Wilfrid Laurier University

Scholars Commons @ Laurier

Theses and Dissertations (Comprehensive)

2019

Structural and Functional Analysis of Three Upregulated Gene Products, TDE0626, TDE1701, and TDE2714 from Treponema denticola During Biofilm Formation

Jonah Nechacov
nech1450@mylaurier.ca

Recommended Citation

Structural and Functional Analysis of Three Upregulated Gene Products, TDE0626, TDE1701, and TDE2714 from *Treponema denticola* During Biofilm Formation

by

Jonah Nechacov

Honours Biochemistry and Biotechnology with Thesis, Wilfrid Laurier University, 2016

THESIS

Submitted to the Department of Chemistry and Biochemistry

in partial fulfillment of the requirements for

Master of Science in Chemistry

Wilfrid Laurier University

Abstract

The progression of human chronic periodontitis within periodontal disease has been often

linked to the presence of key pathogens, such as the presence of *Treponema denticola*, a late colonizer

found in the deepening pockets of the gingival sulcus. This pathogen, as well as its associates

Porphyromonas gingivalis and *Tannerella forsythia*, are classified as the 'red complex' and exist in a

mixed biofilm during infection. It is within this biofilm state that previous transcriptomic analysis

revealed a total of 126 genes that had an increase in their expression by 1.5-fold or greater in *T.

denticola*. Three of these genes, *Tde0626*, *Tde1701*, and *Tde2714*, were further investigated for their

importance in pathogenicity. Initial bioinformatics analyses suggested putative functions of the gene

products to be a polysaccharide lyase for TDE0626, a bacteriocin-like protein for TDE1701, and a large

formylglycine-generating enzyme (FGE) for TDE2714. Each of the targets were recombinantly expressed

and purified for structural and functional analysis. Both TDE1701 and TDE2714 were able to be

successfully crystallized and X-ray diffraction data were and analyzed. The processed data collected for

TDE1701 resulted an I/σI value of 2.2 at 2.4Å and CC$^{1/2}$ value of 0.823. Attempts thus far to obtain a

structure of TDE1701 via molecular replacement have been unsuccessful. Similarly, selenomethionine

data were collected for TDE2714, processed to 2.7Å resolution with an I/σI value of 2.2. However, the

data quality was insufficient to determine appropriate phase information to characterize the structure.

In both cases, further crystallization will be required for definitive structural characterization. In order to

test the hypothesis that TDE0626 was a protein responsible for microbial dispersal, attempts at

visualization of polysaccharide lyase activity for TDE0626 when incubated with isolates of mixed 'red

complex' biofilm was attempted, but unsuccessful, and further investigation will be required. At present,

functional analysis of TDE1701 has yet to be conducted due to the lack of functional hypotheses.

Incubation of TDE2714 with a sulfatase consensus sequence (-LCTPSRA-) revealed an unanticipated

modification to the serine rather than the cysteine residue. The data showed a mass loss of 2.02 g/mol

by ESI-MS/MS on the serine residue suggesting that TDE2714 may be a serine modifying FGE. Further

investigation will be conducted to confirm these results.

Acknowledgments

I gratefully acknowledge my thesis supervisor Dr. Michael Suits for his integral role in the success of my work. I also was to thank him for giving me the opportunity work in his lab as well as training me to be a more confident leader within my field of research. I want to acknowledge Dr. Joel Weadge and Dr. Geoff Horsman for their great help and guidance throughout my thesis. I would also like to thank Dr. Lillian DeBruin for her guidance while on my committee and for her encouraging words during my project. I would like to acknowledge Dr. Dyanne Brewer for her mass spectrometry work on my elusive peptide substrates. I want to thank Maria Ngo and Dr. Sean Liston for their help in training me on the essential laboratory techniques I used during my project. I want to thank Vasu Patel for his excellent work in furthering the TDE0626 project as well as Peter Nguyen for developing the TLC protocol I used in my work. I acknowledge all the Suits Lab, Weadge Lab, and Horsman Lab members past and present who aided me in the daily tasks of my project. And Finally, I want to thank my parents and my sister for always seeing the positives and encouraging me to forge ahead even during the challenging times.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AHL   N-acyl homoserine lactone

APS   Advanced Photon Source Synchrotron

APS   Ammonium persulfate

Bis-Tris   Bis-tris methane

BLAST   Basic Local Alignment Search Tool

CLS   Canadian Light Source Synchrotron

C-PAGE   Carbohydrate-based polyacrylamide gel electrophoresis

Cthe   *Clostridium thermocellum*

C-Type   Calcium-dependent

DTT   Dithiothreitol

DGR   Diversity-generating retroelements

DLS   Dynamic light scattering

EDTA   Ethylenediaminetetraacetic acid

EPS   Extracellular polymeric substance

ESI   Electron-spray ionization

FGE   Formylglycine-generating enzyme

FPLC   Fast protein liquid chromatography

GAG   Glycosaminoglycan

G3P   Glyceraldehyde-3-phosphate

HEPES   4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HHM   Hidden-Markov model

IMAC   Immobilized metal affinity chromatography

IP   Induced pathogenicity

IPTG   Isopropyl β-D-1-thiogalactopyranoside

KEGG   Kyoto Encyclopedia of Genes and Genomes

LB   Luria-Bertani

LC   Liquid chromatography

MAD   Multi-wavelength anomalous diffraction

MDR   Multidrug resistance

MES   2-ethanesulfonic acid

MIC	Minimal inhibitory concentration
MR	Molecular replacement
MS	Mass Spectrometry
MW	Molecular weight
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
OD	Optical Density
PAGE	Polyacrylamide gel electrophoresis
%Pd	Percent polydispersity
PEG	Polyethylene glycol
Phyre^2	Protein Homology/Analogy Recognition Engine 2
pI	Isoelectric point
SAD	Single-wavelength anomalous diffraction
SDS	Sodium dodecyl sulfate
SIC	Sub-inhibitory concentration

TDE	*Treponema denticola*
TEMED	Tetramethylethylenediamine
TLC	Thin-layer chromatography
Tris	Trisaminomethane

1. Introduction

1.1 – Human Periodontal Disease and Biofilm infections

1.1.1 – Biofilm infections

Human Periodontal disease is an oral polymicrobial disease which affects the tissues

surrounding the teeth, including gums, periodontal ligaments, and alveolar bone(1). The global

epidemiology estimates that between 5-20% of the global human population has a form of severe

periodontitis or its several destructive stages(2). In contrast to other prevailing human diseases,

periodontal disease can be categorized as a polymicrobial disease based on the formation of mixed

bacterial biofilms which left untreated can progress the disease(3). The formation of a biofilm on a hard

surface has been identified to have five main stages which are; attachment, microcolony formation,

biofilm formation, maturation, and dispersal of planktonic bacteria(4). Each of these stages are critical

for the survival of pathogens within a host(4). The attachment and development of microcolonies to

hard surfaces such as the tooth surface allows for planktonic bacteria to gain access to nutrients as well

as promote more frequent genetic exchange and promote the sharing of metabolites(5). As more

microcolonies are formed by the aggregation of planktonic cells and colony growth, the metabolism of

nutrients allows for the secretion of extracellular polymeric substances (EPS) consisting of proteins,

polysaccharides, RNA, and DNA(6). This is general considered to be a biofilm. The formation of the early

biofilm may cause for the further incorporation of other species which contribute to the overall diversity

of the biofilm and subsequent infection(7). Further development leads to the maturation of the biofilm

which is characterized by the three-dimensional development of EPS that contains a variety of systems

which are able to aid in the evasion of the immune system(8). The cross-talk between the organisms

within the biofilm allows for horizontal gene transfer in order to induce indirect pathogenicity (IP) in

antibiotic sensitive organisms(9). This is a key exchange component in the proliferation of biofilm

infections as the collective resistance to host defenses allows for the further colonization into the affect

host tissues(10). The recycling of nutrients and by-products within the biofilm also aids in development,

as inaccessible or secondary metabolites can be made readily available to bacteria which in a planktonic

state would have been unavailable. An example of this was shown through the demonstrated utilization

of the lactic acid by-product from *Streptococcal* metabolism by *Aggregatibacter*

actinomycetemcomitans(11). As the maturation occurs, the concentration of available or accessible

nutrients for organisms within the biofilm is decreased, which causes for the detachment and dispersal

of planktonic cells in search of new areas of high nutrient concentrations(12). Colonizing new areas

within the host allows for the increased evasion and decreased chance of eradication by the host

immune system(13). The dispersal stage of biofilm infections is a complex stage, consisting of diverse

processes such as the degradation polymeric substances by secreted enzymes, activation of signal

transduction pathways, and release of chemical effectors. To date, no single mechanism for the

dispersal of a biofilm has been discovered(14).

1.1.2 – Periodontal Disease

Human periodontal disease occurs in two stages of infection based on the formation maturation

of a polymicrobial biofilm(15). Periodontal disease can be classified into two main stages depending on

the severity; gingivitis, and periodontitis(15). The common route of entry for the pathogens is at the free

gingival margin where a host of bacteria attach to the tooth surface and produce a dental plaque or

biofilm(16). It is at this stage that the presence of the bacteria causes for an inflammatory response of

the affected tissues. This inflammation of the gums is referred to a gingivitis and is the beginning stages

of periodontal disease. Consequently, treatments for the early stage gingivitis focus on the removal of

the plaque biofilm which can be accomplished by mechanical displacement through cleaning/flossing

and oral rinses using chlorhexidine or hydrogen peroxide(17). Other methods for removal of the biofilm

involve common dental cleaning practices such as flossing and removal of dental plaque(18). Studies

have shown that the presence of bacteria which produce biologically-active molecules can promote the

proinflammatory response of the host by the increased release cytokines(19). Some common

compounds which are known to cause this are bacterial lipopolysaccharides (endotoxins), chemotactic

peptides, and various protein toxins(19).

If left untreated, the periodontal infection can progress into six more severe forms of the

periodontal disease. The stages of periodontal disease progression are; chronic periodontitis, aggressive

periodontitis, systemic disease periodontitis, necrotizing ulcerative gingivitis/periodontitis,

periodontium abscesses, and combined periodontic-endodontic lesions(20). These latter stages have

been categorized based on their damaging effects on the four main host tissues which are gingiva,

alveolar bone, cementum, and the periodontal ligaments(21). The first stage in the progression of

periodontal disease is chronic human periodontitis which is characterized by the formation of deep

bleeding pockets within the gingival sulcus(22). It is progressed by the colonizing of pathogens which

originated from the dispersal of the surface biofilm and attached to the tooth neck and roots. As the

biofilm has begun to mature within the gingival sulcus, the degradation of the periodontal ligaments as

well as the resorption of the alveolar bone can be observed(23). It is at this stage where the irreversible

damage first occurs. Unique to chronic periodontitis is that the initiation of the disease can be linked to

the presence of gram-negative bacteria which all are suggested to thrive in maturing biofilm(24). At

present it has been shown that there are both anaerobic and aerobic bacteria which are linked to the

progression of the disease(25). Clinical studies conducted on patients with chronic periodontitis showed

that the affected tissues contained 30 species of bacteria which are associated with the progression of

the disease into more severe symptoms(26). Based on the bacterial association such as the localization

within the biofilm as well as related tissue damage, five distinct colored categorizes were developed

which were; the blue, yellow, and green (early colonizers), the purple and orange (bridging colonizer),

and the red complex (late colonizers)(27). Their localization within the periodontal pocket was also

associated, not directly linked to, their relative pathogenicity within the disease(28). The early colonizers

such as *Eikenella corrodens* and *Campylobacter rectus*, located on or near the tooth root surface, were

classified as moderately pathogenic(28). Whereas, the late colonizers at the pocket epithelium such as

Porphyromonas gingivalis and *Tannerella forsythia* were classified as very highly pathogenic(28). The

presence of the red complex bacteria, which are *P. gingivalis* and *T. forsythia*, and *Treponema. denticola*,

within the deep periodontal pockets has been linked to the progression into the later stage of aggressive

periodontitis(29). This later stage has been known to be uncommon in human epidemiology, where 1 in 1000 patents with periodontitis developed the aggressive form of the disease(30). The disease progression is characterized by the localized or the generalized loss of periodontal ligaments which support the teeth to the alveolar bone. These periodontal lesions are linked to the presence of the red complex bacteria namely, *P. gingivalis*, and the spirochete *T. denticola* as they have the ability to deepen the periodontal lesions(31). Understanding the progression of chronic periodontitis into the more severe forms and the development of treatments has been focused around the red complex bacteria and their contributions to the disease.

1.2 – Treatment and Prevention of Chronic Periodontitis

1.2.1 – Conventional Treatment of Periodontitis

Conventional treatments for chronic periodontitis, in comparison to treatment of gingivitis, have been known to be more invasive due to the development of deep periodontal pockets. In these cases, standard brushing and flossing regiments are rendered ineffective when faced with the progressive destruction of periodontal ligaments and surrounding tissues. In turn, the most common therapy for chronic periodontitis has been the use of antibiotics which can be coupled with surgical removal of the developing biofilm and affected tissues. However, due to the wide variety of bacteria present within the biofilms and the surrounding tissues, either one of these treatments alone have been ineffective in the complete eradication of the infection. The overuse of antibiotics to treat bacterial infections has led to the development of widespread resistance in pathogenic bacteria(32). The undesirable development of resistance has been more pronounced in the treatment of biofilm infections, where pathogens are able

5

to utilize the antibiotics as signal molecules for intra- or interspecies interactions to express genes which

are able to aid in the survival of the pathogens(32). One example of enhanced pathogenicity caused by

the use of antibiotics was found in the increased expression of phosphodiesterases within the biofilm

which was shown to increase cell surface adhesion. This was found to be caused by the activation of the

aminoglycoside response factor which could bind to sub-inhibitory concentrations (SICs) of

aminoglycoside antibiotics(33). The ability for pathogens to adapt to treatment against antibiotics within

biofilms has been also linked to the gradient concentrations of nutrients from the surface of the biofilm

to the center(34). This has been shown to be caused by the dormant bacteria which play key roles in the

tolerance of antibiotics(35). Studies have shown that the increased expression of regulatory proteins in

Pseudomonas aeruginosa is linked to the conversion of resistant and sensitive forms and can be

impacted by the concentrations of antibiotics(36). Similarly, the increased IL-17 antibody expression in

diabetes patients showed to affect the pathogenicity of oral pathogens(37). Other factors which can

affect the antibiotic treatment of biofilms has been known to be due to the increased expression of

multidrug resistance (MDR) efflux pumps(38). The efflux pumps are highly active within a biofilm and

can render the treatment of certain antibiotics ineffective in eradicating the total biofilm(39). In

addition, *in vitro* and *in vivo* experimentation on streptococcal biofilms showed an increase in the

minimal inhibitory concentration (MIC) for antibiotics of 10-1000 times relative to planktonic

Streptococcal(35). Although the use of antibiotics alone may be ineffective in eradicating the whole

biofilm, they are able to affect the planktonic bacteria and aid in the decrease of biofilm dispersal(40).

Surgical endodontic treatment of periodontitis has been one of the most effective and widely

used methods for treating the infection. The use of scaling and rooting to remove the deep pocket

biofilms has been employed to treat the periodontal lesion formed during the infection(41). However, a

key challenge in the microsurgery is the requirement of extensive bone and tissue grafting post-

surgery(42). Negative treatment outcomes for the use of endodontic surgery have been the main cause

for concern post-surgery(43). This is due to the re-development of biofilm infections and tissue damage,

resulting in retreatment procedures that can pose irreversible damage to the teeth and tissues(44).

Investigations into the coupling of both surgical and non-surgical treatments have proven to be the most

effective to date(43). However, significant recovery time, due to accidental nerve damage, and potential

relapse of infection has been commonly seen in patients treated(45,46). In short, the conventional

treatment methods for chronic periodontitis require extensive treatment with the potential for

irreversible damage to patients. This has led to alternative treatments to be employed which could

significantly reduce the irreversible damage as well as better eradicate the biofilm infection.

1.2.2 – Alternative Treatments for Disrupting Biofilm Formation

A more recent technique for biofilm infection eradication has been the targeting disruption of

the quorum sensing system present in the biofilms. This system has been found to be employed by the

bacteria within a biofilm to regulate cooperative activities as well as the respond and adapt to the

changing environment surrounding the biofilm. The system hinges on two main categories of molecules,

N-acyl homoserine lactones (AHLs) and oligopeptides, which are expressed and secreted into the

extracellular matrix. These molecules are utilized by other bacteria for "cell-cell communication". This communication allows for the bacteria to coordinate the levels gene expression. The increased expression of quorum effector molecules has been proposed to be key in the division of bacterial labor as well as the repression of growth in certain species of bacteria for nutrient availability. The secreted oligopeptides which are called autoinducers, are used as signaling molecules for the quorum system to enable bacterial competence and virulence(47). The importance of quorum sensing systems for biofilm formation, survival, and maturation has posed as a good target for the development of inhibitory molecules(48). A key target for quorum sensing inhibition is AHL synthase, which synthesizes AHLs for the system. Compounds such as L-S-adenosylcysteine and D/L-S-adenosylhomocysteine, have been used for the inhibition and quenching of the quorum sensing system(49). Structural analogues of the secreted oligopeptides has also been shown to aid in the inhibition of the system(50).

1.3 – Factors Affecting Pathogenicity

1.3.1 – Role of Virulence Factors in Biofilm

The formation and survival of a biofilm requires the need for expressed factors which can aid in the individual survival of the bacterium as well as the maintenance of the biofilm as a whole. These survival factors are commonly secreted into the matrix or the surrounding tissues where they can perform a host of different functions. The main function of a bacterial virulence factor is to increase the pathogenicity of the organism by accomplishing one or more of the following: colonization, immunoevasion, immunosuppression, and nutrient acquisition. Investigations into these systems such as the survivability of the pathogens within a host has allowed for new insight into the field of drug

therapy. This more specified approach to treatment has been shown to be advantageous in eradicating

essential pathogen species while maintaining the host microbiome, in comparison to traditional

antibiotics(51). In order for a successful long-term biofilm infection to form, the pathogens must express

and utilize adhesion molecules and proteins, called adhesins(52). Most commonly found adhesins in

gram-negative dominant biofilms, such as in periodontitis, are the trimeric autotransporter

adhesins(53). These are outer membrane proteins have been known to bind to the epithelial cells by the

attachment of the N-terminal head domain to structural elements such as collagen and fibronectin(54).

These adhesins play a key role in the formation of a bacterial biofilm and may be a target for future drug

development(55). Similarly, the initial survival of the long-term biofilm has been known to require the

appropriate factors which can suppress and evade the host immune system. The activation of the host

immune system produces immunoglobulin antibodies which can bind to surface antigens on the

pathogen for destruction(56). It has been shown that pathogens within a developing biofilm produce

immunoglobulin proteases as virulence factors in order to evade the host immune system(57). The

evasion of the immune system has also been associated with the suppression of the host immune

system by the interactions between the host immune cells and the pathogens. The most studied

immunosuppression methods for pathogens is to prevent phagolysosome biogenesis and the inhibition

of cellular autophagy(58). These factors were first identified in the interaction between *Mycobacterium*

tuberculosis with host macrophages(59). There have also been studies that show some pathogens

exhibit the ability to inhibit antigen presentation to immune cells(60).

Of the host of virulence factors which can contribute to the pathogenicity of the organisms, the most frequently studied proteins/enzymes are critical for the reliable obtaining of nutrients, further colonization and dispersal, as well as contribute to the prolonged evasion of the host immune system(61). These contribute to the progression of destructive infections as they are essential for the long-term survival of the pathogens and biofilm. One means of the survival for bacteria is obtaining inaccessible nutrients as metabolites as the availability of common metabolites is diminished by other present pathogens(62). A class of enzymes commonly associated with the further colonization of pathogens in a biofilm infection is the polysaccharide lyases(63). These enzymes are secreted into the extracellular matrix where N-acetyl-glycosaminoglycans (GAGs) such as hyaluronan and heparin sulfate are present. It is here that the enzymes have been known to degrade the polysaccharides via an anti-β-elimination reaction that cleaves the α-(1,4)-linked uronic acids to produce an unsaturated hexenuronic acid residue and a reducing end(64). The degradation of GAGs within a biofilm has been shown to be a key step in the dispersal of the biofilm. There have also been studies which showed the degradation of GAGs by polysaccharide lyases was involved in bacterial gluconeogenesis(65). This has been shown by using co-expressed ABC-transporter proteins to bring the hexenuronic acid residues into the cells to generate glyceraldehyde 3-phosphate (G3P)(66). This involvement of polysaccharide lyases as well as other lyases reveal the multi-faceted function of these essential factors.

Similarly, access to host nutrients is of importance for pathogenic infection as there are higher concentrations of essential metabolites and nutrients available within the host tissue. This due to the

decreased amount of available nutrients within a biofilm. It has shown that some pathogenic bacteria

such as *Streptococcus agalactiae* and *T. denticola* express large amounts of hemolytic proteins called

hemolysins(67,68). These enzymes are able to agglutinate and lyse erythrocytes in the invaded host. In

order to do so, the hemolysins are extracellularly secreted into the surrounding tissues where the host

cell membranes are degraded via pore formation. Studies showed that there are various methods in

which these hemolytic proteins can function, such as the oligomerization of toxin monomers to make

hydrophilic channels in the membrane(69), as well as the production of hydrogen disulfide (H_2S) which

can produce hemolytic activity(70). An example of the hemolytic activity of H_2S was found in the

expression of cystalysin, which is a 46kDa $C^\beta-S^\gamma$ lyase from *T. denticola*(71). The $C^\beta-S^\gamma$ lyase

demonstrated measurable hemolytic activity on erythrocytes *in vitro*. The data was coupled with prior

information showing that >2 mM concentration of H_2S have been identified within periodontal

pockets(71). This detection leads to the suggestion that hemolytic proteins expressed by pathogens can

be used in order to obtain nutrients from the host cells as well as aid in the further progression of the

disease.

1.3.2 – Development of New Treatments for Chronic Periodontitis

In order to develop new treatments for chronic disease that differ from the conventional

methods, the identification of essential factors for important pathogens is required. Several methods

have been employed over the years to identify new targets for drug design that can be either based on

genomic analysis, transcriptomic analysis, as well as proteomic analysis of the pathogens. These

methods have been used to generate a collective list of the best candidates for drug design. The

parameters for candidates are based on how essential the target is to pathogen survival in the host,

effect on the host immune system, cooperative ability with other pathogens, development of severe

symptoms, as well as general contribution to the progression of disease(72). Once the targets are

identified by the initial techniques, biochemical investigations must be conducted in order to give a

complete characterization of the target as well as an accurate prediction of its role and importance in

the pathogen and disease. Common techniques for analysis include, but are not exclusive to,

bioinformatics, X-ray crystallography, nuclear magnetic resonance, enzymatic kinetic analysis, and

subsequent development and testing of structural substrate analogues(73). Together these methods

were used in this research project for the characterization of specific targets.

For this research project, the organism *T. denticola* was focused on, based on its role in the later

stages of chronic human periodontitis. *T. denticola* is a Gram-negative obligate anaerobic spirochete

that has been found in high quantities within maturing periodontal pockets and periodontal

biofilms(74). The organism is closely associated with the other members of the red complex and has

been found to aid in the progression of periodontitis(74). In order to gain insight into important factors

for pathogen survival function, previous work done by the Reynolds group identified 126 transcripts that

were found to have a 1.5-fold increase in expression when *T. denticola* was grown in a dense

monoculture. This phenomenon is considered to be an upregulation of genes(75). These values were

compared to the expression levels of the planktonic state of *T. denticola*. This information was obtained

by micro array assays to isolate the total RNA of *T. denticola* in the biofilm. Of the 126 genes, a subset

was identified of the upregulated genes which possessed similarities to prophage genes, as well as toxin-

antitoxin protein systems(75). The presence of the prophage genes within *T. denticola* suggested that

the pathogen was associated with bacteriophage infection as the genes were identified as a part of the

genome(76). Furthermore, *T. denticola* contained a vast number of uncharacterized hypothetical

proteins that had similarly increased expression within a dense monoculture. Based on the data shown

by the transcriptomic analysis, three genes further investigated for their role and importance in *T.*

denticola. The selected genes were; *tde0626, tde1701,* and *tde2714*. These genes were not selected by

chance, but rather were selected based on upregulation and the predicted extracellular localization of

the gene products. Each of the genes were recombinantly expressed in an *Escherichia coli* expression

system and purified for biochemical analysis.

1.4 – Research Objective and Hypothesis

       The main objective for each of the targets; TDE0626, TDE1701, and TDE2714 was to determine

the structure and function of the recombinantly expressed protein to gain insight into their role in *T.*

denticola. Based on the lack of structural and functional information of these three gene products,

biochemical techniques, such as X-ray crystallography and functional assays will be conducted. The

annotation of these gene products will be used to investigate if these targets play a specific role in

within *T. denticola* and its involvement in the progression of chronic periodontitis. The identification of a

contribution to the pathogenicity of *T. denticola* will be discovered by comparison of the results to known gene products that play a role in the pathogenicity of other microbes.

The long-term objective would be accomplished primarily by the use of X-ray crystallography to generate accurate structural models as well as developing and testing functional assays and protocols which would be able to give insight into the function of the three targets.

The first short term objective was successful recombinant expression and purification of soluble protein from each of the three targets, as well as structural and functional prediction based on bioinformatics analysis. Once accomplished, the next objective was to successfully crystalize each of the proteins and obtain good quality X-ray diffraction data. This was then used to generated structural models based on molecular replacement and structural refinement. Afterwards the objective was to identify potential substrate binding of each of the protein targets by co-crystallization of mutants generated by site-direct mutagenesis. The last objective was to functional characterize each of the proteins by developing kinetic assays derived from the bioinformatics analysis.

2. Materials and Methods

2.1 – Materials

2.1.1 – Bacterial Transformation and Plasmid Extraction

Custom genes for targets TDE0626, TDE1701, and TDE2714 were codon-optimized for expression *E. coli*, housed in pET21b (+) vectors, were ordered from Bio Basic Canada Inc. (Markham, ON). Calcium chloride dihydrate, tryptone, and yeast extract were also from Bio Basic Inc. Competent

NEB 5-alpha *E. coli* cells and BL21 (DE3) *E. coli* cells were ordered from New England BioLabs Inc.

(Ipswich, MA, USA). Sterile nuclease-free water (nuclease, RNase, DNase, and Protease-free), nutating

mixer: variable speed, and the Fisherbrand minicentrifuge Spout were ordered from Fisher Scientific

(Hampton, NH, USA). Miniprep plasmid extraction kit, QIAprep Spin Minprep Kit was from Qiagen Inc.

(Toronto, ON). The Analog Dry Block Heater and the Clinical 100 Laboratory Centrifuge were from VWR

International LLC. (Mississauga, ON). Reagent grade sodium chloride and ampicillin sodium salt were

ordered from BioShop Canada Inc. (Burlington, ON). 1.5mL microtubes, 92x16mm petri dishes, and

13mL 100x16mm tubes were from Sarstedt Canada (Des Grandes Prairies, QC). The MaxQ 4000

Benchtop orbital shaker was from Thermo Fisher Scientific (Mississauga, ON).

2.1.2 – Bacterial Growth and Recombinant Expression

      Large bacterial cultures were grown in two-liter Pyrex® Shaker flasks with extra deep baffles

from Sigma Aldrich Canada (Oakville, ON). Nalgene centrifuge tubes (50mL) were also ordered from

Sigma Aldrich Canada. Large volume centrifuge Avanti Jxn-26 Centrifuge, the J-Lite® Series JLA-8.1 rotor,

and the J Series JA-20 rotor were from Beckman Coulter Canada Inc. (Mississauga, ON). BioUltraPure

grade Isopropyl β-D-1-thiogalactopyranoside (IPTG), reagent grade trisaminomethane (Tris),

Biotechnology grade Triton® X-100, ACS reagent grade magnesium chloride hexahydrate, hen egg white

lysozyme, sucrose, and reagent grade glycerol were all ordered from BioShop. Deoxycholic acid sodium

salt, DNase I, hydrochloric acid, and ultra-pure DL-dithioreitol were from Bio Basic Inc. The GENESYS™

10S UV-Vis Spectrophotometer was from Thermo Fisher Scientific. The semi-micro visible acrylic disposable cuvettes and the 50mL 114x28mm tubes were from Sarstedt Canada.

2.1.3 – Immobilized Metal Affinity Chromatography

Econo-Column® Chromatography columns and the Econo- Gradient Pump were both ordered from Bio-Rad Laboratories Canada Ltd. (Mississauga, ON) . HisPur™ Ni-NTA resin was from Thermo Fisher Scientific. Reagent grade imidazole, glacial acetic acid, and ethylenediaminetetraacetic acid (EDTA) disodium salt dihydrate were from BioShop. nickel (II) chloride Hexahydrate was ordered from Bio Basic Inc. Solid sodium hydroxide and guanidine hydrochloride were ordered from VWR International LLC. Plastic tubes (15mL) were from Sarstedt Canada.

2.1.4 – SDS-PAGE and Dialysis

BioUltraPure sodium dodecyl sulfate (SDS), 40% acrylamide, 2% bis-acrylamide, tetramethylethylenediamine (TEMED), and Coomassie Brillant Blue G-250 were all purchased from BioShop Inc. Power Pac Basic, Mini-Protean Tetra System, and the precision plus protein™ dual color ladder were from Bio-Rad Laboratories Canada Ltd. Ammonium persulfate (APS) was ordered from Fisher Scientific. Bromophenol blue was from Sigma Aldrich Canada. Glycine was ordered from Bio Basic Inc. 95% ACS methanol was from CFS Custom Chemical Solutions and the Molecular Imager® Versa Doc MP System was purchased from Bio-Rad Laboratories Ltd.

The 50mm Fisher Regenerated cellulose dialysis tubing and the Fisherbrand™ octagon SpinBar magnetic stirring bars were ordered from Fisher Scientific. Reusable plastic 2L graduated beaker was

from Owens Corning Canada (Guelph, ON). The Thermolyne Type 1000 heated stir plate was from Sigma Aldrich Canada.

2.1.5 – Anion Exchange Chromatography and Dynamic Light Scattering

Fast Protein Liquid Chromatography (FPLC) instrument was the ÄKTA Pure from GE Life Sciences Canada (Mississauga, ON) as well as the HiTrap™ Q FF anion exchange chromatography column (5mL). Ethyl alcohol (95% v/v) and the Amicon Ultra-15 centrifugal filter units were ordered from Sigma Aldrich Canada. micro UV 1.2mL disposable cuvettes were purchased from Sarstedt Canada. The Allegra X-14R centrifuge and the SX4750 Swinging-Bucket Rotor were from Beckman Coulter Canada Inc.

Dynamic light scattering (DLS) was conducted with the DynaPro PlateReader-II and Dynamic 7.1.9 program from Wyatt Technology Corporation (Santa Barbara, CA, USA). Black polystyrene microplate (96-well) was from Owens Corning Canada.

2.1.6 – Crystallization and X-Ray Diffraction Analysis

The MCSG crystallization Suite for MCSG-1,2,3, and 4 crystal screens were from Anatrace Inc. (Maumee, OH, USA). PACT Premier™ HT96 crystal screen and the 55L Bench Top Incubator (fully-programmable) were from Molecular Dimensions Ltd. (Maumee, OH, USA). Crystal Screen 1 and 2, Index™ HR2-144 crystal screen, and HR2-110 crystal screen were purchased from Hampton Research (Aliso Viejo, CA, USA). Crystal Gryphon, Intelli-Plate 96-3 3-well sitting drop plates, and 0.2mL protein tubes were purchased from Art Robbins Instruments (Sunnyvale, CA, USA). HD Clear packaging tape was

from ShurTech Brands, LLC (Avon, OH, USA). SteREO Discovery.V8 stereomicroscope was purchased

from Carl Zeiss Canada Ltd. (North York, ON).

TC Plate 24 well cell culture plates were ordered from Sarstedt Canada. Siliconized glass circle

cover slides (18mm) and monodisperse polyethylene glycol (PEG) 3350 were ordered from Hampton

Research. Reagent grade glycerol, ethylene glycol, and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic

acid (HEPES) free-acid were all purchased from BioShop Inc. Polyethylene glycol mono-methyl ether 550

and polyethylene glycol 6000 were ordered from Sigma Aldrich Canada. Sodium nitrate was from Bio

Basic Canada Inc. High vacuum grease was purchased from Owens Corning Canada.

Reusable Goniometer Base B1A and 100, 300, 600, 1000μm Dual-Thickness MicroLoops LD were

purchased from MiTeGen, LLC (Ithaca, NY, USA). Industrial grade liquid nitrogen was ordered from

Praxair Canada Inc. (Kitchener, ON). XT20 Cryogenic Storage Tank was purchased from Taylor-Wharton

(Bolivar, OH, USA). Cassette Loading and shipping kit, ALS Pucks, and shipping cane was purchased from

Crystal Positioning Systems (Jamestown, NY, USA). The 2.9GeV Storage ring Synchrotron was used at the

Canadian Light Source Synchrotron (Saskatoon, SK).

2.1.7 – Functional Assays and Analysis

High purity grade peptide (>95%) – LCTPSRA – was ordered and synthesized from GenScript

(Burlington, ON). Silica gel 60 aluminum thin-layer chromatography (TLC) sheets, Rectangluar TLC

developing tank, ACS reagent ninhydrin, 99.8% pyridine anhydrous were purchased from Sigma Aldrich

Canada. ACS grade sodium phosphate dibasic and 1-butanol were ordered from BioShop Inc. Centivap

concentrator was from Labconco (Kansas City, MO, USA). Precision Model 16 Gravity Convection Lab

Oven from GCA. Bruker AmaZon SL LC-MSn used for electron spray ionization mass spectrometry (ESI-

MS) and liquid chromatography-tandem mass spectrometry (LC-MS/MS) were from the University of

Guelph. Isolated and purified biofilm of mixed culture of red complex bacteria (*T. Forsythia, P. gingivalis,*

and *T. denticola*) were obtained by Christopher Bartlett and Sidney Nechacov from the Slawson Lab

(Dept. of Biology, WLU).

2.2 – Methods and Experimental Design

2.2.1 – Bacterial Transformation and Recombinant Protein Expression

Plasmids housing the three target genes *TDE0626*, *TDE1701*, and *TDE2714* were transformed

into both calcium competent NEB 5-alpha and BL21 (DE3) *E. coli* cells. Each of the genes were

synthesized within a pET21b vector and 40µL of nuclease-free water was added to the 4µg of custom

genes to make a working stock of each plasmid (Figure A5 in Appendix). These were then mixed via a

nutating mixer for 30 minutes prior to transformation. The transformation of the custom genes was

conducted by incubating 1µL of the plasmid with 50µL of the competent cells in sterile 1.5mL

microtubes. The incubation was at 4°C for 30 minutes and then placed at 42°C for one minute for heat

shock using the heat block. The cells were then placed back at 4°C for 3 minutes and then 200µL of

lysogeny broth (LB) (100mL stock containing 1g of tryptone, 0.5g of yeast extract, and 1g of NaCl), was

added to the cells. These were placed in the benchtop orbital shaker for one hour at 240rpm at 37°C.

Once the bacteria were sufficiently grown in the LB, the cells were spread on 100µg/mL ampicillin

containing agar plates and incubated overnight at 37°C. This allowed for selective growth.

The following day, a single colony of bacteria was chosen and grown in 10mL of LB containing

100µg/mL of ampicillin. The 10mL culture of bacteria was grown for 16 hours at 37°C and 240 rpm

overnight. Each of the transformed cell strains for the three targets were made into freezer stocks (for

long term storage) by adding 25% glycerol into a sample of the cultures. These were then frozen and

stored at -80°C.

For the transformed 5-alpha *E. coli*, the cell culture was centrifuged at 12,000xg for 5 minutes

and pelleted for a miniprep plasmid isolation. Each of the three different plasmids were miniprep via the

protocol stated in the QIAprep Spin Minprep Kit. Once the isolated plasmids were collected, the

concentration of dsDNA was found via UV readings at 260nm and 280nm (for purity) using the BioTeck

Cytation 5 image reader. The isolated plasmid DNA for each of the three targets were stored at -20°C.

For transformed BL21 (DE3) *E. coli* cells, the -80°C cell cultures were used to inoculate 1L of LB

broth containing 100µg/mL of ampicillin in 2L shaker flasks. The large cell cultures were then incubated

at 37°C and 240 rpm until an optical density at 600nm (OD$_{600}$) of 0.6 was reached. To test the OD$_{600}$, a

sample of each of the cultures were read using the GENESYS 10S UV-Vis spectrophotometer. Once the

OD$_{600}$ was approximately 0.6, a final concentration of 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG)

was added to each of the large cultures, to induce protein expression. The cultures were the incubated

at 16°C and 160 rpm for 16 hours.

The induced BL21 cells were harvested by centrifugation using the Avanti Jxn-26 centrifuge at

5,000rpm for eight minutes at a temperature of 16°C. The supernatant was decanted, and the cell

pellets were retained for osmotic shock chemical lysis and sonication. It was standardized that two cell

pellets of the same recombinant target would be used in one isolation and purification. Cell pellets

which were not immediately used were stored at -20°C for future experimentation.

For the lysis of TDE0626, ultrasonic frequency sonication was conducted due to the formation of

an large insoluble fraction when chemically lysed. The harvested cell pellets of TDE0626 were suspended

in 100mM sodium phosphate at pH 8.0 and subjected to 20 second pulse intervals at an amplitude of

40% and 30 second off intervals in order to minimize the heat energy given to the lysed cells and protein

contents. The lysate of TDE0626 was centrifuged at 15,000xg for 35 minutes to separate the insoluble

cell components formed during the sonication. All steps of sonication protein isolation were done on ice

to minimize heat denaturation of the recombinant protein.

The harvested cell pellets of TDE1701 and TDE2714 were treated with 730mM sucrose in a

buffered solution containing 50mM Tris-HCl pH 8.0 and 300mM sodium chloride. The salt was used in

order to maintain the solubility of the isolated protein whereas the sucrose was used for hypo-osmotic

shock cell disruption. As the bacteria experience a change in the osmotic pressure, water flows into the

cells which causes for the rupture of the cell membrane and release of cellular contents. Once 25mL of

the sucrose solution was added to the cell pellets, the slurry was stirred using a magnetic stir bar on ice

for 10 minutes. In order to digest the cellular matter from the lysate, 10 milligrams of egg white

lysozyme were added to the stirring lysate for another 10 minutes. After which, two equivalent volumes

of a 24mM deoxycholic acid solution (50mM Tris-HCl pH 8.0 and 300mM NaCl) was added to the

resuspended lysate. The solution also contained 15mM Triton X-100 which was a non-ionic detergent used to solubilize the cellular membranes and other exposed hydrophobic components. The lysate again was stirred for an additional 10 minutes on ice. Lastly, 2.6µg/ml of DNase I was added to the lysate in order to digest the DNA present in the lysate and lower the viscosity of the lysate as well. 5mM of $MgCl_2$ was also added as a cofactor for the DNase I. The lysate was stirred for another 10 minutes to ensure complete digestion of all cellular components which were not the recombinant protein. The lysate was centrifuged at 15,000xg for 35 minutes to separate the insoluble cell debris from the soluble recombinant proteins.

For all three targets, the crude recombinant protein solution was transferred to a 50mL plastic tube and stored at 4^oC. A small sample of the insoluble pellet was taken for each of the targets after centrifugation and resuspended in 50µL of ultra-pure water for analysis. This allowed for the sample to be properly loaded in the polyacrylamide gel. A 50µL sample of the supernatant was also taken for further analysis via SDS-PAGE.

2.2.2 – Nickel Nitrilotriacetic Acid Immobilized Metal Affinity Chromatography

Initial purification of the crude protein extract was conducted via nickel-nitrilotriacetic acid (Ni-NTA) immobilized metal affinity chromatography (IMAC). This was due to the presence of a six-histidine tag on all the recombinant targets located at the C-terminus. Approximately 5mL of the HisPur Ni-NTA resin was loaded into the Econo-Column® Chromatography columns which was stored in 20% (v/v) ethyl alcohol. Once the alcohol was washed out by several column volumes of water (~50mL per CV), the

crude protein extract for each of the targets was added to resin. The resin and protein extract were

incubated on ice for one hour by a shaker rocker instrument. Once the resin was thoroughly mixed, the

slurry was placed into the chromatography column and allowed to flow though via gravity filtration. The

sample collected from the initial flow-though was stored at 4°C for further analysis. One column volume

of ultra-pure water was run through the column in order to wash out any weakly bound proteins or

other contaminants. Two separate buffering solutions were prepared for the IMAC. These were a 5mM

imidazole wash solution, and a 500mM elution solution where each solution also contained 300mM

NaCl and 50mM Tris-HCl pH 8.0. It should be noted that 50mM sodium phosphate buffer pH 8.0 was

used for the purification of TDE0626 instead of Tris. A plunger apparatus was lowered into the column

to a height of 10mm above the resin bed and vacuum sealed for consistent pressure during the IMAC. A

low concentration of imidazole was used as a wash solution as it could compete with the binding of

weakly associated proteins and not the target protein. This was conducted at a volume of 50mL at a flow

rate of 4mL per minute and collected in a conical tube. A step gradient of increasing imidazole

concentration was performed by the mixing of the wash solution and the elution solution. The flow rate

was 4mL/min and fractions were collected every 10mL (in 15mL conical tubes) for corresponding to a

25mM concentration increase of imidazole. The imidazole gradient ranged from 5mM imidazole up to

250mM imidazole. Once the gradient elution was completed, a 10mL fraction was collected of 100%

elution buffer over the column. The gradient and mixing were all conducted by the Econo- Gradient

Pump and all fractions collected were stored at 4°C. In order to clean the column and resin of all non-

23

eluted proteins, a column volume of 10%(v/v) acetic acid was run through the column flowed by a

column volume of 0.5M NaOH. Ultra-pure water was run in-between the cleaning steps. To ensure

proper storage and longevity of the column and resin, the resin was stored under one bed volume (5mL)

of 20% (v/v) ethyl alcohol.

2.2.3 – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Dialysis

In order to evaluate the relative successfulness of the recombinant protein expression, isolation,

and purification each of the collected fractions from the IMAC were subjected to SDS-PAGE and dialysis.

This also included the samples collected from the cell debris pellet and the flow-through sample of the

IMAC. The polyacrylamide gel used consisted of a 12% (w/v) separating gel and 4% (w/v) stacking gel.

The consistency of the gel was made from a mixture of 40% acrylamide and 2% bis-acrylamide. To

solidify the gel, 50μL of 10% ammonium persulfate (APS) and 10μL of tetramethylethylenediamine

(TEMED) were added to the unpolymerized gel layers. Once the gel was cast and polymerized, 20μL

sample from each fraction was transferred to a 1.5mL microtube. Each sample was then treated with

5μL of a 5X SDS sample buffer which contained sodium dodecyl sulfate (SDS), and dithiothreitol (DTT)

and glycerol. These samples were heated to 95°C for five minutes to ensure sufficient protein

linearization and binding of SDS to the protein strands. Once the samples were cooled, they were loaded

beside the precision plus protein^{TM} duel color ladder for protein visualization. Electrophoresis was

conducted in 1X SDS running buffer which contained 25mM Tris-HCl pH 8.0, 200mM glycine, and

3.46mM SDS. The chamber was filled to the equivalent of two gels due to leaking which occurred during

the electrophoresis procedure. Electrophoresis was run at 180 volts and 400mA for 45 minutes which

may have also been run for a shorter period if the dye front was approximately 1cm from the gel

bottom. For staining of the gel, a Coomassie Brillant Blue G-250 staining was used, which was able to

bind to the retained protein bands. The stain also included glacial acetic acid and methanol which

required heating in a microwave to bind to the protein in the acrylamide gel. The gel and stain where

heated for 30 seconds in a microwave before placing destaining solution (containing larger

concentrations of methanol and glacial acetic acid) onto the gel to remove the excess Coomassie which

did not bind to the protein. Once the gel was sufficiently destained, an image of the gel was taken using

the Molecular Imager® Versa Doc MP System at a focal number of 11 and optimized exposure of 1

second on a white background.

Based on the acrylamide gel results, the protein samples which contained a large amount of

"pure" protein were pooled together in order to prepare for dialysis. Dialysis was conducted as the

purity of the isolated protein samples were not sufficient to merit direct crystallization due to the

presence of contaminants at various sizes. A secondary purification step was conducted using anion

exchange chromatography; however, the presence of imidazole in the pooled protein sample would

impede the anion exchange at the pH used (8.0). Therefore, the protein samples were pooled into a

50mm Fisher Regenerated cellulose dialysis tubing which had a molecular weight cut-off at 6-8kDa.

Since all the targets were over 30kDa in molecular weight, it was possible that the protein would remain

within the tubing while the imidazole would be removed to the surrounding solvent through osmosis.

The protein-containing tubing was placed initially into two liters of water and stirred gently for two hours. The tubing was then removed for the water and placed into freshly prepared buffering solution (either 50mM Tris-HCl pH 8.0 for TDE1701 and TDE2714 or 100mM glycine pH 9.6 for TDE0626). The use of 100mM glycine pH 9.6 for TDE0626 was due to the amount of precipitation which occurred when TDE0626 was dialyzed into 50mM Tris-HCl pH 8.0. It was found by multiple samplings of TDE0626 dialyzed in various buffers that the protein solution was most stable and soluble in 100mM glycine-NaOH pH 9.6. Dialysis into the fresh buffer occurred overnight (~16 hours) at 4°C.

2.2.4 – Anion-Exchange Chromatography and Dynamic Light Scattering (DLS)

Further purification of each of the targets was required and was achieved through the use of anion-exchange chromatography. Prior to anion exchange, dialyzed TDE0626 sample was removed from the dialysis tubing and centrifuge at 12,000xg at 10°C for 10 minutes in order to remove any protein aggregates. The decanted protein sample was placed over an ÄKTA Pure FLPC and the UNICORN 6.4 Workstation program from GE which had a HiTrap™ Q FF anion exchange column. For both TDE1701 and TDE2714, the wash buffer contained 50mM Tris-HCl pH 8.0 buffer and the elution buffer contained 50mM Tris-HCl pH 8.0 with 1M NaCl. For TDE0626, the wash buffer contained 100mM glycine-NaOH pH 9.6 and the elution buffer contained 100mM glycine-NaOH pH 9.6 with 1M NaCl. The protein sample line (position A7) was purged in ultra-pure water prior to sample application.

After the column cleaning (using water, high salt concentration, and wash buffer equilibration), a pre-programmed protocol for anion-exchange was initiated for each of the recombinant proteins. The

protocol contained a gradient elution for the bound protein with increasing NaCl concentration which

and elution fractions were collected into 15mL sterile conical tubes. The absorbance increase at 280nm

due to the elution of the purified protein was monitored in real-time based on a gradient program. Once

the fractionation was completed and 100% elution buffer was run over the column, the protocol

finished by removing the excess salt and storing the column in 20% filtered ethyl alcohol.

Prior to the crystallization attempts for the purified recombinant proteins, the selected fractions

from the anion-exchange chromatography were concentrated to 5mg/mL using an Amicon Ultra-15

Centrifugal filter with a molecular cutoff of 10kDa. The fractions were pooled together based on the

peak absorbance reading from the anion-exchange chromatography. The samples were then centrifuged

at 4000xg for 15-minute intervals at 4°C using a Beckman SX4750 swinging-bucket rotor. The flow-

through was collected and the retained protein sample was measured for the concentration. The

concentration was measured based on the absorbance of tryptophan residues in the protein samples at

280nm. The UV-Vis spectrophotometer was blanked using 1mL of ultra-pure water in a disposable UV

cuvette and obtaining a baseline reading for the water. The concentrated protein sample (5μL) was then

added to the water blank and a UV spectra range of 260-300nm was measured. To calculate the protein

concentration, Beer–Lambert law was used using the formula A = ε*C*l (absorbance = molecular

extinction coefficient * concentration * path length). The molecular extinction coefficient for each of the

proteins was calculated via Protparam when all the cysteines form disulfide bridges. The extinction

coefficients at 280nm, for each protein were 16390$M^{-1}cm^{-1}$ for TDE0626, 30620$M^{-1}cm^{-1}$ for TDE1701,

and 170225M^{-1}cm^{-1} for TDE2714. In order to obtain the concentration in mg/mL, the conversion of

mol/L calculated to mg/mL was as follows;

Conc. (mg/mL) = [conc. (mol/l) * MW (g/mol) * (1L/1mL) * (1000mg/1g)] * (1005µL/5µL)

Absorbance readings were conducted for all three of the recombinant proteins to yield

concentrations in the units of mg/mL which was the standard units used for crystallization. Further

analysis prior to crystallization was the use of DLS which was able to determine the

oligomeric/aggregate state of the protein samples. To perform DLS analysis, a 40µL sample of the

5mg/mL protein sample was placed into a 96-well black polystyrene microplate in triplicates (if

possible). Protein samples were carefully added to the wells to ensure no bubbles were present. Once

each sample was added to the plate, the plate was immediately placed into the DynaPro PlateReader-II

at a constant incubation temperature of 25°C. Settings for the DLS reading were programmed into the

Dynamic 7.1.9 program. The plate reading would record samples in succession where each reading of a

well had 20 images collection taken with an acquisition time of 5 seconds. Data were then averaged and

analyzed via generation of a percent mass versus calculated average particle radius in nanometers. Once

sufficient homogeneity of the samples was found for a particular condition, crystallization was

conducted for each of the purified protein samples.

2.2.5 – Crystallization Screening

Initial crystallization for each of the recombinant proteins was accomplished by the use of

sitting-drop vapour diffusion method. Each of the protein samples were concentrated further in a similar

manner as described before to concentrations of 5-30mg/mL. Final concentrations for crystallization

were attempted at a range of 10-15mg/mL. For TDE0626, multiple trials for initial crystallization was

conducted at initial concentrations of 5, 8, 12, 15, 20, 25, 30mg/mL. For TDE1701, trials for

crystallization were conducted at an initial concentration of 20mg/mL. For TDE2714, trials for

crystallization were conducted at an initial concentration of 25mg/mL. Extinction coefficients for each of

the targets found in Table 3. For each of the concentrations the protein sample was centrifuged at

12,000xg at 4°C for 10 minutes in order to remove any aggregates found during the concentration step.

Crystal screening was set up using the Gryphon system with various crystal screens in 96-well

microplates. Crystal screens used were; the MCSG crystallization Suite for MCSG-1,2,3, and 4 crystal

screens, PACT Premier™ HT96 crystal screen, Index™ HR2-144 crystal screen, and HR2-110 crystal

screen. The standard protocol set up for the sitting-drop vapour diffusion was a two-drop protein screen

with the ratio of reservoir to protein sample of 0.3μL:0.3μL as well as 0.3μL:0.2μL. The protein sample

(200μL) was placed in a 0.2mL tube for screening. An initial wash protocol used for the Gryphon prior to

the screening by the use of sterile filtered ultra-pure water which was run through the 96 syringe heads

and the protein needle. Once the initial wash was completed the sample was placed below the protein

needle where all 200μL of protein was aspirated into the needle. All 96 syringe heads drew 60μL of the

reservoir conditions for the crystal screens and a ratio of 3:3 or 3:2 of reservoir to protein was setup for

the sitting drops. Once completed, the microplate was sealed with a layer of HD Clear packaging tape

and incubated at 18°C. Crystal formation was monitored by the use of ZEISS SteREO microscope at a

magnification of 1000X and the conditions which produced protein crystals were recorded and logged

for further optimization.

For the conditions which showed initial crystal formation, optimization was conducted using the

hanging-drop vapour diffusion technique at a larger sample volume. Conditions which produced crystals

were optimized by varying the concentration of the precipitating agents as well as the pH of the

conditions. The ratios used for the hanging protein drops were similar to initial crystallization at either

3μL:3μL or 3μL:2μL of the reservoir to protein sample. The conditions were replicated in TC Plate 24-well

cell culture plates and scaled to a reservoir volume of 500μL. The protein drops were placed on 18mm

siliconized glass cover slides which were cleaned in order to minimize dust particles. The rims of each of

the 24 wells were coated with vacuum grease in order to make a vapour tight seal. Once the conditions

and protein drops were set up, the whole plate was incubated at 18°C until crystal growth was

observed.

2.2.6 – X-Ray Diffraction Analysis and Model Building

Once large bulk crystals were formed, they were subjected to X-ray diffraction analysis in order

to build structural models of each of the targets. From both the initial sitting-drop vapour diffusion and

the 24-well plates were sent for diffraction. In order to do so, freezing of the crystals was conducted by

looping the crystals from the conditions and submerging them in a freshly made cryoprotectant. Each

cryoprotectant was tailored for each different condition as it contained all the same components as the

condition; however, also contained 30% (v/v) glycerol to prevent the formation of ice crystals. Once the

crystals were submerged in the cryoprotectant, flash freezing using liquid nitrogen was used and the

crystals were subsequently stored in liquid nitrogen until shipped to the Canadian Light Source

Synchrotron (CLS) or the Advanced Photon Source Synchrotron (APS).

For X-ray diffraction analysis of each of the sent crystals, an initial test diffraction was conducted

at the angles of 0°, 45°, 90°. Each of the three images were examined for their diffraction quality in order

to determine the optimal angle to begin X-ray diffraction data collection. Once the images for both the

initial diffraction tests and the full datasets of diffraction were collected. iMOSFLM and SCALA were

utilized in order to determine crystallographic statistics from the images collected(77). The generation of

an electron density map was the was attempted after by obtaining phase information(78). This

information coupled with the homology modelling prediction was inputted into Phaser-MR for

molecular replacement and phasing. Once the models were generated, structural refinement was

conducted by both REFMAC-5 refinement software and experimenter input via COOT software. Several

rounds of refinement were conducted, and validation of the structural models produced were analyzed

for correctness via the molecular graphics software COOT(79).

2.2.7 – Selenomethionine Single-wavelength Anomalous Diffraction (SAD)

Selenomethionine derivative crystals for TDE2714 were also formed and utilized for diffraction

analysis. This was conducted by the growth of transformed BL21 *E. coli* cells in minimal media

(Molecular Dimensions) which contained 20mg/mL of L-selenomethionine from Bio Basic Inc. The small

10mL cultures were centrifuged at 12,000xg for five minutes before decanting the LB which contained

methionine. Once the pellets were resuspended in ultra-pure water, the cells were centrifuged again at

12,000xg for five minutes and the water was decanted. The cell pellets were resuspended in the minimal

media which contained the L-selenomethionine and were grown in similar conditions to the native trials.

After cell lysis and subsequent purification, the concentrated selenomethionine containing protein

samples were initial screened via the sitting-drop vapour diffusion and optimized by the hanging drop

vapour diffusion. Large crystals were cryogenically frozen and sent to the CLS for anomalous dispersion

analysis. Using the anomalous signals from the selenium within the protein crystals, phase information

was attempted to be gathered from the data via Phaser-SAD.

2.2.8 – Site-Directed Mutagenesis of TDE2714

During the functional analysis of TDE2714, site-directed mutagenesis was employed in order to

generate a mutant that could be used for co-crystallization attempts with a peptide substrate. This was

conducted by miniprepping DH5α *E. coli* in order to obtain the *tde2714*-pET21b plasmid. A single primer

PCR reaction was conducted with either a forward or reverse primer containing a single point mutation

of the C637 codon from T to a G and G to a C. This mutation would cause for a change from a cysteine to

an alanine residue. The mutation was chosen based on the alignment of an active site cysteine from a

well-characterized FGE (PDB ID: 2AIK). PCR was conducted at using the 10μL of 5X Phusion HF buffer,

5μL of forward/reverse primer (100μM), 1μL of *tde2714*-pET21b (50ng/μL), 1μL of 100mM dNTP, 2μL of

DMSO (100%), 0.5μL of Phusion HF polymerase, and 30.5μL of nuclease-free water. The PCR reaction

was denatured 98°C for 30s and again at 98°C for 10s. The primers were then annealed using touchdown

annealing at -0.4°C starting at 74.7°C. The primers were then extended at 72°C for 3.5 minutes. These

steps were repeated 25 times until a second extension at 72°C for 10 minutes. The completed PCR

reaction was then treated with 1µL of *DpnI* and incubated at 37°C for 1h. After this, 5µL of the reaction

was transformed into DH5α and BL21 *E. coli* for further plasmid production and expression. Samples of

the PCR product were sent for sequencing at Western University. The digested PCR sample was also

loaded on a 1%(w/v) agarose gel that contained 0.5 µg/mL of ethidium bromide and run at 100V for 1h,

prior to imaging.

2.2.9 – Functional Analysis for The Targets

Once the recombinant proteins were purified and shown to be homogeneous, various functional

analysis were conducted to evaluate each of the putative functions. This was based on the

bioinformatics predictions for each of the targets.

For TDE0626, the ability to degrade cellulose was test by incubating 5-50mg/mL of cellulose with

0.5mg/mL of TDE0626 for 1h at 37°C. The samples were the filtered through a 0.2µm filter before

loaded onto a 12% (w/v) acrylamide gel and run at 180V for 45 minutes. Once the electrophoresis

procedure was completed, the gel was treated with a silver nitrate solution for 10 minutes to develop an

image. An experiment was also designed to test the degradation of polysaccharides present in a red

complex biofilm. In order to do so, purified TDE0626 was subjected to crude biofilm isolates that were

tested for degradation. Concentration of biofilm were tested were at 25-500ng/mL at an incubation

temperature of 37°C for 1-3 hours. These reaction samples were at a volume of 20µL in order to allow

for the loading of the samples onto a 12% (w/v) acrylamide gel which was run at 180V for 45 minutes.

Once the gel was finished running, the gel was treated with a silver nitrate solution for 10 minutes to develop an image of the proteins and carbohydrates present in the biofilm. Each of the reactions were run in triplicates and incubated at 37°C. Functional analysis of TDE1701 has not been investigated to date.

For TDE2714, carbohydrate polyacrylamide gel electrophoresis (C-PAGE) was conducted by incubating the purified protein with polysaccharides such as polygalacturonic acid. Purified TDE2714 (5ng/mL) was incubated with 0.5% polygalacturonic acid for 2h at 37°C. The samples were loaded onto a 15%(w/v) acrylamide and 1% (v/v) bis-acrylamide gel with 10% sucrose and 5mg/mL phenol red. The gel was run at 200V for 2h. Staining the gel was done by incubating the gel in 5mg/mL phenol red for 10 minutes then rinsed with water for 1h, prior to treatment with 0.4% silver nitrate. Staining in silver nitrate was done for 10 minutes and rinsed with water before being developed with 0.6M sodium carbonate and 0.08% (v/v) formaldehyde. Development was stopped by 5% (v/v) acetic acid.

For TDE2714, a >95% purity synthetic peptide was ordered from Bio Basic Inc. with the sequence LCTPSRA (sulfatase motif). The powered peptide was dissolved in ultra-pure water to a final concentration of 370.88mM, which was the stock peptide solution. In order to test the function of the enzyme, peptide-based thin layer chromatography (TLC) was employed. The enzymatic reaction consisted of 1mM of the LCTPSRA, 0.1mM purified TDE2714 (7.4mg/mL), 50mM sodium phosphate buffer pH 8.0, and 2mM DTT. Sodium phosphate dibasic pH 8.0 was used instead of Tris-HCl pH 8.0 due

to the amine binding properties of the stain used. The mobile phase (solvent) used was a 3:1:1 ratio of

1-butanol: glacial acetic acid: and water. The stain used was 1% ninhydrin in a 5:1 (v/v) ratio of pyridine

and glacial acetic acid. The reaction volume was set to 100μL in which all the components were added

and TDE2714 was added last in order to begin the reaction. Reactions were then incubated at 37°C for 1-

16 hours before TLC analysis. A half sheet of the silica gel 60 aluminum TLC sheet was marked with a

starting line of 1cm at the bottom for sample application and heated for 10 minutes at 120°C to activate

the silica. Once cooled, the TLC plate was placed into a glass developing chamber with approximately

250mL of the solvent and allowed to run up the plate until 1cm from the top of the plate. The plate was

then removed and dried to prepare for the sample application. The reaction samples were concentrated

using the Centrivap instrument at 60°C until the volume of sample was 10μL. Each sample (3μL) was

then placed onto the TLC plate above the 1cm line and allowed to dry. Once the plate was fully dry, it

was placed into the glass chamber and the solvent was allowed to run up the plate until 1cm from the

top. The plate was removed and dried using a hairdryer prior to staining. The stain solution was poured

directly on the TLC plate until the stain cover the whole of the TLC. This was then placed into an oven

and heated at 100°C for 5 minutes to develop the stain. Once the stain was developed and image was

taken using the Molecular Imager® Versa Doc MP System. Intensity readings for each of the developed

spots were measured and graphed for the visualization of product formation.

In a similar manner, the formation of the potential product of the enzymatic reaction was

conducted for electron spray ionization mass spectrometry (ESI-MS). Reaction samples were set up with

the same components as the TLC protocol. These were incubated for 16 hours before being purified by

filter centrifugation and sent to the ESI-MS/MS at the Advanced Analysis Centre at the University of

Guelph. The samples were frozen for shipment. Samples were analyzed, and their mass spectra

compared by Dr. Dyanne Brewer (University of Guelph) to properly sequence and verify any changes to

the substrate peptide.

In order to better understand the result from the ESI-MS/MS, liquid chromatography-tandem

mass spectrometry (LC-MS/MS) was conducted at the University of Guelph. The peptide samples were

reacted for 16 hours, purified, and sent for LC-MS/MS where the retention times of the substrate

standard and the reaction samples as well as the spectrums were analyzed by Dr. Dyanne Brewer. All

the data from both the ESI-MS/MS and LC-MS/MS were tabulated and graphed for comparison.

3. Results

3.1 – Bioinformatics

3.1.1 – Genomic and Transcriptomic Analysis

In order to gain insight into the role of *Treponema denticola* as a contributor to the progression of human chronic periodontitis, genomic analysis and transcriptomic analysis were conducted. Previous work by Eric C. Reynolds showed a higher expression of genes in *T. denticola* when in a mature biofilm relative to its planktonic state. The analysis was conducted via microarray and hybridization assays in which the total RNA of the bacteria was isolated and measured for changes in the expression levels during its presence in a biofilm, as well as in its planktonic state. This analysis suggested that 126 genes had a fold increase in their expression by 1.5-fold or higher (the threshold considered upregulated), leading to the postulation that these genes are involved in the adaptation from planktonic state to biofilm state. Systems were identified to be upregulated in the biofilm such as toxin-antitoxin systems and expression of virulence factors such as; transposases, cystalysin, and lipoproteins(75). Data analyzed from this study revealed three genes that were shown to have a significant fold increase in expression; TDE0626, TDE1701, and TDE2714. These were functionally classified as putative uncharacterized proteins (Table 1). The identification of three genes which had their expression increased in a biofilm state served as the basis for further analysis into their role in the pathogenicity of *T. denticola*.

Table 1: Transcriptomic analysis of three selected genes from *T. denticola*, which had a fold expression change of 1.5 or higher.

Gene ID	Fold Change (Log$_2$)	Adjusted *P* Value	Uniprot Function
TDE0626	1.26	7.2×10^{-18}	Putative Uncharacterized Protein
TDE1701	0.99	3.5×10^{-14}	Putative Uncharacterized Protein
TDE2714	1.02	1.2×10^{-14}	Putative Uncharacterized Protein

The transcriptomic data collected suggested that genes may have importance for the bacterial

pathogenicity, based solely on the increase in gene expression. This was shown by the small *P* value that

gave strong evidence to the data being true against the regular expression of the genes. However, the

transcriptomic data did not reveal any biophysical characteristics, such as structural elements or

functional properties, and so each gene product was labeled as uncharacterized. The putative function

for the product of the gene *Tde0626* was discovered using reverse transcription PCR (RT-PCR) to be a

putative lipoprotein based on the functional cloning using the generated cDNA library. These are known

to be associated with virulence, as they play a role in the inflammatory response of hosts that is known

to cause tissue damage and cell lysis(80). The data shown in the study allowed for the development of a

hypothesis that these genes may play an important role in the survival of the pathogen as well as its

prevalence in the maturation of the biofilm.

To gain more insight into the characteristics of these genes, further genomic analysis was

conducted by the use of database searching such as; the National Center for Biotechnology Information

(NCBI), as well as Kyoto Encyclopedia of Genes and Genomes (KEGG). NCBI

(https://www.ncbi.nlm.nih.gov/) is a large database that includes information from various other

databases such as; GenBank, PubMed, and 37 other databases that give insight into genetic

characteristics of genes in a particular organism, as well as information about the potential function or

associated function of the gene products(81). The information that NCBI provided about the three genes

gave genetic context about where the genes were located in the genome and the predicted function of

the genes expressed surrounding the three genes. Although these surrounding genes may not have any

relevant associations to the function of each gene, it was important for initial analysis to gain context of

the location of the genes, as well as potentially any associated genes nearby (Figure 1).
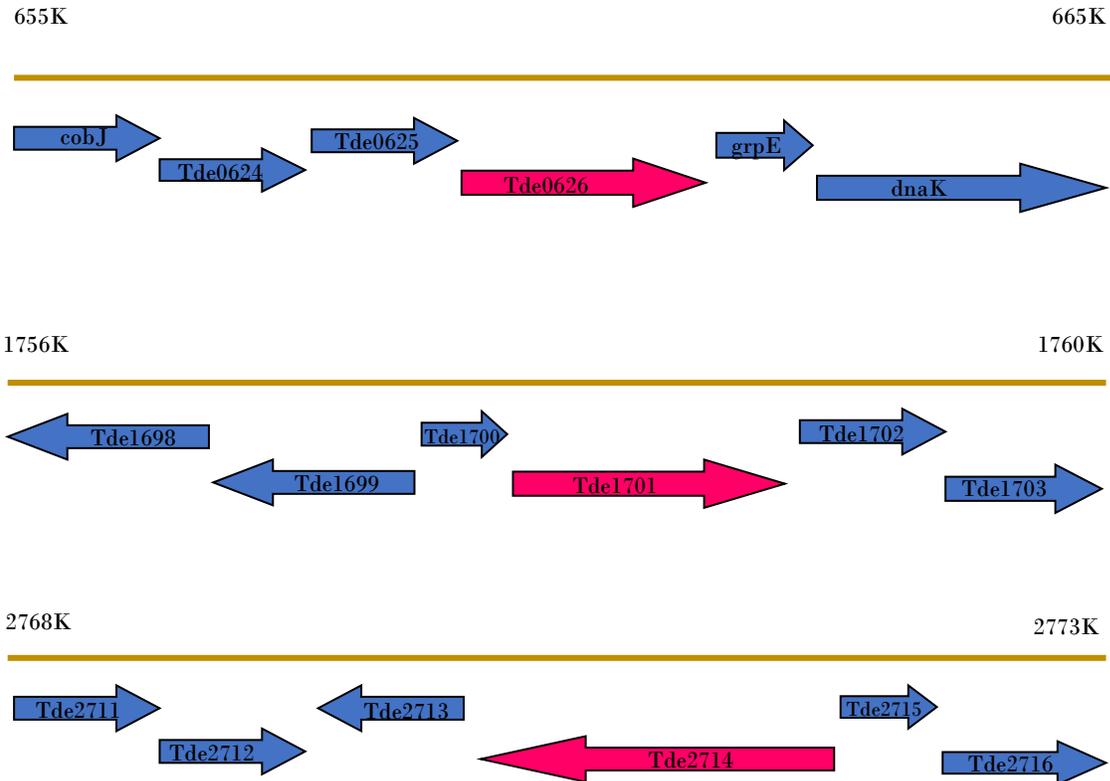
Figure 1: Genomic context of the *Tde0626*, *Tde1701*, and *TDE2714* genes of *Treponema denticola* ATCC 35405. Genomic range represented by yellow bar with nucleotide position within the genome of *T. denticola*. Each gene of interest highlighted in red and surrounding genes shown in blue. For *Tde0626* (hypothetical protein), gene location at 658,230 - 659,744 base pair. For *Tde1701* (hypothetical protein), gene location at 1758280 - 1759407 base pair. For *Tde2714* (hypothetical protein), gene location at 2770717 - 2772708 base pair. Data obtained from NCBI Reference Sequence (RefSeq) in the NCBI database. A putative function was only predicted for TDE0626 operon as being involved in carbohydrate metabolism.

For *Tde0626*, upstream of the gene product is a putative ABC transporter ATP binding protein

(*Tde0625*) and downstream is a co-chaperone protein GrpE. The presence of a chaperone protein may

be important for the functionality of *TDE0626* as it may require additional aid in its folding to be

functional(82). For *TDE1701*, upstream of the gene product is a serine/threonine protein phosphatase

(*Tde1698*). However, directly adjacent to the *TDE1701* upstream and downstream, the gene products

are classified only as hypothetical proteins and give no additional information that could lead to a

prediction of the function of TDE1701. For *TDE2714*, downstream of the gene product is an RNA

methyltransferase (*Tde2713*) and upstream is a haloacid dehalogenase (HAD) family hydrolase

(*Tde2716*). The presence of both of these gene products in close proximality to each other has been

known to be associated with formylglycine residue containing enzymes, as well as formylglycine

generating enzymes themselves in the biosynthesis of ergothioneine, which is a thiourea derivative of

histidine(83). The potential for each of the targets to be in an operon was investigated using the

Database of Prokaryotic Operons (DOOR[2]). For each of the targets, a search of the predicted operons in

T. denticola was conducted that was then complied into a table for all the genes predicted in the operon

as well as the confidence score for the prediction (Table A3 in Appendix). From the database, only

TDE0626 and TDE1701 were predicted to be part of an operon, where TDE0626 was a part of a 4 gene

operon involved in biosynthesis of precorrin-6x and vitamin B12. TDE1701 was identified to be part of a

2 gene operon of unknown function with TDE1700, a hypothetical protein.

Further genomic analysis was conducted via the use of KEGG (https://www.genome.jp/kegg/),

which is a collection of databases that aim to give insight into the higher level gene functions as well as

cellular systems, such as metabolism and signal transduction pathways(84). Each of the genes were

searched in KEGG to identify any known similarities to gene functions in bacterial systems. However, the

data was inconclusive about their involvement in known systems as they are classified as hypothetical

proteins and so other aspects of KEGG were utilized. The KEGG database also contained information

about predicted orthologs and paralogs for each of the gene products that could potentially give

information about their putative function based on homologous proteins. Identifying orthologs each

gene product can reveal similar functions of proteins over a wide range of species and can aid in the

identification of the function of the uncharacterized protein, whereas paralogs can reveal new functions

of a class of proteins within an organism and is useful for discovering multiple homologous proteins in

an organism(85).

Table 2: Summary of predicted ortholog and paralog data from Kyoto Encyclopedia of Genes and Genomes (KEGG) of each of the genes in study.

Gene	Number of Orthologs	Number of Paralogs	Domain(s)/function
TDE0626	1929	0	Cthe_2159 Carbohydrate binding domain
TDE1701	3993	2	None bacteriocin-type signal sequence
TDE2714	2113	16	FGE-Sulfatase Sulfatase modifying factor

From the data collected from KEGG (Table 2), each of the genes had a large amount of orthologs

which were identified, with *TDE1701* having the highest numbers identified. In light of this, it would

have seemed that the function of each gene products could be predicted, however the volume of

different functions found in the orthologs made for difficultly to develop a conclusive function.

Searching for paralogs within *T. denticola* did not reveal a large number of paralogs, in which *TDE0626*

had no identified paralogs. More conclusive functional predictions of the gene products were found by

comparing the gene products to the identified paralogs. The KEGG database revealed that TDE1701 may

be similar to a bacteriocin-type signal sequence, that is known to be a toxic bacterial peptide that is toxic

42

to closely related bacteria(86). An example of this is the expression of enterolysin A, a cell wall-

degrading bacteriocin, which is known to inhibit the growth of Lactobacilli(87). Based on the data

collected from KEGG, TDE2714 may be similar to a FGE-sulfatase modifying factor, that is known to

modify either an active site cystine or serine of a type 1 sulfatase into a formylglycine residue(88).

Functional predictions were generated by searches into the Pfam database that looks for similarities to

known protein motifs. This information was then utilized to generate putative domains and boundaries

that are associated to the function of the uncharacterized protein(89). Using the information from this

database allowed for the hypothesis that a Cthe_2159 carbohydrate binding domain may be present in

TDE0626. The Cthe_2159 domain was identified from *Clostridium thermocellum* as a beta-helix

containing domain that binds to either cellulose or uronic acid containing sugars, such as cellulose,

xylan, or other glycosaminoglycans, and the domain shares similarities with the polysaccharide lyase (PL)

protein families(90). The similarities of the Cthe_2159 domain to TDE0626 suggests that either TDE0626

can bind to cellulose or to GAGs. Based on the genomic analysis, each of gene products may play some

role in the pathogenicity of *T. denticola*, either by aiding in the nutrition of the organism to its survival or

proliferation within a host.

3.1.2 – Predicted Biophysical Properties

Further analysis of the three gene products was conducted via the analysis of their primary

protein sequences. This was accomplished initially by the use of a program Protparam from the ExPASy

Bioinformatics Research Portal (https://web.expasy.org/protparam/). The software uses computational

algorithms to analyze a protein sequence and predict the physiochemical properties. These properties

are based on the structural and functional characteristics of the hypothetical protein, however there is

no information given about the putative function or structural elements, only the parameters such as

the composition of the sequence and the molar absorptivity(91). Each of the primary sequences were

analyzed and the data was tabulated for comparison (Table 3).

Table 3: Predicted biophysical properties of gene subsequent primary sequences generated via
Protparam software from ExPASy Bioinformatics Research Portal.

Gene	Start – End (amino acid position)	Molecular Weight (Da)	Extinction Coefficient (M^{-1}cm^{-1}) at 280nm	Predicted Isoelectric Point (pI)	Number of Methionines
TDE0626	25 - 479	50058	16390	4.98	10
TDE1701	1 - 375	39841	30620	4.60	6
TDE2714	1 - 663	73955	170225	5.31	7

TDE2714 had the highest molecular weight in comparison to the other proteins and had a

theoretical extinction coefficient of 170225 M^{-1}cm^{-1}. This value was calculated by the number of

aromatic residues at 280nm and may not represent the true extinction coefficient of the proteins. The

predicted isoelectric point (pI) was also calculated which is the point at which a certain pH value has a

neutral protein net charge(92). The pI was the lowest for TDE1701 at pH 4.06 and highest at pH 5.31 for

TDE2714. This value is important for further experimental analysis and development, as it allows for the

net protein charge to be changed from positive, neutral, and to negative, depending on buffering

conditions, which can be useful in a variety of applications. One such application could be to advance

crystallization of the expressed and purified protein, where acidic proteins tended to crystalize at a pH

just above their pI value, whereas basic proteins tended to crystallize at a pH below their pI(93). Another

parameter which was important to note was the number of methionine residues per 100 amino acids

within each of the primary sequences. The number of methionine residues may play an important role

in the structural analysis of each of the recombinant proteins, as they may be replaced to the selenium

derivative methionine (selenomethionine) for anomalous diffraction experiments (SAD/MAD), which is

used for the determination of phasing during X-ray diffraction analysis(94). Based on the data, TDE1701

had the lowest number of methionine residues and TDE0626 had the highest methionine residues.

Protein localization or its subcellular localization can be useful information in bacteria as

knowing this information can aid in the understanding of the protein and its role in the organism as a

whole. Studies of subcellular localization of proteins in Gram-negative bacteria revealed that there are

five main localization sites; cytoplasm, periplasm, inner membrane, outer membrane, and

extracellular(95). In order to identify the localization of the proteins in *T. denticola*, various prediction

software programs were used that was condensed into searchable online database called Oralgen

(http://www.oralgen.org/cellular_location/) for the outputs of key localization software programs(96).

The programs which were found the database were Cello (cello.life.nctu.edu.tw/), SubLoc

(https://omictools.com/subloc-tool), subCELL (bioinformatics.ysu.edu/tools/subcell.html), PSORT

(https://www.psort.org/), and SignalP (www.cbs.dtu.dk/services/SignalP/)(97). Each of these programs

used a similar package of algorithms and input information which examined the query protein primary

sequences for characteristics such as composition of amino acids, presence of a signal peptide, similarity

to known protein localization, motifs similar to known localization, and the presence of transmembrane

alpha-helices(95). From each of the programs used, all of the proteins were predicted to be

extracellularly localized, in that SignalP could only identify a signal peptide for extracellular localization

in TDE0626 and not the other two protein sequences. However, PSORT, which is one of the most widely

used localization tools(95), suggested that all of the protein sequences had motifs that were similar to

extracellular protein P-values. All three of the proteins are secreted from the organism into the

extracellular space, based on the localization tools. Studies have shown that the extracellular polymeric

substances that make up the matrix may contain a host of enzymes secreted from the bacteria in order

to aid in the formation and maturation of the biofilm(98).

3.1.3 – Protein Structure and Function Predictions

In order to identify the role of the upregulated gene products in *T. denticola*, additional

information about similarities to known protein structures and functions was generated. For each of the

proteins studied, their primary sequences were examined through software programs that searched for

main signatures from known protein structures; domain searching, homology modelling and motif

identification, known protein-protein interactions, and conserved regions of structural and functional

elements via multiple sequence alignments. To gain insight into the predicted domains or regions in

each of the proteins, their sequences were submitted to UniprotKB. This is a database that the

sequences were examined and evaluated against submitted scientific literature, as well as

computational analysis. The database was able to group each of the sequences to ones with similar

features in order to then identify regions of dissimilarity and generate domain parameters (Table 4)(99).

Similarly, the InterPro database was used, that contained a protein signature recognition method that was cross-linked with Uniprot to aid in the prediction of protein domains and putative functions. InterPro uses Hidden Markov Models (HHMs), which is a statistical model that assumes that the system analyzed is a Markov Model but has unobserved states(100). Using these two databases and processes, additional information about the biophysical characteristics of the proteins were identified. The domain predictions and feature positions shown (Table 4) were based on the FASTA sequence files from codon optimized custom genes (Figure A6 in Appendix) as well as the length of the predicted regions.

Table 4: Domain and structural feature predictions using InterPro and UniprotKB databases for the proteins; TDE0626, TDE1701, TDE2714.

Gene	Domain/features	Position(s)	Length (amino acids)
TDE0626	Signal peptide	1-24	24
	Cthe_2159	81-327	246
TDE1701	No domain prediction	N/A	N/A
TDE2714	FGE-Sulfatase	424-658	235

Once the domain parameters were established using both databases, the data was tabulated for the identification of predicted domains and putative functions(99). The information gathered revealed similar results to the PSORT predictions and KEGG ortholog and paralog predictions. The data showed that TDE0626 contained a signal peptide that was 24 amino acids long as well as a large Cthe_2159 carbohydrate binding domain, which is homologous to the polysaccharide lyase from *C. thermocellum*. Unfortunately, the domain searching for TDE1701 was inconclusive and both were not able to identify and any protein signatures or domains. For TDE2714, contained a large FGE-Sulfatase modifying factor domain of 235 amino acids long. The other large region of the protein may be either structurally or have

47

no inherent importance in the overall structure; however, further analysis was conducted to develop

more conclusive results.

Another method of analysis for the prediction of the structure and function of the proteins was

the use of StringDB database, which used the submitted primary sequences of the proteins to generate

predictions of protein-protein interactions and networks within the identified organism. This was then

able to suggest information about potential function and cellular systems(101). The database and

program worked by analyzing the sequence against scientific texts and genomic features from other

databases. This information was then used to generate interaction between orthologs as protein-

protein interactions (Table 5)(102). String analysis suggested the following information; similar genomic

context (neighborhood), proteins which are expressed together that have similar functions or

occurrence in metabolism (cooccurrence), association of genes based on predicted patterns known to

be in organisms (coexpression)(101)(103).

Table 5: Summary of StringDB data showing only highest scored results for each of the three genes; TDE0626, TDE1701, and TDE2714. Protein-protein interactions generated by similar features found within orthologs.

Gene	Functional partner(s)	Neighborhood	Cooccurrence	Coexpression	Score
TDE0626	- TDE0625 ABC transporter ATP-binding protein	- Yes	- No	- No	0.651
	- TDE0658 ABC transporter permease	- No	-Yes	- No	0.437
TDE1701	- TDE1698 Serine/threonine protein phosphatase	- Yes	- No	- No	0.591
	- TDE1697 (gpmA) Phosphoglyceromutase	- Yes	- No	- No	0.447
TDE2714	- TDE2283 Permease	- Yes	- No	-Yes	0.573
	- TDE2410 Hemolysin	- Yes	-No	-No	0.550

Although the data may not suggest a similar trend as predicted by the prior bioinformatics tools,

the putative associations with known protein structures were shown. This may suggest that the proteins

being investigated share similar expression systems or are involved in similar cellular pathways within *T.*

denticola. The calculated score for each of the proteins is a confidence-based system. Based on the data,

TDE0626 shares similar genomic context with an ABC transporter ATP-binding protein that genetically is

located directly upstream from the gene, but did not show as co-expressed. StringDB also was predicted

for TDE0626 to have either similar function as to an ABC transporter permease or occur in the same

metabolic pathway; however, it does not share similar genomic context. The analysis suggested for

TDE1701 that it shared genomic context with both a serine/threonine protein phosphatase and a

phosphoglyceromutase. This meant that they are within similar regions of the genome and may not

have similar features to each other. TDE2714 has a similar genomic context to a hemolysin, as well as a

permease. However, the database predicted that TDE2714 contained similar patterns in its primary

sequence that were associated with a permease. Based on this, TDE2714 may require a permease to

transport a specific molecule.

For further structural and functional analysis, homology modelling of each of the proteins was

conducted using various software packages. Protein Homology/Analogy Recognition Engine 2 (Phyre^2)

was used, that is a fold recognition server that can generate predicted three-dimensional protein

structures via analysis of the submitted protein sequences. This software uses Hidden Markov Models

similar to InterPro and other algorithms against known protein structures and folds[104]. These protein

structures were used as templates that can be used for intensive sequence alignments. In addition,

Markov models can predict potential mutation-sensitive positions in a submitted sequence in

comparison to known mutations in aligned sequences, which can suggest important catalytic segments

and regions[105]. The primary sequences of the proteins were submitted to an intensive Phyre^2 analysis

and the data which gave the best results were tabulated (Table 6).

Table 6: Data generated from intensive homology modelling via Phyre[2] software for the genes; TDE0626, TDE1701, and TDE2714.

Gene	Template Model	Sequence identity (%)	Confidence (%)	Alignment Coverage (position)	Template Information
TDE0626	4phbA	34	98.8	65-293	Polysaccharide lyase-like protein cthe_2159 from *C. thermocellum*
TDE1701	5j73A	15	76.8	250-296	De novo design of protein homo-oligomers
TDE2714	2y3cA	45	100	394-660	- *T. denticola* variable protein
	2q17C	29	100	402-659	- Formylglycine generating enzyme from *S. coelicolor*
	1z70X	30	100	402-659	- C-type lectin-like, sulfatase modifying factor-like protein

Only highest percent sequence identity models and confidence shown in the table. The fold

recognition software had similar results to the previous prediction tools for the protein TDE0626, that

contained similar signatures with the carbohydrate-binding domain from *C. thermocellum* and was

shown to have a 34% sequence identity at a 98.8% alignment confidence. The protein TDE0626 may

contain the predicted carbohydrate domain based on the presence of a similar beta-helix motif that may

bind polysaccharides. A low sequence identity and low sequence alignment confidence was shown for

TDE1701 that was similar to a *de novo* protein that was designed to be a bundle of homo-oligomers. The

coverage for the sequences and their alignment was small and suggests that only a small part of the

protein may be homologous. The software also predicted that TDE1701 was 50% intrinsically disordered

and therefore may be difficult to characterize further. For TDE2714, there were three high percent

identity models in which the highest was a characterized *Treponema denticola* variable protein with 45%

sequence identity. All the models had an alignment confidence of 100% which seemed suspect. In

further studying the variable protein for its function, the protein PDB:2y3c, which is known to contain a

calcium-dependent lectin fold (C-type lectin). This is known to bind carbohydrates but also is involved in

the anticipatory ligand binding for immune response in vertebrates(106). The PDB:2y3c enzyme may be

similar to phage diversity-generating retroelements (DGRs) which are used by the pathogen to evade

the host immune system, via sequence variability(106)(Figure 2). The homology modelling seemed to

reinforce the prediction of a C-type lectin-like, sulfatase modifying factor-like protein with a 30%

identity. This prediction also contained similar alignment with the Formylglycine generating enzyme

(FGE) from *Streptomyces coelicolor* at 29% sequence identity (Figure 2). From the data for TDE2714 it

was inconclusive without further experimental analysis to determine if the enzyme was a lectin or an

FGE.
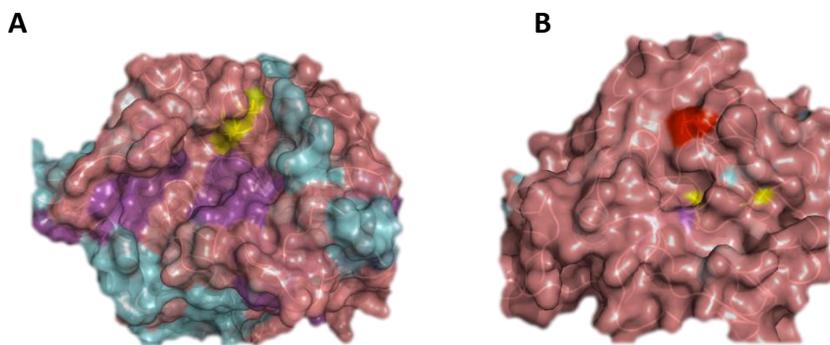
Figure 2: Protein models generated based on homology modelling software, Phyre² for hits with highest sequence identity. Images generated via Pymol with a reduced solvent layer and primary sequence shown. (A) Formylglycine-generating enzyme from *S. coelicolor* (PDB ID: 2q17) which showed a 29% sequence identity to TDE2714. Catalytic cysteine residues 272 and 277 shown in yellow for 2q17. (B) Putative formylglycine-generating enzyme, TDE2714 model based on sequence homology of 2y3c, which showed a 45% sequence identity to TDE2714. Predicted catalytic cysteine residues shown in yellow and alternative catalytic serine shown in red. The blue and purple colors are the peptide backbone colors used from Pymol.

The modelling of TDE2714 onto 2y3c and 2q17 allowed for the identification of the position of

two putative catalytic cysteines for TDE2714 which may be able to bind to a cysteine residue on an

unfolded sulfatase enzyme to generate an active site formylglycine generating enzyme. This data was

gathered from previous work accomplished on the formylglycine-generating enzyme from *S. coelicolor*

which showed two catalytic residues in the active site(107).

Further structural and functional analysis was conducted via the generation of multiple

sequence alignments from homologous primary protein sequences from all domains of life. This was

accomplished by Protein Basic Local Alignment Search Tool (BLASTp) database coupled with Constraint-

based Multiple Alignment Tool (COBALT). BLAST is a program that compares the query primary

sequence with other sequences to find similarities that can be used for seeding analysis of sequences.

This analysis then uses an algorithm to breaks up the sequence into short fragments, called 'words', and

finds all possible combinations of the fragments and aligns them to gain a sequence similarity

score(108). The sequences were aligned with homologous sequences excluding the *Treponema* taxid, in

order to obtain a wider and more diverse alignment. TDE2714 aligned mostly with FGEs compared to all

other enzyme classes. The program COBALT was then able to align the sequences for further conserved

region mapping. ConSurf was used to generate conserved score for each of the amino acids in each of

the proteins studied based on the generated alignments. This was mapped onto the primary sequence

for the identification of conserved regions that may have putative structural and functional information.

The results (Figure 3) for TDE2714 had a highly conserved region (termed RLP region), that is also

conserved among all FGEs from both Eukaryotes and Prokaryotes, at the position at 531-539. Only one

highly conserved cysteine (Cys-625) was discovered and predicted to be functionally important.

```
1          11         21         31         41
MIDSWHLGTD KSAPELKDTD TFYTNAVVFV KSRPDLPDLS SIQGTGDQVK
eeeebbeebe ebeeebeebe ebbeebeebe beeeeeeeee eeeeeeeee

51         61         71         81         91
ITLAVTLEEG GAILGPKSIS VNKGTRWNSV LRTYARAALK VKYGFQGNGW
bebebeeeee eeeeeeeee eeeeeeeee eeeeeeeee eebebeeeeb

101        111        121        131        141
KKNGNTVNET YTFDDDTTIF AELEDMRDNI TVKGDSNVSV PDSTPLVVLS
eeeeeebeee eeeeeeeebe beeeeeeeeb ebeeeebeee eeeeebeebe

151        161        171        181        191
GEKWQEIKEL AAGRIQVSDP SNIAVTAWHR GESASAPVLT DEYEFKGAEG
beeeeebeee beebbbbebe eeebbbbbbe eeebebebbe eebebeeeee
       f                    f    s

201        211        221        231        241
QNRTVYARTG DRRITLTVTY GSGSGTPATA GTITIYDGDE WHNVQRQVAP
eeeeeebeee eeebebbbeb beeeeeeeeb bebbbbbeee beebeeebbe

251        261        271        281        291
LVSVPEDHSI HWYLDNAGGE LINFFYTFKA SDGNARTVYA RVSPPIQLTV
bbebbeebbb bbbbbeeeee bbbbbbbbeb eeeebebbbb ebebebebbb

301        311        321        331        341
TYGRVTGATV TIGTVTTHHR RLWYSVQNEI KYKYPLLSNT EVEWHWNDRN
ebbebeebeb bbbbbbbebb ebbbbbeeeb eeebeebeeb ebebebeeee

351        361        371        381        391
GALISDTYEF DAGNVPSNTV YAFVRPKLIT YYGGLRYIAP IGSNGSYEDY
ebbbbbbbeb ebeebeeebb bbbbbbebbb bbbbbbbbbb bbbeeeeeee
                                                 f

401        411        421        431        441
NMRKIDAVTD GYVGGDAYHY VYNDPHRVSL TAYRIGTTEV TQELYELVMA
beeebebeee eeeeeeeee eeeeeebeb ebbbbeebeb eeeebeebbe
            f    s          s       fs ff

451        461        471        481        491
HNPSYFQGSS HPTASGESQE KRPVEQVSWF DAIAFQNSLT RCCYSLGEAQ
eeeeeeeee eeeeeeeee eeebeebbbb ebbbbeebb ebbeeeeeee
f ff    f              fsf  s     s     f       ff

501        511        521        531        541
CVYTYNGHTY TVEDAQAHNV PVMDMSKKGF RLPTQAEWEW AADSGTAWQK
bbbebeeeee eeeeeeeee bebebeeeee ebeeebebeb bbeeeeeeee
 s                          f ff fsffffsfsf s fff

551        561        571        581        591
WAGPNDEDKL SYYAWYDANE YGRTHQAGRK RANPFGLFDM SGNVSEWCWD
ebeeeeeeeb eebbbbeeee eeeeeeeee eeeebbbbbb eeeebebbee
                 ff          f f    f   s ss   fff fs  f

601        611        621        631        641
WWIQTTPEGG TDPVGPLSGT NRTTCGGGYS FFDSACCCAY RGLEEPDKNK
beeeeeeeee eeeeeeeee eebbebbebe eeeeebebbe eeeeeeeeee
            ff f     f     f  fssf                  f

651        661
DTTGFRIVCR YEF
eebbbebbee eee
   ssf
```

Legend:

The conservation scale:

```
1  2  3  4  5  6  7  8  9
```
Variable   Average   Conserved

e - An exposed residue according to the neural-network algorithm.
b - A buried residue according to the neural-network algorithm.
f - A predicted functional residue (highly conserved and exposed).
s - A predicted structural residue (highly conserved and buried).
X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure 3: Conserved mapping of multiple sequence alignment from COBALT of TDE2714 primary protein sequence. Conserved mapping was generated by ConSurf which was used to show regions of high conservation from a scale of 1-9, where 9 is highly conserved among all FGEs. A large conserved region, RLPTQAEWE, was found in the predicted FGE-sulfatase modifying factor domain. Predicted functional elements label with a red f and structural elements labelled with a blue s.

Similar analysis via conserved mapping was conducted for TDE0626 and TDE1701 (Figures 21

and 22 in Appendix). For TDE0626, there is a highly conserved region at positions 369-375, that is

predicted to be functional in a beta-elimination reaction with polysaccharides (DGIDSNG); however,

further experimental testing of this region being the functional domain has yet to be conducted. The

conserved mapping data for TDE1701 had a relatively low sequence conservation and one region of

average conservation at positions 103-108 (DNGDDE), that was predicted to be the important functional

region for the protein. However, no known domains or homologous proteins had been identified to

have this functional region and so further analysis will need to be conducted to determine the likelihood

of the conserved region being the functional region. It may be that this region is not the functional

region but may play a role in structural integrity of TDE1701.

3.2 – Protein Expression and Purification

3.2.1 – Bacterial Transformation and Recombinant Expression

The first objective was to successfully transform the genes of TDE0626, TDE1701, and TDE2714

into *E. coli* for recombinant expression. In order to do so, the genes housed in pET21b ampicillin-

resistant vectors, were transformed into a DH5α *E. coli* strain to maximize transformation efficiency and

harvest more plasmid. The increased uptake of plasmid is caused by the recA1 and endA1 mutations

that aids in insertion of the plasmids into the *E. coli*(109). Using the Qiagen Miniprep kit, the

concentration of isolated plasmid DNA for each of the genes was 28.8ng/µL of *TDE0626*-pET21b,

46.5ng/µL of *TDE1701*-pET21b, and 70.6ng/µL of *TDE2714*-pET21b. The concentrations were assessed

by measuring the 260nm absorbance for the isolated plasmid samples. Transformation into BL21 *E. coli*

cells allowed for the expression of each of the genes via the induction of the *lac* operon system in the

pET21b plasmids. This was accomplished via the use of a chemical compound called β-D-1-

thiogalactopyranoside (IPTG) that is chemically similar to lactose at concentrations of 1mM IPTG

compared to 14mM of lactose(110). This compound was bound to the *lac* repressor during induction

and allowed for the T7 RNA polymerase to transcribe the three genes for recombinant expression(111).

3.2.2 – Initial Purification via Nickel NTA-IMAC

After centrifugation of 2L cultures of BL21 (DE3) cells, the recombinant proteins were harvested

via osmotic shock chemical lysis with a high sucrose concentration solution. This was accomplished due

to the change in osmotic pressure as the high amount of sucrose is added to the bacterial cultures

allowing for the water to flow into the cells and cause lysis of the cells. All of the lysis and resuspension

of isolated proteins were accomplished using a buffering solution of 50mM Tris-HCl pH 8.0 with 300mM

NaCl. Initially, TDE0626 expressing cells were treated in such a manner. However, upon further

purification steps it was shown that the use of chemical lysis with a buffering solution of Tris-HCl would

lead to the protein being lost to the insoluble portion of the lysate (Figure 4A). A different isolation

method was adopted for the extraction of TDE0626 by the use of ultrasonic frequency sonication of the

resuspended cell pellets containing the desired protein. The cells were resuspended as well in a

buffering solution of 100mM sodium phosphate at pH 8.0 in an attempt to resolubilize any of the

insoluble fraction, that contained TDE0626. The sonication method was used to disrupt the cellular

membrane which can also be called soniporation and did not require the use of a detergent to solubilize

the membrane after lysis(112). Cell lysis was achieved on at 4°C for 10 pulses of 20s, interspersed by 30s

rest periods to prevent temperature elevation. In comparison to chemical lysis of TDE0626, the results

from the ultrasonic sonication revealed a large amount of protein within the buffered solution after the initial purification step of immobilized nickel affinity chromatography (IMAC) (Figure 4B). The results suggested that the dual change to sonication and a buffer of sodium phosphate allowed for the increased solubility of TDE0626, which was able then to be further analyzed in a similar manner to the other two recombinant proteins.

For the initial purification from the crude cell lysate, immobilized nickel affinity chromatography was employed for each of the protein isolates. This was accomplished by the coordination of six-histidine tags (six-His tag) on the recombinant proteins that bound to the nickel ions in IMAC (113). Each of the bound protein samples were also washed with a buffering solution of either Tris-HCl or sodium phosphate containing 5mM of imidazole and collected as wash solution. The wash was used to release any weakly bound proteins or contaminants while the desired protein was coordinating to the nickel resin. An increasing gradient of imidazole from 5mM-500mM in a similarly buffered solution was added to the nickel resin in order to elute the bound protein from the resin. Fractions were analyzed by 12% w/v SDS-PAGE (Figures 4 and 5).
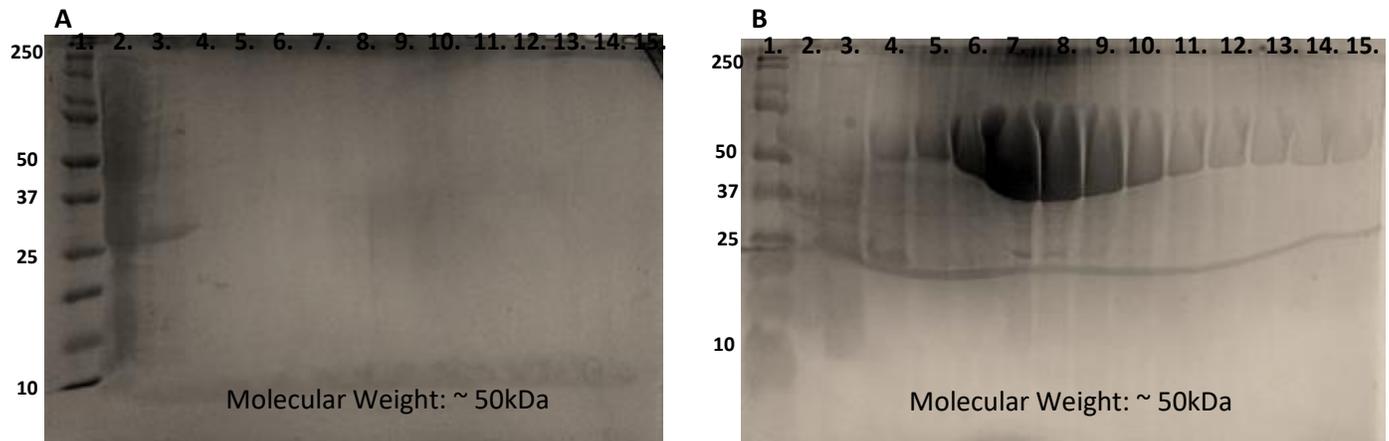
Figure 4: 12% (w/v) SDS-PAGE gels of TDE0626 after nickel-NTA IMAC purification. Lane 1 containing the Precision Plus Protein™ Kaleidoscope ladder for each image. Both wash buffers for lysis method containing 5mM imidazole and 300mM NaCl. The insoluble pellet and flow-through of each method was collected and placed in lanes 2 and 3 of gels. An increasing gradient of imidazole (lanes 4-15) was collected at; 5, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 500mM. (A) Purification of TDE0626 via osmotic shock chemical lysis in 50mM Tris-HCl pH 8.0 in 300mM NaCl. No bands visible after flow-through of crude lysate. (B) Purification of via ultrasonic sonication in 100mM sodium phosphate pH 8.0 in 300mM NaCl. Large protein band at ~50kDa was observed above 75mM imidazole.

Likewise, both TDE1701 and TDE2714 were purified initially via nickel-NTA IMAC. However, they did not require ultrasonication for cell lysis as the use of osmotic shock chemical lysis was sufficient in maintaining the solubility of the isolated protein with relatively low amounts of inclusion bodies. This was revealed during the purification as both proteins showed a large amount of protein within the imidazole gradient during elution (Figure 5). The buffering solution of 50mM Tris-HCl (pH 8.0) with 300mM imidazole was used for the 5mM imidazole wash solution as well as the solutions for the increasing gradient of imidazole.
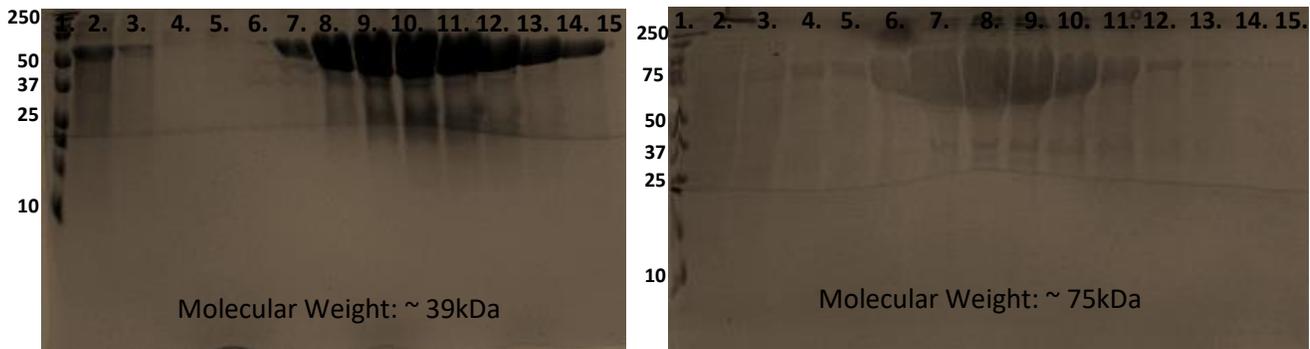
Figure 5: 12% (w/v) SDS-PAGE gels of TDE1701 and TDE2714 after nickel-NTA IMAC purification. Lane 1 containing the Precision Plus Protein™ Kaleidoscope ladder for each image. Both wash buffers for lysis method containing 50mM Tris-HCl pH 8.0 with 5mM imidazole and 300mM NaCl. The insoluble pellet and flow-through of each method was collected and placed in lanes 2 and 3 of gels. An increasing gradient of imidazole (lanes 4-15) was collected at; 5, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 500mM. (A) Purification of TDE1701 via osmotic shock chemical lysis in 50mM Tris-HCl pH 8.0 in 300mM NaCl. A large band (~39kDa) began to show at 100mM imidazole. (B) Purification of TDE2714 via osmotic shock chemical lysis in 50mM Tris-HCl pH 8.0 in 300mM NaCl. A large band (~75kDa) was observed to show at 75mM imidazole.

3.2.3 – Dialysis and Secondary Purification

From the SDS-PAGE analysis, it was important to perform secondary purification to remove

contaminants present in the samples containing purified recombinant protein. A common method for

secondary purification is the use of ion exchange chromatography. The principle is that when the

purified protein is in a buffered solution at a pH different from its pI, it exhibits a net charge that can

then bind to an opposing charged resin. The bulk of the targeted protein is bound to the resin while the

contaminants are repelled or eluted during chromatographic separation(114). For each of the

recombinant proteins, the biophysical results showed that all the pI values were below pH 6. Anion

exchange chromatography was used, where the resin used was positively charged. The buffering

solution used for each purified protein was at 50mM Tris-HCl pH 8.0 for TDE1701 and TDE2714;

however, a 100mM glycine pH 9.6 buffer was used for TDE0626. The choice of buffer was obtained by

screening various common buffers (bis-tris methane (bis-tris), 4-(2-hydroxyethyl)-1-

piperazineethanesulfonic acid (HEPES), glycine, 2-(N-Morpholino)ethanesulfonic acid (MES), sodium

phosphate, as well as sodium acetate). These buffers screened during lysis and were analyzed by SDS-

PAGE. The choice 100mM glycine at pH 9.6 allowed for the net charge on the isolated proteins to be

negatively charged and able to bind to the resin.

However, in working at a pH of 8.0 or 9.6, there was imidazole and salt present in the elution

fractions from the nickel-NTA IMAC that was needed to be removed from the solution. Dialysis was used

for each of the purified protein solutions into an imidazole free and salt free buffer of either 50mM Tris-

HCl pH 8.0 or 100mM glycine pH 9.6. Dialysis was accomplished using a semipermeable membrane

(molecular cutoff of 6-8kDa) which allowed for the imidazole and NaCl to be slowly removed from the

purified protein(115). After a 16-18-hour dialysis into the initial anion exchange buffer, TDE0626 began

to show some precipitation that required centrifugation to remove the insoluble protein from the

solution. Both TDE1701 and TDE2714 did not show visible precipitation after dialysis for 16-18 hours and

so they were directly purified via anion exchange chromatography using the HiTrap™ Q FF anion

exchange column.

Secondary purification via anion exchange was performed similarly for all the freshly dialyzed

protein samples. Appropriately equilibrated columns were loaded with each protein for two full cycles

of their total volume. Columns were washed for 10 column volumes with either 50mM Tris-HCl pH 8.0

for TDE1701 and TDE2714, or 100mM glycine pH 9.6 for TDE0626. In order to elute the bound protein

out of the column, an increasing gradient of NaCl was added to the column using the same buffer which

was in the wash and equilibration step but with 1M NaCl included. The chromatogram visualized the

protein being eluted from the column and was monitored via a constant absorbance reading at 280nm.

For the TDE2714 protein sample a large single peak was shown on the chromatogram (Figure 6).
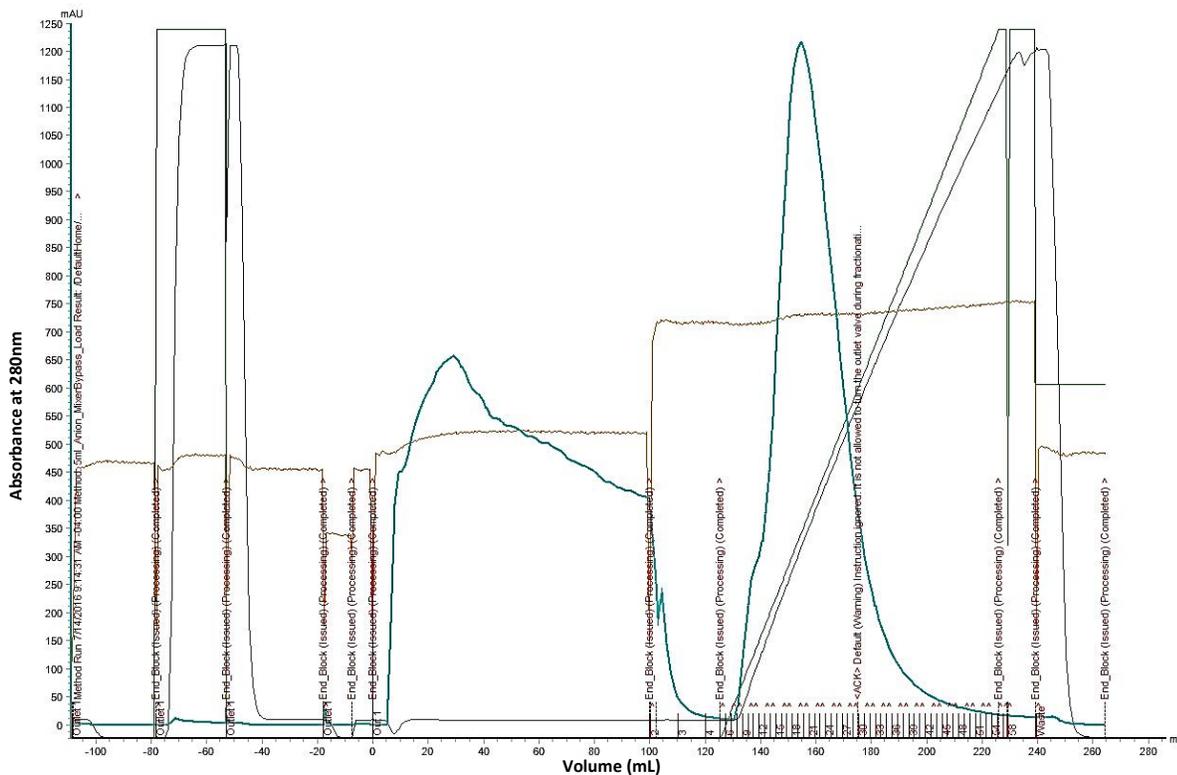
Figure 6: Chromatograph of anion exchange chromatography from purified TDE2714 in 50mM Tris-HCl pH 8.0. Protein sample volume added to the GE HiTrap™ Q FF 5mL column was 100mL based on an initial volume of 50mL, for two full cycles over the column. Absorbance units (mAU) at 280nm and Conductivity (mS/cm) were below a value of 2.0 prior to sample application. Flow rate set to 5 mL/min. The alarm for the delta column pressure maxima was set to 0.5MPa and minimum at 0.0MPa. Wash buffer used was 50mM Tris-HCl pH 8.0 and elution buffer, for salt gradient elution was 50mM Tris-HCl pH 8.0 with 1M NaCl. Peak fractions were at fractions 13-31 (38mL), which corresponded to the NaCl concentration of 357mM. Blue line represents absorbance reading at 280 and dark brown line represents conductivity.

The data from the anion exchange chromatogram for TDE2714 showed a broad peak, at

656.7mAU or 0.6567AU, during the sample application stage that represented the protein that did not

bind to the positive resin. There was a large amount of protein contaminants from the nickel-NTA

chromatography. The chromatogram had a large absorbance peak during the fractionation stage at

1214mAU or 1.214AU that spanned fractions 13-31 (38mL). This was found to be the purified TDE2714

based on the absorbance at 280nm. TDE2714 was successfully purified from the contaminants using the

secondary purification. During the elution stage of the chromatography, the total concentration of salt

in the peak was approximately 357mM NaCl. This was relatively close to the amount used during IMAC.

The anion exchange chromatogram for TDE0626 had a broad peak of contaminants during the

sample application stage with a peak absorbance at 41.02mAU (Figure 7).
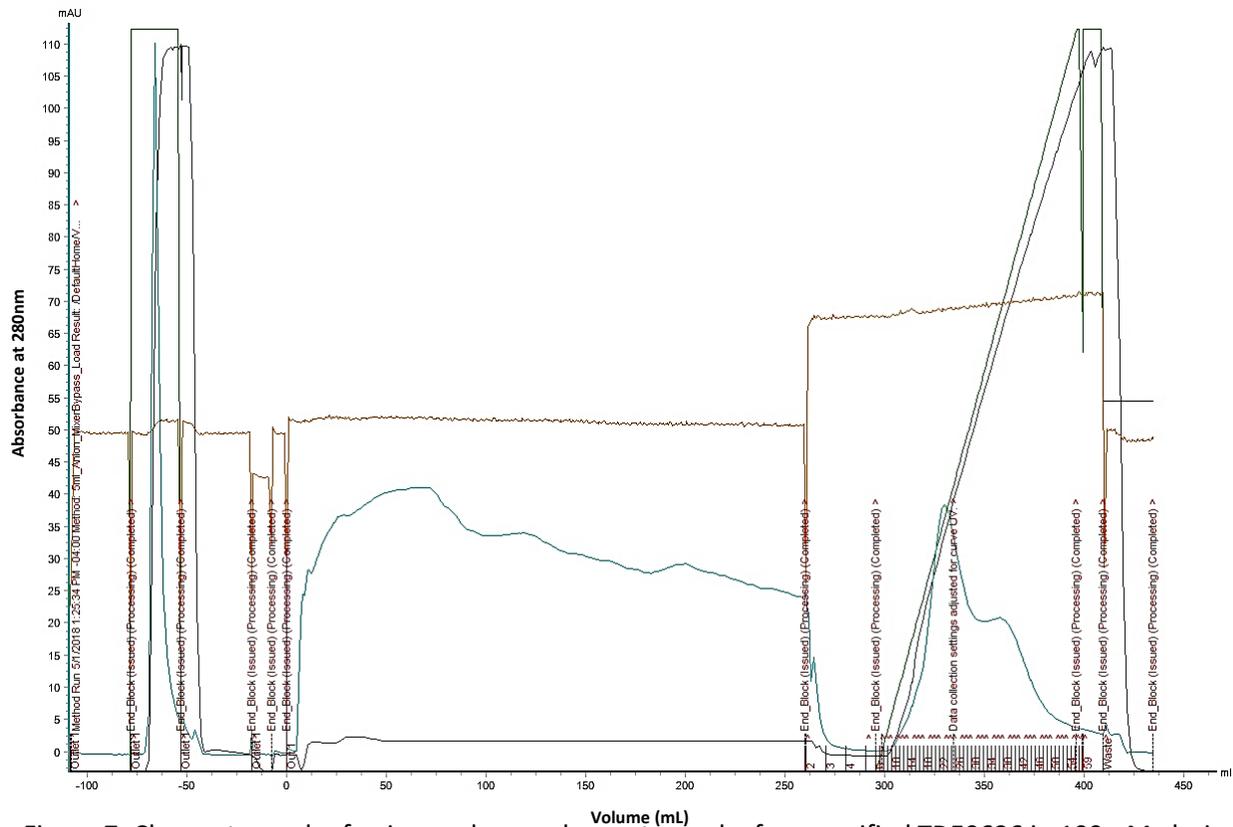
Figure 7: Chromatograph of anion exchange chromatography from purified TDE0626 in 100mM glycine pH 9.6. Protein sample volume added to the GE HiTrap™ Q FF 5mL column was 250mL based on an initial volume of 175mL, for two full cycles over the column. Absorbance units (mAU) at 280nm and Conductivity (mS/cm) were below a value of 2.0 prior to sample application. Flow rate set to 5 mL/min. The alarm for the delta column pressure maxima was set to 0.5MPa and minimum at 0.0MPa. Wash buffer used was 100mM glycine pH 9.6 and elution buffer, for salt gradient elution was 100mM glycine pH 9.6 with 1M NaCl. Peak fractions were at fractions 18-30 (26mL), which corresponded to the NaCl concentration of 238mM. Blue line represents absorbance reading at 280 and dark brown line represents conductivity.

There was also a small and broad peak during the elution stage which had a lower absorbance

reading than the sample application stage at 38mAU. There was also a large shoulder on the peak that

may correspond to another protein eluded from the column. There were more proteins not bound to

the column than the amount that did and or column separation. The results lead to the assumption that

the protein TDE0626 was not able to successfully bind to the column fully and protein was lost from the

purification step. The total concentration of salt in the peak was 238mM NaCl, that was lower than the

amount used for the IMAC.

The results for the anion exchange chromatography for TDE1701 showed a similarly broad peak

during the sample application stage at an absorbance of 134.65mAU (Figure A4 in Appendix). There was

a large single peak observed during the elution stage at 290.01mAU which spanned fractions 15-32

(36mL). TDE1701 was successfully purified from the protein sample with a low number of contaminants

based on the chromatogram. The concentration of the salt in the peak fractions was 337mM that was

similar to the amount of salt during the IMAC.

3.2.4 – Dynamic Light Scattering

In order to optimize the potential for protein samples to crystallize, it was necessary for the

purified samples to contain a low amount of aggregates and contaminants. To analyze purified samples,

dynamic light scattering (DLS) was employed using the protein samples after anion exchange

chromatography in 384-well plates at a final volume of 40μL. The samples were initially concentrated to

5mg/mL, where this concentration was lower than the concentration used during the initial

crystallization trials. The samples were exposed to a monochromatic light source that is then scattered

by the random Brownian motion of the particles in the sample(116). Fluctuations in the scattering

intensity and the subsequent particle movement were time-dependent so were able to give information

about the sample such as the diffusion coefficient, the predicted average radii of the particles, as well as

the polydispersity of the sample(117). The polydispersity measurement (%Pd) was used to for the

distribution of particles in the samples and is a dimensionless parameter that is calculated by the

instrument using the polydispersity index. A polydispersity of 20% or lower indicated a relatively

monodispersed sample and this was used as the standard for comparison for all the protein scattering

measurements(118). This measurement was conducted in an opaque microplate using a small portion of

the concentrated protein sample. The scattering profile was condensed into a series of images for each

of the protein samples and a plot of each sample and the predicted homogeneity was shown (Figures 8

and 9).

Mass Distribution	Radius (nm)	%Pd	Mw-R (kDa)	%Intensity	%Mass
Peak 1	3.433	25.6	60	6.6	96.7
Peak 2	31.212	41.5	10558	91.8	2.9
Peak 3	621.970	8.9	11588800	1.6	0.4
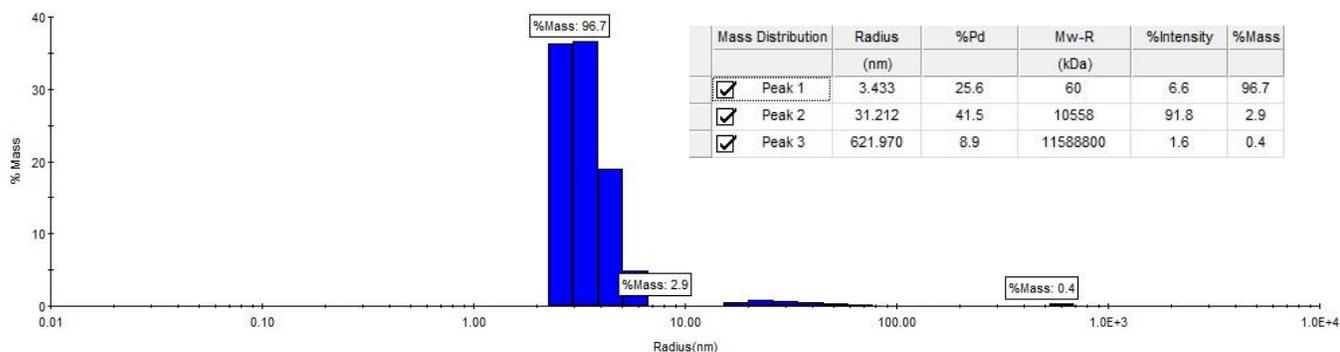
Figure 8: Dynamic light scattering plot for 5mg/mL TDE0626 in 100mM glycine pH 9.6 after anion exchange chromatography. Analysis depicts scattering profile as a measurement of predicted % mass versus the calculated average particle radius (nm). The mass distribution for the protein sample shown as three distinct peaks. Each peak shown with corresponding; radius, percent polydispersity (%Pd), predicted molecular weight based on radius, percent intensity, and percent mass. Twenty images were taken of the scattering profile every 5 seconds and averaged.

For the scattering profile of TDE0626 after anion exchange (Figure 8), three distinct peaks

corresponded to a percent of the total mass of the sample. The most predominant peak (at 96.7% of

mass) had a predicted particle radius of 3.433nm and a polydispersity of 25.6%. TDE0626 samples were

relatively monodispersed as it had a low particle radius and low polydispersity, despite the results by the

anion exchange chromatography, the sample still contained scattering that may be characteristic of

aggregates or other particles. The predicted molecular weight based on the scattering profile was

60kDa, that was similar to the results calculated by Protparam (50kDa). The sample seemed to contain a

few population masses that may correspond other particles or precipitated protein, however they only

accounted for 3.3% of the total sample mass.

Mass Distribution		Radius	%Pd	Mw-R	%Intensity	%Mass
		(nm)		(kDa)		
☑	Peak 1	5.247	30.1	163	97.2	99.8
☑	Peak 2	618.365	19.9	11432300	2.8	0.2
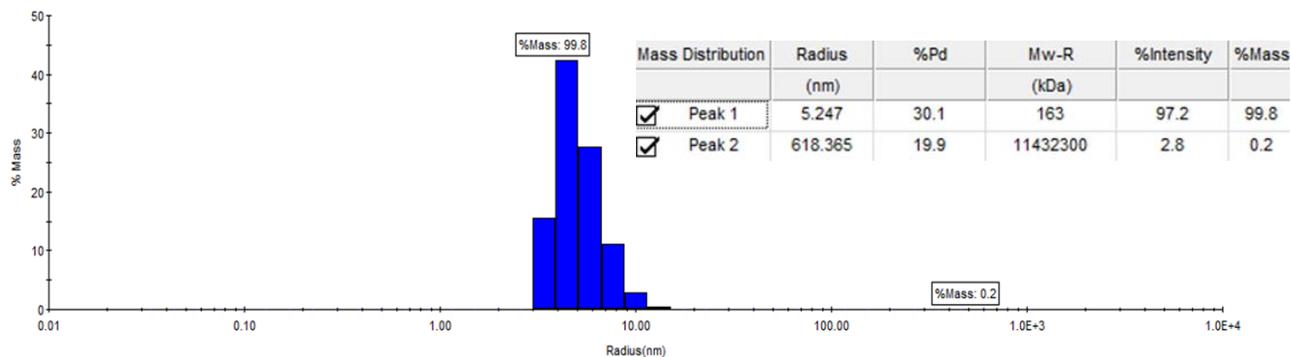
Figure 9: Dynamic light scattering plot for 5mg/mL TDE2714 in 50mM Tris-HCl pH 8.0 after anion exchange chromatography. Analysis depicts scattering profile as a measurement of predicted % mass versus the calculated average particle radius (nm). The mass distribution for the protein sample shown as three distinct peaks. Each peak shown with corresponding; radius, percent polydispersity (%Pd), predicted molecular weight based on radius, percent intensity, and percent mass. Twenty images were taken of the scattering profile every 5 seconds and averaged.

For the scattering profile of TDE2714 after anion exchange (Figure 9), the results showed two

distinct peaks which corresponded to a percent of the total mass of the sample. The most predominant

peak (at 99.8% of mass) had a predicted particle radius of 5.247nm and a polydispersity of 30.1%. This

analysis suggested that TDE2714 samples were relatively monodispersed, although it has a high

polydispersity is was also known to be a relatively large protein at ~75kDa. The protein sample had a

higher particle radius and high polydispersity, that may be due to the size of the protein itself, however

more likely the protein exists as a dimer at a concentration of 5mg/mL. The predicted molecular weight

based on the scattering profile was 163kDa. Further analysis into the possibility of TDE2714 forming a

dimer in solution may be required as the calculated molecular weight was twice the predicted weight of

a monomer of TDE2714. The sample contained a small mass that may correspond to other particles or

precipitated protein, as the predicted molecular radius was over 10 million Daltons, however this only

accounted for 0.2% of the total sample mass and are likely aggregates.

The dynamic light scattering results for TDE1701 after secondary purification (Figure A3 in

Appendix), had two peaks that were not distinct from each other. The most predominant peak (at 99.2%

of mass) had a predicted particle radius of 3.113nm and a polydispersity of 29.9%. The protein sample

was relatively high polydispersity and may be due to the prediction that 50% of the protein could be

intrinsically disordered. The protein sample had a low particle radius and high polydispersity and may

also support the prediction of TDE1701 being intrinsically disordered. The predicted molecular weight

based on the scattering profile was 48kDa, that is very similar to the results calculated by Protparam and

the protein sample. The sample seemed to contain a small mass that may correspond with other

particles or precipitated protein, however, this peak was not distinct from the main peak and this mass

may be denatured protein, however this only accounted for 0.8% of the total sample mass. DLS

summary table (Table A1) of the three targets can be found in the Appendix.

3.3 – Crystallization

3.3.1 – Initial Sitting Drop Crystallization

In order to attempt to obtain structural information about each of the recombinant proteins, X-ray crystallography was conducted. This was accomplished by the attempt for initial formation of protein crystals using sitting-drop batch crystallization. After secondary purification and dynamic light scattering, each purified protein concentration was calculated using the theoretical extinction coefficients from Protparam. For the initial crystallization attempts, TDE0626 was concentrated to 30mg/mL as well as 20mg/mL in 100mM glycine pH 9.6 that reduced the amount of aggregates for the protein during purification. TDE1701 was concentrated to 20mg/mL in 50mM Tris-HCl pH 8.0 that did not produce aggregates at this high concentration. Finally, TDE2714 was concentrated to 25mg/mL in 50mM Tris-HCl pH 8.0 and did not produce aggregates at this concentration. This was deduced for all the proteins based on visual observation of incubated crystallization plates which had a large number of conditions that did not have visible precipitation. The lack of precipitation in the conditions meant that the protein was stable in these conditions and higher concentrations of the protein could be used to promote crystallization.

Purified TDE2714 was initially crystallized using the Molecular Dimensions MCSG-2 Crystallization Screen via sitting-drop vapour diffusion at a reservoir to protein ratio of 1:1. A single crystal formed in the condition C10, which contained 0.2M sodium nitrate, and 20% (w/v) polyethylene glycol 3350 (Figure 10). The temperature for incubation was 18°C for approximately 1 month before

crystallization was recognized. The final concentration of protein for crystallization was 25mg/mL in the

drop.
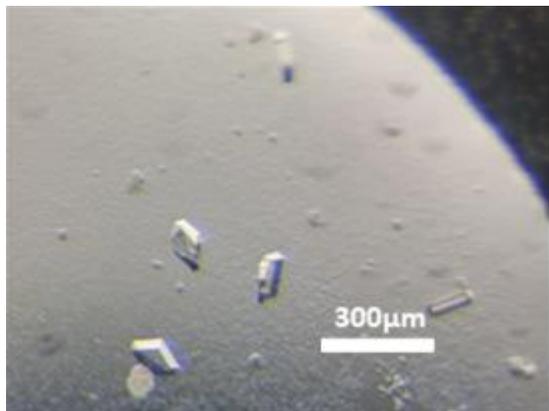
Figure 10: Crystal growth for purified TDE2714 at a final concentration of 25mg/mL in 50mM Tris-HCl pH 8.0. Crystallization occurred using the sitting-drop vapour diffusion technique incubated at a temperature of 18°C for approximately one month. Small plate-like crystals (resembling small diamond-like shapes) were seen in the C10 condition of MCSG-2, which contained 0.2M sodium nitrate and 20% (w/v) polyethylene glycol 3350. Approximate size of crystals was 100μm.

Based on the size and morphology of the crystals, it was determined that further optimization of

the crystal growth was required. This was due to the size of the plate protein crystal being

approximately 100μm which would be relatively small for the subsequent X-ray diffraction analysis. In

addition, the condition for crystallization did not have a buffer present, expect for the buffer in the

protein solution and so the pH of the commercial reservoir solution was taken and determined to be pH

of 8.0.

Initial crystallization of TDE1701 was also successful in two different conditions that produced

crystals. TDE1701 crystallized in condition E3 from MCSG-1 which contained 0.05M magnesium chloride,

0.1M HEPES: NaOH pH 7.5, and 30% (v/v) polyethylene glycol methyl ether 550 (Figure 11A). The ratio

for reservoir to protein was 1:1 and plates were incubated at 18°C. The crystals formed in approximately

3 weeks after incubation.
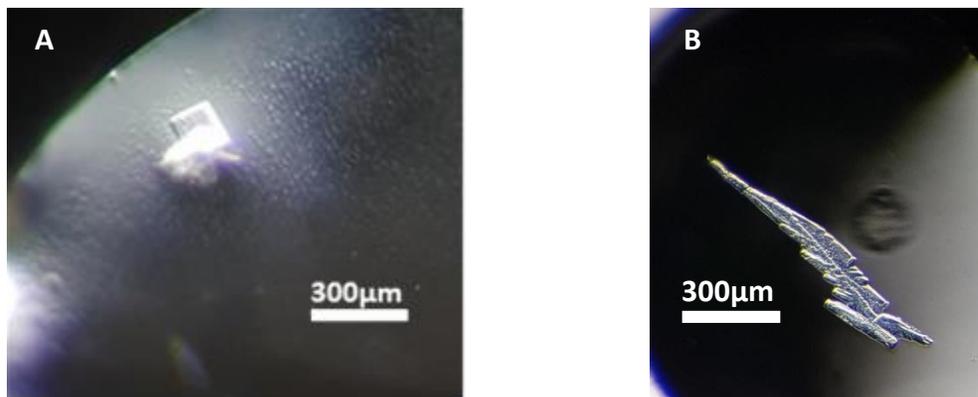
Figure 11: Crystal growth for purified TDE1701 in two different conditions. Final concentration for both conditions was 10mg/mL in 50mM Tris-HCl pH 8.0. Crystallization occurred using the sitting-drop vapour diffusion technique incubated at a temperature of 18°C for approximately three weeks. (A) Small overlapping crystals with one larger crystal were seen in the E3 condition of MCSG-1, which contained; 0.05M Magnesium Chloride, 0.1M HEPES: NaOH pH 7.5, and 30% (v/v) Polyethylene glycol methyl ether 550. Approximate size of the largest crystal was 150µm. (B) A large collection of crystals with multiple large crystals attached was seen in the C10 condition of the PACT screen, which contained; 0.2M magnesium chloride, 0.1M HEPES: NaOH pH 7.5, and 20% (w/v) polyethylene glycol 6000. Approximate size of the largest crystal was 250µm.

Another condition for TDE1701 was found to produce protein crystals (Figure 11B). Crystalline

TDE1701 was grown in the C10 condition of PACT premier™ HT-96 from Molecular Dimensions. The

protein concentration, incubation period, and temperature were the same as previously repeated as

stated above. The C10 PACT condition contained 0.2M magnesium chloride, 0.1M HEPES: NaOH pH 7.5,

and 20% w/v polyethylene glycol 6000. TDE1701 crystallization requires the presence of a low

concentration of magnesium chloride, as well as a HEPES buffer at pH 7.5. Both conditions required

optimization in order to produce better quality crystals for X-ray diffraction analysis. The presence of

Mg^{2+} in the condition may also suggest this to be a cofactor for TDE1701.

Unfortunately, attempts for the crystallization of TDE0626 were unsuccessful even at a variety

of concentrations. Concentrations of 5, 8, 12, 15, 20, 25, and 30mg/mL did not yield any crystals at a

ratio of 1:1 reservoir to protein. Buffers present in concentrated protein were 100mM glycine pH 9.6

and 100mM Na_2PO_4 pH 8.0. After anion exchange chromatography, the protein was further dialyzed to

remove all traces of salt. All available crystallization screens were used for the purified protein such as

MCSG1-4, PACT, Hampton Research Crystal Screen™ HR2-110, Index™ HR2-144, and Molecular

Dimensions Morpheus-HT-96. Each screen was incubated at 18°C until all conditions were shown to be

precipitated protein and no crystal growth.

3.3.2 – Optimization of Crystal Growth

Since the size and morphology of the initial protein crystals were not ideal for X-ray diffraction

analysis, the conditions were expanded by increasing the volume of protein used as well as reservoir

used. Instead of 0.6μL sitting drop to 60μL of the reservoir, a total of 500μL of reservoir solution would

be used and subsequently 6μL of the protein/reservoir drop was used. This larger volume was used in

order to generate large protein crystals by allowing for more protein to be packed into a crystalline

structure. Due to the size of the solutions used, a hanging drop vapour diffusion method was introduced

instead. This method has been known to produce more bulky crystals when conditions that produced

small crystals were used by this method(119). Based off the results for TDE1701 and TDE2714, the

crystal conditions were altered by increasing and decreasing the concentration of polyethylene glycol to

determine a more optimal percentage of the precipitating agent (120). Optimizing conditions for

TDE2714 was conducted using the purified protein in Tris-HCl pH 8.0. Hanging drop vapour diffusion was

setup in a 24-well plate with concentrations of TDE2714 at 12, 14, and 16mg/mL. The concentration of

polyethylene glycol was adjusted from 19.6% to 20.2% (w/v). This was conducted based on the known

principle that PEGs, exert the volume-excluded effect (or increase in the viscosity) as well as they

decrease the dielectric constant of the solvent which allows for the electron densities to be similar to

water and not damaging(119,120). There was an addition of two organic solvents which were glycerol

(2.5% v/v) and ethylene glycol (5% v/v). These solvents were used to impact crystallization to prevent

the formation of a large number of microcrystals. Organic solvents also are used to remove water

localized areas and divert their interactions with the protein, which allowed for better crystal

formation(123). The pH of Tris-HCl ranged from 7-8 based on pH test of commercial reservoir solution.

Plates were incubated at 18°C for 1 month (Table A2 in Appendix). In addition, crystallization of TDE2714

was performed with 2mM cysteine as TDE2714 hypothesized to interact with cysteine.

From the data presented in the table A2, there were 8 separate conditions that yielded large

protein crystals for TDE2714. The results showed that the crystal growth was successfully optimized for

diffraction analysis due to the size and morphology of the protein crystals (Figure 12). These crystals

(Figure 12) were the best conditions thus far for TDE2714.
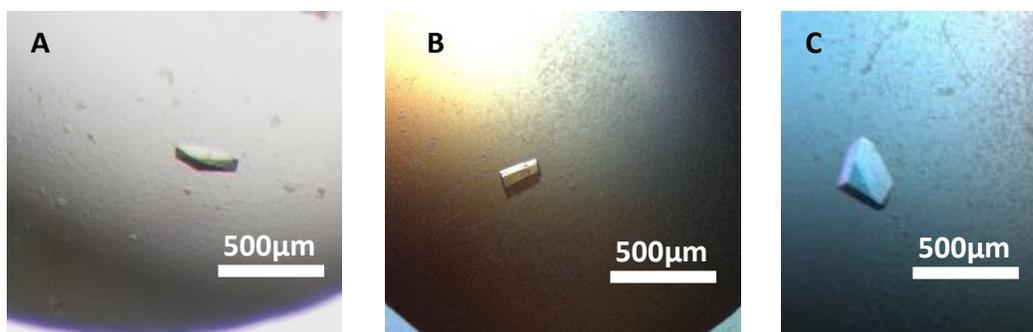
Figure 12: Optimized crystal growth for purified TDE2714 in three different conditions. Final concentration for protein conditions was; 12, 14, 16mg/mL in 50mM Tris-HCl pH 8.0. Crystallization occurred using the hanging drop vapour diffusion technique incubated at a temperature of 18°C for approximately one month. (A) A large 400μm protein crystal was formed from condition A1 of the plate, which contained 2.5% (v/v) glycerol, 50mM Tris-HCl pH 7.0, 19.6% (w/v) PEG 3350, and 0.2M NaNO$_3$. (B) A large 300μm protein crystal was formed from condition C3 of the plate, which contained 2.5% (v/v) glycerol, 50mM Tris-HCl pH 8, 20% (w/v) PEG 3350, and 0.2M NaNO$_3$. (C) A large 600μm protein crystal was formed from condition D5 of the plate, which contained 5% (v/v) ethylene glycol, 50mM Tris-HCl pH 8, 20.2% (w/v) PEG 3350, and 0.2M NaNO$_3$. All stated protein sizes were approximated, and may not be accurate, based on size of loop required for harvesting.

The changes to the polyethylene glycol concentration for the conditions allowed for the increased size and three dimensionality of protein crystals for TDE2714. The morphology of the crystals was bulky diamond shaped. There were also two organic solvents (glycerol and ethylene glycol) added to the conditions which aided growth of the protein crystal. One observation during the incubation of the plate was that the protein seemed to aggregate together and precipitate out initially. However, over the course of approximately 1 month, some of the precipitate dissolved and crystals developed.

Based on the results for the optimization of TDE2714 crystal growth, TDE1701 crystallization was also attempted be optimized in a similar manner. However, the hanging drop method did not seem to produce protein crystals for TDE1701 at incubation temperatures ranging 4°C (to slow the motion within the protein drop) up to 20°C. Changing the concentration of polyethylene glycol by increasing or

decreasing also did not allow for the formation of TDE1701 crystals. The use of hanging drop vapour diffusion for TDE1701 was not ideal for crystal formation and so this was not further explored.

In a similar manner, no optimization of crystal growth was conducted for TDE0626 as initially there were no crystals formed via sitting-drop vapour diffusion. Further crystallization attempts with a range of additives, such as reducing agents, polyamines, volatile and non-volatile organic solvents, and chaotropes, will be added to the protein sample prior to crystal screening. This additive technique has been known to aid in the crystallization of difficult proteins(119,120).

For each of the protein crystals sent, a cryoprotectant was used during the looping process. This was done in order to prevent the formation of ice crystals during the flash-freezing process in liquid nitrogen(126). The buildup of ice crystals during the diffraction analysis produces large dark ice rings in the pattern which decreased the ability for proper indexing of the data(127). A cryoprotectant containing 30% (v/v) glycerol, 0.2M sodium nitrate, 20% (w/v) polyethylene glycol 3350, and 50mM Tris-HCl pH 8.0, was used after each of the TDE2714 crystals were removed from the hanging-drop. Each of the crystals showed no change in morphology when suspended in the cryoprotectant and therefore were flash-frozen with liquid nitrogen for shipment to the CLS.

In a similar manner, the protein crystal from of TDE1701 was suspended in a cryoprotectant that contained; 0.05M magnesium chloride, 0.1M HEPES: NaOH pH 7.5, and 30% v/v polyethylene glycol methyl ether 550. No glycerol was added to the cryoprotectant as the presence of 30% v/v polyethylene

glycol methyl ether 550 served as a sufficient cryoprotectant. Like-wise no changes to the size or

morphology were seen during the suspension of the crystals.

3.4 – X-Ray Diffraction Analysis

3.4.1- Diffraction Quality Crystals and Diffraction Pattern Analysis

For structural characterization of each of the recombinant proteins, TDE1701 and TDE2714,

large crystals were sent to the Canadian Light Source (CLS) Synchrotron for X-ray diffraction. During the

X-ray diffraction of the protein crystals for TDE1701, the X-ray wavelength used was 0.97949 Å at a slit

width of 0.5μm. Images were collected on oscillation-mode, set to 0.2° oscillations over a 720-degree

total rotation in which the 1800 images were indexed, merged and processed via a built-in Mx Live

processing program. The diffraction data key indicators were tabulated for quality analysis (Table 7).

Table 7: Crystallographic statistics for the best collection and refinement of TDE1701. Data collected at
the Canadian Light Source synchrotron at an X-ray wavelength of 0.97949 Å. 1800 images taken over a
720-degree rotation at 0.2° oscillations. Data shown was processed via SCALA. Molecular replacement
used Phaser-MR software and refinement software used was Refmac-5. Ramachandran statistics were
generated via COOT model building software.

Diffraction Collection	
Space group	$P2_12_12_1$
Cell dimensions	
a (Å)	51.09
b (Å)	156.70
c (Å)	158.01
α (degrees)	90.0
β (degrees)	90.0
y (degrees)	90.0
Resolution (Å)	50.0-2.40 (2.53-2.40)
$CC^{1/2}$	0.99 (0.823)
$I/\sigma I$	34.6 (2.2)
Completeness (%)	99.4 (100.0)
Multiplicity	11.9 (12.4)
Average mosaicity	0.2

The data from the diffraction collection of TDE1701 had an outer shell resolution truncated at 2.4 Å and an inner shell resolution of 50Å. This high-resolution measurement had 100% completeness of the data in the highest resolution shell. There was an adequate number of reflections included in the reciprocal space of the asymmetric unit for the calculated space group of $P2_12_12_1$(128). However, from the data processed by SCALA, the R_{merge} value was 1.089 in the outer shell that meant there were discrepancies of the reflections over multiple measurements in the in the data collected. The space group of the protein crystal was orthorhombic, that has been known to be most frequently predicted space group for protein crystals(129). Based on the data collected, the $I/\sigma I$ value, that in terms is the signal to noise ratio at 2.2, suggesting the resolution range was suitable at this resolution. Additionally, the $CC^{1/2}$ indicator was 0.823 indicating that the there was an acceptable correlation of the observed dataset to the unmeasurable true signal, based on a $CC^{1/2}$ cutoff of 1.

Molecular replacement for TDE1701 was attempted in order to obtain phase information in order to build a structural model of the protein. Phaser-MR was used with the input of homologous protein PDB file 5j73A, that had a sequence identity of 15% at the amino acid level, but only had an alignment coverage from positions 250-296 in the sequence. Using the coordinates from the PDB file and the collected X-ray data, a protein model was generated and refined using Refmac-5 (Table 8). Based on the data generated by model refinement, the data was a poor representation of the data collected. Based on the high R_{work}/R_{free} values, the 'model' generated was a poor representation of the true structure analyzed, as this ratio is a valid indication of model quality(130). Furthermore, a Matthew

coefficient generated within CCP4 suggested the presence of two molecules in the asymmetric unit of

TDE1701 based on 53.5% solvent volume. An image of the model of TDE1701 was taken via COOT

however the data should not be interpreted as a correct structural model of TDE1701 (Figure 13).
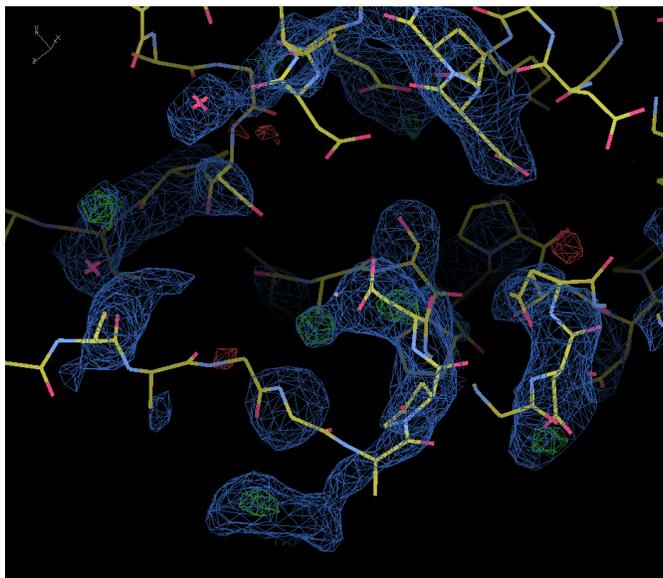
Figure 13: Overall best output from molecular replacement and initial model refinement for TDE1701 via COOT model building software. The generated model shown represented was more consistent with random scattering of water molecules rather than prediction of the true structure based on the values for R_{work}/R_{free}. Molecular replacement conducted via Phaser-MR and model refinement generated by Refmac-5. R_{work}/R_{free} for model was 0.48/0.53. Blue contour of the electron density shown represents the known crystallographic data at 1.57 Å signal cutoff.

In a similar manner, protein crystals of TDE2714 were diffracted for structural characterization.

The X-ray beamline used had a wavelength of 1.000Å and the crystal was imaged at 0.2° oscillations for

a total of 720° and the processed data was tabulated for analysis via SCALA (Table 8).

Table 8: Crystallographic statistics for the best collection and refinement of TDE2714. Data collected at the Canadian Light Source synchrotron at an X-ray wavelength of 1.0000. Images taken at 0.2° oscillations for a total of 720° rotation. Data shown was processed via SCALA. Molecular replacement and Refinement (not included) did not yield sufficient data for model building.

Diffraction Collection	
Space group	$P2_1$
Cell dimensions	
a (Å)	100.02
b (Å)	102.13
c (Å)	113.34
α (degrees)	90.00
β (degrees)	109.34
y (degrees)	90.00
Resolution (Å)	48.41-2.75 (3.03-2.75)
R_{merge}	0.159 (0.819)
$CC^{1/2}$	0.99 (0.81)
$I/\sigma I$	8.4 (2.2)
Completeness (%)	99.9 (99.7)
Multiplicity	7.5 (7.8)
Average mosaicity	0.4

The unmerged data set generated by the program SCALA, revealed a primitive monoclinic $P2_1$ space group that had a resolution of 2.75 Å in the outer shell and a 48.41 Å resolution in the inner shell.

At the resolution cutoff at 2.75 Å, the completeness was 99.7%. In addition to the completeness, the signal to noise ratio was 2.2 in the highest shell. In contrast however, the R_{merge} value was relatively high for the dataset at 0.819 at the 2.75Å resolution. Phaser-MR was attempted to obtain phase information.

Based on the Matthew's coefficient, three molecules were calculated within the asymmetric unit at a cell volume of approximately 50.5% solvent. Molecular replacement was conducted using the homologous protein model (PDB:2y3c) that was the *Treponema denticola* Variable Protein with an amino acid sequence identity of 45% and aligned with positions 394-660 of TDE2714. However, the phases generated via molecular replacement was insufficient to be able to generate a structural model.

3.4.2 – Selenomethionine Crystallization

To attempt the experimental determination of phase information, crystallization of

selenomethionine crystals was conducted using the same optimized conditions for TDE2714. This

procedure was performed in order to obtain better quality phase information via the use of anomalous

diffraction experiments (SAD/MAD). The phase problem is defined as the lack of phase information of

the scattered X-rays which hit the detector after diffraction. This information is crucial as it is required to

calculate the electron density by breaking down the complex waves into simple waves via the Fourier

Transform(131). In order to overcome this challenge Selenomethionine containing crystals were

successfully grown (Figure 14). In anomalous dispersion experiments, selenium (as well as other heavy

metals) are exposed to wavelengths that are at the natural frequency of the metal ions. This causes for

an absorption of the wavelength and a breakdown of Friedel's Law. This phenomenon can then be

identified as distinct atomic anomalous scattering signals, and once the position and amplitude of the

ions is identified, the phase can be calculated of the anomalous scatterer(132). The selenomethionine

crystals were flash-frozen in a similar cryoprotectant for the native TDE2714 crystals and sent to the CLS

for diffraction. The protein crystals were found to have a similar outer shell resolution of 2.71 Å. Three

fluorescent test scans were conducted for the TDE2714 crystal at selenium absorbance peak edge of

12665.7eV and an inflection peak of 12667.7eV (Table A4 in Appendix). Anomalous signals for selenium

ions were unable to be detected when scanned from. With the lack of anomalous signals, the data was

insufficient to aid in the generation of a protein model of TDE2714. For all three of the recombinant

proteins, it was apparent that further crystallization and diffraction analysis will be required in order to

obtain structural information.
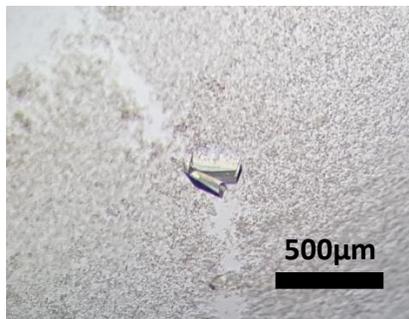
Figure 14: Crystal growth for purified Selenomethionine containing TDE2714. Final concentration for the protein conditions was 12mg/mL in 50mM Tris-HCl pH 8.0. Crystallization occurred using the hanging drop vapour diffusion technique incubated at a temperature of 18°C for approximately one month. This was similar to native conditions. (A) A large 300μm protein crystal was formed which contained; 2mM cysteine, 2.5% (v/v) PEG 400, 0.2M NaNO$_3$, 50mM Tris-HCl pH 8.0, 20% (w/v) PEG 3350. 2.5% (v/v) PEG 400 was added to condition to further increase in the viscosity of the condition.

3.5 – Functional Analysis

3.5.1 – Site-Directed Mutagenesis of TDE2714

Based on the conserved residue map of TDE2714 against 100 different FGEs, there were

residues in the sequence of TDE2714 that were found to be highly conserved across all the FGEs

analyzed. In determining a possible residue to change, a sequence comparison was conducted for

TDE2714 and a cysteine-active FGE from *S. coelicolor* (PDB: 2AIK) that has two catalytically active

cysteines in its active site. One of these cysteines from 2AIK aligned with cysteine-637 of TDE2714 and

this cysteine was chosen to be changed into an alanine via site-directed mutagenesis. After PCR was

conducted an agarose gel was run for the PCR samples (Figure A7 in Appendix). The target PCR product

of the gene was 1998bp and a prominent band can be visualized on the agarose gel at ~ 2kbp. This band

was extracted and sent to Western University for sequencing. The results of the sequencing were able to

confirm a single amino acid change at position 637 from cysteine to alanine. The mutant was then

transformed into BL21 cells and successfully expressed and purified for analysis.

3.5.2 – TDE0626 as a Polysaccharide Lyase

Based on the information gathered by the bioinformatics of TDE0626, the putative

polysaccharide lyase function was further explored. The polysaccharide lyase was predicted to degrade

1-4-α or β uronic acids by a β-elimination reaction based on identified parallel β-helix motif that is

common to this family of enzymes(90,128). Based on the bioinformatics, TDE0626 may be involved in

nutrient uptake and/or dispersal of the biofilm(12). Initially, TDE0626 was test for its ability to degrade

cellulose by incubating purified TDE0626 with cellulose and imaged using silver nitrate (Figure A8 in

Appendix). Since TDE0626 was predicted to be a polysaccharide lyase that degrades GAGs, the enzyme

was unable to degrade cellulose based on the lack of bands visualized from the silver stained acrylamide

gel. The polysaccharide lyase activity was examined by the degradation of a biofilm extract from a co-

culture of the red complex that was developed and conducted by Vasu Patel. Isolated biofilm was

accomplished by Chris Bartlett and Sidney Nechacov. The biofilm degradation assay was conducted via

the addition of purified TDE0626 into the isolated biofilm with an incubation period of 1-3 hours at 37°C.

A silver-stained polyacrylamide gel was used for determination of biofilm degradation (Figure 15).
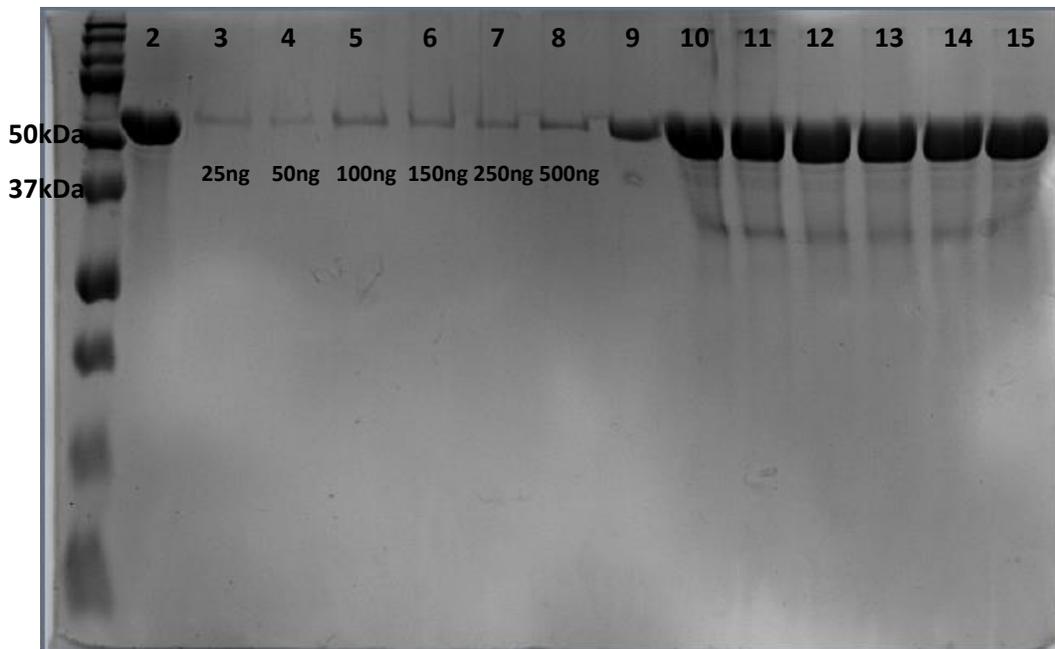
Figure 15: SDS-PAGE gel imaged via silver staining of biofilm degradation for TDE0626. Purified TDE0626 was incubated with an isolated biofilm generated by the red complex bacteria; *T. Forsythia, P. gingivalis,* and *T. denticola*. Treated biofilm samples were incubated for 1-3 hours at 37°C in order to ensure complete degradation of the biofilm. Lane 1 contained the Precision Plus Protein™ Kaleidoscope ladder. Lane 2 was 500ng/mL of TDE0626. Lane 3-8 were isolated biofilm prior to treatment at increasing concentrations from 25-500ng. Lane 9 was TDE0626 supernatant collected after reactions. Lanes 10-15 were subsequent washes of TDE0626 with 1M NaCl samples.

The silver staining assay had an interesting banding pattern from the salt precipitated samples.

In lanes 10-15, there were visible bands that had a lower apparent molecular weight then the untreated

samples. This may have been caused by the degradation of GAGs found in the biofilm however, this may

also be caused by the degradation of the protein sample itself. that the biofilm was not successfully

degraded by TDE0626. This uncertainty was due to the use of a silver nitrate stain as it is an

indeterminate stain and will stain both carbohydrates and proteins, making for resolution of the two

molecules to be unattainable. A crystal violet assay to quantify the degradation of the biofilm by visual

tracking of the degradation was also none resultant(134). In spite of the uncertain results for the biofilm

assay and the unsuccessful results of the crystal violet assay, TDE0626 may still be a polysaccharide lyase, and may be able to successfully degrade the biofilm isolates. Further testing would be required.

3.5.3 – TDE2714 as a Formylglycine-Generating Enzyme

3.5.3.1 – Carbohydrate-PAGE of TDE2714

Initially the carbohydrate-binding function of TDE2714 was examined by carbohydrate-based polyacrylamide gel electrophoresis (C-PAGE). For each of the experiment, TDE2714 was treated with polysaccharides such as polygalacturonic acid, as the long chain carbohydrates may be similar in structure to the natural substrate of TDE2714 in a biofilm. It was hypothesized that if TDE2714 was an extracellular carbohydrate-binding protein, it would possibly bind to various polysaccharides or GAGs found in the mature biofilm. Based on the C-PAGE results (Figure A9 in Appendix), there was no visible change in the mobility of the TDE2714 in C-PAGE when treated with 0.5% (w/v) polygalacturonic acid. The lack of change in the mobility of TDE2714 suggested that there was little to no interaction with the long chain carbohydrate(135).

3.5.3.2 – Peptide Thin-Layer Chromatography of TDE2714

Important also for the nutrient uptake of bacteria is the scavenging of sulfur from sulfate surface glycans(136). One of the main contributors to this has been sulfatases. Interestingly, based in the bioinformatics, TDE2714 contains a FGE (formylglycine-generating enzyme) – sulfatase domain, that is known to modify a cysteine or serine residue on a type I sulfatase via oxidation to a oxoalanine (or formylglycine)(137). Since the Eukaryotic system for FGEs has been well characterized, a synthetic

peptide (-LCTPSRA-) of the sulfatase signature sequence was synthesized and reacted with purified

TDE2714(138). Initial results for the enzymatic reaction showed the formation of a noxious rotten eggs

smell that could indicative of the formation of hydrogen sulfide(139). Based on the purposed enzymatic

mechanism, the target cysteine on the peptide substrate could have been modified to a formylglycine

residue. The hemolytic properties of H_2S have been extensively studied on tissue damage and can be

utilized for pathogens invading hosts for nutrient accessibility(71). As such, it was hypothesized that

TDE2714 may have some hemolytic activity within the host through this modification and subsequent

formation of H_2S.

To further test this hypothesis, a thin-layer chromatography (TLC) protocol was develop with the

aid of Peter Nguyen (Dept. of Chemistry and Biochemistry, WLU) and amino acid/peptide TLC

protocols(140). Ninhydrin (1% (w/v)) was chosen as the developing stain for silica TLC plate and was

selective for the primary amine groups on the peptide used. When TDE2714 was in the presence of the

peptide substrate, there was a separate spot that migrated farther up the plate than samples without

the peptide substrate or enzyme itself. A new non-polar product was formed during the reaction that

may have been the oxoalanine residue, as the mobile phase used contained 60%(v/v) 1-butanol. A full-

time course reaction was conducted to identify a visual rate of the reaction and formation of product

(Figure 16).
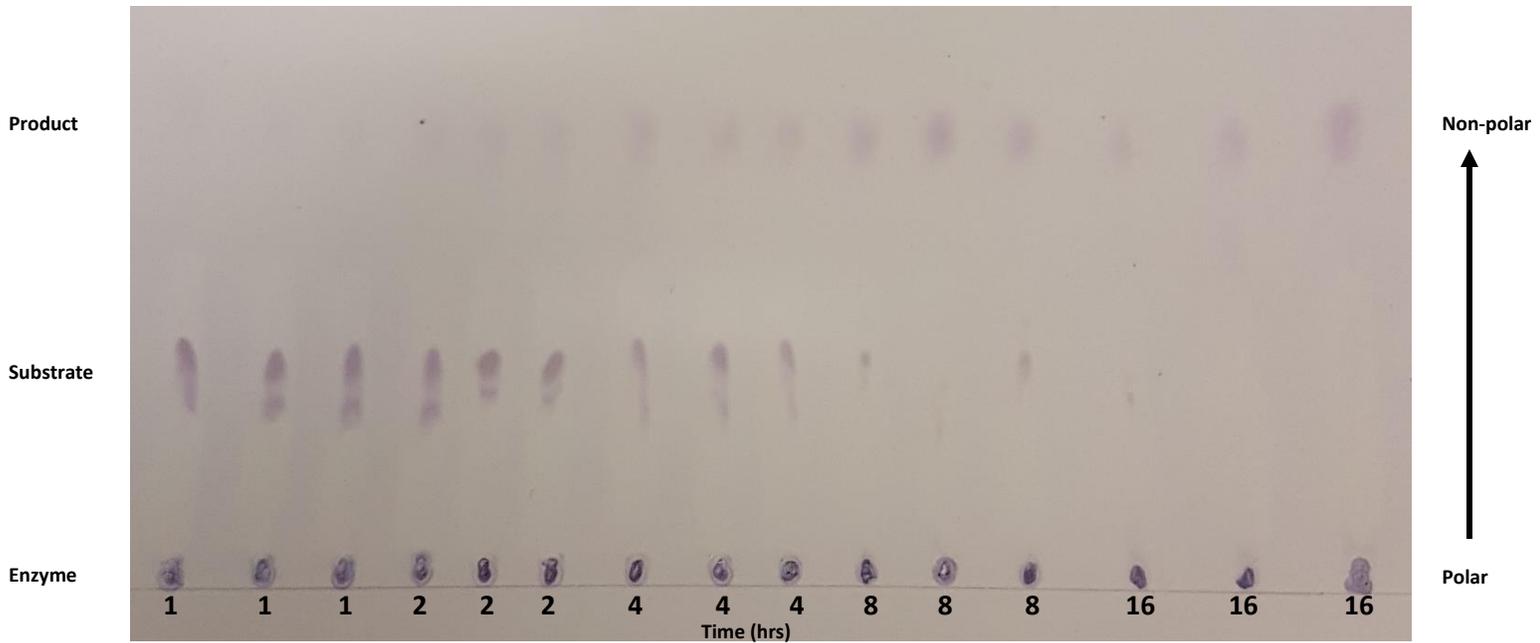
Figure 16: Time-course peptide thin-layer chromatography of TDE2714, a putative formylglycine-generating enzyme. Each of the samples were composed of 0.1mM purified TDE2714 in 50mM sodium phosphate pH 8.0, 1mM of LCTPSRA peptide substrate, and 2mM dithiothreitol (DTT). Samples were incubated at 37°C for time periods of; 1, 2, 4, 8, and 16 hours. The stationary phase used was Millipore TLC Silica gel 60 and the mobile phase was 3:1:1 of 1-butanol: glacial acetic acid: water. The experiment was conducted in triplicates and developed via TLC stain of; 1% (w/v) ninhydrin in 5:1 (v/v) pyridine: glacial acetic acid. Three spots for each of the reactions was observed were denoted as either the enzyme, substrate, or product based on mobility on plate.

From the peptide TLC results, three distinct spots were visualized. The spots that were located

at the bottom the plate corresponded to the enzyme in the reaction based on prior testing in the

absence of the peptide substrate, where the enzyme alone lacked mobility on the plate. There was a

consistent concentration of enzyme added to each of the reactions (0.1mM). It was also observed that a

second less polar spot or smear was stained, this was assumed to be the unmodified peptide substrate

in each of the samples. This assumption based on prior testing with the peptide substrate in the absence

of the enzyme (Figure A10A in Appendix). There was also a third non-polar spot that appeared during

the second hour of the enzymatic reaction and increased in intensity over the course of time. The non-

polar product was possibly caused by the formation of the oxoalanine/formylglycine. In addition to the

formation of the potentially modified product, there was also a disappearance of the peptide substrate

over the time elapsed that may be caused by the substrate was being converted to the non-polar

product. In order to better visual the data, an image was taken, and the relative intensities could be

measured and plotted against the time elapsed (Figure 17).
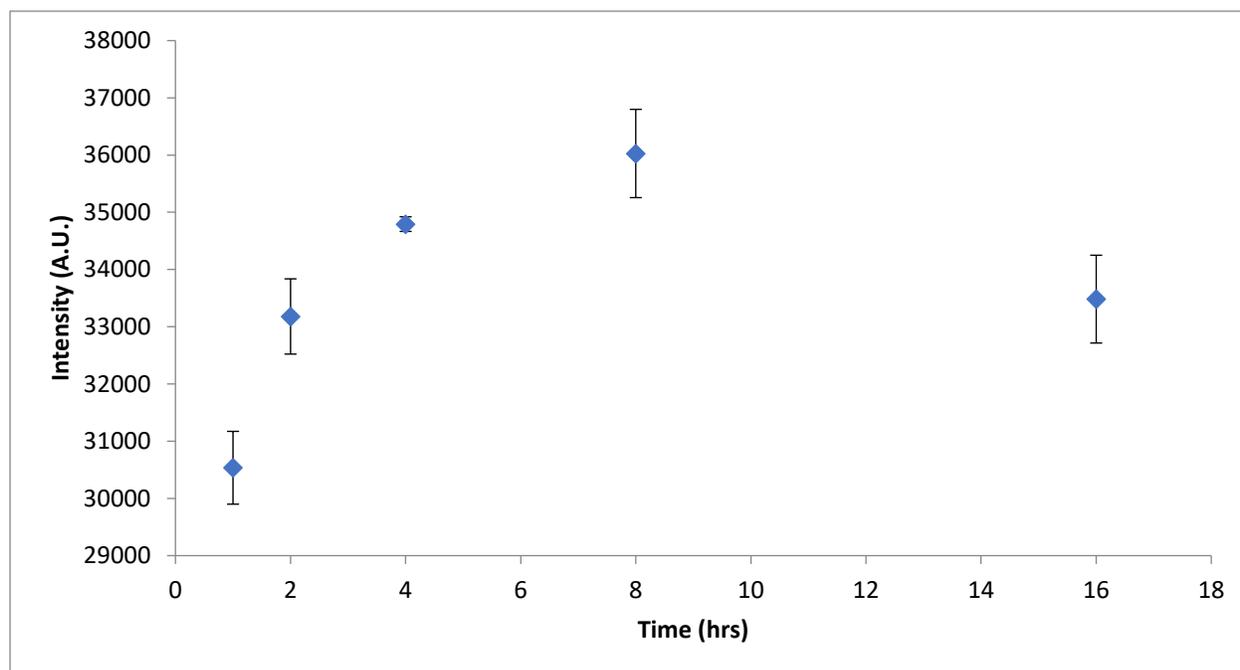
Figure 17: Relative intensity over time elapsed of the third most polar spot from the time course thin-layer chromatography of TDE2714 reaction. The TDE2714 reaction contained; 0.1mM purified TDE2714 in 50mM sodium phosphate pH 8.0, 1mM of LCTPSRA peptide substrate, and 2mM dithiothreitol (DTT) and were incubated at 37°C for time periods of; 1, 2, 4, 8, and 16 hours prior to staining. Intensities were measured via the VersaDoc densitometry function, where each spot's intensity was measured with the same cell volume. Each measurement was conducted in triplicates.

There was a measurable increasing in the intensity for the potential peptide product formation,

that reached a peak intensity of approximately 37000 A.U. at 8hrs before decreasing to 33000 A.U. The

data had similarities to a Michaelis-Menten saturation curve. This gave an indication that there was an

enzymatic reaction that occurred and so further analysis via the characterization of the potentially

modified peptide product was conducted.

Using the C637A mutant in place of the wild-type TDE2714, a similar peptide TLC experiment

was conducted. This was used to determine if cysteine-637, which aligns with a known catalytic cysteine

in an FGE from *S. coelicolor*, is important for the function of TDE2714. The TLC experiment with the

C637A mutant was similar to the TLC experiment with the wild-type (Figure A10B in Appendix). The non-

polar spot that was assumed to be the modified peptide product was still visible suggesting that the

mutant of TDE2714 remained catalytically active. The retention of activity of the mutant suggests that

cysteine-637 is not catalytically integral for TDE2714.

3.5.3.3 – Electro-Spray Ionization Mass Spectrometry of Peptides

For the characterization of the putative product from the enzymatic reaction, electron spray

ionization mass spectrometry was conducted. Analysis of aqueous samples was done at the Advanced

Analysis Center at University of Guelph by Dr. Dyanne Brewer. Samples of the unmodified and 16h

(modified) reactions were flash-frozen and sent directly to the facility. This was done in order to

determine the mass change due to the predicted loss of a thiol group on the modified cysteine of the

peptide. The calculated molecular weight of the unmodified peptide sample was 746.37Da and based on

the deconvoluted mass spectrum a peak of 746.37Da was noted (Figure 18A and B). Dr. Dyanne Brewer

was then able to further analyze the collection of mass peaks and generate a correct control sequence

of LCTPSRA (Figure 18C).
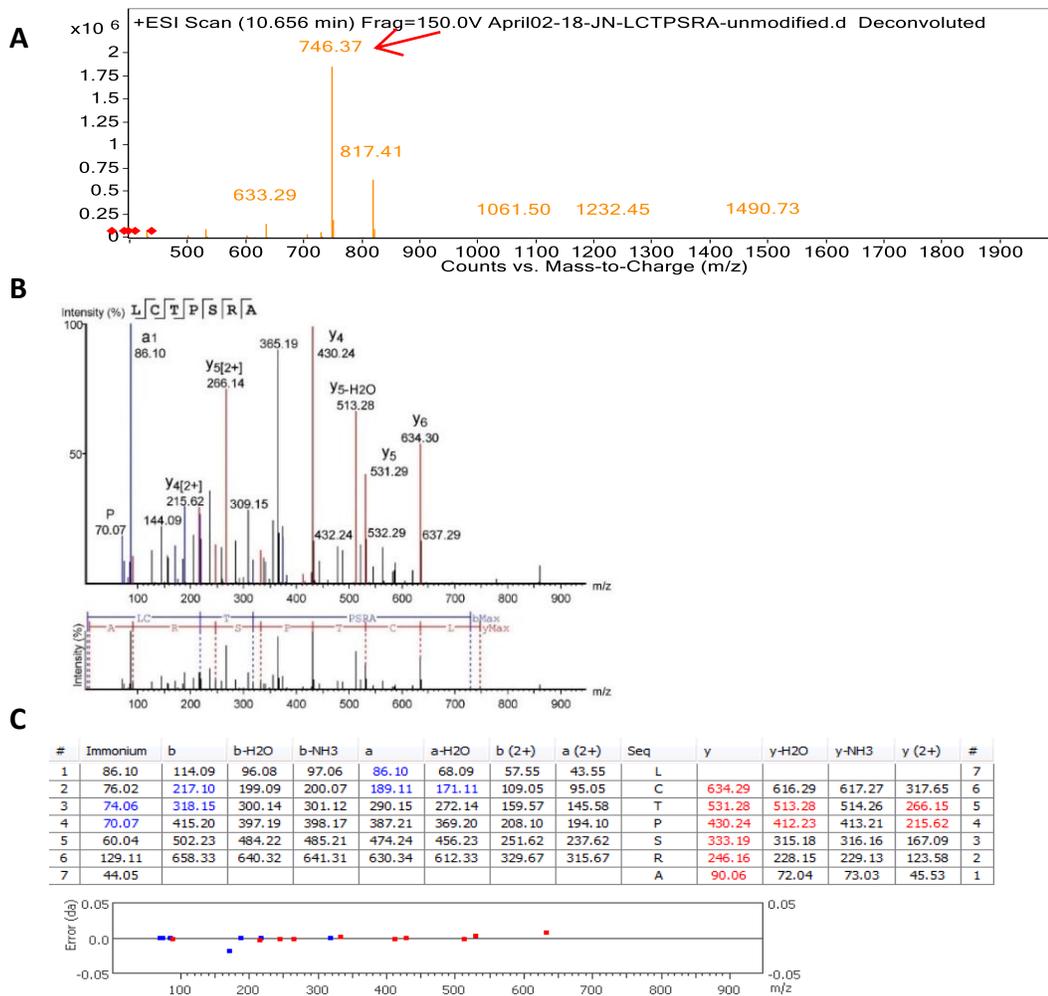
Figure 18: Electron Spray Ionization Mass Spectrometry (ESI-MS) of control peptide LCTPSRA at a concentration of 2mM. Peptide suspended in water. Spectroscopic analysis of sample conducted at the Advanced Analysis Center in the University of Guelph by Dr. Dyanne Brewer. Ionization polarity for sample was positive. (A) Mass spectrum for the control peptide -LCTPSRA- generated via time of fight measurement (TOF) which was deconvoluted to show a peak mass signal at 746.37Da. (B) Mass spectra of control peptide for each of the amino acids overlaid on a single spectrum for sequencing. Each peak was further labeled based on its intensity and charge, shown as ymax value and b value in (C).

Since the data was sufficient for the characterization of the known peptide substrate, a similar condition and analysis was conducted for the unknown putative product of the reaction. A mass spectrum was generated, and subsequent sequencing was accomplished (Figure 19).

#	Immonium	b	b-H2O	b-NH3	a	a-H2O	b (2+)	a (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	86.10	114.09	96.08	97.06	86.10	68.09	57.55	43.55	L					7
2	76.02	217.10	199.09	200.07	189.11	171.11	109.05	95.05	C	634.29	616.29	617.27	317.65	6
3	74.06	318.15	300.14	301.12	290.15	272.14	159.57	145.58	T	531.28	513.28	514.26	266.15	5
4	70.07	415.20	397.19	398.17	387.21	369.20	208.10	194.10	P	430.24	412.23	413.21	215.62	4
5	60.04	502.23	484.22	485.21	474.24	456.23	251.62	237.62	S	333.19	315.18	316.16	167.09	3
6	129.11	658.33	640.32	641.31	630.34	612.33	329.67	315.67	R	246.16	228.15	229.13	123.58	2
7	44.05								A	90.06	72.04	73.03	45.53	1

Figure 19: Electron Spray Ionization Mass Spectrometry (ESI-MS) of putative modified peptide product from the enzymatic reaction. Peptide sample was flash-frozen directly after 16-hour reaction. Spectroscopic analysis of sample conducted at the Advanced Analysis Center in the University of Guelph by Dr. Dyanne Brewer. Ionization polarity for sample was positive. (A) Mass spectrum for the putative modified peptide generated via time of fight measurement (TOF) which was deconvoluted to show a peak mass signal at 744.36Da. (B) Mass spectra of the modified peptide for each of the amino acids overlaid on a single spectrum for sequencing. Each peak was further labeled based on its intensity and charge, shown as ymax value and b value in (C). Gray background for better clarity.

Surprisingly, the data from the mass spectrometry of the modified peptide sample had a peak

mass signal at 744.36Da. This was a mass loss of 2.01Da that was attributed to the loss of two hydrogen

atoms from the serine residue of the peptide substrate. The reaction of an FGE with a serine residue can

produce an oxoalanine or formylglycine residue by oxidation to the carbonyl group of the serine in the

peptide. This was not anticipated as the removal of a thiol group from the cysteine residue would result

in a molecular weight of 728.77Da, however the mass calculated from the cysteine remained

unchanged. This result caused for further investigation into the potential for a serine modifying FGE

instead of the previously assumed cysteine modifying function.

3.5.3.4 – Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

Further analysis of the enzymatic reaction for TDE2714 was conducted via TLC and was then

analyzed using liquid chromatography-tandem mass spectrometry (LC-MS/MS). The technique was used

in order to determine any changes in polarity already demonstrated by the peptide TLC. LC-MS/MS

analyzes small molecules and their polarity differences by liquid chromatography that is then coupled

directly to a mass spectrometer for peak analysis(141). This allowed for the direct mass analysis of peaks

found in the liquid chromatography. ESI-MS/MS was chosen for the analysis of samples that were

isolated from the TLC plate. The samples analyzed included a control sample (LCTPSRA), a crude directly

flash-frozen modified peptide sample, and a purified modified peptide sample. Each of these samples

were run through a Poroshell C18 LC-column coupled with a positive ionization electron spray mass

spectrometer. The subsequent mass spectrum was generated for all three samples (Figure 20).
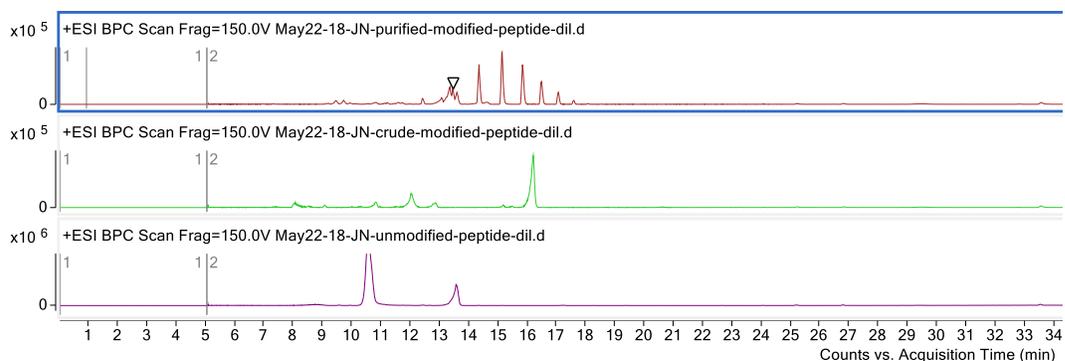
Figure 20: Mass spectrum of liquid chromatography-tandem mass spectrometry (LC-MS/MS) for the three peptide samples after the TDE2714 enzymatic reaction. Each sample was flash-frozen prior to analysis and placed through a Poroshell C18 LC-column for liquid chromatography. Mass peaks of samples were analyzed via ESI-TOF-MS and compared for retention time as well as calculated mass. Top spectrum was from the purified modified peptide sample, which had a peak of interest at 16.00 minutes. The middle spectrum was from the crude unfiltered modified peptide sample, which also showed a peak at 16.00 minutes. The bottom spectrum was from the control peptide LCTPSRA which showed a peak at 13.25 minutes which corresponded to the mass of the peptide itself.

From the LC-MS/MS data shown, the control peptide had two peaks were visualized by the spectrum. The unmodified spectra had a peak at 13.25 minutes with a calculated mass of 746.37Da, that correctly corresponded to the mass of the control peptide. Sequencing analysis also had the correct LCTPSRA sequence. Therefore, the retention time was used as a standard for the other peptides. For each of the modified peptide samples there were numerous peaks that, when analyzed, were made up of various peptide fragments and a range of masses. However, further analysis revealed that a prominent peak at 16.00 minutes had similar results to the mass of the modified peptide that had a mass loss of 2.01Da. The longer shift in the retention time agreed with the TLC results as the change in polarity would cause of the oxoalanine-containing peptides to slower out of the column. The increased retention of the modified sample was due to the interaction with the hydrophobic stationary phase in reverse-phase LC. However, the sequencing of the peaks for each of modified peptides were

inconclusive as the oxoalanine was not found on the serine residue and also an extra carbon atom was

accounted for. In light of these results, it was determined that further analysis and replication for each

of the reactions is required and will be conducted in the near future.

4. Discussion

4.1 – Thesis Objective and Goals

The objective of the research project was to characterize three targets from *T. denticola* that

showed a significant fold increase in their transcription when in a dense monoculture (or biofilm)(75).

These targets were TDE0626, TDE1701, and TDE2714 and they were previously classified as hypothetical

proteins. The research focused on identifying their role in the pathogenicity of *T. denticola* by

structurally and functionally characterizing each of the targets. This was conducted based on the

premise that since these targets were found to be more prominent in *T. denticola* when in a disease

state, they may play an important role its nutrition and colonization. For each of the targets isolation,

expression, and purification were required in order to obtain structural and functional characteristics.

The general procedure conducted for structural characterization was the use of X-ray diffraction

analysis, while the functional characterization varied for each of the targets. The data and information

gathered was able to give insight into the purpose for the upregulation of these targets in the biofilm.

4.2 – Characterization of the Target TDE0626

4.2.1 – Bioinformatics

Since TDE0626 was classified as a hypothetical protein in various databases (NCBI, UniprotKB,

etc.), the establishment of a reasonable functional hypothesis was required. This was accomplished by

the use of various bioinformatics tools that showed a homologous function to known polysaccharide

lyases, specifically ones containing a β-helix motif. However, prior to the development of the functional

hypothesis, the genomic context revealed that within the neighborhood of TDE0626 is a putative ABC

transporter ATP binding protein and a co-chaperone protein GrpE, from the NCBI database and StingDB.

The identifications of these proteins within close proximity to TDE0626 may indicate its involvement in

the degradation system of polysaccharides in *T. denticola*. Studies into polysaccharide lyases, specifically

uronic acid active enzymes, showed close association with ABC-transporter proteins in the bacterial

system(142). The degradation of N-acetyl glycosaminoglycans has been known to be highly active in

pathogenic bacteria when invading and colonizing host tissues(143). Pathogens employ the use of ABC-

transporter proteins to import the degraded host glycosaminoglycans (GAGs), that are then utilized for

pathogen nutrition, such as *T. denticola*(142). GAGs, such as heparin sulfate and hyaluronan, are

degraded via anti-β-elimination reactions that cleave the α-(1,4)-linked uronic acids to produce an

unsaturated hexuronic acid residue and a reducing end, by cell surface or extracellular polysaccharide

lyases(144). The putative function of TDE0626 was determined by the identification of a homologous

Cthe_2159 domain from *C. thermocellum*. Studies into the function of this domain revealed an ability to

bind to polysaccharides such as cellulose and polygalacturonic acid. To accomplish this binding, the

domain contained a parallel β-helix motif with the catalytic residues of lysine as the Brønstead base and

several aspartate residues for calcium binding(145). These features were reported by the crystal

structure to be similar to the features in the PL9 lyase family that is known to be able to bind to GAGs

for degradation. However, the ability for the Cthe_2159 domain to bind and hydrolyze GAGs has yet to be determined. In light of the domain identified, the conserved region mapping of TDE0626 had three putative catalytic lysine residues (Lys-140, 198, and 460) that could be involved in GAG degradation. The conserved mapping also had a number of highly conserved aspartate residues (Asp-54, 176, 221, 308, 309, 396, and 372), that in other polysaccharide lyases are known to coordinate one or more calcium ions(146). In order for TDE0626 to be either a carbohydrate binding protein or a polysaccharide lyase, the localization should be extracellular as the both cellulose and GAG degradation would be in the extracellular matrix. The discovery of a signal peptide sequence, by PSORT and SignalP, led to the hypothesis that TDE0626 was extracellularly localized. This supported the hypothesis that TDE0626 is either an upregulated carbohydrate binding protein or a polysaccharide lyase involved in the carbohydrate metabolism by *T. denticola*.

4.2.2 – Experimental Analysis of Functional Hypothesis

Due to the information gathered via bioinformatics that suggested TDE0626 to be either carbohydrate binding protein or a putative polysaccharide lyase, functional characterization was required and conducted. Prior to functional assays, TDE0626 was successfully recombinantly expressed and purified to concentrations of 10-30mg/mL with relatively low amounts of precipitation. The hypothesis of binding to polysaccharides, was tested initially by the binding of cellulose to the purified enzyme. However, based on the previous work that showed some binding to cellulose and polygalacturonic acid the conclusion made was that the Cthe_2159 domain may be more likely involved

in the degradation of uronic acids. Furthermore, in testing TDE0626 with cellulose, there was no

noticeable change in degradation. If TDE0626 is a polysaccharide lyase, similar to the PL9 family, the

cellulose linkages would not be able to be cleaved as proper binding would not occur as polysaccharide

lyases catalyze the hydrolysis of the 1-4 glyosidic bond of GAGs. The cellulose would not be able to

effectively fit into the active site of TDE0626 and therefore would be unable to be hydrolyzed. The use

of cellulose as a potential substrate was due to it being an architectural element in the biofilm matrix

and was previously suggested to have a role in the dispersal or degradation of the biofilm itself(147).

However, based on the cellulose assay, there was no indication that TDE0626 is able to degrade

cellulose. This led to the further investigation of TDE0626 being more similar to the PL9 lyase family.

A more suitable experimental analysis for TDE0626 would be measurement of the potential

substrate binding with GAGs such as hyaluronan and heparin sulfate, that contain uronic acid subunits

and may be cleaved by TDE0626. Physiologically, the presence of these GAGs would be found in higher

concentrations in host mammalian tissues and so pathogenic bacteria would be able to access the host

polysaccharides for the use in biosynthesis of essential compounds and nutrition(148). Like-wise, studies

showed that these GAGs can increase the formation of biofilms which then can progress the infection of

the pathogen(149). Within mixed biofilms produced by the pathogens such as the red complex, there

are GAGs also present called extracellular polymeric substances (EPS) that can be broken down by

polysaccharide lyases in order to promote dispersal as well as the spread of toxins deeper into the

tissues(12). In order to test this alternative function of TDE0626, the putative enzyme was treated with

isolated samples of biofilm produced by the mixed culture of the red complex. The results from the

experiment were inconclusive was there did not appear to be any detected degradation of the biofilm

by TDE0626. From the experiment, there appears to be no activity on the GAGs within the biofilm;

however, there may not have been a large number of GAGs present within the biofilm samples used.

The number of GAGs within the isolates were not measured prior to treatment with TDE0626. Further

analysis for the lyase function of TDE0626 with either purified hyaluronan or heparin sulfate will be

conducted. Likewise, treatment of biofilm isolates from the red complex with known compositions of

GAGs and other polysaccharides will be conducted.

4.2.3 – Structural Characterization of TDE0626

Unfortunately, attempts to crystallize TDE0626 were unsuccessful despite the large amounts of

protein produced. During purification, low amounts of precipitation did occur, but it was possible to

obtain high concentrations upwards to 30mg/mL. Initial crystal screening using the sitting-drop vapour

diffusion was not able to generate crystal formation in a wide range of crystallization conditions. Further

attempts to crystalize the protein was conducted by varying conditions where the protein drop seemed

to be in a metastable condition. The state of protein called the metastable zone, it is a supersaturated

state of protein in a solution in which no nucleation points may occur. It is only when the optimal

conditions are found that the condition shifts to a nucleation state and forms nuclei. This can then be

grown with more protein in the metastable zone to make larger protein crystals(150). There arose a few

conditions that were believed to be within the metastable state and so the concentrations of the

components were fluctuated. These conditions did not produce crystals, however, and so a different

approach to crystallization will need to be conducted. The addition of GAGs at low concentrations may

improve crystallization as this may allow for the rigid binding of the enzyme to a substrate and decrease

its motion within the solution(151). In addition, the addition of organic solvents may be required to aid

in the crystallization of TDE0626 by decreasing the amount of aggregation within the protein drops.

4.3 – Characterization of Target TDE1701

4.3.1 – Bioinformatics Analysis

Initial information gathered of the transcription of TDE1701 showed that there was no predicted

association to the neighbouring genes of *T. denticola*, based on StringDB. The only association with a

gene was with an upstream serine/phosphate phosphatase (*TDE1698*) however, there was no predicted

operon or gene cluster identified from the NCBI or KEGG database. Further investigation into the KEGG

database revealed two predicted paralogs in *T. denticola* which were bacteriocin-type signal sequences.

This was an interesting discovery as these sequences are known to aid in the pathogenicity of bacteria

within a biofilm infection(152). The use of bacteriocins by an organism has been known to suppress the

growth and proliferation of competitive species within a biofilm, such as the oral *streptococci ssp*(153).

Pathogens such as *T. denticola* produce these bacteriocins and secrete them into the biofilm in order to

increase the colonization of the pathogen into the subgingival plaque of a host(154). The production of

bacteriocins in pathogens require three main gene products; bacteriocin, a bacteriocin ABC-transporter,

and a bacteriocin immunity protein(154). The prediction of TDE1701 identified that it potentially could

be part of the bacteriocin ABC-transporter as a bacteriocin precursor sequence was discovered that

contained a small bacteriocin signature of two glycine residues beside each other. This signature is often

referred to as a double-glycine-type leader peptide and is involved in the secretion of the

bacteriocin(155). It is with this sequence that bacteriocin can be produced and further exported from *T.*

denticola via the use of an associated ABC-transporter protein(156). Further evidence to support the

putative bacteriocin function of TDE1701 was shown in the transcriptomic analysis, which suggested

that when *T. denticola* is in a mature biofilm, the expression of TDE1701 is increased(75). This increase

in expression of a putative bacteriocin would allow for the suppression of competitive bacteria and

therefore allow *T. denticola* to have an increased pathogenicity within the host.

The concerning information gathered from the bioinformatics analysis was that TDE1701 was

predicted to be 50% intrinsically disordered based on Phyre^2 homology modelling. This was also

highlighted by the lack of homologous proteins predicted or similar identified motifs, using other

domain searches. TDE1701 was predicted to either contain novel folds or motifs, however Phyre^2

suggested TDE1701 to be a disordered protein, and for this to be true, the protein would be difficult to

crystallize. TDE1701 as a bacteriocin-like protein could be used in bacteriocin activity that could explain

the prediction that this protein is secrete into the biofilm. The prediction was caused by the

identification of motifs that were synonymous with secreted extracellular proteins by localization

software such as CELLO and PSORT. Based on the conserved sequence mapping onto protein models, it

seemed that there was a low conservation of residues within TDE1701. However, further comparison to

known bacteriocin associated proteins suggested the double-glycine-leader sequence (used to produce

lantibiotic bacteriocins) present in TDE1701 may be able in the Gram-negative *T. denticola* in targeting

Gram-positive pathogens that can be present in the periodontal pocket(151,152).

4.3.2 – Structural Characterization of TDE1701

In spite of the prediction that TD1701 was a disordered protein, recombinant expression and

purification was successfully accomplished. The purity of the isolated protein was shown to be relatively

high and was able to yield a concentration upwards to 10mg/mL. This was an interesting result as a

disordered protein should not give such a good reading from DLS. The results of the DLS that yielded a

monodispersed sample could be caused by the aggregation of the protein in the solution by the high

concentration of 5mg/mL. This meant that the prediction of a 50% intrinsically disordered TDE1701 was

most likely not the case.

Initial crystallization of the protein was successful using sitting-drop vapour diffusion and

screened over a variety of conditions. There were two conditions which showed the formation of

protein crystals after approximately three weeks of incubation at 18°C. The first of these conditions was

found in solution E3 of MCSG-1, which contained; 0.05M magnesium chloride, 0.1M HEPES: NaOH pH

7.5, and 30% (v/v) polyethylene glycol methyl ether 550. The morphology of the protein crystals were

spherulites, which are small crystals that originate from the same nucleation point(159). These protein

crystals can yield a higher mosaicity during the X-ray diffraction analysis due to the difference in planes

from the crystals(160). However, there was also a larger protein crystal which formed separate from the

spherulites which would yield better quality and more complete diffraction data.

Similarly, there was another condition from the initial crystal screening, which yielded a large amount of protein crystals. The condition of the second crystallization was found in C10 of the PACT screen, which contained; 0.2M magnesium chloride, 0.1M HEPES: NaOH pH 7.5, and 20% w/v polyethylene glycol 6000. The formation of these crystals seemed to also be a form of spherulite as all the crystals stemmed from a single extended line. Based on the assumed diffraction quality yielded by the protein crystals, only one of the crystals was selected from the condition which was hypothesized to have a low mosaicity. This was based solely on the morphology of the crystals, however this did not mean there would be good quality diffraction.

Further attempts to replicate the crystallization of TDE1701 were unsuccessful when each condition's components were varied in a hanging drop vapour diffusion crystallization method. The larger volume and adjustments of the precipitating agents (polyethylene glycol) were not able to crystallize TDE1701. This may have been due to the difference in the diffusion rate from the hanging drop instead of the original sitting drop, which has previously been reported to affect the crystallization of a number of proteins(161). In order to further crystallize TDE1701, it was imperative that optimization of conditions be conducted using the sitting drop vapour diffusion.

For definitive functional identification, X-ray diffraction analysis of the crystalized TDE1701 was conducted using the two best selected protein crystals. The diffraction data collected by the Canadian Light Source synchrotron revealed a high-resolution dataset that was truncated at 2.40Å and a completeness of 100%. This data was processed and showed a $P2_12_12_1$ orthorhombic space group.

Molecular replacement was conducted to generate a structural model of TDE1701. Initial refinement

used Refmac5 software to generate the model TDE1701. However, based on the data generated during

molecular replacement and refinement, the R_{work}/R_{free} values for the model was 0.48/0.53. This was very

high for a refined structural model and shows that the data generated by the model building did not

match the diffraction data(130). It was therefore determined that the structural model for TDE1701 was

not an accurate representation of the true protein structure. Another method which will be explored is

the use of selenomethionine containing TDE1701 crystals for MAD experimental phase determination,

rather than molecular replacement.

Functional analysis of the purified TDE1701 sample has not yet been conducted. However, it has

been hypothesized that testing for the presence of TDE1701 within a mature mixed culture biofilm from

the red complex may give indication that it is a bacteriocin-associated factor. The testing will be

conducted similar to the testing conducted by tailocins, which are high molecular weight bacteriocins

from phages(162).

4.4 – Characterization of Target TDE2714

4.4.1 – Bioinformatics

The development of a working hypothesis for the function and role of TDE2714 in *T. denticola*

was accomplished via bioinformatics. The initial data gathered revealed three putative roles for the

function of TDE2714 that were a calcium-dependent carbohydrate binding protein, a large diversity-

generating retroelements (DGR), or a hemolytic formylglycine-generating enzyme. In order to best

identify the true function of TDE2714, more context would be required via further use of bioinformatics tools.

Based on the data collected from the KEGG database, there were 16 paralogs identified in *T. denticola* that all had a putative formylglycine-generating enzyme (FGE) function. Since there are quite a few predicted FGEs in *T. denticola*, it is imperative to look to this as being the potential function of TDE2714, as sequence alignments with the paralogs revealed a number of conserved signatures in TDE2714 and the 16 others that are unique to FGEs. These signatures include, the RLPTXAEWE region, highly conserved GN region, and a cluster of cysteines near the C-terminal. The known role of FGEs within both Eukaryotes and Prokaryotes is to modify either an active site cysteine or serine residue within a type I sulfatase(88). These sulfatases are important for bacteria due to their involvement in scavenging sulfates as a source of sulfur(137). This has been known to occur within the endoplasmic reticulum of Eukaryotes and either the cytoplasm or periplasm in Prokaryotes(163). However, based on the results from CELLO and PSORT for TDE2714, the predicted localization was extracellular. This means that there may exist other functions for FGEs in *T. denticola*.

Cystalysin is a hemolytic protein expressed in *T. denticola* that has been known to generate H_2S extracellularly as a by-product of producing pyruvate from cysteine residues of blood cells(164). This generation of H_2S within infected host tissues has been shown to be in high concentrations and exhibits hemolytic properties similar to hemolysins(164). The hemolysis of tissues can be important for *T. denticola* as it may aid in the further colonization into host tissues, as well as the access to nutrients.

Interestingly, the formation of a formylglycine residue from a cysteine is also known to generate H_2S by

the removal of a thiol group(165). Based on the predicted FGE function of TDE2714, it may be possible

that this enzyme also exhibits hemolytic activity on host tissues, as it is exported from the bacteria into

the surrounding tissues. Further evidence for this was the identification of an encoded hemolysin

(*Tde2410*) that shares genetic context with *Tde2714* based on the StringDB database. This suggests that

similar associated functions may be present or similar involvement in the pathogenicity of the bacteria.

Homology modelling of TDE2714 using software, such as Phyre^2, suggested two alternative

functions for the protein that were either a diversity-generating retroelement (DGR) or a calcium-

dependent carbohydrate binding protein. The sole link between these two predictions was the *T.*

denticola Variable Protein 1, which has been found to contain a large C-type lectin domain which bears

resemblance to bacteriophage DGRs(166). The primary sequence identity was high at 45% identical

which represents a high likelihood that either TDE2714 possesses a similar function and role or contains

similar structural elements such as motifs. The identification of these DGRs within bacteria have been

known to aid in the mutations of surface proteins which can lead to massive biophysical changes in the

proteins. This utilized by the bacteria to avoid detection from the host immune system and increase the

pathogenicity of the bacteria(167). The C-type lectin domain can be used as a scaffold for massive

mutations for immunoevasion(106). This may be explained in TDE2714 as it was predicted to be

extracellularly localized which can then be mutated to aid in immunoevasion. However, based on

conserved residue mapping of closely related proteins from all domains of life showed a highly

conserved region that is present in all FGE-sulfatase modifying enzymes (RLPTQAEWE). The putative

function of TDE2714 was a hemolytic formylglycine generating enzyme that bears structural

resemblance to diversity generating retroelements.

The results from the site-directed mutagenesis of TDE2714 to the C637A mutant were

perplexing. Based on multiple sequence alignments and conserved residue mapping, the active site

cysteines should have been located within the same proximity as previously identified FGEs, such as

2AIK from *S. coelicolor*. For all cysteine modifying FGEs, the catalytic residues have been reported to be

two cysteines that form a disulfide bridge and are involved in the oxidative removal of a thiol from an

active site cysteine in a type I sulfatase(107). The change of the Cys-637 to Ala-637 should have reduced

or eliminated the function of TDE2714 on the peptide substrate -LCTPSRA-. However, the same non-

polar product was visualized for the mutant after incubation and TLC. This meant that the C637 residue

is not catalytically important and the other cysteines are potentially involved in catalysis. Investigation

into generating mutants that have all the cysteines changed to alanine residues is underway.

4.4.2 – Functional Characterization of TDE2714

The functional analysis of TDE2714 was conducted with two different cases in mind, either

TDE2714 is a carbohydrate-binding protein, or it is an formylglycine-generating enzyme. These two cases

for the function were determined by the fold-recognition software and homology modeling of TDE2714.

The first case was tested that TDE2714 is a carbohydrate-binding protein as it was homologous to a C-

type lectin protein with a sequence identity of 45%. The initial experimentation of TDE2714 were unable

to confirm the first case. This was seen in the C-PAGE results were there was no identified mobility change in TDE2714 when incubated with polygalacturonic acid. If TDE2714 were able to bind to the polysaccharides, there would have been a conformational change in the enzyme that would had affected its mobility through the gel, by making a more rigid condensed structure. This was not seen in the results and so the putative carbohydrate binding of TDE2714 was dismissed. Further investigation may be conducted for this function however, the results point towards TDE2714 being a formylglycine-generating enzyme.

For testing the hypothesis of TDE2714 being an FGE, a peptide substrate was generated that contained the sulfatase motif sequence of -LCTPSRA-(168). This peptide sequence was designed based on the most studied and well-characterized type of FGE, which is a cysteine-modifying FGE in both Eukaryotes and Prokaryotes(169). The cysteine at position 2 in the peptide was anticipated to be converted to a 3-oxoalanine residue (formylglycine) by oxidation with TDE2714. Initial testing in a reduced environment generated a noxious rotten eggs odor, which was suspected to be from the production of H_2S gas. This however was not able to be confirmed as the gas and the smell could have been contributed by DTT within the reaction. Visual quantification of the potential enzymatic reaction was conducted by a peptide TLC protocol in which the same testing with the peptide substrate was done and spotted on a TLC plate. The anticipated spots of the enzyme and the substrate revealed a third spot was identified to be less polar than the substrate. Based on this, it was plausible to assert that the oxidation reaction of a cysteine to a formylglycine was successful based on the product being less polar.

The removal of the two hydrogens or thiol group of the peptide would have yielded a less polar product

or a formyl group its place. The third spot most likely was product generated by the enzyme and this was

due to the results of a time-course TLC. Over the course of time, there was an increase in the intensity of

the non-polar spot and gave indications that this may be caused by the enzyme turning over more

product. However, enzymatic kinetics were unable to be calculated based on the intensities of the spots

on the TLC plate, but these results were more encouraging in comparison to the carbohydrate binding

assays.

The results from the ESI-MS for both the modified and the unmodified peptides were able to

give an expected substrate molar mass of 746.37g/mol was detected and sequenced to show the correct

LCTPSRA sequence. This was encouraging as the methodology was able to give a reproducible control.

However, using the same parameters of mass spectrometry for the modified peptide product, a new

molar mass was detected of 744.36g/mol which was a 2.01g/mol difference from the substrate. The loss

of the equivalent of two protons could have been anticipated that there were two types of mechanisms

for Prokaryotic FGEs(170). The widely studied cysteine modifying that removes a thiol group from an

active site cysteine, as well as a metal dependent serine modifying FGE similar to AtsB from *K.*

pneumoniae. This enzyme uses two auxiliary iron-sulfur sites and a radical s-adenosylmethionine (SAM)

to catalyze the oxidative removal of two protons of a serine residue in a sulfatase. The sequencing

results of the mass spectrum for the modified peptide showed modified sequence of -LCTPSRA- as the

serine residue on modified peptide substrate was shown to have lost 2.01g/mol. These results were not

anticipated as the experimental design was to modify the cysteine residue and was based on the

similarities of TDE2714 to cysteine modifying enzymes. A cysteine modification would have yielded a

mass difference of 18.11g/mol by the removal of a thiol group. Based on the data, TDE2714 as a cysteine

modifying FGE was dismissed.

As stated before, Prokaryotic FGEs can also modify active site serine residues within type I

sulfatases(171)(170). The reaction has the ability to produced H_2 gas instead of H_2S during the oxidation

of the serine(172). The resultant mass difference would be 2.02g/mol. Using iron-sulfur prediction

software such as MetalPredator(173), it was determined that TDE2714 did not contain the known

residues required for iron-sulfur binding as TDE2714 lacked the sufficient residues to coordinate with

iron-sulfur clusters. This caused for the hypothesis, that TDE2714 is not an iron-sulfur containing FGE

however it has the capability to modify serine residues. However, H_2 gas has been detected in affected

periodontal pockets and has been known to exhibit similar hemolytic activity(174)(175). This give

credence to the potentially new claim that there exists a seine modifying FGE that operates without the

use of iron-sulfur clusters and a radical SAM. Based on the information gather thus far, TDE2714 may be

an extracellular hemolytic serine FGE in *T. denticola*.

Since *T. denticola* is an obligate anaerobe, the proposed mechanism for cysteine modifying FGEs

cannot be used for TDE2714 as there would be no oxygen available for the enzyme to use. The proposed

mechanism for TDE2714 would potentially be a hybrid of the two known mechanisms, such as the thiol

or hydroxyl binding of a substrate cysteine or serine, followed by a nucleophilic attack on the proton of

side chain carbon by an adenosine-based free radical. This then could be utilized to make 5'-

deoxyadenosine and new radical can either form directly into the 3-oxoalanine or a thioaldehyde

intermediate, then to the 3-oxoalanine (Figure A11 in Appendix).

In light of this new hypothesis of TDE2714 that there may have been errors in the data analyzed,

as the experiment was designed for cysteine active FGEs and not serine active, and so further attempts

were conducted using liquid chromatography-tandem mass spectrometry (LC-MS/MS). However, the

data was only conclusive for the peptide substrate that showed a similar mass of 746.37g/mol to the

ESI-MS experiments. The modified peptide samples sent showed a peak at had a longer shift in the

elution time compared to the unmodified substrate. This was expected for the conversion of a cysteine

or serine to an oxoalanine, as the peptide would become less polar and be able to elute slower out of

the reverse-phase column. However, the MS/MS spectra and sequencing revealed an extra carbon atom

that was calculated for the modified peptide sample at the 3-oxoalanine site of the serine. This unusual

addition of a carbon atom is a cause of concern as there may have been more than one unexpected

reaction that may have occurred during either the enzymatic reaction or the ESI-MS. Based on this,

further analysis will be conducted to correctly identify the potentially modified peptide product via 2-

dimensional nuclear magnetic resonance (2-D NMR) such as employing correlated spectroscopy (COSY)

NMR.

4.4.3 – Structural Characterization of TDE2714

In order to determine the correct function of TDE2714, structural characterization was also required to generate a working model that could be used in comparison to the predicted homologous proteins. This was accomplished by the successful recombinant expression and purification of TDE2714 which was able to yield concentrations ranging from 10mg/mL to 25mg/mL with high purity. The initial crystallization showed the formation of small plate crystals which were determined not to be ideal for X-ray diffraction analysis based on the assumed high mosaicity during the diffraction. This protein crystallization condition (0.2M $NaNO_3$, 20% (w/v) polyethylene glycol, and 50mM Tris-HCl pH 8.0) was optimized by varying the concentration of PEG from 19.6-20.2%. The resultant crystals were shown to be larger and bulkier, which was found to be promising for X-ray diffraction analysis. The diffraction analysis for the optimized protein crystals resulted in a 2.71Å resolution with a calculated primitive monoclinic $P2_1$ space group for the asymmetric unit. However, based on the statistics generated during processing of the diffraction data, the I/σI value (intensity/error in intensity) was shown to be 2.2 in the outer shell which was low for the calculated space group. This meant that further diffraction analysis was required such phase determination using selenomethionine crystals. These crystals were successfully grown in the same optimized conditions for TDE2714 and also sent for selenium-MAD analysis. The data obtained did not successfully give sufficient phase information for the generation of a working structural model of TDE2714 and so further crystallization and diffraction will be required.

5. Conclusion

For this research project, three upregulated gene products, TDE0626, TDE1701, and TDE2714, from *T. denticola* were investigated for their role in aiding the pathogen to progress chronic periodontal disease. The main objectives of the research were to structurally and functionally characterize these gene products in order to better understand how *T. denticola* is able to progress chronic periodontitis. Not all of the main objectives were accomplished for all of the targets in the study, however the results of each of the targets was able to satisfy some portions of the objectives.

Initial bioinformatics analysis suggested the putative functions of the targets to be a N-acetyl-glycosaminoglycan degrading polysaccharide lyase, a bacteriocin-like protein, and a formylglycine-generating enzyme, respectfully. Each of these suggested functions have been known to play a role in the pathogenicity of various microorganisms and this was investigated for *T. denticola*. Based on the mixed biofilm degradation assays for TDE0626, no change from the treated and untreated samples was found. Further assays will be conducted to confirm whether TDE0626 can degrade N-acetyl-glycosaminoglycans solely or within a biofilm. The mass spectrometry for the reaction of TDE2714 was able to modify a sulfatase consensus sequence (-LCTPSRA-) to an oxoalanine residue at the serine site. This was not anticipated based on the general design of the assay however further validation of the results will be conducted via LC-MS/MS in the absence of salts. Furthermore, TDE2714 will be treated

with a variety of peptide substrates that lack either cysteine or serine residues or both completely. The

analysis will also be conducted by the generation of mutants at the conserved '-RLPTQAEWE-' region of

TDE2714. The development of a functional assay for TDE1701 will be conducted after more structural

data can found. Based on the functional results, the function of TDE0626 was able to be better

hypothesized as a polysaccharide lyase that may be involved in the nutrition of *T. denticola*. Whereas,

the function of TDE2714 was able to be identified as a formylglycine-generating enzyme that can

catalyze the oxidative conversion of either a cysteine or serine residue. Unfortunately, the function of

TDE1701 has yet to be determined however, the initial results from bioinformatics suggests a

bacteriocin function.

Two of the three targets (TDE1701 and TDE2714), were successfully crystallized, and initial X-ray

diffraction analysis was conducted. Both of the targets showed relatively good crystallographic data

based on data collection and allowed for attempts for generate structural models. For TDE1701, the

data showed a high resolution of 2.40Å in the outer shell with an acceptable I/σI of 2.2 and CC$^{1/2}$ of

0.823. For TDE2714, the data showed a high resolution of 2.75Å in the outer shell and I/σI of 2.2.

However, in both cases the R_{merge} values were close to 1.0 (1.089 for TDE1701 and 0.819 for TDE2714)

which seemed to be of concern for the datasets. Molecular replacement was unable to generate an

acceptable structure which could be representative of the 'true' protein structure. Selenomethionine

anomalous diffraction experiments was successfully conducted for TDE2714; however, the data was

unable to improve the prior collected data for TDE2714. Attempts to grow selenomethionine containing

crystals or improve the quality of crystals will be conducted for both TDE1701 and TDE2714.

Crystallization attempts will also be conducted for TDE0626 for structural characterization. The structure

of each of the enzymes has yet to be determined however for TDE1701 and TDE2714, they are on track

to being structurally characterized.

6. References

1. Loe, H. (1993) Periodontal Disease. *Periodontal Dis. sixth Complicat. diabetes Mellit.* **16**, 329–334

2. Dye, B. A. (2012) Global periodontal disease epidemiology. *Periodontol. 2000*. 10.1111/j.1600-0757.2011.00413.x

3. Flemmig, T. F. (1999) Periodontitis. in *Annals of periodontology / the American Academy of Periodontology*, pp. 32–38, **4**, 32–38

4. Monroe, D. (2007) Looking for chinks in the armor of bacterial biofilms. *PLoS Biol.* 10.1371/journal.pbio.0050307

5. Palmer, J., Flint, S., and Brooks, J. (2007) Bacterial cell attachment, the beginning of a biofilm. *J. Ind. Microbiol. Biotechnol.* 10.1007/s10295-007-0234-4

6. Xavier, J. B., and Foster, K. R. (2007) Cooperation and conflict in microbial biofilms. *Proc. Natl. Acad. Sci.* 10.1073/pnas.0607651104

7. Gupta, P., Sarkar, S., Das, B., Bhattacharjee, S., and Tribedi, P. (2016) Biofilm, pathogenesis and prevention—a journey to break the wall: a review. *Arch. Microbiol.* 10.1007/s00203-015-1148-6

8. Solano, C., Echeverz, M., and Lasa, I. (2014) Biofilm dispersion and quorum sensing. *Curr. Opin. Microbiol.* **18**, 96–104

9. O'Connell, H. A., Kottkamp, G. S., Eppelbaum, J. L., Stubblefield, B. A., Gilbert, S. E., and Gilbert, E. S. (2006) Influences of biofilm structure and antibiotic resistance mechanisms on indirect pathogenicity in a model polymicrobial biofilm. *Appl. Environ. Microbiol.* **72**, 5013–5019

10. Glaros, and Laffey (2012) Situating The Residents. *J. Film Video*. 10.5406/jfilmvideo.64.1-2.0072

11. Ramsey, M. M., Rumbaugh, K. P., and Whiteley, M. (2011) Metabolite cross-feeding enhances virulence in a model polymicrobial infection. *PLoS Pathog.* 10.1371/journal.ppat.1002012

12. Kaplan, J. B. (2010) Biofilm Dispersal. *J. Dent. Res.* 10.1177/0022034509359403

13. Landini, P., Antoniani, D., Burgess, J. G., and Nijland, R. (2010) Molecular mechanisms of compounds affecting bacterial biofilm formation and dispersal. *Appl. Microbiol. Biotechnol.* **86**, 813–823

14. Karatan, E., and Watnick, P. (2009) Signals, Regulatory Networks, and Materials That Build and Break Bacterial Biofilms. *Microbiol. Mol. Biol. Rev.* 10.1128/MMBR.00041-08

15. Pihlstrom, B. L., Michalowicz, B. S., and Johnson, N. W. (2005) Periodontal diseases. in *Lancet*, 10.1016/S0140-6736(05)67728-8

16. Kinane, D. F. (2001) Causation and pathogenesis of periodontal disease. *Periodontol. 2000*. 10.1034/j.1600-0757.2001.22250102.x

17. Van Strydonck, D. A. C., Slot, D. E., Van Der Velden, U., and Van Der Weijden, F. (2012) Effect of a chlorhexidine mouthrinse on plaque, gingival inflammation and staining in gingivitis patients: A systematic review. *J. Clin. Periodontol.* 10.1111/j.1600-051X.2012.01883.x

18. Masanja, I. M., and Mumghamba, E. G. S. (2004) Knowledge on gingivitis and oral hygiene practices among secondary school adolescents in rural and urban Morogoro, Tanzania. *Int. J. Dent. Hyg.* 10.1111/j.1601-5037.2004.00096.x

19. Okada, H., and Murakami, S. (1998) Cytokine expression in periodontal health and disease. *Crit. Rev. Oral Biol. Med.* 10.1177/10454411980090030101

20. Armitage, G. C. (1999) Development of a Classification System for Periodontal Diseases and Conditions. *Ann. Periodontol.* 10.1902/annals.1999.4.1.1

21. Kinane, D. F., and Mombelli, A. (2011) *Periodontal disease*, 10.1159/isbn.978-3-8055-9834-7

22. Armitage, G. C., Cullinan, M. P., and Seymour, G. J. (2010) Comparative biology of chronic and aggressive periodontitis: Introduction. *Periodontol. 2000*. 10.1111/j.1600-0757.2010.00359.x

23. Takedachi, M., Sawada, K., Yamamoto, S., Ozasa, M., Shimabukuro, Y., Kitamura, M., and Murakami, S. (2013) Periodontal tissue regeneration by transplantation of adipose tissue-derived stem cells. *J. Oral Biosci.* 10.1016/j.job.2013.04.004

24. Sunde, P. T., Olsen, I., Debelian, G. J., and Tronstad, L. (2002) Microbiota of periapical lesions refractory to endodontic therapy. *J. Endod.* 10.1097/00004770-200204000-00011

25. Fujii, R., Saito, Y., Tokura, Y., Nakagawa, K. I., Okuda, K., and Ishihara, K. (2009) Characterization of bacterial flora in persistent apical periodontitis lesions. *Oral Microbiol. Immunol.* 10.1111/j.1399-302X.2009.00534.x

26. Socransky, S. S., Haffajee, a D., Cugini, M. a, Smith, C., and Kent, R. L. (1998) Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* **25**, 134–144

27. Ximénez-Fyvie, L. a, Haffajee, a D., and Socransky, S. S. (2000) Comparison of the microbiota of supra- and subgingival plaque in health and periodontitis. *J. Clin. Periodontol.* **27**, 648–57

28. López, N. J., Socransky, S. S., Da Silva, I., Japlit, M. R., and Haffajee, A. D. (2004) Subgingival Microbiota of Chilean Patients With Chronic Periodontitis. *J. Periodontol.* 10.1902/jop.2004.75.5.717

29. Walley, S. (2012) Essential microbiology for dentistry, 4th edition. *Br. Dent. J.* 10.1038/sj.bdj.2012.309

30. Needleman, I. (2016) The Good Practitioner's Guide to Periodontology. *Br. Soc. Periodontol.*

31. Suzuki, N., Yoneda, M., and Hirofuji, T. (2013) Mixed red-complex bacterial infection in periodontitis. *Int. J. Dent.* 10.1155/2013/587279

32. Romero, D., Traxler, M. F., López, D., and Kolter, R. (2011) Antibiotics as signal molecules. *Chem.*

Rev. 10.1021/cr2000509

33. Hoffman, L. R., D'Argenio, D. A., MacCoss, M. J., Zhang, Z. Y., Jones, R. A., and Miller, S. I. (2005) Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature*. **436**, 1171–1175

34. Stewart, P. S. (2002) Mechanisms of antibiotic resistance in bacterial biofilms. *Int. J. Med. Microbiol.* 10.1078/1438-4221-00196

35. Høiby, N., Ciofu, O., Krogh Johansen, H., Song, Z.-J., Moser, C., Østrup Jensen, P., Molin, S., Givskov, M., Tolker-Nielsen, T., and Bjarnsholt, T. (2011) The clinical impact of bacterial biofilms. *Int J Oral Sci Int. J. Oral Sci.* **3**, 55–65

36. Drenkard, E., and Ausubel, F. M. (2002) Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature*. 10.1038/416740a

37. Xiao, E., Mattos, M., Vieira, G. H. A., Chen, S., Corrêa, J. D., Wu, Y., Albiero, M. L., Bittinger, K., and Graves, D. T. (2017) Diabetes Enhances IL-17 Expression and Alters the Oral Microbiome to Increase Its Pathogenicity. *Cell Host Microbe*. 10.1016/j.chom.2017.06.014

38. De Kievit, T. R., Parkins, M. D., Gillis, R. J., Srikumar, R., Ceri, H., Poole, K., Iglewski, B. H., and Storey, D. G. (2001) Multidrug efflux pumps: Expression patterns and contribution to antibiotic resistance in Pseudomonas aeruginosa biofilms. *Antimicrob. Agents Chemother.* 10.1128/AAC.45.6.1761-1770.2001

39. Kvist, M., Hancock, V., and Klemm, P. (2008) Inactivation of efflux pumps abolishes bacterial biofilm formation. *Appl. Environ. Microbiol.* 10.1128/AEM.01310-08

40. Banin, E., Brady, K. M., and Greenberg, E. P. (2006) Chelator-induced dispersal and killing of Pseudomonas aeruginosa cells in a biofilm. *Appl. Environ. Microbiol.* 10.1128/AEM.72.3.2064-2069.2006

41. Graeber, J. J. (2015) Scaling and root planing. *J. Am. Dent. Assoc.* 10.1016/j.adaj.2015.10.007

42. Sharma, R., Hegde, V., Siddharth, M., Hegde, R., Manchanda, G., and Agarwal, P. (2014) Endodontic-periodontal microsurgery for combined endodontic-periodontal lesions: An overview. *J. Conserv. Dent.* 10.4103/0972-0707.144571

43. Cohn, S. A. (2005) Treatment choices for negative outcomes with non-surgical root canal treatment: non-surgical retreatment vs. surgical retreatment vs. implants. *Endod. Top.* 10.1111/j.1601-1546.2005.00163.x

44. Del Fabbro, M., Corbella, S., Sequeira-Byron, P., Tsesis, I., Rosen, E., Lolato, A., and Taschieri, S. (2016) Endodontic procedures for retreatment of periapical lesions. *Cochrane Database Syst. Rev.* 10.1002/14651858.CD005511.pub3

45. Pogrel, M. A. (2007) Damage to the inferior alveolar nerve as the result of root canal therapy. *J. Am. Dent. Assoc.* 10.14219/jada.archive.2007.0022

46. Allen, R. K., Newton, C. W., and Brown, C. E. (1989) A statistical analysis of surgical and nonsurgical endodontic retreatment cases. *J. Endod.* 10.1016/S0099-2399(89)80221-3

47. Wu, H., Song, Z., Hoiby, N., and Givskov, M. (2004) Quorum sensing in Gram-negative bacteria. *Prog. Nat. Sci.* 10.1080/10020070412331343661

48. Li, Y. H., and Tian, X. L. (2016) Quorum Sensing and Bacterial Social Interactions in Biofilms: Bacterial Cooperation and Competition. in *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*, 10.1002/9781119004813.ch116

49. Chan, K. G., Liu, Y. C., and Chang, C. Y. (2015) Inhibiting N-acyl-homoserine lactone synthesis and quenching Pseudomonas quinolone quorum sensing to attenuate virulence. *Front. Microbiol.* 10.3389/fmicb.2015.01173

50. Kalia, V. C. (2013) Quorum sensing inhibitors: An overview. *Biotechnol. Adv.* **31**, 224–245

51. Clatworthy, A. E., Pierson, E., and Hung, D. T. (2007) Targeting virulence: a new paradigm for antimicrobial therapy. *Nat. Chem. Biol.* **3**, 541–548

52. Absalon, C., van Dellen, K., and Watnick, P. I. (2011) A communal bacterial adhesin anchors biofilm and bystander cells to surfaces. *PLoS Pathog.* 10.1371/journal.ppat.1002210

53. Szczesny, P., Linke, D., Ursinus, A., Bär, K., Schwarz, H., Riess, T. M., Kempf, V. A. J., Lupas, A. N., Martin, J., and Zeth, K. (2008) Structure of the head of the bartonella adhesin BadA. *PLoS Pathog.* 10.1371/journal.ppat.1000119

54. Nummelin, H., Merckel, M. C., Leo, J. C., Lankinen, H., Skurnik, M., and Goldman, A. (2004) The Yersinia adhesin YadA collagen-binding domain structure is a novel left-handed parallel β-roll. *EMBO J.* 10.1038/sj.emboj.7600100

55. Lazar Adler, N. R., Dean, R. E., Saint, R. J., Stevens, M. P., Prior, J. L., Atkins, T. P., and Galyov, E. E. (2013) Identification of a predicted trimeric autotransporter adhesin required for biofilm formation of Burkholderia pseudomallei. *PLoS One*. 10.1371/journal.pone.0079461

56. Morrison, S. L., Johnson, M. J., Herzenberg, L. A., and Oi, V. T. (1984) Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc. Natl. Acad. Sci.* 10.1073/pnas.81.21.6851

57. Shannon, O., Mörgelin, M., and Rasmussen, M. (2010) Platelet activation and biofilm formation by Aerococcus urinae, an endocarditis-causing pathogen. *Infect. Immun.* 10.1128/IAI.00469-10

58. Monari, C., Kozel, T. R., Paganelli, F., Pericolini, E., Perito, S., Bistoni, F., Casadevall, A., and Vecchiarelli, A. (2006) Microbial Immune Suppression Mediated by Direct Engagement of Inhibitory Fc Receptor. *J. Immunol.* 10.4049/jimmunol.177.10.6842

59. Deretic, V., Singh, S., Master, S., Harris, J., Roberts, E., Kyei, G., Davis, A., de Haro, S., Naylor, J., Lee, H. H., and Vergne, I. (2006) Mycobacterium tuberculosis inhibition of phagolysosome biogenesis and autophagy as a host defence mechanism. *Cell. Microbiol.* 10.1111/j.1462-5822.2006.00705.x

60. Mariotti, S., Teloni, R., Iona, E., Fattorini, L., Romagnoli, G., Gagliardi, M. C., Orefici, G., and Nisini, R. (2004) Mycobacterium tuberculosis diverts alpha interferon-induced monocyte differentiation from dendritic cells into immunoprivileged macrophage-like host cells. *Infect. Immun.* 10.1128/IAI.72.8.4385-4392.2004

61. Huang, R., Li, M., and Gregory, R. L. (2011) Bacterial interactions in dental biofilm. *Virulence*. 10.4161/viru.2.5.16140

62. Barkal, L. J., Theberge, A. B., Guo, C. J., Spraker, J., Rappert, L., Berthier, J., Brakke, K. A., Wang, C.

C. C., Beebe, D. J., Keller, N. P., and Berthier, E. (2016) Microbial metabolomics in open microscale platforms. *Nat. Commun.* 10.1038/ncomms10610

63. Sutherland, I. W. (1995) Polysaccharide lyases. *FEMS Microbiol. Rev.* 10.1111/j.1574-6976.1995.tb00179.x

64. Michaud, P., Da Costa, A., Courtois, B., and Courtois, J. (2003) Polysaccharide lyases: Recent developments as biotechnological tools. *Crit. Rev. Biotechnol.* 10.1080/07388550390447043

65. Jeon, J. H., Suh, H. N., Kim, M. O., Ryu, J. M., and Han, H. J. (2014) Glucosamine-Induced OGT Activation Mediates Glucose Production Through Cleaved Notch1 and FoxO1, Which Coordinately Contributed to the Regulation of Maintenance of Self-Renewal in Mouse Embryonic Stem Cells. *Stem Cells Dev.* 10.1089/scd.2013.0583

66. Frey, J., and Falquet, L. (2015) Patho-genetics of Clostridium chauvoei. *Res. Microbiol.* 10.1016/j.resmic.2014.10.013

67. Karunakaran, T., and Holt, S. C. (1994) Cloning and expression of hemolysin genes from Treponema denticola strains ATCC 35404 (TD-4) and human clinical isolate GM-1 in Escherichia coli. *Microb. Pathog.* 10.1006/mpat.1994.1034

68. Spellerberg, B. (2000) Pathogenesis of neonatal Streptococcus agalactiae infections. *Microbes Infect.* 10.1016/S1286-4579(00)01328-9

69. Song, L., Hobaugh, M. R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J. E. (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science*. **274**, 1859–66

70. Greabu, M., Totan, A., Miricescu, D., Radulescu, R., Virlan, J., and Calenic, B. (2016) Hydrogen Sulfide, Oxidative Stress and Periodontal Diseases: A Concise Review. *Antioxidants*. 10.3390/antiox5010003

71. Chu, L., Ebersole, J. L., Kurzban, G. P., and Holt, S. C. (1997) Cystalysin, a 46-kilodalton cysteine desulfhydrase from Treponema denticola, with hemolytic and hemoxidative activities. *Infect. Immun.*

72. Bennedsen, M., Stuer-Lauridsen, B., Danielsen, M., and Johansen, E. (2011) Screening for antimicrobial resistance genes and virulence factors via genome sequencing. *Appl. Environ. Microbiol.* **77**, 2785–2787

73. Wu, H. J., Wang, A. H. J., and Jennings, M. P. (2008) Discovery of virulence factors of pathogenic bacteria. *Curr. Opin. Chem. Biol.* **12**, 93–101

74. Sela, M. N. (2001) Role of Treponema denticola in periodontal diseases. *Crit. Rev. Oral Biol. Med.* 10.1177/10454411010120050301

75. Mitchell, H. L., Dashper, S. G., Catmull, D. V., Paolini, R. A., Cleal, S. M., Slakeski, N., Tan, K. H., and Reynolds, E. C. (2010) Treponema denticola biofilm-induced expression of a bacteriophage, toxin-antitoxin systems and transposases. *Microbiology*. **156**, 774–788

76. Bertani, L. E. (1968) Abortive induction of bacteriophage P2. *Virology*. 10.1016/0042-6822(68)90119-0

77. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect.*

D Biol. Crystallogr. 10.1107/S0907444910048675

78. Odersky, M., Spoon, L., and Venners, B. (2011) Scala. *URL http//blog. typesafe. com/why-scala (last accessed 2012-08-28)*. 10.1017/S0967199411000700

79. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 10.1107/S0907444910007493

80. Rosen, G., Sela, M. N., Naor, R., Halabi, A., Barak, V., and Shapira, L. (1999) Activation of murine macrophages by lipoprotein and lipooligosaccharide of Treponema denticola. *Infect. Immun.*

81. NCBI Resource Coordinators (2017) Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 10.1093/nar/gkw1071

82. Hartl, F. U. (1996) Molecular chaperones in cellular protein folding. *Nature*. 10.1038/381571a0

83. Sheridan, K. J., Lechner, B. E., Keeffe, G. O., Keller, M. A., Werner, E. R., Lindner, H., Jones, G. W., Haas, H., and Doyle, S. (2016) Ergothioneine Biosynthesis and Functionality in the Opportunistic Fungal Pathogen, Aspergillus fumigatus. *Sci. Rep.* 10.1038/srep35306

84. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 10.1093/nar/27.1.29

85. Koonin, E. V. (2005) Orthologs, Paralogs, and Evolutionary Genomics. *Annu. Rev. Genet.* 10.1146/annurev.genet.39.073003.114725

86. Schillinger, U., and Lücke, F. K. (1989) Antibacterial activity of Lactobacillus sake isolated from meat. *Appl. Environ. Microbiol.* 10.1016/S0168-1605(03)00051-5

87. Nilsen, T., Nes, I. F., and Holo, H. (2003) Enterolysin A, a cell wall-degrading bacteriocin from Enterococcus faecalis LMG 2333. *Appl. Environ. Microbiol.* 10.1128/AEM.69.5.2975-2984.2003

88. Dierks, T., Lecca, M. R., Schmidt, B., and Von Figura, K. (1998) Conversion of cysteine to formylglycine in eukaryotic sulfatases occurs by a common mechanism in the endoplasmic reticulum. *FEBS Lett.* **423**, 61–65

89. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2014) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-30

90. Close, D. W., D'angelo, S., and Bradbury, A. R. M. (2014) A new family of β-helix proteins with similarities to the polysaccharide lyases. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 10.1107/S1399004714015934

91. Garg, V. K., Avashthi, H., Tiwari, A., Jain, P. A., Ramkete, P. W. R., Kayastha, A. M., and Singh, V. K. (2016) MFPPI – Multi FASTA ProtParam Interface. *Bioinformation*. **12**, 74–77

92. Kozlowski, L. P. (2017) Proteome-pI: Proteome isoelectric point database. *Nucleic Acids Res.* 10.1093/nar/gkw978

93. Kirkwood, J., Hargreaves, D., O'keefe, S., and Wilson, J. (2015) Analysis of crystallization data in the Protein Data Bank. *Acta Crystallogr. Sect. Struct. Biol. Commun.* 10.1107/S2053230X15014892

94. Walden, H. (2010) Selenium incorporation using recombinant techniques. *Acta Crystallogr. Sect.*

D Biol. Crystallogr. 10.1107/S0907444909038207

95.   Yu, C.-S., Lin, C.-J., and Hwang, J.-K. (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 10.1110/ps.03479604

96.   Nakano, Y., Shibata, Y., Kawada, M., Kojima, M., Fukamachi, H., Shibata, Y., Okano, S., Matsushima, K., Abiko, Y., and Yamashita, Y. (2005) A searchable database for proteomes of oral microorganisms. *Oral Microbiol. Immunol.* 10.1111/j.1399-302X.2005.00235.x

97.   Yu, C.-S., Chen, Y.-C., Lu, C.-H., and Hwang, J.-K. (2006) Prediction of protein subcellular localization. *Proteins*. **64**, 643–51

98.   Romaní, A. M., Fund, K., Artigas, J., Schwartz, T., Sabater, S., and Obst, U. (2008) Relevance of polymeric matrix enzymes during biofilm formation. *Microb. Ecol.* 10.1007/s00248-007-9361-8

99.   The UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res.* 10.1093/nar/gku989

100.   Blunsom, P. (2004) Hidden Markov Models. *Lect. notes, August*. 10.1007/978-1-61779-400-1_22

101.   Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447-52

102.   Schwartz, A. S., Yu, J., Gardenour, K. R., Finley, R. L., and Ideker, T. (2009) Cost-effective strategies for completing the interactome. *Nat. Methods*. 10.1038/nmeth.1283

103.   Müller, H., and Mancuso, F. (2008) Identification and analysis of co-occurrence networks with netcutter. *PLoS One*. 10.1371/journal.pone.0003178

104.   Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858

105.   Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., Day, I. N. M., and Gaunt, T. R. (2013) Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Hum. Mutat.* 10.1002/humu.22225

106.   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., and Ghosh, P. (2005) The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat. Struct. Mol. Biol.* 10.1038/nsmb992

107.   Carlson, B. L., Ballister, E. R., Skordalakes, E., King, D. S., Breidenbach, M. A., Gilmore, S. A., Berger, J. M., and Bertozzi, C. R. (2008) Function and structure of a prokaryotic formylglycine-generating enzyme. *J. Biol. Chem.* 10.1074/jbc.M800217200

108.   Mahram, A., and Herbordt, M. C. (2015) NCBI BLASTP on High-Performance Reconfigurable Computing Systems. *ACM Trans. Reconfigurable Technol. Syst.* 10.1145/2629691

109.   Bethesda Research Laboratories (1986) E. coli DH5 alpha competent cells. *Focus - Bethesda Res. Lab.* **8**, 9

110.   Lim, H. K., Lee, S. U., Chung, S. Il, Jung, K. H., and Seo, J. H. (2004) Induction of the T7 promoter

using lactose for production of recombinant plasminogen kringle 1-3 in Escherichia coli. *J. Microbiol. Biotechnol.*

111. Marbach, A., and Bettenbrock, K. (2012) Lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J. Biotechnol.* 10.1016/j.jbiotec.2011.10.009

112. Liu, D., Zeng, X. A., Sun, D. W., and Han, Z. (2013) Disruption and protein release by ultrasonication of yeast cells. *Innov. Food Sci. Emerg. Technol.* 10.1016/j.ifset.2013.02.006

113. Sulkowski, E. (1985) Purification of proteins by IMAC. *Trends Biotechnol.* 10.1016/0167-7799(85)90068-X

114. Ion Exchange Chromatography & Chromatofocusing - Principles and Methods (2004) Ion Exchange Chromatography & Chromatofocusing - Principles and Methods. *GE Healthc.*

115. Andrew, S. M., Titus, J. A., and Zumstein, L. (2002) Dialysis and concentration of protein solutions. *Curr. Protoc. Toxicol.* 10.1002/0471140856.txa03hs10

116. Goldburg, W. I. (1999) Dynamic light scattering. *Am. J. Phys.* 10.1119/1.19101

117. Arzenšek, D., Podgornik, R., and Kuzman, D. (2010) Dynamic light scattering and application to proteins in solutions. in *Seminar*

118. D'Arcy, A. (1994) Crystallizing proteins - a rational approach? *Acta Crystallogr. D. Biol. Crystallogr.* 10.1107/S0907444993014362

119. Saridakis, E., and Chayen, N. E. (2008) Improving protein crystal quality by decoupling nucleation and growth in vapor diffusion. *Protein Sci.* 10.1110/ps.9.4.755

120. Whon, T. W., Lee, Y. H., An, D. S., Song, H. K., and Kim, S. G. (2009) A simple technique to convert sitting-drop vapor diffusion into hanging-drop vapor diffusion by solidifying the reservoir solution with agarose. *J. Appl. Crystallogr.* **42**, 975–976

121. Nada, A. M. A., Dawy, M., and Salama, A. H. (2004) Dielectric properties and ac-conductivity of cellulose polyethylene glycol blends. *Mater. Chem. Phys.* 10.1016/S0254-0584(02)00418-2

122. Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol.* 10.1016/0022-2836(91)90499-V

123. Sato, S., Gondo, D., Wada, T., Kanehashi, S., and Nagai, K. (2013) Effects of various liquid organic solvents on solvent-induced crystallization of amorphous poly(lactic acid) film. *J. Appl. Polym. Sci.* 10.1002/app.38833

124. Liu, X., Huettner, S., Rong, Z., Sommer, M., and Friend, R. H. (2012) Solvent additive control of morphology and crystallization in semiconducting polymer blends. *Adv. Mater.* 10.1002/adma.201103097

125. Song, R. Q., and Cölfen, H. (2011) Additive controlled crystallization. *CrystEngComm*. 10.1039/c0ce00419g

126. Warkentin, M., and Thorne, R. E. (2009) Slow cooling of protein crystals. *J. Appl. Crystallogr.* 10.1107/S0021889809023553

127. Parkhurst, J. M., Thorn, A., Vollmar, M., Winter, G., Waterman, D. G., Fuentes-Montero, L., Gildea, R. J., Murshudov, G. N., and Evans, G. (2017) Background modelling of diffraction data in the presence of ice rings. *IUCrJ*. 10.1107/S2052252517010259

128. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 10.1016/S0076-6879(97)76066-X

129. Schmit, J. D., and Dill, K. A. (2010) The stabilities of protein crystals. *J. Phys. Chem. B*. 10.1021/jp9107188

130. Wang, J. (2015) Estimation of the quality of refined protein crystal structures. *Protein Sci.* 10.1002/pro.2639

131. Taylor, G. (2003) The phase problem. in *Acta Crystallographica - Section D Biological Crystallography*, pp. 1881–1890, **59**, 1881–1890

132. González, A. (2007) A comparison of SAD and two-wavelength MAD phasing for radiation-damaged Se-MET crystals. in *Journal of Synchrotron Radiation*, 10.1107/S0909049506041045

133. Garron, M. L., and Cygler, M. (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology*. 10.1093/glycob/cwq122

134. Shukla, S. K., and Rao, T. S. (2017) An Improved Crystal Violet Assay for Biofilm Quantification in 96-Well Microtitre Plate. *bioRxiv*. 10.1101/100214

135. Olajos, M., Hajós, P., Bonn, G. K., and Guttman, A. (2008) Sample preparation for the analysis of complex carbohydrates by multicapillary gel electrophoresis with light-emitting diode induced fluorescence detection. *Anal. Chem.* 10.1021/ac8002598

136. Sogi, K. M., Gartner, Z. J., Breidenbach, M. A., Appel, M. J., Schelle, M. W., and Bertozzi, C. R. (2013) Mycobacterium tuberculosis Rv3406 Is a Type II Alkyl Sulfatase Capable of Sulfate Scavenging. *PLoS One*. 10.1371/journal.pone.0065080

137. Bojarová, P., and Williams, S. J. (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination. *Curr. Opin. Chem. Biol.* **12**, 573–581

138. Jonas, S., van Loo, B., Hyvönen, M., and Hollfelder, F. (2008) A New Member of the Alkaline Phosphatase Superfamily with a Formylglycine Nucleophile: Structural and Kinetic Characterisation of a Phosphonate Monoester Hydrolase/Phosphodiesterase from Rhizobium leguminosarum. *J. Mol. Biol.* 10.1016/j.jmb.2008.08.072

139. Kemp, P. J., and Telezhkin, V. (2014) Oxygen sensing by the carotid body: is it all just rotten eggs? *Antioxid. Redox Signal.* 10.1089/ars.2013.5377

140. Dolowy, M., and Pyka, A. (2014) Application of TLC, HPLC and GC methods to the study of amino acid and peptide enantiomers: A review. *Biomed. Chromatogr.* 10.1002/bmc.3016

141. Xiao, J. F., Zhou, B., and Ressom, H. W. (2012) Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC - Trends Anal. Chem.* 10.1016/j.trac.2011.08.009

142. Oiki, S., Mikami, B., Maruyama, Y., Murata, K., and Hashimoto, W. (2017) A bacterial ABC transporter enables import of mammalian host glycosaminoglycans. *Sci. Rep.* 10.1038/s41598-017-00917-y

143. Kim, D. H., Kim, B. T., Park, S. Y., Kim, N. Y., Han, M. J., Shin, K. H., Kim, W. S., and Kim, Y. S. (1998) Degradation of acharan sulfate and heparin by Bacteroides stercoris HJ-15, a human intestinal bacterium. *Arch. Pharm. Res.* 10.1007/BF02975378

144. Linhardt, R. J., Galliher, P. M., and Cooney, C. L. (1987) Polysaccharide lyases. *Appl. Biochem. Biotechnol.* 10.1007/BF02798420

145. Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes (2018) *Glycobiology*. 10.1093/glycob/cwx089

146. Zheng, Y., Huang, C. H., Liu, W., Ko, T. P., Xue, Y., Zhou, C., Guo, R. T., and Ma, Y. (2012) Crystal structure and substrate-binding mode of a novel pectate lyase from alkaliphilic Bacillus sp. N16-5. *Biochem. Biophys. Res. Commun.* 10.1016/j.bbrc.2012.02.148

147. Gunathilake, K. M. D., Ratnayake, R. R., Kulasooriya, S. A., and Karunaratne, D. N. (2013) Evaluation of cellulose degrading efficiency of some fungi and bacteria and their biofilms. *J. Natl. Sci. Found. Sri Lanka*. 10.4038/jnsfsr.v41i2.5710

148. Sasisekharan, R., Raman, R., and Prabhakar, V. (2006) GLYCOMICS APPROACH TO STRUCTURE-FUNCTION RELATIONSHIPS OF GLYCOSAMINOGLYCANS. *Annu. Rev. Biomed. Eng.* 10.1146/annurev.bioeng.8.061505.095745

149. Savage, J. R., Pulsipher, A., Rao, N. V., Kennedy, T. P., Prestwich, G. D., Ryan, M. E., and Lee, W. Y. (2016) A modified glycosaminoglycan, gm-0111, inhibits molecular signaling involved in periodontitis. *PLoS One*. 10.1371/journal.pone.0157310

150. Vekilov, P. G. (2016) Nucleation of protein crystals. *Prog. Cryst. Growth Charact. Mater.* 10.1016/j.pcrysgrow.2016.04.007

151. Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S. J. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R., and Shewchuk, L. M. (2006) Crystallization of protein-ligand complexes. in *Acta Crystallographica Section D: Biological Crystallography*, 10.1107/S0907444906047020

152. Bucci, V., Nadell, C. D., and Xavier, J. B. (2011) The Evolution of Bacteriocin Production in Bacterial Biofilms. *Am. Nat.* 10.1086/662668

153. Tait, K., and Sutherland, I. W. (2002) Antagonistic interactions amongst bacteriocin-producing enteric bacteria in dual species biofilms. *J. Appl. Microbiol.* 10.1046/j.1365-2672.2002.01692.x

154. Tanaka-Kumazawa, K., Kikuchi, Y., Sano-Kokubun, Y., Shintani, S., Yakushiji, M., Kuramitsu, H. K., and Ishihara, K. (2016) Characterization of a potential ABC-type bacteriocin exporter protein from Treponema denticola. *BMC Oral Health*. 10.1186/s12903-016-0243-7

155. Van Belkum, M. J., Worobo, R. W., and Stiles, M. E. (1997) Double-glycine-type leader peptides direct secretion of bacteriocins by ABC transporters: Colicin V secretion in Lactococcus lactis. *Mol. Microbiol.* 10.1046/j.1365-2958.1997.3111677.x

156. Grenier, D. (1996) Antagonistic effect of oral bacteria towards Treponema denticola. *J. Clin. Microbiol.*

157. Arora, N., Mishra, A., and Chugh, S. (2014) Microbial role in periodontitis: Have we reached the top? Some unsung bacteria other than red complex. *J. Indian Soc. Periodontol.* 10.4103/0972-

124X.128192

158. Lobos, O., Padilla, A., and Padilla, C. (2009) In vitro antimicrobial effect of bacteriocin PsVP-10 in combination with chlorhexidine and triclosan against Streptococcus mutans and Streptococcus sobrinus strains. *Arch. Oral Biol.* 10.1016/j.archoralbio.2008.11.007

159. Krebs, M. R. H., MacPhee, C. E., Miller, A. F., Dunlop, I. E., Dobson, C. M., and Donald, A. M. (2004) The formation of spherulites by amyloid fibrils of bovine insulin. *Proc. Natl. Acad. Sci.* 10.1073/pnas.0405933101

160. Bellamy, H. D., Snell, E. H., Lovelace, J., Pokross, M., and Borgstahl, G. E. O. (2000) The high-mosaicity illusion: Revealing the true physical characteristics of macromolecular crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 10.1107/S0907444900007356

161. Growth, C. (2017) Hanging Drop Vapor Diffusion Crystallization Hanging Drop Vapor Diffusion Crystallization. *Hapt. Res.*

162. Ghequire, M. G. K., and De Mot, R. (2015) The Tailocin Tale: Peeling off Phage Tails. *Trends Microbiol.* 10.1016/j.tim.2015.07.011

163. Dierks, T., Schmidt, B., and von Figura, K. (1997) Conversion of cysteine to formylglycine: a protein modification in the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.94.22.11963

164. Krupka, H. I., Huber, R., Holt, S. C., and Clausen, T. (2000) Crystal structure of cystalysin from Treponema denticola: a pyridoxal 5'-phosphate-dependent protein acting as a haemolytic enzyme. *EMBO J.* 10.1093/emboj/19.13.3168

165. Roeser, D., Preusser-Kunze, A., Schmidt, B., Gasow, K., Wittmann, J. G., Dierks, T., von Figura, K., and Rudolph, M. G. (2006) A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.0507592102

166. Handa, S., Paul, B. G., Miller, J. F., Valentine, D. L., and Ghosh, P. (2016) Conservation of the C-type lectin fold for accommodating massive sequence variation in archaeal diversity-generating retroelements. *BMC Struct. Biol.* 10.1186/s12900-016-0064-6

167. Medhekar, B., and Miller, J. F. (2007) Diversity-generating retroelements. *Curr. Opin. Microbiol.* 10.1016/j.mib.2007.06.004

168. Dierks, T., Lecca, M. R., Schlotterhose, P., Schmidt, B., and von Figura, K. (1999) Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. *EMBO J.* **18**, 2084–2091

169. Von Figura, K., Schmidt, B., Selmer, T., and Dierks, T. (1998) A novel protein modification generating an aldehyde group in sulfatases: Its role in catalysis and disease. *BioEssays*. 10.1002/(SICI)1521-1878(199806)20:6<505::AID-BIES9>3.0.CO;2-K

170. Grove, T. L., Lee, K. H., St. Clair, J., Krebs, C., and Booker, S. J. (2008) In vitro characterization of AtsB, a radical SAM formylglycine-generating enzyme that contains three [4Fe-4S] clusters. *Biochemistry*. **47**, 7523–7538

171. Dierks, T., Miech, C., Hummerjohann, J., Schmidt, B., Kertesz, M. A., and Von Figura, K. (1998) Posttranslational formation of formylglycine in prokaryotic sulfatases by modification of either

cysteine or serine. *J. Biol. Chem.* **273**, 25560–25564

172. Benjdia, A., Leprince, J., Guillot, A., Vaudry, H., Rabot, S., and Berteau, O. (2007) Anaerobic sulfatase-maturating enzyme [radical SAM cat in vitro sulfatase post-translational mod]. *J. Am. Chem. Soc.*

173. Valasatava, Y., Rosato, A., Banci, L., and Andreini, C. (2016) MetalPredator: A web server to predict iron-sulfur cluster binding proteomes. *Bioinformatics*. 10.1093/bioinformatics/btw238

174. Zhao, J., Zhang, M., Li, Y., Zhang, Z., Chen, M., Liu, T., Zhang, J., and Shan, A. (2017) Therapeutic effect of hydrogen injected subcutaneously on onion poisoned dogs. *J. Vet. Res.* 10.1515/jvetres-2017-0068

175. De Geest, S., Laleman, I., Teughels, W., Dekeyser, C., and Quirynen, M. (2016) Periodontal diseases as a source of halitosis: A review of the evidence and treatment approaches for dentists and dental hygienists. *Periodontol. 2000*. 10.1111/prd.12111

Appendix

Table A1: Summary table of Dynamic Light Scattering for all three targets, TDE0626, TDE1701, TDE2714. Table showing peak scattering with highest percent mass. Each sample was concentrated to 5mg/mL prior to DLS analysis. Target proteins all suspended in anion exchange buffer (either 50mM Tris-HCl pH 8.0 or 100mM glycine-NaOH pH 9.6).

Target	Radius (nm)	%Pd	Mw-R (kDa)	%Intensity	%Mass
TDE0626	3.433	25.6	60	6.6	96.7
TDE1701	3.113	29.9	48	22	99.2
TDE2714	5.247	30.1	163	97.2	99.8

Table A2: Crystallization conditions for the optimization of TDE2714 using purified protein in Tris-HCl pH 8.0. Hanging drop vapour diffusion was in a 24-well tissue culture plate with concentrations of TDE2714 at 12, 14, and 16mg/mL. Adjusted polyethylene concentration from 19.6% to 20.2% (w/v). There was an addition of two organic solvents (glycerol and ethylene glycol). Buffering pH ranged from 7-8 based on pH test of condition. Plates were incubated at 18°C for one month. Star beside well number indicates large protein crystals.

Well	Organic solvent	Concentration (% v/v)	Buffer	Concentration (M)	Polyethylene Glycol	Concentration (% w/v)	Salt	Concentration (M)
*A1	Glycerol	2.5	Tris pH 7	0.05	3350	19.6	NaNO₃	0.2
A2	Glycerol	2.5	Tris pH 7	0.05	3350	19.8	NaNO₃	0.2
A3	Glycerol	2.5	Tris pH 7	0.05	3350	20.0	NaNO₃	0.2
A4	Glycerol	2.5	Tris pH 7	0.05	3350	20.0	NaNO₃	0.2
A5	Glycerol	2.5	Tris pH 7	0.05	3350	20.0	NaNO₃	0.2
A6	Glycerol	2.5	Tris pH 7	0.05	3350	20.2	NaNO₃	0.2
*B1	Glycerol	2.5	Tris pH 7.5	0.05	3350	19.6	NaNO₃	0.2
*B2	Glycerol	2.5	Tris pH 7.5	0.05	3350	19.8	NaNO₃	0.2
B3	Glycerol	2.5	Tris pH 7.5	0.05	3350	20.0	NaNO₃	0.2
B4	Glycerol	2.5	Tris pH 7.5	0.05	3350	20.0	NaNO₃	0.2
B5	Glycerol	2.5	Tris pH 7.5	0.05	3350	20.0	NaNO₃	0.2
B6	Glycerol	2.5	Tris pH 7.5	0.05	3350	20.2	NaNO₃	0.2
*C1	Glycerol	2.5	Tris pH 8	0.05	3350	19.6	NaNO₃	0.2
*C2	Glycerol	2.5	Tris pH 8	0.05	3350	19.8	NaNO₃	0.2
*C3	Glycerol	2.5	Tris pH 8	0.05	3350	20.0	NaNO₃	0.2
*C4	Glycerol	2.5	Tris pH 8	0.05	3350	20.0	NaNO₃	0.2
C5	Glycerol	2.5	Tris pH 8	0.05	3350	20.0	NaNO₃	0.2
C6	Glycerol	2.5	Tris pH 8	0.05	3350	20.2	NaNO₃	0.2
D1	Ethylene Glycol	5.0	Tris pH 8	0.05	3350	19.6	NaNO₃	0.2
D2	Ethylene Glycol	5.0	Tris pH 8	0.05	3350	19.8	NaNO₃	0.2
D3	Ethylene Glycol	5.0	Tris pH 8	0.05	3350	20.0	NaNO₃	0.2
D4	Ethylene Glycol	5.0	Tris pH 8	0.05	3350	20.0	NaNO₃	0.2
*D5	Ethylene Glycol	5.0	Tris pH 8	0.05	3350	20.0	NaNO₃	0.2
D6	Ethylene Glycol	5.0	Tris pH 8	0.05	3350	20.2	NaNO₃	0.2

Table A3: Predicted operons for TDE0626 and TDE1701 from DOOR^2 database. Each target operon was predicted a data mining classifier that identifies parameters such as intergenic distance, neighborhood conservation and phylogenetic distance. A confidence score was given for each operon however no additional information about the operons were provided.

TDE0626 Operon				
Gene	Product	Start (bp)	End (bp)	Confidence (%)
TDE0623	Precorrin-3B C17-methyltransferase	655923	656645	100
TDE0624	Precorrin-6X reductase	656642	657400	
TDE0625	ABC-transporter protein	657393	658163	
TDE0626	Hypothetical protein	658230	659744	
TDE1701 Operon				
Gene	Product	Start (bp)	End (bp)	Confidence (%)
TDE1700	Hypothetical protein	1758180	1758287	47
TDE1701	Hypothetical protein	1758280	1759407	

Table A4: Fluorescence test scans for selenomethionine-containing TDE2714 crystals. Scans tested at CLS for identification of anomalous signals. The anomalous scattering coefficients for both the peak and inflections shown for all three test scans.

Scan	Peak (eV)	Inflection (eV)	f'' peak	f' peak	f'' inf	f' inf
1	12665.7	12667.7	3.62	-5.74	1.77	-8.01
2	12659.7	12656.7	4.53	-7.53	1.65	-8.62
3	12716.7	12656.7	3.10	-5.39	0.91	-6.48

```
1          11         21         31         41
MKIN KLVKT VCMFFLALSL LNCRSEDRFP IGRMEVNLKT PVQDTSKLF
eeeeeebbbb bbbbbbbbbb ebbeeeeee  eeeeeeeeee ebeeebeeeb
                                                     s
51         61         71         81         91
SDKDFKTEYG SSKVVSVNLS ALSGISAKGL TVNGNTVTIN EGGQYIISGS
eeeeeeeeee eeebbebebe eebebeeeeb ebeeeebebe eebbbbbbbe
     f    ff                            fs   fs s    s
101        111        121        131        141
LNDGQIIVDA PDNDEVHLIL DNADISNSSM PVIYAKKAGK MLITLAKGSK
beebebbbeb eeeeebebbb ebbebebeee bbbbbeebee bbbbbbeeee
   s     s        f f s              f  sff      sss
151        161        171        181        191
NKLNVNGKFA DSDAGKTNAV IFSQTDLTLN CTGELNIESK YGSGIVSKKD
eebeeeeeeb eeeeeebebb bbbeeebebe bebebebeee eeebbebeee
f                f  f   sf f f f  s s s        fs fsffff
201        211        221        231        241
LRVTGGSFTV SASKHALKGC DNVSIADGKF TLTAGKDGIH SENEENAESG
bebbbbebeb ebbeebbeee ebbebeebeb ebebeeebbe beeeeeeeeb
s                 f f             f  f    f          s
251        261        271        281        291
NIYIKNGEFT INAASEALDA INDITIDGGY INIAKADEGM EALTININGG
bbbbebbebe bebeeebbeb ebbbbbeebe bebeebeeeb eeeebebebb
        s        f         s           ff  ff        s
301        311        321        331        341
KIMVVSSDDG LNASYSDKGE IEAKLSGTAL TDNSSKKTEK KGPPVLSESA
ebebebeeeb bebbeeeeee eeeeeeeeee eeeeeeeeee eeeeeeeeee
        ff      fs
351        361        371        381        391
ASTYVNITGG EVTINSQADG IDSNGSVYVS GGKVNILGPV SDGDAALDYD
eebbbebebb ebebebeeee eeeebebebe ebebbbebee eeeebbbebe
      s ss         fff  fffs      fs    sf    f    f
401        411        421        431        441
LTALISGGEF IASGSRGVVQ GFSDKSTCAS FIANFSKTVK GEVIVSDSSG
eebebebbeb bbbbbbbbbe bbeeeeebb bbbebeeeee  eebbbeeeee
    s  s            s f       f f                 f  f
451        461        471        481        491
AVILKTNQDK DFQSIVVSSK GLKVGETYKI TAGGQTLTVK MDSISVGDDM
eebbebebee ebeebbbbbb ebeeeebeb  ebeeeeebe  beeebeeeee
        f        f    s        f  sf        f
501
KIR
eeee
```

Legend:

The conservation scale:

```
1  2  3  4  5  6  7  8  9
Variable     Average    Conserved
```

e - An exposed residue according to the neural-network algorithm.
b - A buried residue according to the neural-network algorithm.
f - A predicted functional residue (highly conserved and exposed).
s - A predicted structural residue (highly conserved and buried).
X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure 21: Conserved mapping of multiple sequence alignment from COBALT of TDE0626 primary protein sequence. Conserved mapping was generated by ConSurf which was used to show regions of high conservation from a scale of 1-9, where 9 is highly conserved among all polysaccharide lyases. Data showing a highly conserved region (DGIDSNG) at positions 369-375 for TDE0626. Predicted functional elements label with a red f and structural elements labelled with a blue s.

```
1          11         21         31         41
MDDLMNVVXX DTTINTRGID VCNPSPEPTY IIDDDKNNHH ARSPSPTNEK
eeebbebbbb eeebebebbe bbeeeeeeeb bbbeeeebeb bbeeeeeeee
  ff         f    fsf          f         f        f   f

51         61         71         81         91
KAPSTSHPGN PPNTGNDCDG GCDSKDPTHD TPSDPSNPVN DDGSMGSSGG
eeeeeeeeee eeeeeeeeee eeeeeeeeee eeeeeeeeee eeeeeeeeee
 f   f                          f        f             f

101        111        121        131        141
TPDNSDDETG GNTNLPTDDD GDIENESVGN GSSQTTEPAA NINTTNTTQA
eeeeeeeeee eeeeeeeeee eebeebeebe eeeeeeeebb ebeeeeeeeb
  ff fff          f          f         f  f          f   fs

151        161        171        181        191
ADVSTKPVSN AERARSILAN LTNPIEQTVN AKTPCKNDML AAYYDDKLTN
eeeeeeeeee eeebebbbbe beeebeeeee beeebeeebb ebbbeeeeee
     ff             f        f        f       f f

201        211        221        231        241
KKLSFDAEYK PVQQLLPMQK DKNLKVIYPN TCKTNSTISA VRLSGYTYKD
eebebeebbe eeeebbeeee eeeeeebbbe bbebeeeeeb bebbbeebee

251        261        271        281        291
DITKRDVEKI TNKEVSSKKC KLHTDVIKEV FGIDAVYYEI PAGLSEKELK
ebeeeebeeb beebeeeeeb eebeeebeeb bbbbbbbbeb eeebeeeebe
  f                    f    f   f

301        311        321        331        341
ALVGDNDAHI REPQRAFWGN DKLPEDGEHG IGYVMGGPIF SYMSRHGESW
eeeeeeeebb ebeeeeeeee eeeeeeeeeb bbeeeebebb beebeeeeeb
     s                         f       s        s       f

351        361        371
KDVRPARSTA DFTENLGGLY FEIKK
eebeeeebeb ebebebebbe beeee
      s       f
```

Legend:

The conservation scale:

```
1  2  3  4  5  6  7  8  9
Variable    Average    Conserved
```

e - An exposed residue according to the neural-network algorithm.
b - A buried residue according to the neural-network algorithm.
f - A predicted functional residue (highly conserved and exposed).
s - A predicted structural residue (highly conserved and buried).
X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure 22: Conserved mapping of multiple sequence alignment from COBALT of TDE1701 primary protein sequence. Conserved mapping was generated by ConSurf which was used to show regions of high conservation from a scale of 1-9, where 9 is highly conserved among all homologous proteins. Data showing relatively low sequence conservation and one region of average conservation at positions 103-108 (DNGDDE) for TDE1701. Predicted functional elements label with a red f and structural elements labelled with a blue s.

Mass Distribution	Radius (nm)	%Pd	Mw-R (kDa)	%Intensity	%Mass
☑ Peak 1	3.113	29.9	48	22.0	99.2
☑ Peak 2	27.337	51.4	7742	78.0	0.8
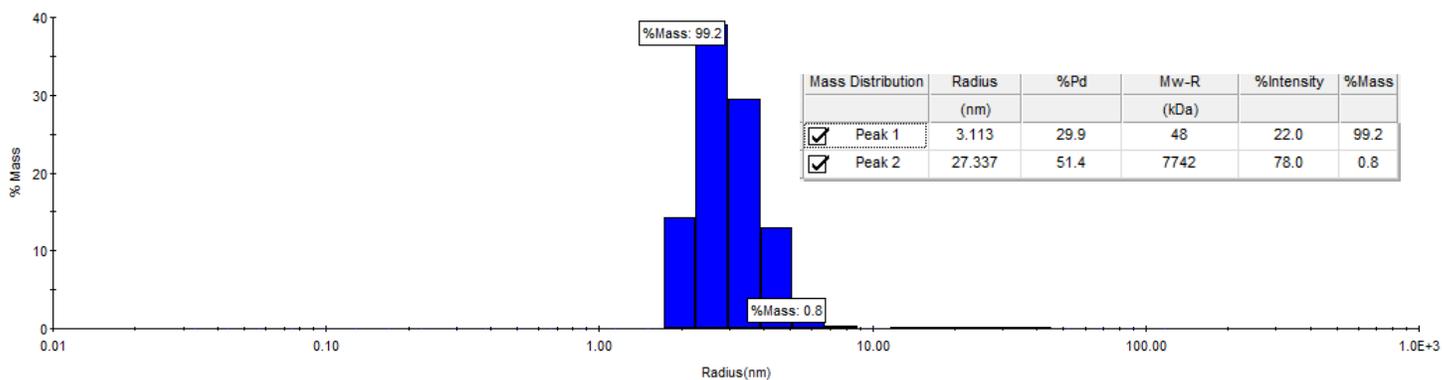
Figure 23: Dynamic light scattering plot for 5mg/mL TDE1701 in 50mM Tris-HCl pH 8.0 after anion exchange chromatography. Analysis depicts scattering profile as a measurement of predicted % mass versus the calculated average particle radius (nm). The mass distribution for the protein sample shown as three distinct peaks. Each peak shown with corresponding; radius, percent polydispersity (%Pd), predicted molecular weight based on radius, percent intensity, and percent mass. 20 images were taken of the scattering profile every 5 seconds and averaged.
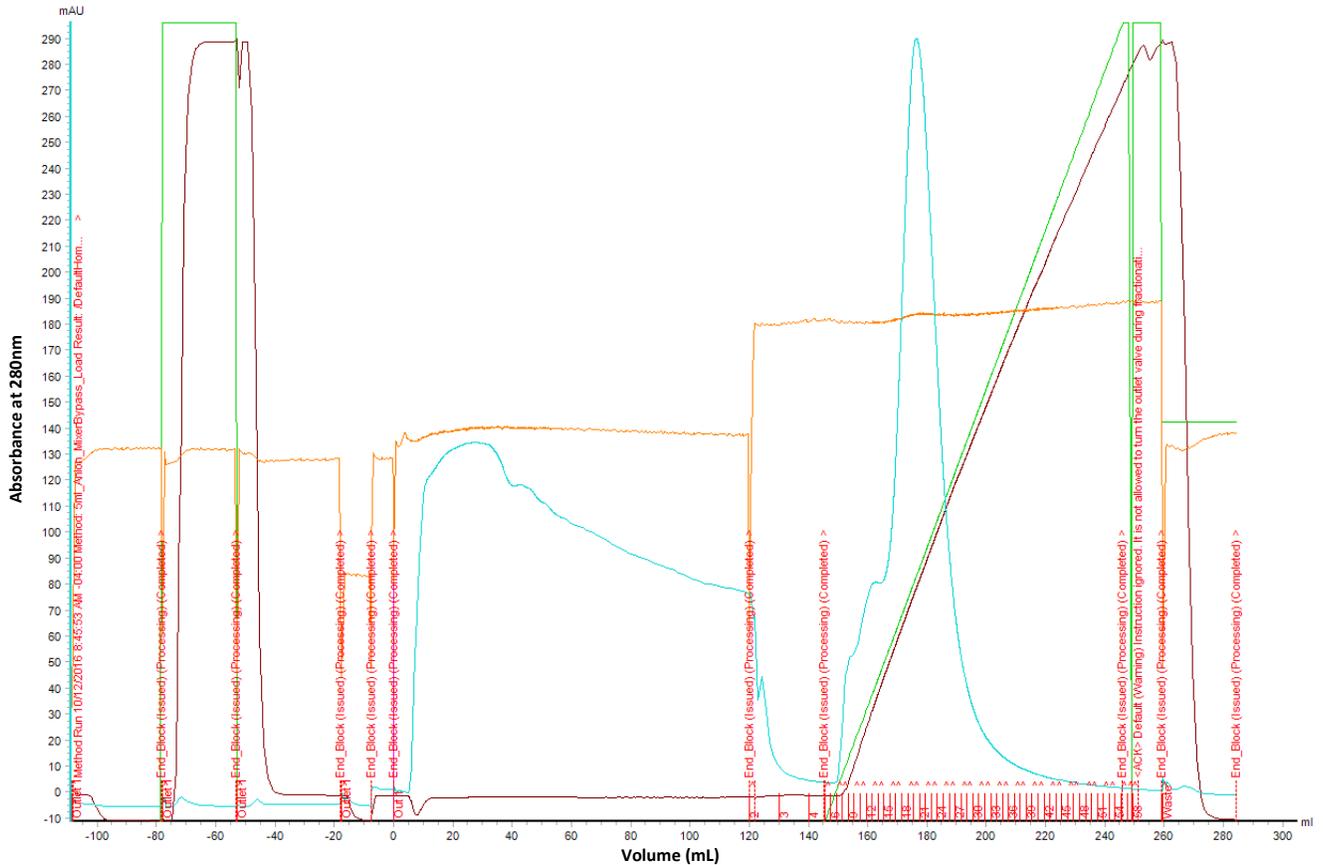
Figure 24: Chromatograph of anion exchange chromatography from purified TDE1701 in 50mM Tris-HCl pH 8.0. Protein sample volume added to the GE HiTrap™ Q FF 5mL column was 250mL based on an initial volume of 175mL, for two full cycles over the column. Miliabsorbance units (mAU) at 280nm and Conductivity (mS/cm) were below a value of 2.0 prior to sample application. Flow rate set to 5 mL/min. The alarm for the delta column pressure maxima was set to 0.5MPa and minimum at 0.0MPa. Wash buffer used was 50mM Tris-HCl pH 8.0 and elution buffer, for salt gradient elution was 50mM Tris-HCl pH 8.0 with 1M NaCl. Peak fractions were at fractions 15-32, which corresponded to the NaCl concentration of 337mM. Blue line represents absorbance reading at 280 and dark brown line represents conductivity.
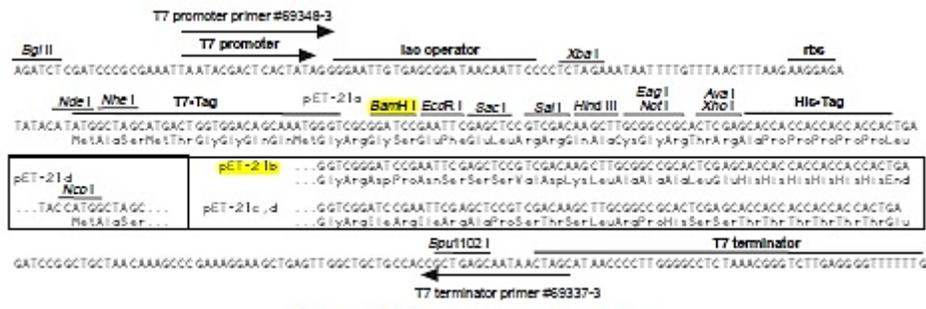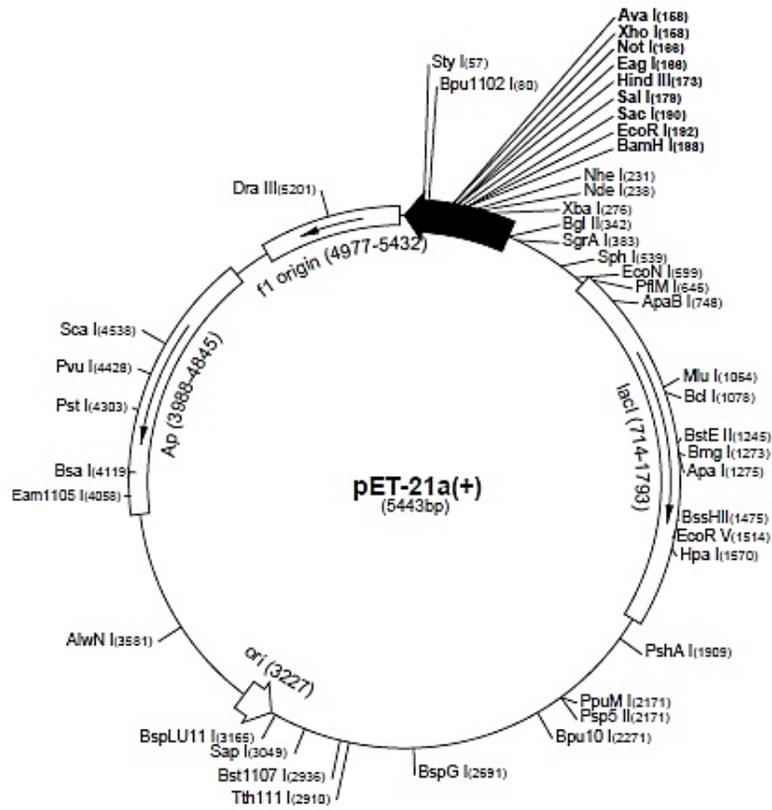
Figure 25: pET21a-d(+) vector sequence map. Plasmid map showing all restriction enzyme cut sites as well as inducible *lac* operon and multiple cloning site. Cloning and expression region shown with pET21b sequence highlighted

>TDE0626

EDRFPIGRMEVNLKTPVQRDTSKLFSDKDFKTEYGSSKVVSVNLSALSGISAKGLTVNGNTVTINEGGQYIISGSLNDGQII
VDAPDNDEVHLILDNADISNSSMPVIYAKKAGKMLITLAKGSKNKLNVNGKFADSDAGKTNAVIFSQTDLTLNGTGELN
IESKYGSGIVSKKDLRVTGGSFTVSASKHALKGCDNVSIADGKFTLTAGKDGIHSENEENAESGNIYIKNGEFTINAASEAL
DAINDITIDGGYINIAKADEGMEALTININGGKIMVVSSDDGLNASYSDKEEIEAKLSGTALTDNSSKKTEKKGPPVLSESA
ASTYVNITGGEVTINSQADGIDSNGSVYVSGGKVNILGPVSDGDAALDYDLTALISGGEFIASGSRGMVQGFSDKSTQA
SFIANFSKTVKGEVIVSDSSGAVILKTNQDKDFQSIVVSSKGLKVGETYKITAGGQTLTVKMDSISVGDNMHKRR

>TDE1701

MDDLMNVVGGDTTINTRGIDVCNPSPEPDYIIDDDGNHHHAPSPSPTNPKPAPGTGNPGNPPNTGNDGDGGGDSK
DPTEDTPSDPGNPVNDDGGNGSSGGTPDNGDDETGGNTNLPTDDDGDIGNGSVGNGGQTTTPPAANINTTQTTQA
ADVSPKPVSPAERARSILANLTNPIEQTVNAKTGCKNDMLAAYYDDKLTNKKLSFDASYKEVQDDLPMQKDKNGKVIY
PNTCKYNSTIGAVRLSGYTYKDDLTKEDVEKITNKAVSSKKCKLHTDVIKEVFGLDAVYYEIPAGLSEKELKALVGDNPAMI
RFPQRAFWGNDKLPEDGEHGIGYVNGGFIEPYMGRHGESWKDVRGARSTADFTSNLGGLYFEIKK

>TDE2714

MIDSWHLGTDESAPELKDTDTFYTNAVVFVKSRPDLPDLSSIQGTGDQVKITLAVTLEEGGAILGPKSISVNKGTRWNSV
LKTYAKAALKVKYGFQCNGWKKNGNTVNDTYTFDDDTTIFAELEDMRINITVKGDSNVSVPDSTPLVVLSGRKWQDIK
ELAAGRIQVSDPSNIAVTAWHRGESASAPVLTDEYEFKKAEGQNRTVYAKTGDRRITLTVTYGSGSGTPATAGTITIYDG
DEWHNVQKQVAPLVSVPEDHSIHWYLDNAGGELINFFYTFKASDGNARTVYARVSPPIQLTVTYGRVTGATVTIGTVT
THHRRLWYSVQNEIKYKYPLLSNTEVEWHWNDKNGALISDTYEFDAGNVPSNTVYAFVRPKIITYYGGLRYIAPIGSNGS
YEDYNMKKIDAVTDGYVGGDAYHYVYNDPHKVSLTAYRIGTTEVTQELYELVMAHNPSYFQGSSHPTASGESQEKRPV
EQVSWFDAIAFCNELTRCCYSLGEAQCVYTYNGHTYTVEDAQAHNVPVMDMSKKGFRLPTQAEWEWAAQSGTAW
QKWAGTNDEDKLSYYAWYDANEYGRTHQAGRKKANPFGLFDMSGNVSEWCWDWWIQTTPEGGTDPVGPLSGTN
RTTCGGGYSFFDSACCCAYRGLEEPDKNKDTTGFRIVCRYEF

Figure 26: FASTA protein sequences for all targets; TDE0626, TDE1701 and TDE2714. Custom genes and optimized sequences ordered from GenScript in pET21b vector. Affinity tags not shown for any of the sequences.
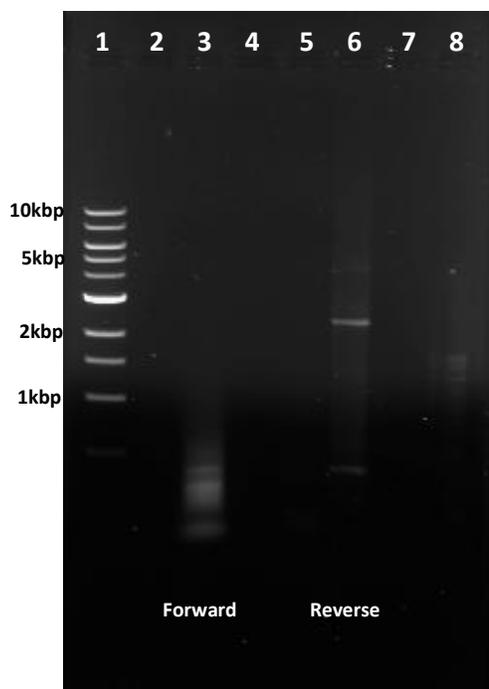
Figure 27: Site-directed mutagenesis of C637A mutant of TDE2714 visualized by a 1%(w/v) agarose gel and imaged with 0.5µg/mL of ethidium bromide. Lane 1 containing NEB 1kb DNA ladder. Lane 3 containing the PCR reaction with 5'-forward primer. Lane 6 containing the PCR reaction with 3'-reverse primer. Lane 8 containing the PCR reaction of both forward and reverse primers.
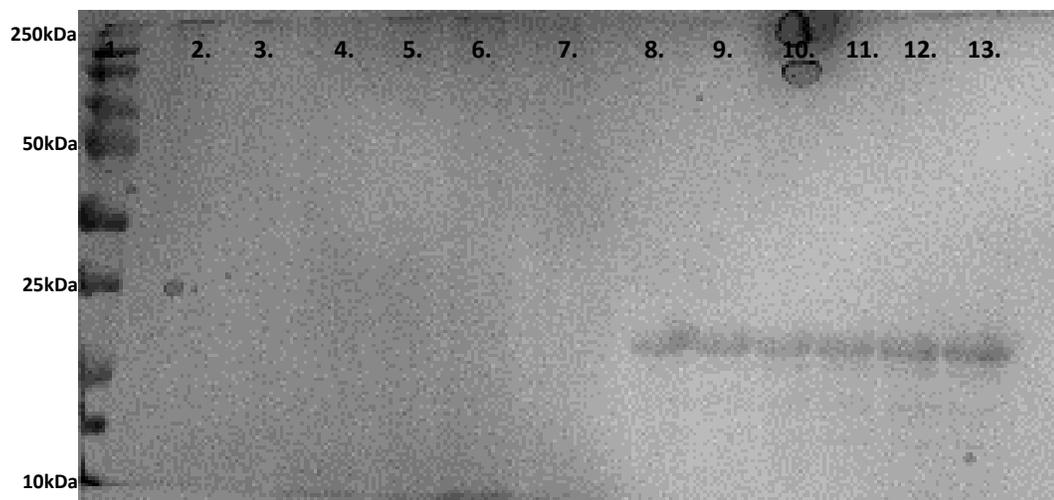
Figure 28: Cellulose binding assay for purified TDE0626 at 0.5mg/mL in 50mM glycine pH 9.6. Six samples of cellulose were incubated with TDE0626 for 1h at 37°C. The concentrations of cellulose were 5, 10, 15, 20, 25, 50mg/mL. The filtered samples were visualized on a 12%(w/v) polyacrylamide gel and stained with silver nitrate. Lane 1 containing the Precision Plus Protein™ Kaleidoscope ladder. Lanes 2-7 containing the supernatant after filtration for each of the samples. Lanes 8-13 containing the resuspended pellets of cellulose after filtration.

Figure 29: C-PAGE gel of TDE2714 at 5ng/mL incubated with 0.5%(w/v) polygalacturonic acid. Polyacrylamide gel contained 15%(w/v) acrylamide and 1%(v/v) bis-acrylamide. Spot 1 contained TDE2714. Spot 2 contained 5ng/mL BSA standard + polygalacturonic acid. Spot 3 contained TDE2714 and polygalacturonic acid. Gel was imaged via silver nitrate, formaldehyde, and sodium carbonate.
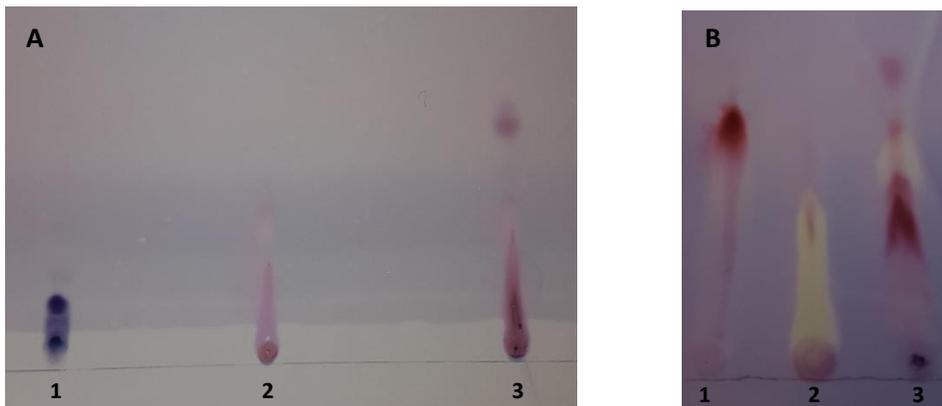
Figure 30: Thin-layer chromatography tests for wild-type TDE2714 and C637A mutant when incubated with the sulfatase consensus sequences – LCTPSRA – at 37°C for 16h. A) wild-type TDE2714 peptide TLC. Spot 1 contained 1mM peptide, 1mM DTT, and 50mM sodium phosphate pH 8.0. Spot 2 contained 0.1mM WT-TDE2714, 1mM DTT, and 50mM sodium phosphate pH 8.0. Spot 3 contained 1mM peptide,0.1mM WT-TDE2714, 1mM DTT, and 50mM sodium phosphate pH 8.0. B) C637A mutant peptide TLC. Spots 1-3 contained same components.
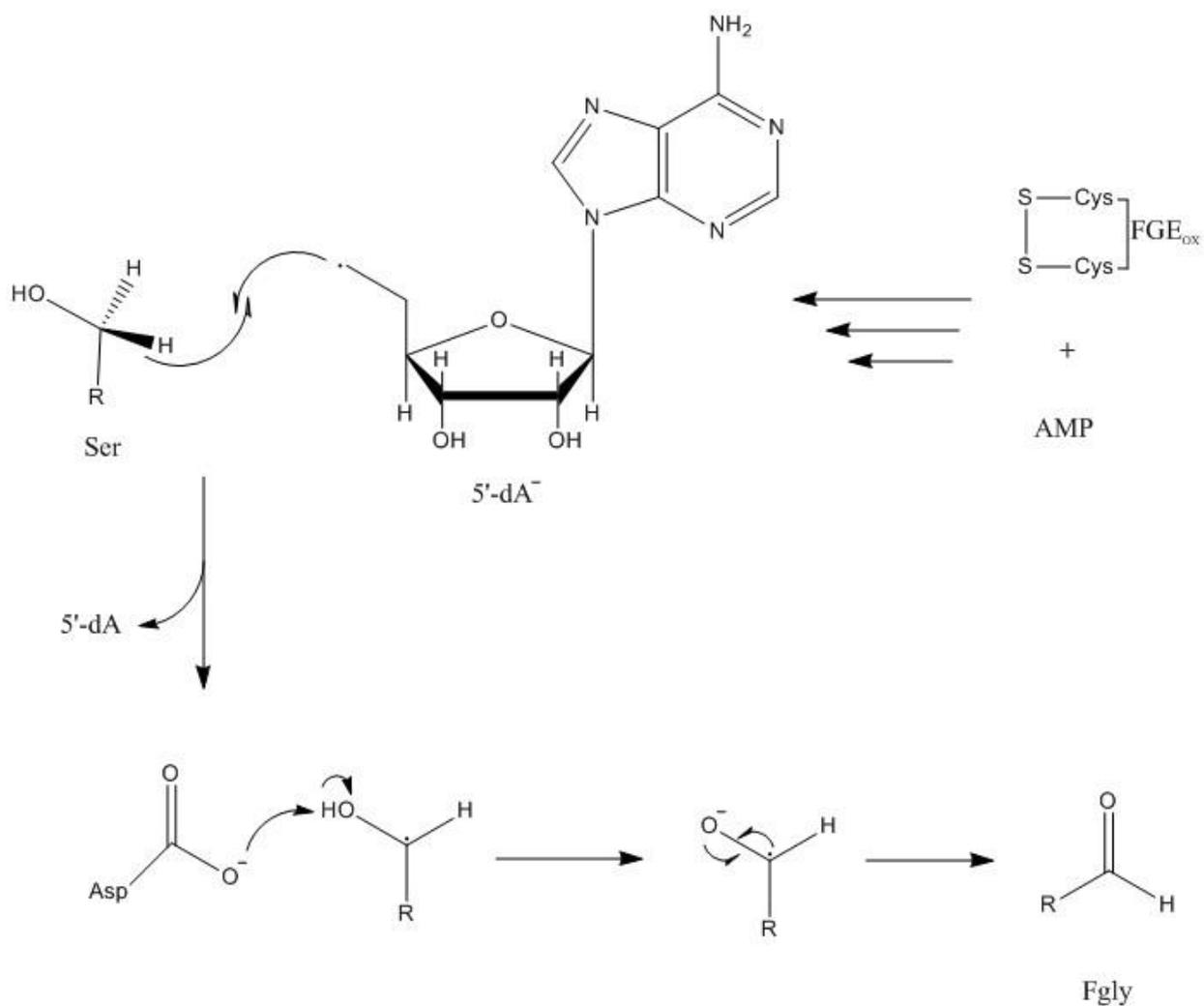
Figure 31: Proposed mechanism of anaerobic FGE lacking iron-sulfur clusters with either a radical s-adenosylmethionine or a similar compound. A precursor electron donor is required for the reaction to commence. A catalytic aspartate would be required for the formation of the formylglycine (Fgly).

135
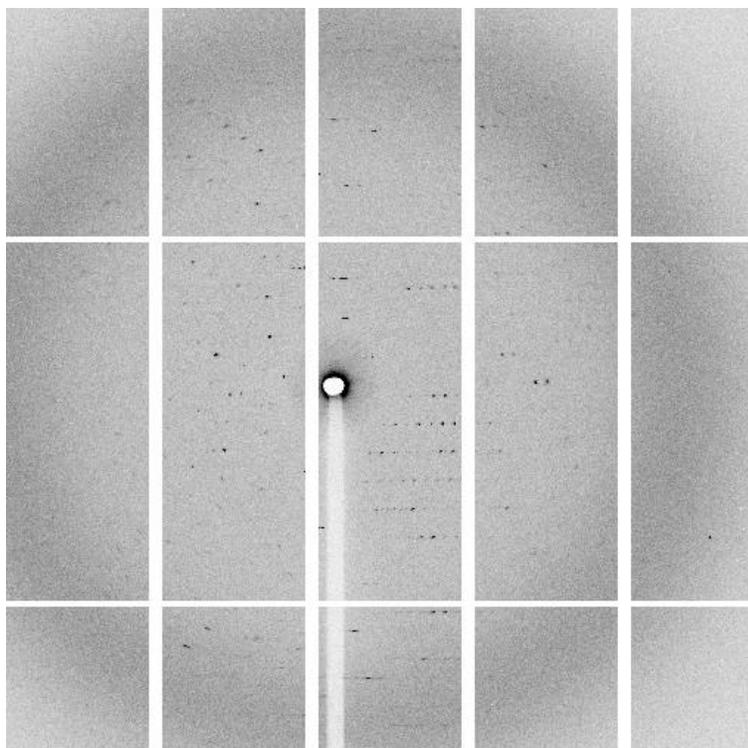
Figure 32: Diffraction image of TDE2714 collected at CLS and visualized by iMosfilm. Image shown at 90°
during 0.2° oscillations over 720° rotation.