limit of detection and limit of quantification for each metal tested based on the standard curve concentrations produced for each metal.

3.5.2 DSF of Tde1511

To discern whether the interacting metals identified in ICP-OES formed stable ligand complexes, DSF was employed to quantify the interactions based on the increase in melting point observed for the overall protein structure. Table 6 outlines the observed melting point shifts when EDTA-treated Tde1511 (80μM) was incubated with various divalent metal ions at a ratio of 1:7 protein to metal. It also includes Cu-Tde1511 (50μM) incubated with various period 4 transition metals at a ratio of 1:7 protein to metal. Results indicate that EDTA-treated Tde1511 forms many transient protein-metal complexes (with Ca^{2+} , Ni^{2+} , Mg^{2+} , Mn^{2+} , Fe^{2+} , and Fe^{3+}) that do not result in a significant increase in melting temperature. Interactions with Zn^{2+} and Co^{2+} are shown to create stable metal complexes due to the significant increase in melting temperatures compared to the apo protein. As expected, Cu^{2+} creates the most significant metal complex,

resulting in a melting temperature increase from 43.3°C to 85.0°C, a difference of 41.7°C. The analysis of Cu-Tde1511 also shared expected results. Once copper is bound, no other metals create significant changes in melting temperature except for manganese and iron. Both Fe²⁺ and Mn²⁺ produced similar increases in melting temperature, likely due to their similar mass-to-charge ratios at play in the coordination sphere. Outcompeting all other metal ions, Fe³⁺ produced the largest increase in melting temperature from 83.6°C to 90.3°C, an increase of 6.7°C.

Table 6. DSF analysis of the shift in melting temperature after incubating EDTA-treated and copper loaded Tde1511 with various biological divalent metal cations

EDTA-treated				Cu-loaded			
Metal	T _m (°C)	Standard Deviation	ΔT _m (°C)	Metal	T _m (°C)	Standard Deviation	ΔT _m (°C)
Apo (Initial)	43.3	0.077	/	Start (Initial)	83.6	0.333	/
Co ²⁺	82.1	0.378	38.8	Co ²⁺	86.0	0.604	2.4
Cu²⁺	85.0	0.630	41.7	Zn ²⁺	87.1	0.360	3.5
Zn ²⁺	78.6	1.548	35.3	Ni ²⁺	84.9	0.554	1.3
Ca ²⁺	46.9	0.941	3.6	Mn ²⁺	88.4	0.723	4.8
Mg ²⁺	48.0	0.939	4.7	Fe³⁺	90.3	0.497	6.7
Ni ²⁺	51.8	1.527	8.5	Fe ²⁺ (βME)	88.7	0.430	5.1
Mn ²⁺	56.9	2.199	13.6				
Fe ³⁺	46.8	0.508	3.5				
Fe ²⁺ (βME)	49.4	0.218	6.1				

3.5.3 Peptide TLC Analysis of Tde2714

To identify potential FGE catalytic activity, Tde2714 was incubated in a 1:10 ratio with four peptides that contained different FGE recognition sequences with the addition of a negative control peptide. Resultant reaction mixtures were concentrated and separated via TLC to identify novel product formation. Figure 31 depicts the TLC analysis of each peptide. With the expectation that only one recognition sequence would produce a unique TLC spotting pattern, it

can be seen that all five peptides produced very similar patterns after their reaction with Tde2714.

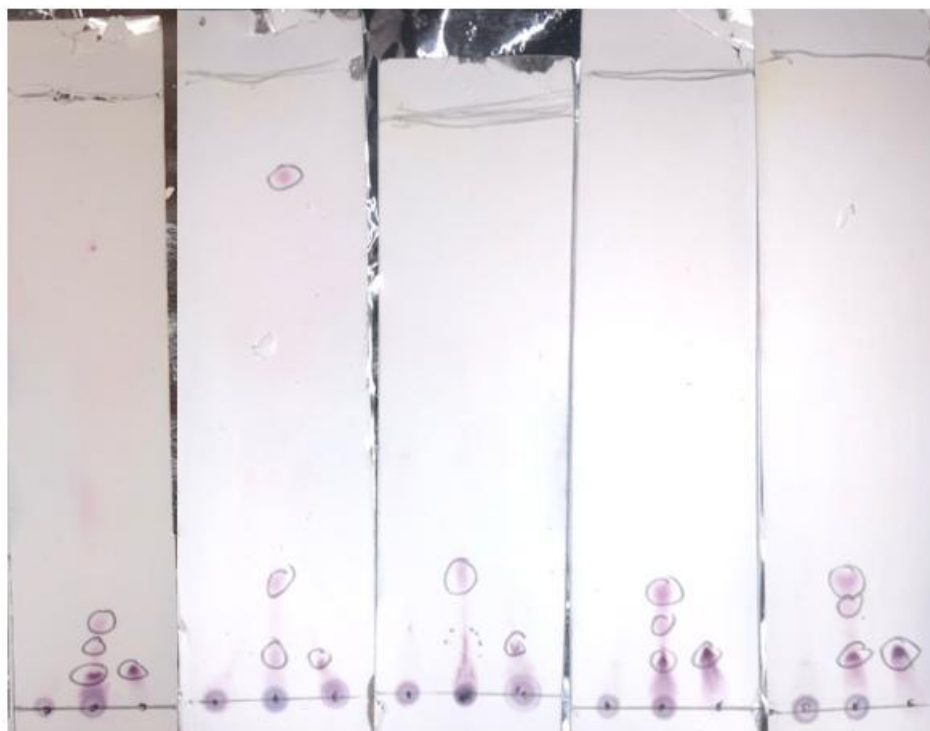


Figure 31. Peptide modification reactions using Tde2714 with peptides containing FGE recognition sequences and a negative control analyzed through TLC. In a buffer containing 50mM sodium phosphate pH 8 and 1mM DTT, 10 μ M Tde2714 was incubated with 100 μ M of each peptide in duplicate trials and left to react for 8h at 37°C. Reaction mixtures were concentrated and blotted onto prepared TLC strips in the order of control enzyme on the left, reaction mixture in the middle, and control peptide on the right. TLC strips were then placed into development chambers and the solvent was allowed to run until it reached 1cm from the top of the strip. Strips were stained using a pyridine and acetic acid stain and baked in an oven at 80°C for 10min to visualize spotting patterns.

With each peptide reacting in a similar manner with Tde2714, proteolytic activity was assayed using a similar method. One peptide (SxAxR) was used during the analysis with an increasing concentration of a protease inhibitor cocktail (PIC) tablet. Figure 32 outlines the resultant TLC analysis following this reaction. It is observed that as the concentration of PIC increases, the intensity of the product spot decreases.

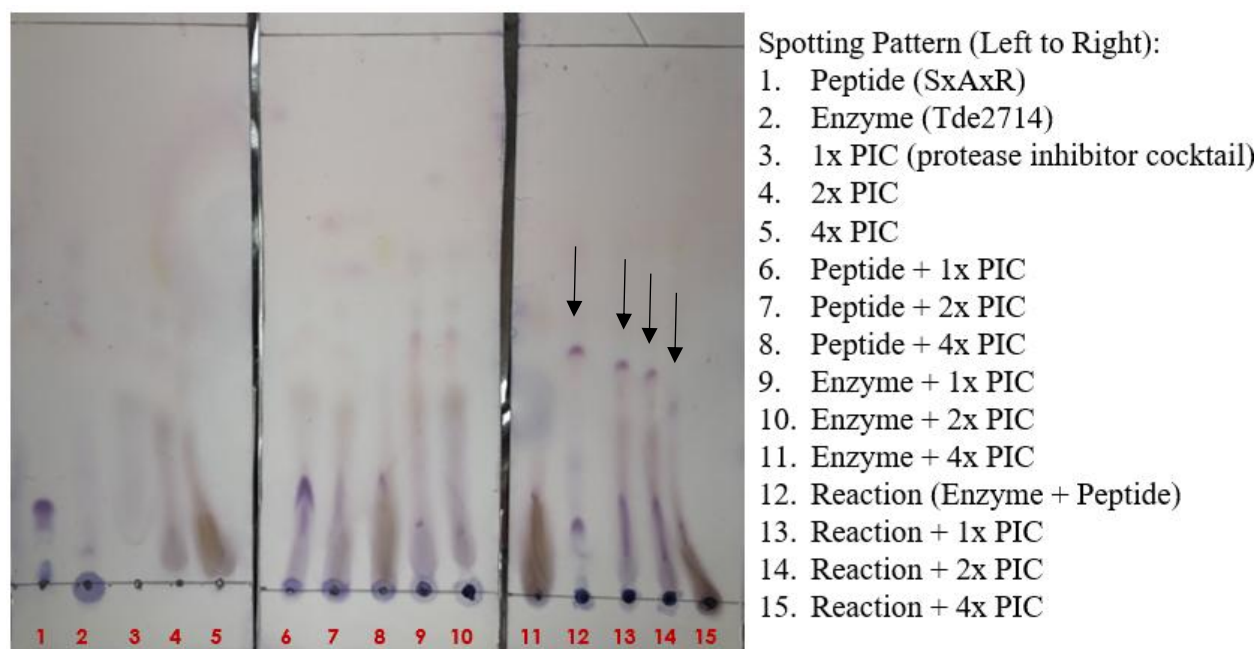


Figure 32. TLC analysis of the reaction of Tde2714 with a peptide containing the SxAxR FGE recognition sequence in the presence of increasing concentrations of a protease inhibitor cocktail tablet. In a buffer containing 50mM sodium phosphate pH 8 and 1mM DTT, 10 μ M Tde2714 was reacted with 100 μ M of a peptide containing the SxAxR FGE recognition sequence in duplicate trials for 8h at 37°C. Identical reactions were ran in parallel which contained increased concentrations of a protease inhibitor cocktail tablet. Many negative controls were run as well. Following reaction times, samples were concentrated and blotted onto prepared TLC strips and placed into development chambers until the solvent front reached 1cm from the top of the strip. These strips were then stained using a pyridine and acetic acid stain and baked in an oven at 80°C for 10min to visualize spotting patterns. Black arrows indicate product spots of interest.

3.5.4 Proteolytic Degradation Assay with Tde2714

To further verify potential proteolytic functions of Tde2714, a degradation assay was designed and performed. Tde2714 was incubated with 3 proteins with and without the presence of a PIC tablet. Samples were taken from each mixture at time points of 0 days, 7 days, and 14 days and analyzed via SDS-PAGE to visualize potential protein degradation products. It is observed in Figures 33, 34, and 35 that over the two-week period, none of the protein targets

used in the experiment showed any significant degradation products. Interestingly, it was Tde2714 which exhibited varying banding patterns with and without the addition of PIC. In samples containing Tde2714 and PIC, intense bands are observed at its expected molecular weight of about 74 kDa. In samples containing Tde2714 and no PIC, the accumulation of a species banding at about 34 kDa is observed to grow in intensity during the time of analysis. A species at this molecular weight aligns perfectly with the expected molecular weight of the proposed catalytic domain of Tde2714 (residues 378-663) of 34.2 kDa. This result alludes to a potential autocatalytic processing taking place.

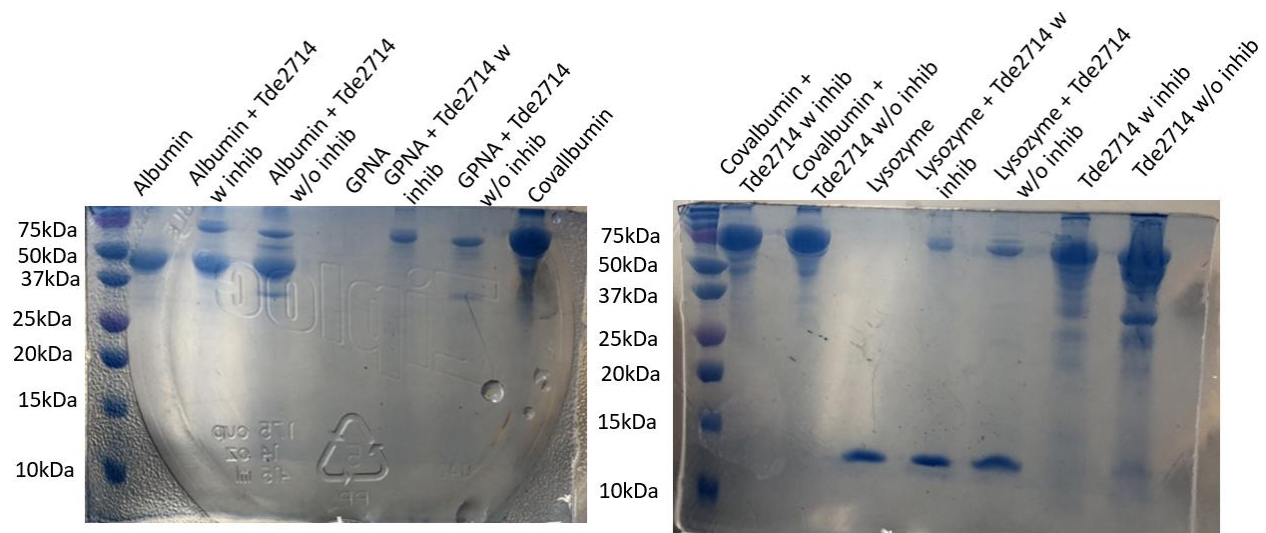


Figure 33. Day 0: SDS-PAGE analysis of the proteolytic degradation using Tde2714 with various protein substrates in the presence and absence of PIC. In a buffer containing 50mM sodium phosphate pH 8 and 1mM DTT, 10 μ M Tde2714 was reacted with 100 μ M of albumin, covalbumin, and lysozyme in the presence or absence of 5x PIC. When samples were created, aliquots were removed right away and analyzed via SDS-PAGE. After sample preparation and gel loading, the gel was run at 180 volts for 55min. It was stained using Coomassie brilliant blue and subsequently destained over night. The far most left lane on both gels represents the standard molecular weight ladder with each molecular weight marked on the left side of each image.

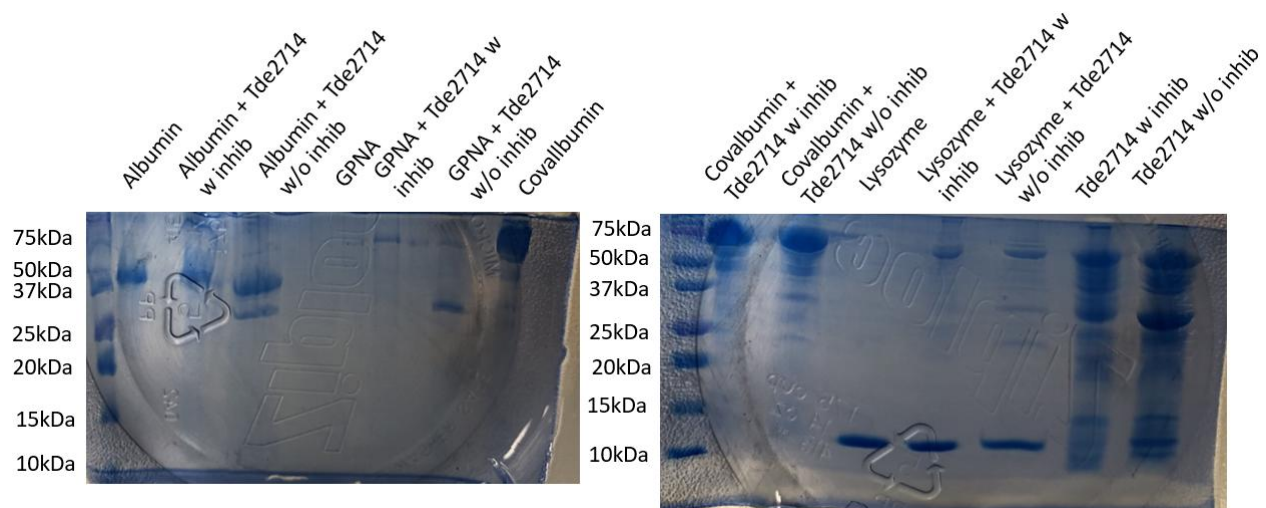


Figure 34. Day 7: SDS-PAGE analysis of the proteolytic degradation using Tde2714 with various protein substrates in the presence and absence of PIC. In a buffer containing 50mM sodium phosphate pH 8 and 1mM DTT, 10 μ M Tde2714 was reacted with 100 μ M of albumin, covalbumin, and lysozyme in the presence or absence of 5x PIC. After 7 days from sample creation, aliquots were removed and analyzed via SDS-PAGE. After sample preparation and gel loading, the gel was run at 180 volts for 55min. It was stained using Coomassie brilliant blue and subsequently destained over night. The far most left lane on both gels represents the standard molecular weight ladder with each molecular weight marked on the left side of each image.

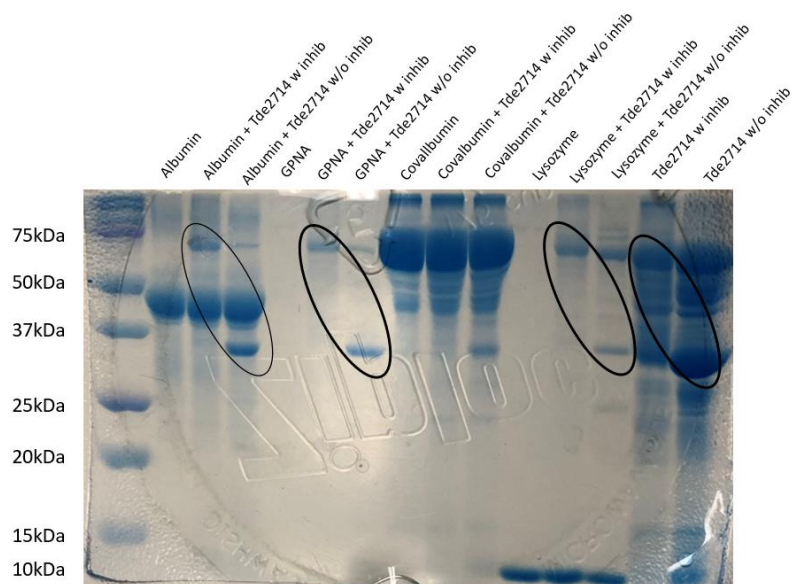


Figure 35. Day 14: SDS-PAGE analysis of the proteolytic degradation using Tde2714 with various protein substrates in the presence and absence of PIC. In a buffer containing 50mM sodium phosphate pH 8 and 1mM DTT, 10 μ M Tde2714 was reacted with 100 μ M of albumin, covalbumin, and lysozyme in the presence or absence of 5x PIC. After 14 days from sample creation, aliquots were removed and analyzed via SDS-PAGE. After sample preparation and gel loading, the gel was run at 180 volts for 55min. It was stained using Coomassie brilliant blue and subsequently destained over night. The far most left lane on both gels represents the standard molecular weight ladder with each molecular weight marked on the left side of each image. Black ovals highlight two species expected to be Tde2714 at its initial molecular weight of about 74 kDa in the presence of PIC and its proposed catalytic domain at about 34 kDa in the absence of PIC.

Chapter 4. Discussion

4.1 Analysis of Tde1414

4.1.1 *Tde1414 and the Family of Thiamine Diphosphate Requiring Enzymes*

The protein target of the gene *Tde1414* identified in *T. denticola* has been bioinformatically annotated as a putative phosphonopyruvate decarboxylase. This family of enzymes uses Pnpy as a substrate and performs a non-oxidative decarboxylation to create PnAA and carbon dioxide with the help of thiamine diphosphate (ThDP) and a magnesium ion as cofactors^{39,44,96}. To date, only two Ppd enzymes have been functionally characterized, but in all attempts the enzyme has been refractive to crystallization, resulting in a lack of structural information for this enzyme family^{41,96,97}. On the contrary, the diverse family of ThDP-dependent enzymes has many functionally and structurally characterized members, such as benzoylformate decarboxylase, pyruvate decarboxylase, and transketolase^{44,45,48}. This allows for insightful discussions and comparisons to still be made. All ThDP requiring enzymes contain a pyrophosphate (PP) domain and a pyrimidine (Pyr) binding domain⁹⁶. PP and Pyr domains show independent evolution from one another but are fused together into one polypeptide chain⁹⁶. Ppd is an evolutionary forerunner to the subset of decarboxylase enzymes in the ThDP family,

allowing for meaningful comparisons to be made to other ThDP-dependent enzymes ⁹⁷. This is exemplified in the phylogenetic tree shown in Figure 36. Indicated with the red arrow, the common ancestor which gave rise to the provided Ppd sequences most closely resembles the most distant common ancestor which gave rise to all the provided ThDP-dependent decarboxylase sequences, as well as all other ThDP-dependent enzymes incorporated into the phylogenetic tree.

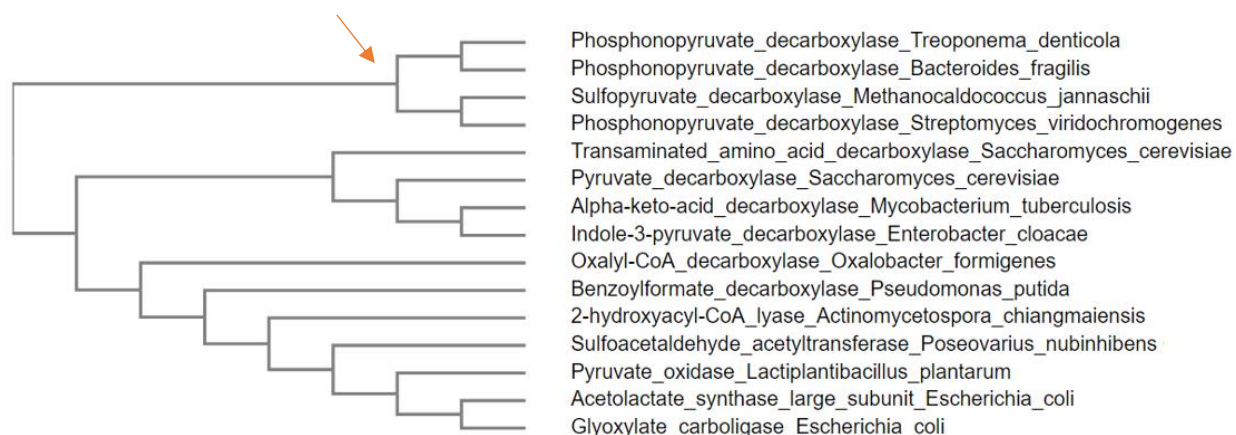


Figure 36. Phylogenetic tree comparing the sequence divergence between thiamine diphosphate dependent enzymes. After creating a multiple sequence alignment between maximally diverse ThDP-dependent enzymes, a phylogenetic tree was then produced to visualize the relatedness of each sequence. The red arrow indicates the common ancestor which gave rise to each of the phosphonopyruvate decarboxylase used to create this diagram. This image was created using Clustal Omega.

When comparing other ThDP requiring decarboxylases, the only similarities in sequences are observed in the proposed cofactor binding residues, while active site/ catalytic residues show the greatest diversity ^{43,98}. When expanding the search to all ThDP-dependent enzymes, the same sequence similarity is observed, concentrated on a highly conserved GDG-X₂₅-NN motif proposed for ThDP cofactor binding ^{46,98}. Predicted secondary structures of the GDG and NN motifs forms a striking resemblance to the active site of other dinucleotide binding enzymes,

allowing some functional annotations to be hypothesized ⁹⁸. It is also hypothesized that the GDG recognition motif may facilitate interactions with the diphosphate moiety and the magnesium ion while the NN motif may provide hydrogen bonding to the nitrogens within the thiamin moiety of ThDP ^{43,98}.

4.1.2 Bioinformatics

Initial bioinformatic efforts were pursued to determine putative functions of Tde1414. All searches conducted using various programs converged in strong agreement on a single function as a phosphonopyruvate decarboxylase. This enzyme plays a critical role in the production and diversification of phosphonates by facilitating the first committed irreversible reaction in their biosynthetic route ³⁹. A significant number of phosphonate products are created through the decarboxylation of Pnpy to PnAA, adding to the significance of the reaction performed. Using the STRINGS database, it was predicted that Tde1414 interacts with Tde1413 (Ppm) and Tde1415 (an aminotransferase), which are two other established enzymes in the phosphonate biosynthetic pathway. BLASTp was helpful in identifying homologous sequences spanning various bacterial species and different organisms, all of which were annotated to be putative phosphonopyruvate decarboxylases. In addition, the central annotations made by InterPro were for domains involved in thiamine diphosphate binding.

With this knowledge, different MSA's were created to tease out potential conserved motifs to further support the function of a phosphonopyruvate decarboxylase. Literature stated that very few, if any, catalytic residues are conserved in the family of ThDP-dependent decarboxylases. Only residues used to bind the required cofactors, ThDP and Mg^{2+} , are typically found to hold high conservation ⁹⁸. This resulted in two different MSAs: one using characterized and putative phosphonopyruvate decarboxylase sequences and another using sequences of

previously characterized ThDP-requiring enzymes. The MSA comparing characterized and putative Ppd enzymes with Tde1414 provided a significant number of conserved residues, the majority of which are used to bind ThDP and Mg^{2+} . This includes E49, Q72, S74, M235, D260, N288, H291, and S293. Residues D260 and N288 are also from the conserved GDG-X₂₅-NN motif found in Tde1414 and shown to coordinate the magnesium ion^{46,98}. When analyzing the MSA created with Tde1414 against other ThDP-requiring enzymes, much less conservation was observed. The strongest region of conservation occurred for the GDG-X₂₅-NN motif. Therefore, with all the information gathered, it was reasonably substantiated that the putative function of Tde1414 was that of a phosphonopyruvate decarboxylase.

4.1.3 Structural Characterization of Tde1414

When attempting to crystallize Tde1414, the apo structure was concentrated and used in crystal screenings at concentrations of 10, 15, 20, and 25 mg/mL with no success. A literature search of crystallizations of various ThDP-dependent enzymes revealed a pattern of the requirement for ThDP, Mg^{2+} , and often DTT to be present. Thus, magnesium was supplemented into the lysis and dialysis buffers, and purified protein preparations of Tde1414 were dialyzed against 0.2mM $MgCl_2$, 0.2mM ThDP, and 1mM DTT for 24h prior to final protein concentrations. Crystal screens of this reconstituted protein occurred at concentrations of 15 and 20 mg/mL without success. A last attempt was tried at a concentration of 30 mg/mL where the crystallization was successful.

Due to Tde1414 being the first Ppd to report successful crystallization, initial efforts for selenomethionine (SeMet) incorporation into the structure was attempted to allow for downstream SeMet phasing of crystallization data. This approach was unsuccessful. With the

rise in artificial intelligence protein structure prediction confidence, an AlphaFold2 predicted model of Tde1414 was used as the molecular replacement model for the crystallographic data obtained from Tde1414 crystals. With this, a confident solution was identified and the first structural solution for a Ppd enzyme was found. Due to the absence of Ppd structural characterizations found in literature, this marks the first annotations provided for the binding interactions that occur between a Ppd (Tde1414) and its ThDP and Mg^{2+} cofactors.

Although one ThDP was modeled with moderate density visualization, the other was modeled with high agreement to the electron density and will be used for the rest of the analysis. ThDP can be seen in its strained (“V”) conformation that can be identified due to the conformation of the carbon bond that connects the pyrimidine ring to the thiazolium ring^{45,46}. This conformation brings the 4' imino of the pyrimidine ring less than 3.5 Å away from the C2 carbon of the thiazolium ring, permitting proton transfer and catalysis to occur⁴⁶. Analyzing the MSAs between Tde1414, characterized and putative Ppd enzymes, and other ThDP-requiring enzymes revealed the heavily conserved motif GDG-X₂₅-NN motif proposed to be required for ThDP cofactor binding and Mg^{2+} coordination. It was found that the aspartic acid of the GDG moiety and the latter asparagine of the NN moiety are what directly interact and coordinate the Mg^{2+} . It is through this magnesium cation that the pyrophosphate moiety of the ThDP cofactor is anchored into the binding site. This allows the thiazolium and pyrimidine ring to interact with the protein binding pocket in the appropriate orientation. ThDP is seen with the pyrophosphate group stabilized within the C-terminal domain of Tde1414, while the pyrimidine ring stretches into the N-terminal domain of Tde1414 where it is hydrogen bonded with various amino acid residues. Thus, the C-terminal domain of Tde1414 is categorized as the PP domain while the N-terminal domain is classified as the Pyr domain^{45,96}.

The structure also shows the pyrophosphate group stabilized by other amino acid residues, such as arginine and serine. Along with the magnesium cation and peptide backbone providing additional coordination, these metal and residue interactions are expected to occur and are consistently found when phosphate groups interact with proteins ⁹⁹. Following hard and soft acid/base (HSAB) ligand theory, hard acids stabilize hard bases ¹⁰⁰. With magnesium characterized as a hard acid due to its relatively small, non-polarizable structure and the oxygens from phosphate groups classified as a hard base, this metal-ligand interaction is expected to occur. Hydrogen bonding of the thiazolium and pyrimidine rings is shown to occur mainly through interactions of the heteroatoms of ThDP and neighbouring amino acid residues of Tde1414. The thiazolium sulfur interacts with NE1 of H291 and its nitrogen interacts with the δ -sulfur of M235. These interactions orient the thiazolium ring to have its C2 carbon in the closest proximity to the pyrimidine ring. The pyrimidine ring is then stabilized via oxygen ligand interactions from residues E49, Q72, and S74. Residue E49 is also significant as it interacts with N1 of the pyrimidine ring, allowing tautomer stabilization to occur to support the formation of the imino group at the C4 position. These interactions are necessary to promote the 4' imino group to be within the appropriate distance to perform catalysis with the C2 carbon of the thiazolium ring. Thus, the first structure of a Ppd has been structurally solved and characterized on its binding interactions with its required cofactors, ThDP and Mg^{2+} .

In the future, steps should be taken towards confirming Ppd catalytic activity of Tde1414 and uncovering the interactions of this enzyme and its cofactors with its substrate phosphonopyruvate and various substrate intermediates. It is seen that the potential ylide formation at the C2 carbon of the thiazolium ring is oriented towards the protein chain that binds the same ThDP and not toward the other monomer in the dimeric complex. Thus, Pnp binding

may also occur within the same monomer. With both cofactor and substrate potentially binding within the same monomer, this brings into question if a dimeric complex is required for catalytic activity. Uncovering this will allow a complete mechanism of decarboxylation catalysis to be determined for the conversion of Pnpy to PnAA. This level of mechanistic understanding will allow mutant Ppd enzymes to be synthesized that have a higher affinity for its substrate and perform catalysis at a faster rate of reaction. This will result in useful, commercialized phosphonate products to be synthesized in greater yields with higher efficiencies, providing an impactful benefit to the global community.

4.2 Analysis of Tde1511

4.2.1 Bioinformatics

Bioinformatic efforts to characterize the target Tde1511 provided strong evidence that this protein was capable of binding metal ions, more specifically, ferrous iron. When running the initial BLASTp search, all results obtained while excluding similar sequences from *T. denticola* ATCC 35405 were unanimously annotated as iron transport proteins. To complement this, domain homologies identified through InterPro had identified similarities to Tp34-type metal binding domains, as well as a domain specifically for the transport of ferrous iron. It was also able to identify a potential membrane lipoprotein lipid attachment site. Lipoproteins are known to outnumber every other protein expressed in spirochetes ¹⁰¹. Acyl groups that covalently modify these proteins are used to provide a hydrophobic moiety that aids in the anchoring of these proteins into the cell membrane ⁶³. The EggNOG analysis indicating Tde1511 to be apart of COG3470 then sparked the search to identify previously characterized members of this protein family. This was successful in finding a small pool of characterized members in this family that

share very significant sequence similarities to Tde1511^{35,54,59,60,62}. All characterized proteins of this family contain a copper cation cofactor as well as transport iron as its primary substrate.

These findings led to the generation of an MSA, which compared the primary sequence of Tde1511 against other characterized members from COG3470. Through this, a handful of conserved residues were identified, but the most significant finding was the shared conservation of residues identified to coordinate both copper and iron from the previously characterized members and Tde1511. This generated a hypothesis that Tde1511 uses residues H94, M141, H148, and H178 to coordinate copper, while residues E96, D98, and D145 are used to facilitate iron binding. This was also supported by using HSAB theory that states iron (a hard acid) coordination in proteins preferentially occurs through interactions of Glu/Gln and Asp/Asn (hard bases) while copper (a soft acid) coordination favours interactions with His, Cys, and Met ligands (His: borderline soft base, Cys, Met: soft bases)¹⁰².

With strong bioinformatic convergence that Tde1511 may have a putative function as a metal-binding iron transport protein, a STRINGS analysis had identified various protein partners hypothesized to interact with Tde1511. This included Tde1508-1510, Tde1512, and Tde1514-1518. Using <https://microbesonline.org/> database, it was then suggested that these gene products form an operon. They are all translated in the same direction and each gene is positioned within a relatively close distance to one another, prompting the prediction of an operonic system. Determining bioinformatic data on each of these targets allowed for the construction of a potential iron transport system. Many of the proteins involved contained MacB-like and FtsX domains, characterized for their macrolide export functions¹⁰³. Proteins within this superfamily of ABC transporters with homologous domains that lack macrolide export functions also show functions in trafficking lipoproteins from the inner membrane to the periplasm¹⁰³. Thus, it was

expected to then find other proteins in this putative operon to contain ABC transporter-related functions such as proteins with ATP-binding domains and ABC-type permease proteins. This left Tde1510 (iron membrane permease), Tde1512 (FeS containing protein), Tde1517 (FMN-binding domain containing protein), and Tde1511 possessing putative non-ABC transporter related functions.

With this knowledge, two potential iron import systems were proposed using Tde1510, Tde1511, Tde1512, and Tde1517 as shown in Figure 37. What distinguishes these two systems from one another is the mechanism in which iron is oxidized before its translocation into the cytoplasm. Drawing on similarities to the homologous iron import system FtrABCD, naming conventions were given to these four members. FtrA showed the highest homology to Tde1511; it is a periplasmic lipoprotein, which binds copper as a cofactor and facilitates iron binding and translocation to its membrane-bound permease ^{104,105}. Therefore, Tde1511 will be identified as HftA (**H**igh-affinity **f**errous **t**ransport protein **A**). FtrC shows significant sequence similarities to Tde1510 and is a membrane bound ferric iron permease, showing homology with the oxidase-dependent ferrous transport (OFeT) superfamily of membrane permeases ¹⁰⁶. These membrane permeases contain seven transmembrane helices (TMH) in which TMH1 and TMH4 contain the highest sequence conservation due to the presence of iron binding motifs REGLE and/or REGAE ^{106,107}. Upon further characterization of Tde1510 through the generation of MSAs and AlphFold2 predicted structures, Tde1510 also contains the required iron-binding motif REGAE on both TMH1 and TMH4 ¹⁰⁸. Interestingly, Tde1510 also showed the presence of an Fth1 domain proposed to function in ferroxidase activity, which will become significant later on. Therefore, Tde1510 will be identified as HftC. FtrD is an FeS-containing protein that showed the greatest similarity to Tde1512. This protein acts as an electron sink during the redox reactions

that take place to allow iron to be translocated into the cytoplasm¹⁰⁴. Thus, Tde1512 will be identified as HftD. Unique to *T. denticola* is the use of a flavin mono-nucleotide (FMN)-binding protein (Tde1517) within this operon. When flavinylation genes are colocalized with iron transporter genes, they often function in iron reduction and assimilation¹⁰⁹. In the Ftr system, FtrB is a multicopper oxidase responsible for oxidizing ferrous iron bound to FtrA into ferric iron before being passed to FtrC. Due to the anaerobic environment surrounding *T. denticola*, facilitating oxygen-dependent oxidation reactions is unlikely. Using a potential Fth1 domain on HftC for oxidation reactions and allowing electron transfer chemistry to occur through an FMN-containing protein forms a much more likely route for iron import to occur through. Following suit, the role of this protein putatively engaging in electron transfer and redox chemistry, similar to FtrB, has afforded it the name HftB.

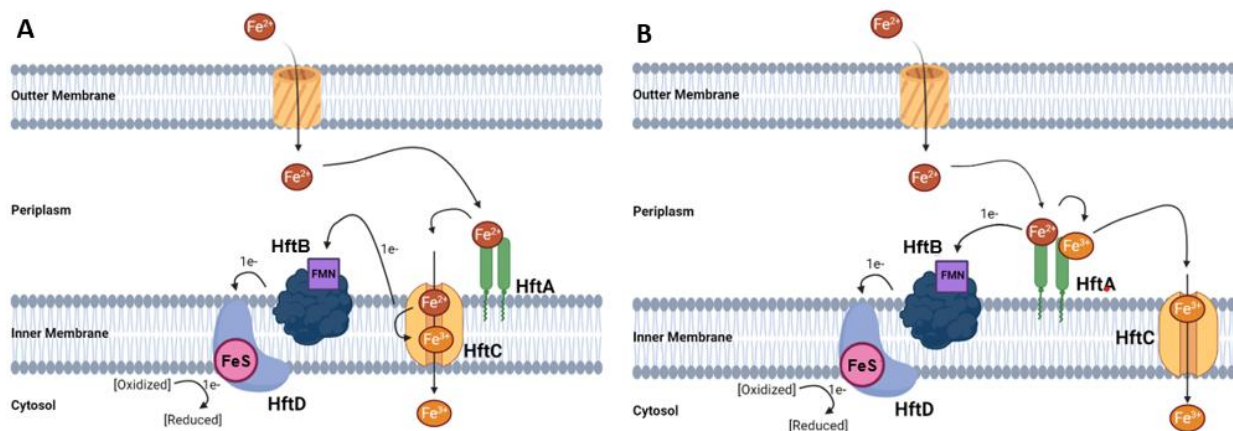


Figure 37. Proposed iron import operonic system, HftABCD, identified in *T. denticola*. Through bioinformatic characterization, a putative iron import operon has been identified. Until the ferroxidase activity of HftA can be identified, two proposed models of iron import are given. A) ferrous iron is passively transported across the outer membrane through nonspecific pores. In the periplasm, membrane anchored HftA binds to free ferrous iron to then transfer to the membrane permease, HftC. Putative ferroxidase activity of HftC oxidizes ferrous iron into ferric iron, where it can be passed through the permease successfully. An electron is then accepted by HftB from HftC. This electron is then shuttled to HftD, resetting the redox potential of HftB. From here, a terminal electron acceptor within the cytoplasm is reduced. B) ferrous iron is passively transported across the outer membrane through nonspecific pores. In the periplasm, membrane anchored HftA binds to free ferrous iron, oxidizes it into ferric iron, which is then given to HftC to translocate into the cytoplasm. An electron is then accepted by HftB to allow HftA to oxidize subsequent ferrous iron ions. This electron is then shuttled to HftD, resetting the redox potential of HftB. From here, a terminal electron acceptor within the cytoplasm is reduced.

Through this bioinformatic effort, the iron import system High-affinity ferrous transport (Hft) ABCD operon was hypothesized. HftA (Tde1511) is hypothesized to bind ferrous iron that passively diffused through porins in the outer membrane. HftA then traffics this ferrous iron to HftC (Tde1510). The putative ferroxidase activity of HftC would then oxidize ferrous iron to ferric iron and translocate it across the inner membrane. HftB (Tde1517) would accept this electron from HftC and pass it along to HftD where a cytoplasmic electron acceptor would pick up this electron. This is the expected pathway for *T. denticola* to acquire iron from its environment. The only obstacle lies in the fact that a homolog of Tde1511, Fet3 from *Saccharomyces cerevisiae*, performs its own ferroxidase activity using the copper cofactor as a redox partner to oxidize bound ferrous iron¹⁰⁶. Although sequence annotations of Tde1511 did not show the presence of ferroxidase domains, this avenue of iron oxidation cannot be dismissed until Tde1511 is proven to not perform this function. Thus, HftA would then bind free ferrous iron, oxidize it itself into ferric iron, and transfer this substrate to HftC to translocate across the inner membrane. HftB would then oxidize HftA, and HftD would then pick up this electron to donate to a cytoplasmic terminal electron acceptor. Nonetheless, current bioinformatic analysis of this putative iron import system shows the strongest agreement for the first proposed method of iron import as the most likely to occur.

4.2.2 Structural Characterization

Initial crystallization of Tde1511 was quite successful on the first attempt, allowing greater time to be spent on crystal growth optimization for structural determination. During the first round of crystallographic data collection at Cornell High Energy Synchrotron Source (CHESS), the monomeric structure was solved. In other data sets collected at that time,

unintentional spherical electron density was observed in the bioinformatically proposed metal binding site containing the ligands H94, M141, H148, and H178* (* denotes ligand originating from another monomer). This was observed prior to identifying Tde1511 as a member of COG3470, sparking its further bioinformatic characterization. This led to the discovery that homologous proteins to Tde1511 require copper to coordinate first, causing a conformational shift which opens the iron binding site ^{61,62}. With the apo, monomeric structure of Tde1511 solved, crystal growth was then tested after treating Tde1511 exhaustively with EDTA and reconstituting the structure with cupric cations. Crystal production was successful, and these were harvested for data collection. A handful of these crystals were then subsequently soaked in the same mother liquor supplemented with ferric chloride for 2h. Ferrous chloride supplemented with DTT was avoided due to no secondary measures on hand that could verify the reduction of iron throughout the soaking time frame. Without this, confident modelling of ferrous iron in the binding site could not occur. Native data sets were collected, and X-ray fluorescence (XRF) scans were conducted to verify the elemental composition of the crystals.

As a result, three successful structures of Tde1511 were obtained: apo, copper reconstituted, and copper/iron reconstituted. Due to the high sequence similarity of Tde1511 with Tp34 (PDB 3JPL) at 69%, successful molecular replacement solutions were found. Apo Tde1511 is a beta sandwich protein, possessing an H-type Ig-like fold. Ig-like proteins typically show functions in molecular recognition and substrate binding and occur in both Gram positive and negative bacteria ^{110,111}. The H-type subcategorization is noted due to the number of strands within the structure and the location of strands C' and D ¹¹¹. Due to the presence of 8 beta strands within the overall structure, and strands C' and D connect directly while spanning from one beta sheet to the next, an H-type Ig-like fold is created. To categorize this structure further, many Ig-

like fold containing proteins also contain a Greek Key motif ^{111,112}. The Greek Key motif is presented when four local beta strands coordinate with a topology order of strands 1-4-3-2 or 4-1-2-3 ^{112,113}. The order of these beta strands follows the pattern of either +1, +1, -3 or -1, -1, +3. The Greek Key motif can then be subdivided into (1, 3), (3, 1), or (2, 2) categories based on how many strands are hydrogen bonded and in what sheet of the beta sandwich ¹¹⁴. Using strands B, C, C'/D, and E, Tde1511 is shown to hold a (2, 2) Greek Key motif. Strands B and E (1 and 4) hydrogen bond in one beta sheet in a pattern of +1, +1, -3 while strands C and C'/D form hydrogen bonds in the other beta sheet.

Inspection of the metal binding sites reveals the bioinformatically identified residues for metal coordination were correct. A cupric ion was identified within the spherical electron density coordinated in a tetrahedral geometry via interactions with NE2 of H94, H148, and H178*, and through the δ -sulfur of M141. The presence of copper within the copper and copper/iron structures of Tde1511 was also confirmed through XRF scanning. The modelling of copper into these structures can also be supported through HSAB theory. Copper is a soft acid, and thus shows a preference to being coordinated by soft bases, such as sulphur and imidazole nitrogen of histidine side chains. Furthermore, it was identified that Tde1511 contains a conserved MxM motif typically found in copper binding proteins in the PcoC family ⁵⁹. The “x” residue can either be glycine, alanine, or proline, and the latter methionine is what coordinates with the metal ⁵⁹. In Tde1511, “x” is proline, and the latter methionine is in fact coordinating with the copper cation. On top of the fact the modelled copper ion shares similar B-factors to its surrounding local residues, many lines of information were used to confidently model copper into the observed electron density. Modelling iron into the structure was also confidently supported using similar lines of evidence. Iron is typically coordinated with acidic residues, and HSAB highlights that

iron, a hard acid, preferentially coordinates with hard bases, such as oxygen^{100,102}. After identifying conserved residues in a previous MSA, it was expected to find iron coordinated in an acidic pocket in an octahedral geometry containing E96, D98, D145, and a water molecule. What was unanticipated is the iron coordination using a histidine side chain from the affinity tag. This may be explained due to the binding site for iron being surface exposed, its relative short distance from the N-terminal affinity tag, and the fact that iron coordination through histidine side chains is a common interaction found in biological systems¹⁰². With XRF scans confirming the presence of iron in the protein crystal as well as the comparative B-factors shared between the iron ion and local residues within the environment, the decision to model iron in this second spherical electron density identified is confidently supported.

4.2.3 *Functional Properties in Metal Binding*

Both ICP-OES and DSF were used to identify metals Tde1511 preferentially interacts with and how these metals influence the overall structural stability of the protein. ICP-OES revealed many associations of Tde1511 with various divalent metal ions across a range of concentrations. Relevant metals were identified when their measured concentration surpassed their respective LOQ, as this value represents the concentration of analyte that can be consistently measured with reproducibility. The conclusion of this analysis resulted in the observation that Tde1511 can associate with metals not exclusive to just copper and iron. Noted interactions were observed with copper, iron, nickel, zinc, cobalt, and manganese. This was unsurprising as homologous proteins to Tde1511, such as Tp34, have been co-crystallized using zinc, cobalt, and nickel³⁶. Therefore, the idea that this family of proteins is involved with divalent metal transport as well as iron transport has been proposed^{36,54}.

To further quantify the interactions observed from the ICP-OES results, DSF was employed to determine if interacting metals conferred any structural stability throughout the protein structure. The thought process was that if meaningful metal coordination interactions were formed, the protein structure would require greater energy inputs to overcome these coordinating bonds and denature the protein structure. The EDTA-treated sample of Tde1511 showed impressive melting temperature increases when the addition of zinc or cobalt were used. Ultimately, copper still provided the largest increase in melting temperature, which was also identified in the homolog FtrA⁶². Next, Cu-Tde1511 was incubated with other period 4 transition metals and positive increases in melting temperature were observed for all. Slight raises in melting temperature were observed when Cu-Tde1511 interacted with cobalt, zinc, and nickel. The largest increases in melting temperature came from the addition of manganese, ferrous iron, and ferric iron produced the largest increase in melting temperature. With copper and iron being the proposed cofactor and substrate, respectively, it came to no surprise that the combination of these two metals produced the largest increase in melting temperature for Tde1511. With copper occupying the primary binding site, a hypothesis of Tde1511 functioning as a divalent metal transporter can be substantiated through these results. The only obstacle then becomes identifying permease complexes capable of trafficking zinc, cobalt, nickel, or manganese across the inner membrane. Nonetheless, the main result gathered from these functional analyses is that Tde1511 promotes meaningful interactions in its coordination of copper in its primary binding site and iron in its secondary binding site.

Future directions in Tde1511 characterization would include the use of isothermal titration calorimetry (ITC) to quantify the dissociation constants for each metal and establish quantified metal binding affinities. It would also be advised to characterize members of

HftABCD to have experimental evidence of their functional properties. This will strengthen the validity of the proposed iron import systems. To determine which model of iron import to use, ferroxidase activity of Tde1511 also needs to be confirmed or denied going forward. Iron has shown to be a required element for this pathogen's growth promotion and infection. To quantify how significant the putative HftABCD operon is to the survival of this organism will allow for novel therapeutics to be designed to limit the reproductive success of *T. denticola* during periodontal infections.

4.3 Analysis of Tde2714

4.3.1 Bioinformatics

The lack of experimental functional annotations for Tde2714 required the use of bioinformatics to gather information and begin to form functional hypotheses. The EggNOG database identified Tde2714 as a member of COG1262, housing FGEs for sulphatase activity. Using BLASTp, an overwhelming number of results annotated for FGE-domain containing proteins or previously characterized FGEs were produced. This was then coupled by the results found through InterPro which stated the presence of an FGE-domain in the proposed catalytic domain of Tde2714. A study quantifying proteins that *T. denticola* enriches in OMVs had identified that Tde2714 is highly concentrated specifically on the outer membrane surface of these vesicles³³. It had also identified six other FGE-domain containing proteins to also be enriched on the outer membrane surface of OMVs. With this result, it was found that one of these FGE-domain containing proteins, Tde2269, had been characterized as TvpA, a diversity generating retroelement (DGR)⁶⁹. Following the same principle as antibodies, DGRs can accommodate massive sequence variation, thought to be used in novel ligand binding

interactions ^{69,73}. The authors that characterized TvpA also note a significant sequence similarity to FGEs due to the FGE-fold being a subtype of the larger C-type lectin fold that TvpA uses to accommodate its sequence variation ⁶⁹. An analysis of DGR elements in *T. denticola* identified that Tde2714 does not contain the correct sequence to function as a DGR ⁷¹. The C-type lectin fold, also found in Tde2714, typically facilitates glycan binding through heavily conserved motifs “EPN” or “WND” ⁷⁴. These motifs are not present in Tde2714. Finally, the program DeepFRI had identified that the proposed catalytic domain of Tde2714 performed cysteine-type endopeptidase activities ⁹³.

This new result was quickly supported through functional annotations produced through the I-TASSER suite of programs ⁹⁴. An interaction of a peptide substrate was expected as FGEs act on protein substrates. The identification of a proposed interaction with a protease inhibitor then sparked the search for proteolytic functions. SDS-PAGE separation of Tde2714 showed multiple degradation products being produced, with the greatest intensity bands correlating to the molecular weights of the predicted repeating subunits and putative catalytic domain identified in the AlphFold2 predicted structure of Tde2714. A search for proteases capable of cleaving Tde2714 at these points resulted in the identification of clostripain, a cysteine-type endopeptidase.

Generated MSAs comparing Tde2714 against FGEs and clostripain proteases also support the putative proteolytic function. Comparisons with FGEs resulted in the observation that Tde2714 does not contain the required catalytic residues to perform FGE activity. Upon comparison to clostripain and clostripain-like primary sequences, striking conservations were seen. Unfortunately, Tde2714 does not have its catalytic cysteine align with the catalytic cysteine from clostripain. Interestingly, the histidine conserved from this MSA is the same histidine

conserved when Tde2714 was compared against putative clostripain-like enzymes. The cysteine that was identified as conserved was six residues away from a cysteine that was heavily conserved in the clostripain-like enzymes. It is this association that prompted mutant sequences of Tde2714 to be created which mutate this His and Cys into alanine. When sequences are ordered and arrived, a spectroscopic assay using clostripain substrate N-benzoyl-L-arginine ethyl ester (BAEE) can be conducted to confirm clostripain-type proteolytic activity.

4.3.2 *Structural Characterization*

Crystals of Tde2714 were difficult to consistently obtain, but the seeding technique used significantly aided in the volume of crystals produced. Due to these conditions being previously identified, it was thought that a fresh set of hands may be able to produce a Tde2714 crystal appropriate for complete structural determination. It was determined that a suitable crystal was created, which diffracted to an upper resolution limit of 2.5 Å. As this was better than prior data collections, the AlphaFold2 predicted model of Tde2714 was used to identify a structural solution. Through this, residues 126-375 had a successful structural solution identified, despite that a full-length construct of the protein is what was submitted for crystallization. Due to these residues comprising part of the protein sequence with no apparent functional properties and no annotated domain predictions, structure refinement was not pursued. The AlphaFold2 predicted structure of Tde2714 and identified conserved residues identified through MSAs were both used to create truncated mutants of Tde2714. These mutants only contain residues 378-663, and conserved residues were mapped onto the predicted catalytic domain of Tde2714 to identify conserved His/Cys diads localized within close proximity from one another that could promote proteolytic catalysis. This identified H452, C487, and C493 as potential catalytic residues for site

directed mutagenesis. For future structural determinations, residues 378-663 should be focused on. A native, H452A mutant, C487A mutant, and C493A mutant of Tde2714 should be screened for ideal protein crystallization solutions.

4.3.3 *Functional Properties in Proteolysis*

Due to previous peptide reactions using one of the four FGE recognition sequences with Tde2714 resulting in an identified peptide modification, all FGE recognition sequences were synthesized in peptide substrates with a negative control. FGEs modify a catalytic cysteine or serine residue into a formylglycine residue. They recognize this catalytic residue in 1 of 4 specific recognition motifs: CxPxR, SxPxR, CxAxR, and SxAxR ¹¹⁵. In addition to these, a negative control peptide containing the sequence AxPxR was also analyzed. Initial peptide reactions using all available peptide sequences produced TLC spotting patterns that were highly similar to each other, and each peptide produced a similar spotting pattern to the peptide reactions conducted by a prior student. With the expectation that only one FGE recognition sequence should be recognized and modified, it is apparent that no discrimination occurred between each of the five peptide substrates including the negative control.

With functional results that are unaligned to what was expected for FGE activity, the hypothesis to support a proteolytic function was strengthened. Peptide TLC analysis with the incorporation of increasing concentrations of PIC produced exciting initial results. After analyzing each component of the reaction mixtures through TLC to visualize the various spotting patterns produced, PIC introduction into the reaction mixtures putatively produced expected results that indicate proteolysis is occurring. Until these product spots can be quantified and characterized through analytical methods, such as liquid chromatography-tandem mass

spectrometry (LC-MS/MS), definitive functional annotations cannot be assigned. As shown in Figure 32, spotting products identified with black arrows appear to decrease in their intensity as the concentration of PIC increases. With many inhibitors inhibiting clostripain following a reversible mechanism, it is expected that the introduction of PIC will not abolish all proteolytic activity of Tde2714 ⁷⁷.

The results from the TLC analysis prompted a protein degradation assay using protein substrates found on hand and incubating them with Tde2714 in the presence or absence of PIC. Unfortunately, none of the protein substrates evaluated exhibited degradation over the two-week period. Interestingly, it was Tde2714 itself that saw a change in its SDS-PAGE banding pattern when PIC was incorporated or not. After two weeks, the sample of Tde2714 supplemented with 5x PIC saw an intense band on the SDS-PAGE gel that aligned with a molecular weight of about 74 kDa. This is the expected molecular weight of the full-length peptide sequence of Tde2714. When Tde2714 was stored without PIC, an intense band was observed on the two-week SDS-PAGE gel that aligned with a molecular weight of about 34 kDa. This is the expected molecular weight of the proposed catalytic domain of Tde2714. If this was caused through unintentional degradation of the protein structure, the same intensity band at 34 kDa would be seen in samples of Tde2714 with PIC as well. Clostripain recognizes and binds to arginine in its P1 site and performs catalysis on the carboxyl end of Arg residues ⁷⁵⁻⁷⁷. Using the AlphaFold2 predicted structure of Tde2714, it is observed that an arginine residue is located in the short random coil region that connects each smaller domain throughout the peptide chain. To couple this with the fact that each FGE peptide synthesized contained more than one arginine residue, the function of an arginine-directed cysteine endopeptidase begins to be substantiated.

Future directions for this gene target outside of the previously mentioned mutant generation and truncations would be to confirm potential proteolytic activity. Unequivocal results to support a function of a clostripain-like enzyme would be to perform reactions using the substrates BAEE and diethyl pyrocarbonate. BAEE contains a structure that is recognized by clostripain and its proteolytic degradation can be followed using spectroscopic methods. Additionally, diethyl pyrocarbonate binds to and inhibits histidine residues that may be required for catalysis to occur ⁷⁷. If these two substrates produce their expected positive results, clostripain functionality can be confidently assigned to Tde2714. This will allow a higher understanding of the mechanisms of infection used by *T. denticola* to attack and invade its host. Placing enzymes like this on the outside of OMVs may facilitate their trafficking through the ECM, allowing successful navigation through proteinaceous regions of the biofilm until other DGRs on the surface bind and anchor the OMV to its targeted surface.

Chapter 5. Conclusion

The goal of this research was to structurally and functionally characterize three gene products which putatively contribute to the virulence of *T. denticola* and promote the progression and severity of periodontal disease. Tde1414 was initially annotated as a putative phosphonopyruvate decarboxylase, Tde1511 was characterized as a putative pathogen-specific surface antigen with functions in iron transport, and Tde2714 showed domain homologies that coincided with FGE catalysis. This research was successful in its structural efforts to characterize Tde1414, Tde1511, and about 42% of the total structure of Tde2714. It was also successful in its

functional analyses, strongly substantiating the expected function of Tde1511 and providing a new functional hypothesis for Tde2714.

Due to *T. denticola* containing gene products for phosphonate biosynthesis resulting in the formation of AEP, initial hypotheses for Tde1414 were that this enzyme aided in the production of AEP, resulting in the modification of cell membrane phospholipids to incorporate this head group throughout its surface. Phosphonates are known to be resistant to several hydrolysis methods, thus, incorporation of AEP into the phospholipid bilayer may strengthen membrane integrity during infection and allow *T. denticola* to invade and persist within host tissues^{31,38,39,42}. Bioinformatic results converged unanimously on a function as a phosphonopyruvate decarboxylase, as well as functional residues required for cofactor binding. This enzyme family is known to require thiamine diphosphate and Mg^{2+} as cofactors to facilitate catalysis. As a result, the first ever reported structural characterization of a phosphonopyruvate decarboxylase was completed while complexed with its required cofactors at a resolution of 1.95 Å. Due to difficulties in organic synthesis of phosphonopyruvate and the high financial burden this substrate carries to commercially purchase, functional annotations for Tde1414 are ongoing. Confirmation of Pnpy decarboxylation activity as well as crystal structures of Tde1414 complexed with Pnpy will solidify the complete characterization of Tde1414 as a phosphonopyruvate decarboxylase.

Bioinformatic characterization of Tde1511 also provided strong convergence on the expected activity of this protein to be involved in iron binding and transport. Iron is known to be a significant mineral that aids in *T. denticola* growth promotion, and pathogenicity as a cofactor in virulent metalloproteins^{20,35,116}. This knowledge allowed for accurate structural modelling which produced structures of apo Tde1511 at 1.79 Å (PDB 8T7K), copper reconstituted Tde1511

at 1.89 Å (PDB 8T7L), and copper and iron reconstituted Tde1511 at 1.82 Å (PDB 8T7M). Functional analysis provided reassurance in the expected function of Tde1511. Although this protein may interact with various divalent metal cations, Tde1511 forms the most significant interactions with both copper and iron. The similarities identified between proteins found in *T. denticola* to the FtrABCD iron import system allowed for a putative iron uptake operon, HftABCD, to be bioinformatically identified and substantiated. Iron-limited environments have shown to increase the expression of HftA, HftC, and other MacB-domain like ABC transporters identified in Table 2⁶⁴. Further characterization of the other Hft members will provide accurate functional annotations to validate their role in the proposed mechanism of iron import in Figure 17. Growth studies for cell cultures of *T. denticola* lacking all Hft members will also provide answers to the significance this putative iron import operon holds in the growth, survival, and infection processes of this bacterium.

With initial bioinformatic results of Tde2714 converging on a FGE related function, this proved to be the most challenging protein target to assign a functional annotation to. Initial generation of MSAs and functional analyses using peptides containing FGE recognition motifs provided the results needed to abandon the hypothesis that Tde2714 is an FGE. Further functional work incorporating PIC tablets into reaction mixtures began to mirror new bioinformatic results that Tde2714 may hold proteolytic functions. Characterization as a clostripain-like protease began to make sense, as all peptides (including Tde2714) that were monitored for protease activity all included arginine residues, which these proteases discriminate for. Although not perfect, generated MSAs provided an initial guide to putative catalytic residues required for clostripain-like proteolysis. Coupled with the lack of significant structural characterization for the full-length sequence of Tde2714, truncations that include just the

proposed catalytic domain were created, and sequences were generated which mutated putative catalytic residues. In the future, the native truncation of Tde2714 should be crystallized and structurally determined. Mutant truncations should then be assayed for proteolytic activity using BAEE and ethyl pyrocarbnate substrates. If all these experiments produce positive results, Tde2714 can be confidently annotated as a clostripain-like protease. This will provide new mechanistic insights into how *T. denticola* OMVs traverse the biofilm matrix and effectively reach their intended destination. It may also uncover proteolytic activity directed to cell membrane proteins, revealing hidden cryptic epitopes to bind to. The virulent capacity of this protein can be quantified, and potential therapeutics may be designed to cause a greater difficulty in *T. denticola* promoting infection from within the biofilm.

In conclusion, the results presented within reveal new insights into the structural and functional capabilities of three identified gene targets thought to be leveraged by *T. denticola* during infection, resulting in the progression of periodontal disease. To date, periodontal disease still ranks as one of the most prevalent issues negatively impacting society. Due to the high costs of professional dental help and its inaccessibility to many populations, novel solutions are required to mitigate irreversible damage caused to patients experiencing periodontal disease. Through this work, characterizing essential virulence factors used by *T. denticola* for infection may lead to the generation of affordable, highly accessible therapeutics incorporated into toothpastes and mouth washes to provide immediate relief for all patients suffering from the effects of periodontal disease.

References

- (1) Pihlstrom, B. L.; Michalowicz, B. S.; Johnson, N. W. Periodontal Diseases. *The Lancet* **2005**, *366* (9499), 1809–1820.
- (2) Marsh, P. D. Dental Plaque as a Biofilm and a Microbial Community - Implications for Health and Disease. *BMC Oral Health* **2006**, *6* (Suppl. 1), S14.
- (3) Yu, O. Y.; Zhao, I. S.; Mei, M. L.; Lo, E. C. M.; Chu, C. H. Dental Biofilm and Laboratory Microbial Culture Models for Cariology Research. *Dentistry Journal* **2017**, *5* (2), 21.
- (4) Ramírez-Larrota, J. S.; Eckhard, U. An Introduction to Bacterial Biofilms and Their Proteases, and Their Roles in Host Infection and Immune Evasion. *Biomolecules* **2022**, *12* (2), 306.
- (5) Flemming, H. C.; Wingender, J. The Biofilm Matrix. *Nat. Rev. Microbiol.* **2010**, *8*, 623–633.
- (6) Roshna, T.; Nandakumar, K. Generalized Aggressive Periodontitis and Its Treatment Options: Case Reports and Review of the Literature. *Case Rep. Med.* **2012**, *2012*, 535321.
- (7) Trindade, D.; Carvalho, R.; Machado, V.; Chambrone, L.; Mendes, J. J.; Botelho, J. Prevalence of Periodontitis in Dentate People between 2011 and 2020: A Systematic Review and Meta-Analysis of Epidemiological Studies. *J. Clin Periodontol.* **2023**, *50* (5), 604-626.
- (8) Dye, B. A. Global Periodontal Disease Epidemiology. *Periodontol 2000* **2012**, *58* (1), 10-25.
- (9) Eke, P. I.; Dye, B. A.; Wei, L.; Slade, G. D.; Thornton-Evans, G. O.; Borgnakke, W. S.; Taylor, G. W.; Page, R. C.; Beck, J. D.; Genco, R. J. Update on Prevalence of Periodontitis in Adults in the United States: NHANES 2009 to 2012. *J. Periodontol* **2015**, *86* (5), 611–622.
- (10) Botelho, J.; Machado, V.; Leira, Y.; Proença, L.; Chambrone, L.; Mendes, J. J. Economic Burden of Periodontitis in the United States and Europe: An Updated Estimation. *J Periodontol* **2022**, *93* (3), 373–379.
- (11) Jain, N.; Dutt, U.; Radenkov, I.; Jain, S. WHO's Global Oral Health Status Report 2022: Actions, Discussion and Implementation. *Oral Dis.* **2023**.
- (12) Russell, A. L. A System of Classification and Scoring for Prevalence Surveys of Periodontal Disease. *J Dent Res* **1956**, *35* (3), 350–359.
- (13) Dhingra, K.; Vandana, K. L. Indices for Measuring Periodontitis: A Literature Review. *Int Dent J.* **2011**, *61* (2), 76–84.
- (14) Beltrán-Aguilar, E. D.; Eke, P. I.; Thornton-Evans, G.; Petersen, P. E. Recording and Surveillance Systems for Periodontal Diseases. *Periodontol 2000* **2012**, *60* (1), 40–53.
- (15) Nomura, Y.; Okada, A.; Kakuta, E.; Gunji, T.; Kajiura, S.; Hanada, N. A New Screening Method for Periodontitis: An Alternative to the Community Periodontal Index. *BMC Oral Health* **2016**, *16* (1), 64.
- (16) Graziani, F.; Karapetsa, D.; Alonso, B.; Herrera, D. Nonsurgical and Surgical Treatment of Periodontitis: How Many Options for One Disease? *Periodontol 2000* **2017**, *75* (1), 152–188.

- (17) Mohanty, R.; Asopa, S. J.; Joseph, M. D.; Singh. Bhupender; Rajguru, J. P.; Saidath, K.; Sharma, U. Red Complex: Polymicrobial Conglomerate in Oral Flora: A Review. *J Family Med Prim Care* **2019**, 8 (11), 3480-3486.
- (18) Sahni, J.; Talegaonkar, S.; Tariq, M.; Ahmad, Z.; Ali, J.; Baboota, S.; Iqbal, Z. Treatment Modalities and Evaluation Models for Periodontitis. *Int J Pharm Investig* **2012**, 2 (3), 106.
- (19) Socransky, S. S.; Haffajee, A. D.; Cugini, M. A.; Smith, C.; Kent Jr., R. L. Microbial Complexes in Subgingival Plaque. *J Clin Periodontol* **1998**, 25 (2), 134-144.
- (20) Sela, M. N. Role of *Treponema denticola* in Periodontal Diseases. *Crit Rev Oral Biol Med* **2001**, 12 (5), 399-413.
- (21) Kumawat, R.; Ganvir, S.; Hazarey, V.; Qureshi, A.; Purohit, H. Detection of Porphyromonas Gingivalis and Treponema Denticola in Chronic and Aggressive Periodontitis Patients: A Comparative Polymerase Chain Reaction Study. *Contemp Clin Dent* **2016**, 7 (4), 481-486.
- (22) Mitchell, H. L.; Dashper, S. G.; Catmull, D. V.; Paolini, R. A.; Cleal, S. M.; Slakeski, N.; Tan, K. H.; Reynolds, E. C. Treponema Denticola Biofilm-Induced Expression of a Bacteriophage, Toxin-Antitoxin Systems and Transposases. *Microbiology* **2010**, 156 (3), 774-788.
- (23) Ishihara, K. Virulence Factors of *Treponema Denticola*. *Periodontol 2000* **2010**, 54 (1), 117-135.
- (24) Subramanian, G.; Koonin, E. V; Aravind, L. Comparative Genome Analysis of the Pathogenic Spirochetes *Borrelia Burgdorferi* and *Treponema Pallidum*. *Infect Immun* **2000**, 68 (3), 1633-1648.
- (25) Seshadri, R.; A Myers, G. S.; Tettelin, H.; Eisen, J. A.; Heidelberg, J. F.; Dodson, R. J.; Davidsen, T. M.; DeBoy, R. T.; Fouts, D. E.; Haft, D. H.; Selengut, J.; Ren, Q.; Brinkac, L. M.; Madupu, R.; Kolonay, J.; Durkin, S. A.; Daugherty, S. C.; Shetty, J.; Shvartsbeyn, A.; Gebregeorgis, E.; Geer, K.; Tsegaye, G.; Malek, J.; Ayodeji, B.; Shatsman, S.; McLeod, M. P.; majs, D. S.; Howell, J. K.; Pal, S.; Amin, A.; Vashisth, P.; McNeill, T. Z.; Xiang, Q.; Sodergren, E.; Baca, E.; Weinstock, G. M.; Norris, S. J.; Fraser, C. M.; Paulsen, I. T. Comparison of the Genome of the Oral Pathogen *Treponema Denticola* with Other Spirochete Genomes. *PNAS USA* **2004**, 101 (15), 5646-5651.
- (26) Sharma, A. K.; Dhasmana, N.; Dubey, N.; Kumar, N.; Gangwal, A.; Gupta, M.; Singh, Y. Bacterial Virulence Factors: Secreted for Survival. *Indian J Microbiol* **2017**, 57 (1), 1-10.
- (27) Martínez, O. F.; Cardoso, M. H.; Ribeiro, S. M.; Franco, O. L. Recent Advances in Anti-Virulence Therapeutic Strategies with a Focus on Dismantling Bacterial Membrane Microdomains, Toxin Neutralization, Quorum-Sensing Interference and Biofilm Inhibition. *Front Cell Infect Microbiol* **2019**, 9, 74.
- (28) Barczak, A. K.; Hung, D. T. Productive Steps toward an Antimicrobial Targeting Virulence. *Curr Opin Microbiol* **2009**, 12 (5), 490-496.
- (29) Rasko, D. A.; Sperandio, V. Anti-Virulence Strategies to Combat Bacteria-Mediated Disease. *Nat Rev Drug Discov* **2010**, 9, 117-128.
- (30) Heras, B.; Scanlon, M. J.; Martin, J. L. Targeting Virulence Not Viability in the Search for Future Antibacterials. *Br J Clin Pharmacol* **2015**, 79 (2), 208-215.

- (31) Rice, K.; Batul, K.; Whiteside, J.; Kelso, J.; Papinski, M.; Schmidt, E.; Pratasouskaya, A.; Wang, D.; Sullivan, R.; Bartlett, C.; Weadge, J. T.; Van der Kamp, M. W.; Moreno-Hagelsieb, G.; Suits, M. D.; Horsman, G. P. The Predominance of Nucleotidyl Activation in Bacterial Phosphonate Biosynthesis. *Nat Commun* **2019**, *10* (1), 3698.
- (32) Veith, P. D.; Dashper, S. G.; O'Brien-Simpson, N. M.; Paolini, R. A.; Orth, R.; Walsh, K. A.; Reynolds, E. C. Major Proteins and Antigens of *Treponema Denticola*. *Biochim Biophys Acta Proteins Proteom* **2009**, *1794* (10), 1421–1432.
- (33) Veith, P. D.; Glew, M. D.; Gorasia, D. G.; Chen, D.; O'Brien-Simpson, N. M.; Reynolds, E. C. Localization of Outer Membrane Proteins in *Treponema Denticola* by Quantitative Proteome Analyses of Outer Membrane Vesicles and Cellular Fractions. *J Proteome Res* **2019**, *18* (4), 1567–1581.
- (34) Schwechheimer, C.; Kuehn, M. J. Outer-Membrane Vesicles from Gram-Negative Bacteria: Biogenesis and Functions. *Nat Rev Microbiol* **2015**, *13*, 605–619.
- (35) Koch, D.; Chan, A. C. K.; Murphy, M. E. P.; Lilie, H.; Grass, G.; Nies, D. H. Characterization of a Dipartite Iron Uptake System from Uropathogenic *Escherichia Coli* Strain F11. *J Biol Chem* **2011**, *286* (28), 25317–25330.
- (36) Deka, R. K.; Brautigam, C. A.; Tomson, F. L.; Lumpkins, S. B.; Tomchick, D. R.; Machius, M.; Norgard, M. V. Crystal Structure of the Tp34 (TP0971) Lipoprotein of *Treponema Pallidum*: Implications of Its Metal-Bound State and Affinity for Human Lactoferrin. *J Biol Chem* **2007**, *282* (8), 5944–5958.
- (37) Yu, X.; Doroghazi, J. R.; Janga, S. C.; Zhang, J. K.; Circello, B.; Griffin, B. M.; Labeda, D. P.; Metcalf, W. W. Diversity and Abundance of Phosphonate Biosynthetic Genes in Nature. *PNAS USA* **2013**, *110* (51), 20759–20764.
- (38) Kafarski, P. Phosphonates: Their Natural Occurrence and Physiological Role. In Churchill, D.; Sikiric, M. D.; Čolović, B.; Milhofer, H. F. Contemporary Topics about Phosphorus in Biology and Materials. IntechOpen, 2019.
- (39) Li, S.; Horsman, G. P. An Inventory of Early Branch Points in Microbial Phosphonate Biosynthesis. *Microb Genom* **2022**, *8* (2), 000781.
- (40) Ju, K. S.; Doroghazi, J. R.; Metcalf, W. W. Genomics-Enabled Discovery of Phosphonate Natural Products and Their Biosynthetic Pathways. *J Ind Microbiol Biotechnol* **2014**, *41* (2), 345–356.
- (41) Zhang, G.; Dai, J.; Lu, Z.; Dunaway-Mariano, D. The Phosphonopyruvate Decarboxylase from *Bacteroides Fragilis*. *J Biol Chem* **2003**, *278* (42), 41302–41308.
- (42) Jia, H.; Chen, Y.; Chen, Y.; Liu, R.; Zhang, Q.; Bartlam, M. Structural Characterization of a 2-Aminoethylphosphonate:Pyruvate Aminotransferase from *Pseudomonas Aeruginosa* PAO1. *Biochem Biophys Res Commun* **2021**, *552*, 114–119.
- (43) Hasson, M. S.; Muscate, A.; Mcleish, M. J.; Polovnikova, L. S.; Gerlt, J. A.; Kenyon, G. L.; Petsko, G. A.; Ringe, D. The Crystal Structure of Benzoylformate Decarboxylase at 1.6 Å Resolution: Diversity

- of Catalytic Residues in Thiamin Diphosphate-Dependent Enzymes. *Biochemistry* **1998**, 37 (28), 9918-9930.
- (44) Dyda, F.; Furey, W.; Swaminathan, S.; Sax, M.; Farrenkopf, B.; Jordanl, F. Catalytic Centers in the Thiamin Diphosphate Dependent Enzyme Pyruvate Decarboxylase at 2.4-Å Resolution. *Biochemistry* **1993**, 32 (24), 6165-6170.
 - (45) Mitschke, L.; Parthier, C.; Schröder-Tittmann, K.; Coy, J.; Lüdtke, S.; Tittmann, K. The Crystal Structure of Human Transketolase and New Insights into Its Mode of Action. *J Biol Chem* **2010**, 285 (41), 31559–31570.
 - (46) Nemeria, N. S.; Chakraborty, S.; Balakrishnan, A.; Jordan, F. Reaction Mechanisms of Thiamin Diphosphate Enzymes: Defining States of Ionization and Tautomerization of the Cofactor at Individual Steps. *FEBS J* **2009**, 276 (9), 2432–3446.
 - (47) Kluger, R. Thiamin Diphosphate: A Mechanistic Update on Enzymic and Nonenzymic Catalysis of Decarboxylation. *Chem Rev* **1987**, 87 (5), 863-876.
 - (48) Polovnikova, E. S.; McLeish, M. J.; Sergienko, E. A.; Burgner, J. T.; Anderson, N. L.; Bera, A. K.; Jordan, F.; Kenyon, G. L.; Hasson, M. S. Structural and Kinetic Analysis of Catalysis by a Thiamin Diphosphate-Dependent Enzyme, Benzoylformate Decarboxylase. *Biochemistry* **2003**, 42 (7), 1820–1830.
 - (49) Carpenter, C.; Payne, S. M. Regulation of Iron Transport Systems in Enterobacteriaceae in Response to Oxygen and Iron Availability. *J Inorg Biochem* **2014**, 133, 110–117.
 - (50) Lau, C. K. Y.; Krewulak, K. D.; Vogel, H. J. Bacterial Ferrous Iron Transport: The Feo System. *FEMS Microbiol Rev* **2015**, 40 (2), 273–298.
 - (51) Ma, Z.; Jacobsen, F. E.; Giedroc, D. P. Coordination Chemistry of Bacterial Metal Transport and Sensing. *Chem Rev* **2009**, 109 (10), 4644–4681.
 - (52) Fenton, H. J. H. Oxidation of Tartaric Acid in Presence of Iron. *J Chem Soc, Trans* **1894**, 65, 899-910.
 - (53) Haber, F.; Weiss, J. The Catalytic Decomposition of Hydrogen Peroxide by Iron Salts. *Proc R Soc Lond A* **1934**, 147 (861), 332-351.
 - (54) Brautigam, C. A.; Deka, R. K.; Ouyang, Z.; Machius, M.; Knutsen, G.; Tomchick, D. R.; Norgard, M. V. Biophysical and Bioinformatic Analyses Implicate the *Treponema Pallidum* Tp34 Lipoprotein (Tp0971) in Transition Metal Homeostasis. *J Bacteriol* **2012**, 194 (24), 6771–6781.
 - (55) Koppenol, W. H. The Haber-Weiss Cycle - 70 Years Later. *Redox Rep* **2001**, 6 (4), 229–234.
 - (56) Takahashi, T.; Shiku, H. Cell Surface Antigens: Invaluable Landmarks Reflecting the Nature of Cells. *Cancer Immun* **2012**, 12, 2.
 - (57) Lawrence, M. C.; Pilling, P. A.; Chandana Epa, V.; Berry, A. M.; Ogunniyi, A. D.; Paton, J. C. The Crystal Structure of Pneumococcal Surface Antigen PsaA Reveals a Metal-Binding Site and a Novel Structure for a Putative ABC-Type Binding Protein. *Structure* **1998**, 6 (12), 1553-1561.

- (58) Ogunniyi, A. D.; Paton, J. Chapter 4 - Vaccine Potential of Pneumococcal Proteins. In Brown, J.; Hammerschmidt, S.; Orihuela, C. *Streptococcus Pneumoniae: Molecular Mechanisms of Host-Pathogen Interactions*. Academic Press, 2015; pp 59–78.
- (59) Chan, A. C. K.; Doukov, T. I.; Scofield, M.; Tom-Yew, S. A. L.; Ramin, A. B.; MacKichan, J. K.; Gaynor, E. C.; Murphy, M. E. P. Structure and Function of P19, a High-Affinity Iron Transporter of the Human Pathogen *Campylobacter Jejuni*. *J Mol Biol* **2010**, *401* (4), 590–604.
- (60) Dubbels, B. L.; DiSpirito, A. A.; Morton, J. D.; Semrau, J. D.; Neto, J. N. E.; Bazylnski, D. A. Evidence for a Copper-Dependent Iron Transport System in the Marine, Magnetotactic Bacterium Strain MV-1. *Microbiology* **2004**, *150* (9), 2931–2945.
- (61) Chan, A. C. K.; Lin, H.; Koch, D.; Grass, G.; Nies, D. H.; Murphy, M. E. P. A Copper Site Is Required for Iron Transport by the Periplasmic Proteins P19 and FetP. *Metallomics* **2020**, *12*, 1530.
- (62) Steunou, A. S.; Vigouroux, A.; Aumont-Nicaise, M.; Plancqueel, S.; Boussac, A.; Ouchane, S.; Moréra, S. New Insights into the Mechanism of Iron Transport through the Bacterial Ftr System Present in Pathogens. *FEBS J* **2022**, *289* (20), 6286–6307.
- (63) Kovacs-Simon, A.; Titball, R. W.; Michell, S. L. Lipoproteins of Bacterial Pathogens. *Infect Immun* **2011**, *79* (2), 548–561.
- (64) Paula Goetting-Minesky, M.; Godovikova, V.; Fenno, J. C.; Wooten, R. M.; Munro, R.; Kreth, J. Approaches to Understanding Mechanisms of Dentilisin Protease Complex Expression in *Treponema Denticola*. *Front Cell Infect Microbiol* **2021**, *11*, 668287.
- (65) López-Otín, C.; Bond, J. S. Proteases: Multifunctional Enzymes in Life and Disease. *J Biol Chem* **2008**, *283* (45), 30433–30437.
- (66) Buttle, D. J.; Mort, J. S. Cysteine Proteases. In William, J.; Lennarz, M.; Lane, D. *Encyclopedia of Biological Chemistry: Second Edition*. Academic Press, 2013; pp 589–592.
- (67) Mótýán, J.; Tóth, F.; Tózsér, J. Research Applications of Proteolytic Enzymes in Molecular Biology. *Biomolecules* **2013**, *3* (4), 923–942.
- (68) Ishihara, K.; Okuda, K. Molecular Pathogenesis of the Cell Surface Proteins and Lipids from *Treponema Denticola*. *FEMS Microbiol Lett* **1999**, *181* (2), 199–204.
- (69) Le Coq, J.; Ghosh, P. Conservation of the C-Type Lectin Fold for Massive Sequence Variation in a *Treponema* Diversity-Generating Retroelement. *PNAS USA* **2011**, *108* (35), 14649–14653.
- (70) Handa, S.; Shaw, K. L.; Ghosh, P. Crystal Structure of a *Thermus Aquaticus* Diversity-Generating Retroelement Variable Protein. *PLoS One* **2019**, *14* (1), e0205618.
- (71) Nimkulrat, S.; Lee, H.; Doak, T. G.; Ye, Y. Genomic and Metagenomic Analysis of Diversity-Generating Retroelements Associated with *Treponema Denticola*. *Front Microbiol* **2016**, *7*, 852.
- (72) Wu, L.; Gingery, M.; Abebe, M.; Arambula, D.; Czornyj, E.; Handa, S.; Khan, H.; Liu, M.; Pohlschroder, M.; Shaw, K. L.; Du, A.; Guo, H.; Ghosh, P.; Miller, J. F.; Zimmerly, S. Diversity-Generating Retroelements: Natural Variation, Classification and Evolution Inferred from a Large-Scale Genomic Survey. *Nucleic Acids Res* **2018**, *46* (1), 11–24.

- (73) McMahon, S. A.; Miller, J. L.; Lawton, J. A.; Kerkow, D. E.; Hodes, A.; Marti-Renom, M. A.; Doulatov, S.; Narayanan, E.; Sali, A.; Miller, J. F.; Ghosh, P. The C-Type Lectin Fold as an Evolutionary Solution for Massive Sequence Variation. *Nat Struct Mol Biol* **2005**, *12* (10), 886–892.
- (74) Cummings, R. D.; McEver, R. P. C-type Lectins. In Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E. *Essentials of Glycobiology*, 2nd ed; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 2009; Chapter 34.
- (75) Labrou, N. E.; Rigden, D. J. The Structure-Function Relationship in the Clostripain Family of Peptidases. *Eur J Biochem* **2004**, *271* (5), 983–992.
- (76) Mitchell, W. M.; Harrington, W. F. Purification and Properties of Clostridiopeptidase B (Clostripain). *J Biol Chem* **1968**, *243* (18), 4683–4692.
- (77) Kembhavi, A. A.; Buttle, D. J.; Rauber, P.; Barrett, A. J. Clostripain: Characterization of the Active Site. *FEBS Lett* **1991**, *283* (2), 277–280.
- (78) McLuskey, K.; Grewal, J. S.; Das, D.; Godzik, A.; Lesley, S. A.; Deacon, A. M.; Coombs, G. H.; Elsliger, M. A.; Wilson, I. A.; Mottram, J. C. Crystal Structure and Activity Studies of the C11 Cysteine Peptidase from Parabacteroides Merdae in the Human Gut Microbiome. *J Biol Chem* **2016**, *291* (18), 9482–9491.
- (79) Chen, J. M.; Rawlings, N. D.; Stevens, R. A. e.; Barrett, A. J. Identification of the Active Site of Legumain Links It to Caspases, Clostripain and Gingipains in a New Clan of Cysteine Endopeptidases. *FEBS Lett* **1998**, *441* (3), 361–365.
- (80) Roncase, E. J.; Moon, C.; Chatterjee, S.; González-Páez, G. E.; Craik, C. S.; O'Donoghue, A. J.; Wolan, D. W. Substrate Profiling and High Resolution Co-Complex Crystal Structure of a Secreted C11 Protease Conserved across Commensal Bacteria. *ACS Chem Biol* **2017**, *12* (6), 1556–1565.
- (81) Lithgow, K. V.; Buchholz, V. C. H.; Ku, E.; Konshuh, S.; D'Aubeterre, A.; Sycuro, L. K. Protease Activities of Vaginal Porphyromonas Species Disrupt Coagulation and Extracellular Matrix in the Cervicovaginal Niche. *NPJ Biofilms Microbiomes* **2022**, *8* (1), 8.
- (82) Nawaz, M.; Shah, N.; Zanetti, B. R.; Maugeri, M.; Silvestre, R. N.; Fatima, F.; Neder, L.; Valadi, H. Review Extracellular Vesicles and Matrix Remodeling Enzymes: The Emerging Roles in Extracellular Matrix Remodeling, Progression of Diseases and Tissue Repair. *Cells* **2018**, *7* (10), 167.
- (83) Foschi, F.; Izard, J.; Sasaki, H.; Sambri, V.; Prati, C.; Muller, R.; Stashenko, P. *Treponema Denticola* in Disseminating Endodontic Infections. *J Dent Res* **2006**, *85* (8), 761–765.
- (84) Nakai, K.; Horton, P. PSORT: A Program for Detecting Sorting Signals in Proteins and Predicting Their Subcellular Localization. *Trends Biochem Sci* **1999**, *24* (1), 34–35.
- (85) Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S. K.; Cook, H.; Mende, D. R.; Letunic, I.; Rattei, T.; Jensen, L. J.; Von Mering, C.; Bork, P. EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res* **2019**, *47* (D1), D309–D314.

- (86) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bursteinas, B.; Bye-A-Jee, H.; Coetzee, R.; Cukura, A.; da Silva, A.; Denny, P.; Dogan, T.; Ebenezer, T. G.; Fan, J.; Castro, L. G.; Garmiri, P.; Georghiou, G.; Gonzales, L.; Hatton-Ellis, E.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Jokinen, P.; Joshi, V.; Jyothi, D.; Lock, A.; Lopez, R.; Luciani, A.; Luo, J.; Lussi, Y.; MacDougall, A.; Madeira, F.; Mahmoudy, M.; Menchi, M.; Mishra, A.; Moulang, K.; Nightingale, A.; Oliveira, C. S.; Pundir, S.; Qi, G.; Raj, S.; Rice, D.; Lopez, M. R.; Saidi, R.; Sampson, J.; Sawford, T.; Speretta, E.; Turner, E.; Tyagi, N.; Vasudev, P.; Volynkin, V.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M. C.; Bolleman, J.; Boutet, E.; Breuza, L.; Casals-Casas, C.; de Castro, E.; Echioukh, K. C.; Coudert, E.; CuChe, B.; Doche, M.; Dornevil, D.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Keller, G.; Kerhornou, A.; Lara, V.; Le Mercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T. B.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Pozzato, M.; Pruess, M.; Rivoire, C.; Sigrist, C.; Sonesson, K.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Garavelli, J. S.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Yeh, L. S.; Zhang, J.; Ruch, P.; Teodoro, D. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res* **2021**, *49* (D1), D480–D489.
- (87) Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets. *Nucleic Acids Res* **2021**, *49* (D1), D605–D612.
- (88) McGinnis, S.; Madden, T. L. BLAST: At the Core of a Powerful and Diverse Set of Sequence Analysis Tools. *Nucleic Acids Res* **2004**, *32*, W20–W25.
- (89) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res* **2003**, *31* (13), 3784–3788.
- (90) Hunter, S.; Apweiler, R.; Attwood, T. K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L.; Finn, R. D.; Gough, J.; Haft, D.; Hulo, N.; Kahn, D.; Kelly, E.; Laugraud, A.; Letunic, I.; Lonsdale, D.; Lopez, R.; Madera, M.; Maslen, J.; Mcanulla, C.; McDowall, J.; Mistry, J.; Mitchell, A.; Mulder, N.; Natale, D.; Orengo, C.; Quinn, A. F.; Selengut, J. D.; Sigrist, C. J. A.; Thimma, M.; Thomas, P. D.; Valentin, F.; Wilson, D.; Wu, C. H.; Yeats, C. InterPro: The Integrative Protein Signature Database. *Nucleic Acids Res* **2009**, *37* (Suppl. 1), D211–D215.
- (91) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.
- (92) Akdel, M.; Pires, D. E. V.; Pardo, E. P.; Jänes, J.; Zalevsky, A. O.; Mészáros, B.; Bryant, P.; Good, L. L.; Laskowski, R. A.; Pozzati, G.; Shenoy, A.; Zhu, W.; Kundrotas, P.; Serra, V. R.; Rodrigues, C. H. M.;

- Dunham, A. S.; Burke, D.; Borkakoti, N.; Velankar, S.; Frost, A.; Basquin, J.; Lindorff-Larsen, K.; Bateman, A.; Kajava, A. V.; Valencia, A.; Ovchinnikov, S.; Durairaj, J.; Ascher, D. B.; Thornton, J. M.; Davey, N. E.; Stein, A.; Elofsson, A.; Croll, T. I.; Beltrao, P. A Structural Biology Community Assessment of AlphaFold2 Applications. *Nat Struct Mol Biol* **2022**, *29* (11), 1056–1067.
- (93) Gligorijevic, V.; Douglas Renfrew, P.; Kosciolk, T.; Leman, J. K.; Cho, K.; Vatanen, T.; Berenberg, D.; Taylor, B.; Fisk, I. M.; Xavier, R. J.; Knight, R.; Bonneau, R.; Chandler, C.; Vlamakis, H. Structure-Based Protein Function Prediction Using Graph Convolutional Networks. *Nat Commun* **2021**, *12* (1), 3168.
- (94) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat Methods* **2015**, *12* (1), 7–8.
- (95) Widmann, M.; Radloff, R.; Pleiss, J. The Thiamine Diphosphate Dependent Enzyme Engineering Database: A Tool for the Systematic Analysis of Sequence and Structure Relations. *BMC Biochem* **2010**, *11* (9).
- (96) Johnen, S.; Sprenger, G. A. Characterization of Recombinant Thiamine Diphosphate-Dependent Phosphonopyruvate Decarboxylase from *Streptomyces Viridochromogenes* Tü494. *J Mol Catal B Enzym* **2009**, *61* (1–2), 39–46.
- (97) Pallitsch, K.; Rogers, M. P.; Andrews, F. H.; Hammerschmidt, F.; McLeish, M. J. Phosphonodifluoropyruvate Is a Mechanism-Based Inhibitor of Phosphonopyruvate Decarboxylase from *Bacteroides Fragilis*. *Bioorg Med Chem* **2017**, *25* (16), 4368–4374.
- (98) Hawkins, C. F.; Borges, A.; Perham, R. N. A Common Structural Motif in Thiamin Pyrophosphate-Binding Enzymes. *FEBS Lett* **1989**, *255* (1), 77–82.
- (99) Gruber, M.; Greisen, P.; Junker, C. M.; Hélix-Nielsen, C. Phosphorus Binding Sites in Proteins: Structural Preorganization and Coordination. *J Phys Chem B* **2014**, *118* (5), 1207–1215.
- (100) Pearson, R. G. Hard and Soft Acids and Bases- the Evolution of a Chemical Concept. *Coord Chem Rev* **1990**, *100*, 403–425.
- (101) Christodoulides, A.; Boyadjian, A.; Kelesidis, T. Spirochetal Lipoproteins and Immune Evasion. *Front Immunol* **2017**, *8*, 364.
- (102) Harding, M. M. The Architecture of Metal Coordination Groups in Proteins. *Acta Cryst* **2004**, *60* (Pt. 5), 849–859.
- (103) Crow, A.; Greene, N. P.; Kaplan, E.; Koronakis, V. Structure and Mechanotransmission Mechanism of the MacB ABC Transporter Superfamily. *PNAS USA* **2017**, *114* (47), 12572–12577.
- (104) Brickman, T. J.; Armstrong, S. K. Iron and PH-Responsive FtrABCD Ferrous Iron Utilization System of *Bordetella* Species. *Mol Microbiol* **2012**, *86* (3), 580–593.
- (105) Elhassanny, A. E. M.; Anderson, E. S.; Menscher, E. A.; Roop II, R. M. The Ferrous Iron Transporter FtrABCD Is Required for the Virulence of *Brucella Abortus* 2308 in Mice. *Mol. Microbiol* **2013**, *88* (6), 1070–1082.

- (106) Stearman, R.; Yuan, D. S.; Yamaguchi-Iwai, Y.; Klausner, R. D.; Dancis, A. A Permease-Oxidase Complex Involved in High-Affinity Iron Uptake in Yeast. *Science* **1996**, *271* (5255), 1552-1557.
- (107) Fang, H.-M.; Wang, Y. Characterization of Iron-Binding Motifs in *Candida Albicans* High-Affinity Iron Permease CaFtr1p by Site-Directed Mutagenesis. *Biochem J* **2002**, *368* (Pt. 2), 641-647.
- (108) Hallgren, J.; Tsigos, K. D.; Pedersen, M. D.; Armenteros, J. J. A.; Marcatili, P.; Nielsen, H.; Krogh, A.; Winther, O. DeepTMHMM Predicts Alpha and Beta Transmembrane Proteins Using Deep Neural Networks. *bioRxiv* **2022**.
- (109) Méheust, R.; Huang, S.; Rivera-Lugo, R.; Banfield, J. F.; Light, S. H. Post-Translational Flavinylation Is Associated with Diverse Extracytosolic Redox Functionalities throughout Bacterial Life. *Elife* **2021**, *10*, e66878.
- (110) Bodeló, G.; Palomino, C.; Fernández, L. A. Immunoglobulin Domains in Escherichia Coli and Other Enterobacteria: From Pathogenesis to Applications in Antibody Technologies. *FEMS Microbial Rev.* **2012**, *37* (2), 204-250.
- (111) Bork, P.; Holm, L.; Sander, C. The Immunoglobulin Fold: Structural Classification, Sequence Patterns and Common Core. *J. Mol. Biol.* **1994**, *242* (2), 309–320.
- (112) Zhang, C.; Kim, S.-H. A Comprehensive Analysis of the Greek Key Motifs in Protein-Barrels and Sandwiches. *Proteins* **2000**, *40* (3), 409-419.
- (113) Woolfson, D. N.; Evans, P. A.; Hutchinson, E. G.; Thornton, J. M. Topological and Stereochemical Restrictions in Beta-Sandwich Protein Structures. *Protein Eng* **1993**, *6* (5), 461-470.
- (114) Richardson, J. S. Beta-Sheet Topology and the Relatedness of Proteins. *Nature* **1977**, *286*, 495-500.
- (115) Rush, J. S.; Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J Am Chem Soc* **2008**, *130* (37), 12240–12241.
- (116) Scott, D.; Chan, E. C. S.; Siboo, R. Iron Acquisition by Oral Hemolytic Spirochetes: Isolation of a Hemin-Binding Protein and Identification of Iron Reductase Activity. *Can J Microbiol* **1996**, *42* (10), 1072-1079.