2018

# INITIAL FUNCTIONAL CHARACTERIZATION OF BFO 2291 AND BFO 2294 FROM Tannerella forsythia; DEGRADERS OF CHONDROITIN SULFATE A

Rony Eshaque
*Wilfrid Laurier University*, esha4510@mylaurier.ca

Rony Eshaque
*Wilfrid Laurier*, rony.eshaque@gmail.com

# INITIAL FUNCTIONAL CHARACTERIZATION OF BFO 2291 AND BFO 2294 FROM *Tannerella forsythia*; DEGRADERS OF CHONDROITIN SULFATE A

By:
Rony Eshaque

A thesis
presented to the University of Wilfrid Laurier
in fulfillment of the
thesis requirement for the degree of
Master of Science
in Chemistry

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

# ABSTRACT

Periodontitis is characterized by the inflammation of the periodontal tissues in response to bacterial action. The Global Burden of Disease (GBD) 2010 study shows that periodontitis ranks 6$^{th}$ in the global prevalence of oral health conditions and affects 11% of the world population. It is initiated by the formation of biofilms containing different types of bacteria. These biofilms exist as dental plaque and contain three types of bacteria called *Tannerella forsythia, Porphyromonas gingivalis, and Treponema denticola*. These three types of bacteria are strongly associated with periodontitis and they are termed as the Red complex. The main focus of the research presented herein is on *T. forsythia* which secrets different types of enzymes that degrade the periodontal tissues. These enzymes are transcribed and translated together from the genes clustered together regulated by an operon system. Periodontal tissues are primarily composed of glycosaminoglycans or (GAGs). These GAGs are linear chains of polysaccharides composed of different disaccharide units. GAGs can exist in conjunction with proteoglycans which are a protein core appended with different types of GAGs The degradative enzymes from *T. forsythia* degrades the GAGs connected to the protein core, thus contributing to the destabilization and eventual destruction of the periodontal tissues such as alveolar bone and ligament causing tooth loss. Based on the bioinformatics and enzyme kinetics, it was found that BFO 2291 (chondroitin AC lyase) breaks down chondroitin sulfate A into oligosaccharides containing glucuronic acid and galactosamine. Chondroitin sulfate A is (a type of GAG) one of the main constituents of many periodontal tissues. Enzyme activity was initially analyzed by doing a pH profile using various ranges of buffers and pH values. By doing pH profile, we determined the pH at which this enzyme is the most active which is pH 6.5. Chondroitin AC lyase along with the other enzymes are responsible for the degradation of chondroitin sulfate A using a GAG degradation

pathway. The other enzyme analyzed herein was BFO 2294 (KDPG aldolase) which breaks down KDPG (2-keto-3-deoxy-6-phosphogluconate) aldolase into pyruvate and D-glyceraldehyde-3-phosphate. Using a coupled enzymatic assay, it was shown the BFO2294 catalyzed conversion to pyruvate was subsequently converted into lactic acid by lactic dehydrogenase using cofactor NADH which in turn was oxidized to $NAD^+$. In the last step of GAG degradation pathway, KDPG is broken down by KDPG aldolase into pyruvate which can enter into the Krebs cycle and electron transport chain for production of ATP. The final outcome is the production of ATP which is used by the bacteria to drive its cellular functions. Another part of our research was structural characterization of BFO2294 via X-ray diffraction analysis. The BFO2294 enzyme, or KDPG aldolase was successfully crystallized and the diffraction data was collected. KDPG Aldolase catalyzes the reversible cleavage of 2-keto-3deoxy-6-phosphogluconate (KDPG) into pyruvate and D-glyceraldehyde-3-phosphate by retro aldol cleavage. Using various research articles, it was found that the active site of this enzyme contains zwitterionic pair of Glu-49 and Lys-141 which are involved in the catalytic reaction between the enzyme (KDPG aldolase) and the substrate (KDPG).

# ACKNOWLEDGEMENTS

First of all, I am extremely grateful and thankful to my supervisor, Dr. Michael Suits, for providing me with the opportunity to work in his laboratory. His valuable suggestions and direction and the time and effort he devoted to my project are greatly appreciated.

I want to express my gratitude to my committee members, Dr. Geoff Horsman and Dr. Lillian DeBruin for their valuable suggestions and advice.

I am also grateful to the past and present members of Dr. Suits' lab for their help, consideration and time. I want to give special thanks to Jonah Nechacov for teaching me how to operate the instruments as well as teaching me some valuable lab techniques.

Finally, I would like to thank my mother Sarah Eshaque, my sister Bithi Eshaque, my brother in law Shamsul Arefeen, my nephew Sabahat and my niece Zaafirah for their unconditional love and affection, without whom my life would be unimaginable.

I would like to acknowledge the financial support received for my work in the form of scholarships and teaching assistantships from the Department of Chemistry, and research assistantships from Dr. Suits.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| GAGs | Glycosaminoglycans |
| *T. forsythia* | *Tannerella forsythia* |
| *P. gingivalis* | *Porphyromonas gingivalis* |
| *T. denticola* | *Treponema denticola* |
| CS | Chondroitin Sulfate |
| DS | Dermatan Sulfate |
| KS | Keratan Sulfate |
| KDPG | 2-dehydro-3-deoxy-phosphogluconate aldolase |
| *E. coli* | *Escherichia coli* |
| IPTG | Isopropyl β-D-1 thiogalactopyranoside |
| IDA | Iminodiacetic acid |
| NTA | Nitrilotriacetic acid |
| pI | Isoelectric point |
| $V_{max}$ | Maximum velocity |
| $K_M$ | Michaelis constant |
| $k_{cat}$ | Catalytic constant |

# 1. INTRODUCTION

## 1.1 What Is Periodontitis?

Periodontitis is characterized by the infection and inflammation of the periodontal tissues that often leads to alveolar bone loss (1). It is characterized by the inflammation of the periodontal tissues in response to bacterial infection and initiated by biofilms or dental plaque containing different types of bacteria (2). The periodontal tissue is made up of different types of glycosaminoglycans (GAGs) (2). The three pathogens that are also known as the **"Red Complex"** (*Tannerella forsythia, Porphyromonas gingivalis, and Treponema denticola*) are strongly associated with periodontitis (2). There are several steps that lead to periodontitis which is shown in figure 1. The first step is the formation of dental plaque which is a biofilm containing different types of bacteria. In the second step, biofilm is hardened to form tartar or calculus. In the third step, if the calculus is not treated then gingivitis occurs. Gingivitis is the inflammation of gingivae. If the gingivitis is not treated, then one is affected by periodontitis. In periodontitis, a packet develops between the gum and the surface of the tooth. This packet contains different types of bacteria which destroy the periodontal tissue.

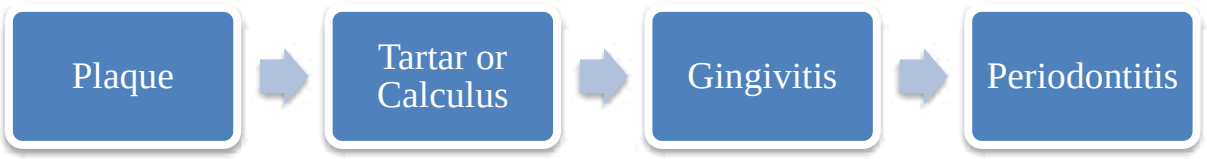



**Figure 1:** Steps leading to periodontitis. At first, plaque is formed and if not treated then tartar builds up. If the tartar is not treated then gingivitis occurs and eventually a more aggressive form which is periodontitis destroys the gum and teeth.

## 1.2 Red Complex

The three microorganisms that are known as "Red Complex" are *T. forsythia, P. gingivalis, and T. denticola* (3). This complex is the interest of study because it is associated with bleeding on probing that is an important parameter of destructive periodontal disease (3). According to the study, 13000 subgingival plaque samples were examined and after that DNA hybridization methodology and community ordination techniques were used to identify the specific microbial groups within dental plaque (4). The presence of 40 different subgingival species were determined using checkerboard DNA-DNA hybridization technique (4). These organisms were grouped into five complexes according to the presence of an organism and the related periodontal status which are presented in the figure 2 (4). Among these five groups, yellow, green and purple complexes precede the multiplication of gram negative red and orange complexes (4). The red and orange complexes are thought to be the main etiologic agents of the periodontal disease (4). The bacteria in the "Red Complex" region together produce virulence factors and in turn initiate an immune response that starts the process of inflammation and the destruction of periodontal tissues (2). Neutrophils are the white blood cells that are secreted by

the host defense system in inflamed gingival tissues (5). Plaque embedded bacteria and their by-products could initiate neutrophils recruitment to the area of bacterial invasion in the periodontal tissues that would result in stimulation of free radical generation (6). Usually, reactive oxygen species (ROS) produced by phagocytes are used for the killing of the invading pathogens but prolonged release of ROS and increased matrix metalloproteinases activity cause bone resorption and degradation of connective tissue surrounding the teeth (6). *T. forsythia* is a gram-negative anaerobic bacterium isolated from gingival sulci and periodontal pockets from the patients with periodontitis (2). *T. forsythia* contains some putative virulence factors, such as a trypsin-like protease, a sialidase, hemagglutinin, and a cell surface-associated and secreted protein (BspA) (7). The protein BspA is considered as a virulence factor that is important for alveolar bone loss in mice (7). The second bacterium called *P. gingivalis* produces many virulence factors such as fimbriae, lipopolysaccharides, and proteases (7). Fimbriae are surface appendages protruding from the outer membrane of the bacterial cell (8). A study showed that fimbriae have major roles in the binding as well as invasion of the host cells (8). The lipopolysaccharides are located in the outer membrane of *P. gingivalis* and are responsible for the activation of the host inflammatory response as well as disruption of bone remodeling process (8). This bacterium also secrets different types of proteases such as trypsin-, thiol-, caseinolytic-proteinases, and peptidases and they are responsible in destruction of the periodontal tissues (8). Additionally, proteases also make the microorganisms highly resistant to the host defense system (8). The third bacterium called *T. denticola* resembles a small oral spirochete, is frequently found with *P. gingivalis* in progressing periodontitis lesions (7).

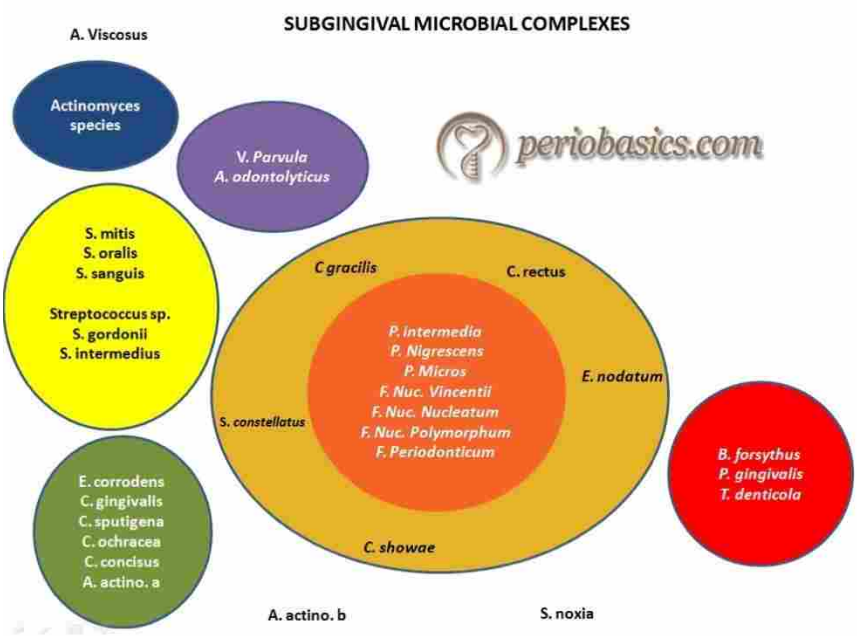



**Figure 2**: Five complexes. According to the presence of an organism and the related periodontal status, organisms are grouped into five complexes and they are red, orange, yellow, green and purple. Figure is taken from reference (4).

## 1.3 Glycosaminoglycans (GAGs)

The extracellular matrix (ECM) underlies all the epithelia, endothelia and surrounds the connective tissues that provides mechanical and physical support (9). The connective tissues of periodontium are composed of two soft tissues and they are gingival connective tissues and periodontal ligament (9). They are also composed of two hard tissues that are alveolar bone and cementum (9). The connective tissues can be classified into as fibrous (collagens and elastins) and non-fibrous (proteoglycans) (9). Proteoglycans are generally composed of a core protein attached with one or more glycosaminoglycans (GAGs) (10). The GAGs chains are linear polysaccharides composed of repeating disaccharide units that are negatively charged due to the occurrence of sulfate and carboxyl groups (10). The nature of the GAGs found in periodontal

tissues depends on their origin, with heparan sulfate occupying some 60% of gingival epithelium, dermatan sulfate occupying some 60% of the connective tissue and chondroitin-4-sulphate comprising the major GAGs (90%) of alveolar bone (11). The enzymatic degradations of GAGs involve two enzymes which are hydrolases and lyases (12). The hydrolase cleaves the glycosidic bond by the addition of the water molecule, whereas the lyase acts by depolymerizing the GAGs through beta-elimination reaction which is shown in figure 3 (12). The GAG degrading enzymes depolymerize GAGs through a beta-elimination mechanism characterized by the removal of a relatively acidic proton from the C5 carbon (the chiral center) of the uronic acid, and the release of the 4-linked hexosamine with the generation of a C4- C5 double bond at the uronic acid ring.

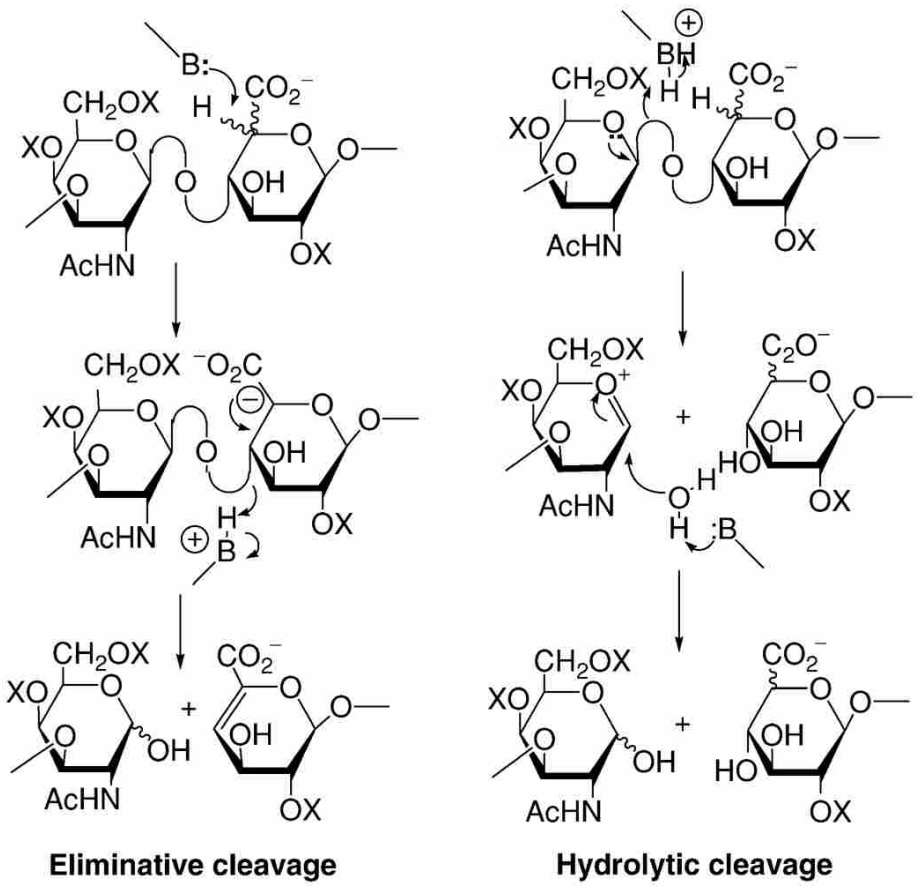



**Figure 3:** Degradation of GAGs by two different mechanisms. Panel 1(a) shows eliminative cleavage using beta-elimination reaction to break down the glycosidic bond. Panel 2(b) shows hydrolytic cleavage using water molecule to break down the glycosidic bond. Figures are taken from reference (12).

# 1.4 Structures of Glycosaminoglycans (GAGs)

GAGs are linear chains of polysaccharides composed of different repeating units of disaccharide (10). The GAGs are divided into four classes and they are: 1. Heparan sulfate (HS)/Heparin; 2. Chondroitin sulfate (CS) and Dermatan sulfate (DS); 3. Keratan sulfate (KS); 4. Hyaluronan.

HS/heparin are composed of repeating disaccharide units of hexuronic acid (HexA) and glucosamine (GlcN). The HexA can be either D-glucuronic acid (GlcA) or L-iduronic acid (IdoA), and the GlcN residues can be N-acetylated (GlcNAc), N-sulfated (GlcNS) or N-unsubstituted (GlcNH2) (10). CS and DS contain repeating disaccharide units of HexA and N-acetylgalactosamine (GalNAc), CS only contains GlcA, while some of the GlcA can be epimerized to IdoA in DS (10). Both HexA and GalNAc can be sulfated (10). KS is composed of alternating partially sulfated galactose (Gal) and GlcNAc residues (10). Hyaluronan is a non-sulfated GAG composed of repeating disaccharide units of GlcA and GlcNAc (10). All the different types of glycosaminoglycans are presented in the table 1 below and some of the structures are shown in figure 4.

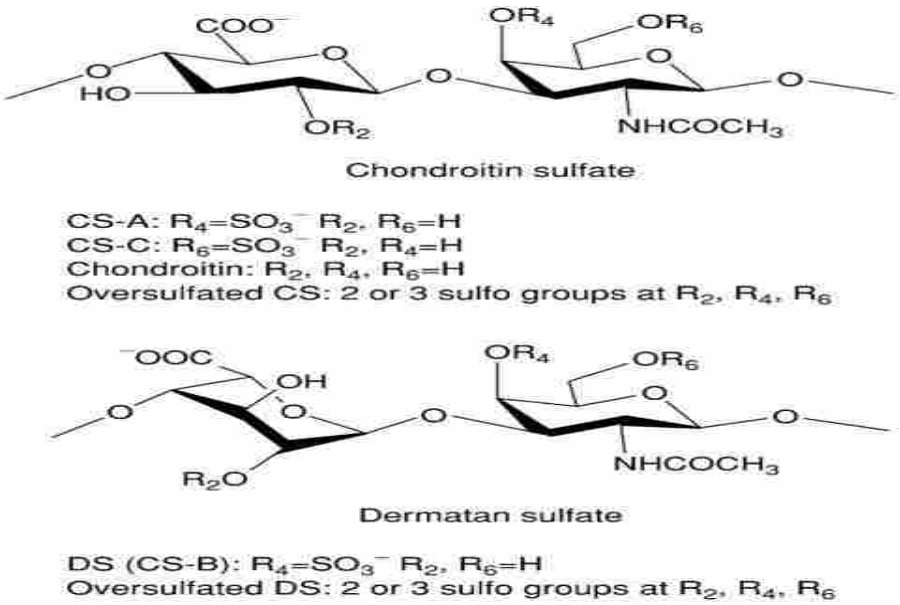



**Figure 4:** Structures showing Chondroitin sulfate A and C and Dermatan sulfate. Based on the positions of the sulfate groups, different types of GAGs are produced. Figures are taken from reference (12).

**Table 1:** Compositions of different types of glycosaminoglycans. Table 1 is showing different glycosaminoglycans having sulfate groups at varying positions.

| Glycosaminoglycans | Disaccharide units | Sulfate | Sulfate position |
|---|---|---|---|
| Chondroitin Sulfate A | D-Glucuronic acid, D- Galactosamine | O-sulfate | 4 |
| Chondroitin Sulfate B or Dermatan Sulfate | L-iduronic acid, D-Galactosamine | O-sulfate | 4 |
| Chondroitin Sulfate C | D-Glucuronic acid, D-Galactosamine | O-sulfate | 6 |
| Heparan Sulfate | D-Glucuronic acid/ L-iduronic acid, D-glucosamine | N-, O- sulfate | 2, 3 and 6 |
| Heparin | D-Glucuronic acid/ L-iduronic acid, D-glucosamine | N-, O- sulfate | 2, 3 and 6 |
| Keratan Sulfate | D-galactose, D-glucosamine | O-sulfate | 6 |
| Hyaluronan | D-Glucuronic acid, D-glucosamine | None | |

# 1.5 BFO 2285-2294 GAG -Degrading Operon System

The gene products we are working on are part of an operon system. Operons are groups of adjacent genes that are each transcribed into a single mRNA (13). These operons usually code for the genes that are in the same functional pathway (13). They are usually compact and the genes in the same operon are separated by less than twenty base pairs of DNA (13). Why several genes are placed in the same operon? The explanation is that the genes placed in the same operon will have similar expression at the same time together if induced under certain conditions. This also explains why the operons are believed to contain functionally related genes (13). We are currently working with two putative GAGs degrading enzymes which are BFO 2291 and 2294 based on bioinformatics analysis. BFO 2291 has putative heparinase II, III, chondroitin AC lyase and alginate lyse activities, while BFO 2294 is predicted to be similar to the enzyme KDPG/KHG aldolase. These gene products are predicted to be operated by an operon system and

together they are involved in the degradations of GAGs (Figure 5).

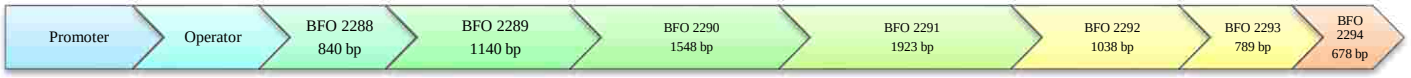



**Figure 5**. Putative Operon Structure. The figure is showing the seven gene products in an operon system which will all be transcribed together by T7 RNA polymerase bound to the promoter when an inducer binds to the operator.

**Table 2:** Bioinformatics analysis showing putative functions of the seven gene products

| Gene Products | Putative Functions |
|---|---|
| BFO 2288 | 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase |
| BFO 2289 | Lipoprotein |
| BFO 2290 | Arylsulfatase |
| BFO 2291 | GAG lyase |
| BFO 2292 | Carbohydrate kinase |
| BFO 2293 | Gluconate-5-dehydrogenase |
| BFO 2294 | KDPG/KHG aldolase |

## 1.6 Significance of Research

The main question that one might ask is why we are conducting research on periodontitis or in general oral health? The table 3 shows the global prevalence of oral health conditions (14). It highlights that periodontitis ranks 6$^{th}$ in the global prevalence of oral health conditions and affects 11% of the world population, while the untreated decay (caries) of permanent teeth ranks 1$^{st}$ in the global prevalence and affects 35% of the world population (14). The treatment options are currently limited and there is a need for effective treatment for this disease (1). In a healthy state, the periodontium maintains local tissue homeostasis by balancing its immune response to the local microbial ecosystem but in a deceased state this homeostasis is broken (1). This process forces the host immune system to destroy its own tissue, proliferates the pathogenic microorganisms and also further aggravates the host immune system response (1). This vicious cycle is the key to periodontitis and the main aim should be to develop therapeutics to break this cycle and to restore the tissue homeostasis in periodontium (1). Research into characterizing the following gene products BFO 2291 and BFO 2294 can guide us in understanding the degradation of periodontal tissues during periodontitis. By initially determining the structure and function of each enzyme, we can understand how each of the enzymes are involved in the GAGs degradations during periodontitis. By studying the kinetics of GAGs degrading enzymes, one can understand the nature of the enzyme and how they behave. By using X-ray crystallography, we can elucidate the structure of the enzyme which can lead to enzyme-substrate specificity and based on this we can also develop drugs which can inhibit these enzymes to slow down or halt the progression of periodontitis.

**Table 3:** Global prevalence of oral health Conditions

| Conditions | World ranking of disease in prevalence | World Prevalence (%) |
|---|---|---|
| Untreated decay (caries) of permanent teeth | 1st | 35% |
| Severe periodontal (gum) disease | 6th | 11% |
| Untreated caries of deciduous teeth | 10th | 9% |
| Severe tooth loss | 36th | 2% |

# 2. DEGRADATION OF CHONDROITIN SULFATE A

Our main hypothesis can be summarized by the figure 6 below. The enzymes identified from *T. forsythia* are believed to be responsible for degrading chondroitin sulfate A, one of the main components of periodontal tissues. Chondroitin sulfate A is broken down by BFO 2291 (GAG lyase) into disaccharides or oligosaccharides made up of glucuronic acid and N-acetyl galactosamine linked by glycosidic bond. Then, several other enzymes in the operon further break down the oligosaccharides into 2-keto-3-deoxy-6-phosphogluconate (KDPG). Then BFO 2294 (Aldolase) acts on (KDPG) to produce pyruvate that enters into Kreb cycle and electron transport chain for the production of ATP. In conclusion, we can summarize the hypothesis by saying that *T. forsythia* produces certain degradative enzymes that are involved in the GAGs degradation pathway with the production of ATP used to drive its cellular function.

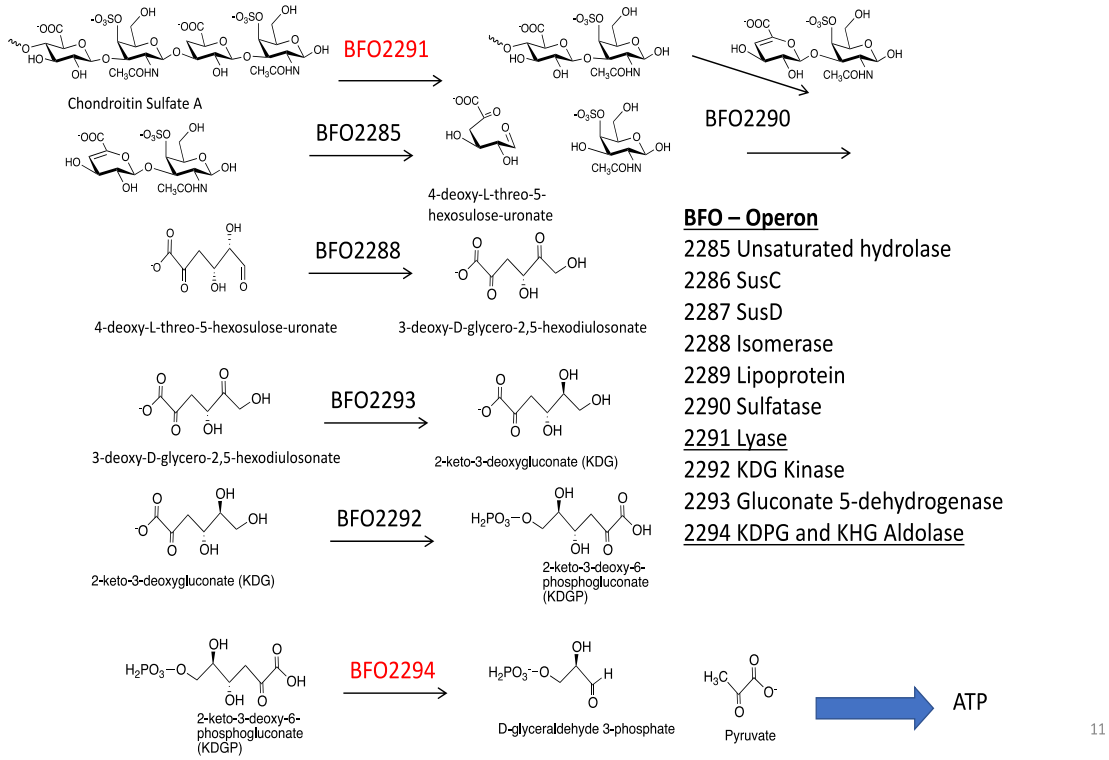



**Figure 6.** BFO operon system showing different enzymes responsible for the degradation of Chondroitin Sulfate A. First, Chondroitin Sulfate A is broken down by BFO 2291 into disaccharides or oligosaccharides made up of glucuronic acid and N-acetyl galactosamine. Down the road, glucuronic acid is detached from the oligosaccharide by unsaturated hydrolase. This molecule is further broken down by various enzymes into KDPG which is broken down into D-glyceraldehyde-3-phosphate and pyruvate. Finally, ATP is produced using Krebs cycle and electron transport chain to drive the cellular functions of the bacteria.

# 3. METHODS AND MATERIALS

## 3.1 Bioinformatics and Construct Design

As previously discussed, BFO 2290 and BFO 2294 are part of the operon system which are suggested via bioinformatics analysis to be involved in GAGs degradation. Using the sequences provided by National Centre for Biotechnology Information (NCBI), a series of bioinformatics resources will help to predict the functions and structures of the members of the operon. Some of the main programs to be used include: *Interpro* (15), *PyMOL* (16), *Phyre*$^2$ (17), *STRING-10* (18), and *SignalP* (19).

## 3.2 Construction of Recombinant Plasmids for BFO 2294

The first step of cloning was to design the primers considering any sequence tags. The restriction enzymes used were *Nde*I and *Xho*I. The recognition sequence of *Nde*I is 5'-CATATG-3' and *Xho*I is 5'-CTCGAG-3'. The overhang used for *Nde*I is 5'-CTAGCG-3' and for *Xho*I is 5'-TGCATG-3'. The forward primer (5'-CTAGCGCATATGATAAATATGGCAAGATTCGATAAGAAAG-3') contained recognition sequence of *Nde*I. The reverse primer (5'-TGCATGCTCGAGCTCACGGATAACACCTAATGCC-3') contained the recognition sequence of *Xho*I. These two primers ordered from IDT along with the JumpStart Taq Ready Mix from sigma were used to construct the gene of interest BFO 2294. The PCR mixtures included 1 μL of both the forward and reverse primers, 25 μL of JumpStart Taq Ready Mix, 5 μL of genomic DNA (*Tannerella forsythia* 92A2 ATCC 43037) and 18 μL of nuclease free water to make up volume of 50 μL. The construct was amplified by PCR using the annealing temperature of $60^0$ C and DNA gel electrophoresis was done to check the size of the fragment which was 681 bp. The

pET-21b plasmid was used for the recombinant expression of the gene products (see appendix I). This plasmid is 5442 base pairs long and encodes for ampicillin resistance. Both the pET-21b and gene of interest were cut by the restriction enzymes *Nde*I and *Xho*I. The mixtures included 25 µL of PCR product or 20 µL of the plasmid, 2.5 µL of *Nde*I and *Xho*I, 5 µL of NEB cut smart buffer and 20 µL or 15 µL of nuclease free water. The mixtures were incubated at $37^0$ C for 1.5 hours and then inactivated at $65^0$ C for 20 minutes. After this process, both the digested PCR product and the plasmid were cleaned by Qiagen PCR Purification kit. The amplicon was then inserted between *Nde*I and *Xho*I sites in pET-21b plasmid. Then the fragments were joined by DNA ligase which is termed as ligation. The ligation mixtures included 5 µl of PCR product, 1 µL of plasmid, 1 µL of NEB DNA ligase, 2 µL NEB T4 DNA ligase buffer and 1 µL of nuclease free water to make up the volume of 10 µL. The mixture was incubated at $4^0$ C for the overnight and then inactivated at $65^0$ C for 10 minutes. The resulting designated plasmid was termed as pET-21b-BFO2294.

## 3.3 Custom Gene for BFO 2291

For BFO 2291, custom gene was ordered from Bio Basic Inc., ligated into pET-28a expression vector was codon-optimized for expression *Escherichia coli*. At first, the lyophilized powder was re-suspended in 40 µL of nuclease free water. This was the original stock. In a new centrifuge tube, 1:10 dilution of the stock was made with the nuclease free water.

## 3.4 Transformation and Protein Production

The plasmids (pET-21b) containing the amplicons (BFO 2294) were transformed and plated into an agar plate. The agar plates were made by mixing 5 grams of Tryptone from Biobasic, 2.5 grams of Yeast extract from Biobasic, 5 grams of NaCl from Bioshop and 7.5 g of Agar A from

Biobasic with 500 mL of Milli-q water. Ampicillin concentration of 50 μg/mL was used since the plasmid encodes for ampicillin resistance. The plasmid and amplicon underwent transformation into a cell line with high efficiency uptake such as the DH5α strain of *Escherichia coli* containing many copies of plasmids with the amplicons. The competent cells (DH5α strain of *E. coli*) were made competent by treating with calcium chloride. The competent cells were taken out from the freezer ($-80^0$ C) and allowed to thaw in the ice for 10 minutes. After 10 minutes, 1 to 2 μL of the plasmid containing the amplicon were transformed into the competent cell and allowed to cool in the ice for 30 minutes. For the custom gene (BFO 2291), 2 μL of the original stock DNA and 2 μL of 1:10 diluted stock DNA were transformed into the competent cell. After 30 minutes, the cells were heat shocked at $42^0$ C for 42 seconds and allowed to cool in the ice for 2 to 3 minutes. Then, 200 μL of lysogeny broth (LB) was added to a 1.5 mL micro-centrifuge tube containing the competent cell (DH5α strain of *E. coli*). It was then allowed to shake in a shaker at $37^0$ C at 240 rpm for an hour. After an hour, the solution was spread in an agar plate containing the correct antibiotic and incubated in an incubator at $37^0$ C overnight. One of the single colonies from many colonies were picked and inoculated in liquid media (5 mL) containing the appropriate antibiotic. The minipreparation using QIAprep Spin Miniprep Kit (50) was done from DH5α cells and subsequently transformed into BL21 (DE3) strain of *E. coli*. One of the single colonies from many colonies were picked and inoculated in liquid media (5 mL) containing the appropriate antibiotic. Then 5 mL of liquid culture was added to 1 L of LB media composing of 10 grams of tryptone from Biobasic, 5.0 grams of yeast extract from Biobasic, 10 grams of NaCl from Bioshop. The LB media containg the culture was allowed to shake at $37^0$ C at 240 rpm until the $OD_{600}$ reached approximately ~0.6. After reaching the appropriate $OD_{600}$, 500 μM of 1 mL of isopropyl β-D-1 thiogalactopyranoside (IPTG) was introduced for protein

induction at $16^0$ C at 160 rpm for overnight. The cells were spin down at 5000 rpm for 8 minutes at $16^0$ C and the pellets were collected for chemical lysis. The isolation of the protein from the cells can involve various forms of cell lysis techniques including chemical lysis, sonication, and French press (20). Chemical lysis involves introduction of 730 mM of 25 mL sucrose solution to the pellets and stirred until pellets were smooth. Then, 10 mg of lysozyme was added and stirred for 10 minutes. After this step, 25 mM of 50 mL deoxycholate solution was added and stirred for 10 minutes. The deoxycholate solution included 25 mM deoxycholic Acid, 20 mL of Triton X-100 (2% final), 50 mM of 50 mL Tris at pH 7.5, and 300 mM NaCl to make up a 1 L of solution. The lysozyme, deoxycholate and Triton X-100 were used to disrupt the cell walls and deoxycholate is effective in disrupting and dissociating many types of protein interaction. After this stage, 375 μL of $MgCl_2$ and 75 μL of DNase were added. The addition of DNase is responsible for cleaving up single-stranded and double-stranded DNA to decrease viscosity from the chemical lysate (21). DNase solution contained 666 μg/mL DNase from Biobasic, 6 mL of 100% glycerol and 4 mL distilled water to make up to 30 mL of solution. At this stage, the solution should be smooth. The solution was spun down at 15000xg at $4^0$ C for 35 minutes and the supernatant was collected.

## 3.5 IMAC, Anion Exchange Chromatography and Dialysis

Immobilized metal affinity chromatography (IMAC) followed chemical lysis which is a method to purify recombinant protein with a short affinity tag consisting of polyhistidine residues. IMAC is based on the interaction between a metal ion ($Cu^{2+}$, $Ni^{2+}$ and $Co^{2+}$) and specific amino acids which is histidine in this case (22). Subsequently, nickel nitrilotriacetic acid (Ni-NTA) agarose was used in order to purify the protein which was bought from ThermoFisher Scientific. Ni-NTA agarose uses nitrilotriacetic acid (NTA), a chelating ligand that binds the $Ni^{2+}$ ions coupled to a

solid support resin usually agarose (23). The supernatant was loaded on $Ni^{2+}$ ions coupled to a solid support resin (6% crosslinked agarose) and allowed to equilibrate on a rocker for 1 hour at $4^0$ C. At first, flow through was collected and then using wash buffer (5 mM imidazole, 300 mM NaCl and 50 mM Tris-HCl), low affinity bound proteins were removed. The elution buffer contained 50 mM Tris-HCl, 300 mM NaCl and 500 mM imidazole. Using imidazole gradient at constant pH 8, different fractions were collected. The imidazole gradient used were 5 mM, 25 mM, 50 mM, 75 mM, 100 mM, 125 mM, 150 mM, 175 mM, 200 mM, 225 mM and 250 mM Imidazole and the samples were run on SDS gel. All of the protein fractions were pooled together and were prepared for dialysis for buffer exchange using Fisherbrand dialysis tubing (50mm in width). For BFO 2291, the dialysis tube containing the protein was buffer exchanged in 20 mM sodium phosphate buffer at pH 7.8 for 1 hour and then it was switched into fresh 20 mM sodium phosphate buffer at pH 7.8 for overnight buffer exchange. For BFO 2294, the dialysis tube containing the protein was buffer exchanged in 50 mM Tris-HCl buffer at pH 8 for 1 hour and then it was switched into fresh 50 mM Tris-HCl buffer at pH 8 for overnight buffer exchange.

Further purification was done using an anion exchange chromatography where anion exchanger was used to remove the contaminants (24). For BFO 2294, the buffer used was 50 mM Tris-HCl at pH 8 and for the protein BFO 2291, the buffer used was 20 mM sodium phosphate at pH 7.8 because this protein is soluble in the sodium phosphate buffer. The concentration of salt used to elute the protein bound to the column was 1 M NaCl. After this step, the protein was concentrated using Amicon Ultra-15 centricons in a centrifuge at $4^0$ C until the desired protein concentration was achieved. The centricons have a filter with a cut-off limit (10 kDa) and particles such as salt or smaller proteins that fall below the limit will flow through,

while the proteins above the limit (including the expressed protein of interest) will collect above the filter and concentrate to a lower volume (25). Before working with the purified protein, the concentration of the protein must be determined. The reading of protein concentration was achieved using a Thermo Scientific Genesys 10S UV-VIS Spectrophotometer to scan the absorbance from 300nm to 250nm. Rather than reading from only 280nm to indicate the presence of aromatic rings in protein structure, a range can depict peak broadening or if contamination is present at wavelengths not expected to read for protein. The absorbance values at 280nm would apply to Beer's Law to calculate protein concentration given the path length of the cuvette and extinction coefficient. The extinction coefficient for BFO 2291 was estimated via ProtParam to be 140525 $M^{-1}$ $cm^{-1}$ and BFO 2294 was 28585 $M^{-1}$ $cm^{-1}$.

## 3.6 Substrate Specificity For BFO 2291

Different substrates were tested for substrate specificity for the enzyme BFO 2291 (GAG lyase). Heparan sulfate as well as various types of chondroitin sulfate (chondroitin sulfate A, chondroitin sulfate B and shark chondroitin sulfate) were tested spectrophotometrically for the activity. The buffer used was 50 mM sodium phosphate at pH 6.5. A constant amount of each substrate (1 mg/mL) and 50 μg/mL of the enzyme were taken and lyase activity was measured based on the increase in absorbance by the double bond containing glucuronic acid at 232 nm at room temperature. (See figure 1A) (26).

## 3.7 pH Profile For BFO 2291

Different buffers were used for determining the optimum pH that gave the maximum enzymatic activity. For the enzyme, BFO 2291 (GAG lyase), different buffers at different pH were used in order to establish the maximum enzymatic activity. The enzyme was concentrated to 50 μg/mL.

The substrate used was chondroitin sulfate A and the concentration used was 1 mg/mL. The different buffers at 50 mM concentration were used. The buffers used were sodium phosphate at pH 6.5, 7.0, 7.5 and 8.0, Glycine-NaOH at pH 9.0 and 9.5, Hepes-NaOH at pH 7.0 and 7.5, Tris-HCl at pH 7.5, 8.0 and 8.5, Pipes-NaOH at pH 6.0 and 6.5, Sodium Citrate-NaOH at pH 4.5, 5.0 and 5.5 and Sodium Acetate-NaOH at pH 4.0 and 4.5.

## 3.8 Enzyme Kinetics For BFO 2291

The kinetic analysis of GAG lyase (BFO 2291) was done spectrophotometrically using the substrate chondroitin sulfate A. The GAG lyase activity was measured based on the increase in absorbance at 232 nm of the double bond containing glucuronic acid (Product) (26). A stock solution of 3 mg/ml of chondroitin sulfate A was prepared by dissolving 30 mg/ml of chondroitin sulfate A in 10 mL of 50 mM sodium phosphate buffer at pH 6.5. A range of concentrations were used ranging from 0.1 mg/mL, 0.2 mg/mL, 0.3 mg/mL, 0.5 mg/mL, 0.7 mg/mL, 1.0 mg/mL, 1.5 mg/mL, 2.0 mg/mL, 2.5 mg/mL. The enzyme concentration was 50 μg/mL. The total volume of the solution was 1 mL and they were added to a quartz cuvette with 1 cm path length. The absorbance readings were recorded for one minute at 232 nm for each of the concentration at room temperature. Each reaction was done in triplicates. The molar extinction coefficient, ε, of double bond containing glucuronic acid used was 3800 $M^{-1}$ $cm^{-1}$ to determine how much product was formed in μM/min (26).

## 3.9 Enzyme Kinetics for BFO 2294

BFO 2294 (Aldolase) activity was measured using coupled assay with L-lactic dehydrogenase. The reagents used were 50 mM Hepes buffer at pH 7.0, 6mM NADH, 500 mM stock concentration of the substrate 2-keto-3-deoxy-6-phosphogluconate (KDPG), L-lactic

dehydrogenase from rabbit muscle: Type II, ammonium sulfate suspension, 800-1,200 units/mg protein and they were added to a quartz cuvette with 1 cm path length (27). The total volume of the solution was 1 mL and different concentrations of the substrate were used (100 mM, 120 mM, 130 mM, 140 mM, 150 mM and 160 mM). The enzyme aldolase (50 μg/mL) was added last to initiate the reaction and the disappearance of NADH was monitored for an interval of 1 minute by observing the decrease in absorbance at 340 nm at room temperature (27). The molar extinction coefficient of NADH used was 6220 $M^{-1}$ $cm^{-1}$ (27). The enzymatic activity of aldolase was also checked by negative controls such as not adding aldolase or LDH at room temperature.

## 3.10 X-ray Crystallography

Commercially available sparse-matrix conditions were screened for positive crystallization conditions which require a wide selection of buffers, precipitating agents and additives with varying conditions including pH, temperature, and concentration. Vapour diffusion method can be performed in either the sitting-drop or hanging-drop format and both are feasible techniques that can be prepared in 24 or 96-well plates (28). Some suggested pre-made crystallization screens that may be used for from Microlytic Anatrace MCSG-1-4 crystal screen HT (1.7mL), Hampton Research Index Solution Set 1 reagents 1-48 HR2-144 and Solution Set 2 reagents 49-96 HR-144. The protein BFO 2294 was dialyzed in 50 mM Tris-HCl buffer at pH 8.0. The protein was concentrated using Amicon Ultra-15 centricons in a centrifuge at $4^0$ C until the desired protein concentration (14 mg/mL) was achieved. For the protein BFO 2294, an expansion plate was set up (24 wells) using different reagents at varying concentration including 40 M of Tacsimate (TM) at pH 7.0 and 11 M of Tacsimate (TM) at pH 9.0 from Hampton, 0.75 M to 6.25 M of glycerol at pH 8.2 as reservoir solution. Tacsimate is a unique crystallization reagent developed exclusively by Hampton Research. Tacsimate is composed of a mixture of

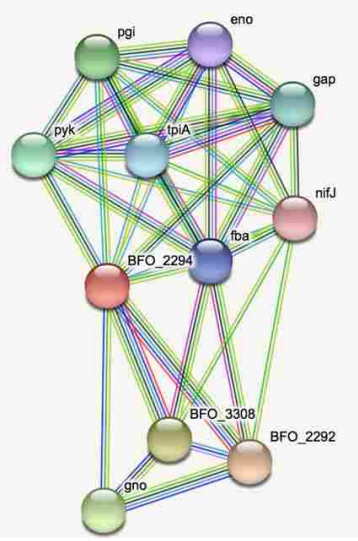
titrated organic acid salts. Tacsimate contains 1.8305 M Malonic acid, 0.25 M Ammonium citrate tribasic, 0.12 M Succinic acid, 0.3 M DL-Malic acid, 0.4 M Sodium acetate trihydrate, 0.5 M Sodium formate, and 0.16 M Ammonium tartrate dibasic. This mixture is titrated to the appropriate pH using sodium hydroxide and is available in pH 4, 5, 6, 7, 8, or 9 reagent formulations. The crystals were grown in the condition containing 6.25 M of glycerol, 400 μl or 40 M of Tacsimate 7.0 and 106 μl or 11 M of 9.0 at pH 8.2 as reservoir solution. Then a plastic coverslip was used where different ratios of reservoir solution (1-2 μL) and protein (2-3 μL) was mixed. The final concentration of the protein was 14 mg/mL. Then the crystal plate was kept in the incubator at $16^0$ C. After three weeks, the crystals were obtained and the shapes of the crystals were cubic. The sizes of the crystals were approximately 0.4 μm and the cryoprotectant used to soak the crystals was 30% Tacsimate 7.0 and 9.0. For the protein BFO 2291, 96 well plate was set up using 200 mM of imidazole as reservoir solution. The protein BFO 2291 was dialyzed in 20 mM phosphate buffer at pH 7.8. Then a plastic coverslip was used where different ratios of reservoir solution (1-2 μL) and protein (2-3 μL) was mixed. The final concentration of the protein was 19 mg/mL. Then the crystal plate was kept in at room temperature. After one day, the crystals were obtained and the shapes of the crystals were cubic. The sizes of the crystals were approximately 60 μm and the cryoprotectant used to soak the crystals was 30% ethylene glycol. All the crystals were sent to The Canadian Light Source for X-ray diffraction data collection and analysis (29).

# 4. RESULTS And DISCUSSION

## 4.1 Bioinformatics Analyses

The first step was to do bioinformatics analyses of the gene products in order to find out the putative functional characteristics. *STRING-10* provides a visual representation and interactive map of associated proteins (Figure 7) with protein-protein interactions surrounding the target protein thus leading to the understanding of the pathways within the species' genome (30). To figure out more specific details about the targeted sequence, InterPro is useful for predicting domains and important sequence-based features (Figure 8) that might be present (31). InterPro is classified as a database of protein families, domains and functional sites in which known features found in known proteins can be applied to new protein sequences in order to functionally characterize the new protein (15)

Additionally, *Phyre*$^{2}$ can help determine potential truncation sites when designing primers for recombinant proteins by comparing the structure of similar proteins based on sequence homology (32). *SignalP 4.1* is another useful bioinformatics tool that predicts whether the input sequence possesses a signal peptide (Figure 9) for cellular targeting (33). This tool is useful when cloning certain genes by cleaving the signal peptide.




**Your Input:**

| | | |
|---|---|---|
| BFO_2294 | *KDPG and KHG aldolase (226 aa)* | |

**Predicted Functional Partners:**

| | | Neighborhood | Gene Fusion | Cooccurrence | Coexpression | Experiments | Databases | Textmining | [Homology] | Score |
|---|---|---|---|---|---|---|---|---|---|---|
| BFO_2292 | *Carbohydrate kinase, PfkB family (346 aa)* | ● | ● | ● | | | ● | | | 0.998 |
| BFO_3308 | *Carbohydrate kinase, PfkB family (355 aa)* | | | ● | | | ● | | | 0.986 |
| gno | *Gluconate 5-dehydrogenase (263 aa)* | ● | | | | | | | | 0.903 |
| pgi | *Glucose-6-phosphate isomerase (448 aa)* | | | | | | ● | | | 0.887 |
| pyk | *Pyruvate kinase (481 aa)* | | | | | | ● | | | 0.836 |
| gap | *Glyceraldehyde-3-phosphate dehydrogenase, type I (336 aa)* | | | | | | ● | | | 0.829 |
| tpiA | *Triose-phosphate isomerase; Involved in the gluconeogenesis. Catalyzes stereospecifically the conversion of dihydrox...* | | | | | | ● | | | 0.825 |
| fba | *Fructose-1,6-bisphosphate aldolase (353 aa)* | | | | | | ● | | | 0.824 |
| eno | *Phosphopyruvate hydratase; Catalyzes the reversible conversion of 2- phosphoglycerate into phosphoenolpyruvate. It i...* | | | | | | ● | | | 0.819 |
| nifJ | *Pyruvate synthase; Oxidoreductase required for the transfer of electrons from pyruvate to flavodoxin (1180 aa)* | | | | | | ● | | | 0.817 |

**Figure 7:** Interactive map and predicted functional partners of BFO 2294 provided by *STRING-10*. Other associated proteins selected for the project based off this bioinformatics tool are BFO 2292 and BFO 3308 (30). It suggests that BFO 2292 and BFO 3308 are in the same pathway as BFO 2294 and they are involved in the metabolism of carbohydrate.

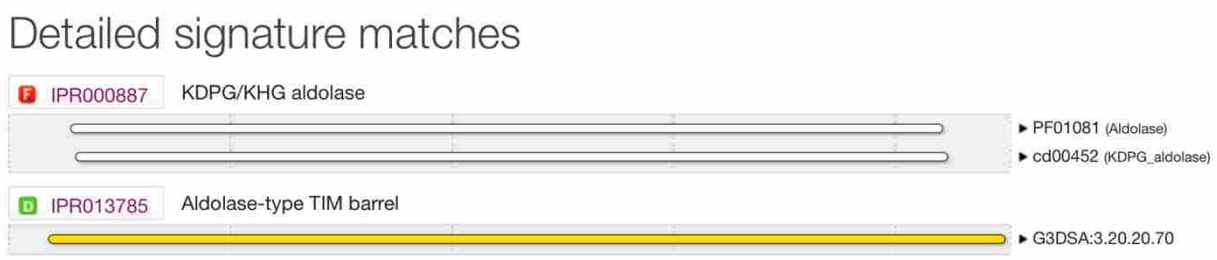



**Figure 8:** InterPro results for BFO 2294 suggest signature matches to a KDPG/KHG aldolase domain eluding to a TIM barrel structure. BFO 2294 corresponds to the KDPG/KHG aldolase family (31).

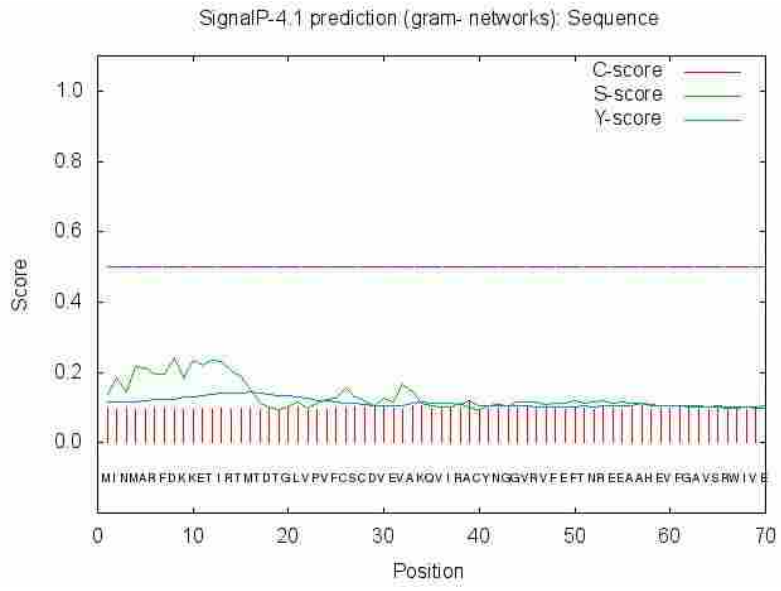



**Figure 9:** SignalP 4.1 graph displaying signal peptide scores to distinguish predicted localization of BFO 2294 as a putative secretory protein. C-score (red) represents the raw cleavage site score; S-score (green) represents the signal peptide score; and Y-score (blue) represents the combined effects of the C- and S- scores for a more accurate predicted cleavage site. In the case of BFO 2294, no signal peptide exists because it is targeted towards bacterial cytoplasm.

## 4.2 IMAC and Protein Purification

Immobilized metal affinity chromatography (IMAC) was used to purify the recombinant proteins BFO 2291 and 2294. IMAC utilizes the interaction between a transition metal ion $Ni^{2+}$ immobilized on a matrix and a specific amino acid called histidine (22). Histidine forms the strongest interaction with immobilized matrix because electron donor group on the histidine imidazole ring form coordination bond with immobilized transition metal (22). Affinity tags consisting of six polyhistidine residues are commonly used in IMAC and they can be incorporated either at the N or C terminus of the recombinant protein (22). The gene encoding the polyhistidine tag was incorporated in the N terminus of the pET-28a and pET-21b plasmids (22). There are various immobilized metal matrices available for use in IMAC and the choices include iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA) as metal ligands (22). The problem with IDA is that it can bind metal ion by three-coordination sites (22). Hence, the metal ions bind weakly to that three-coordination matrix resulting in metal leaching which causes lower yields and impure products (22). In our case, Ni-NTA was used because it can bind $Ni^{2+}$ ions by four-coordination sites, leaving two of the transition metal coordination sites free to interact with the histidine residues (22). As described in the Methods and Materials, wash buffer (50 mM Tris, 300mM NaCl and 5mM imidazole, pH 8.0) was used to wash away any non-specific associated protein and the protein of interest with polyhistidine tag will be tightly bound to the Ni-NTA beads (22). The tagged protein usually has higher binding affinity to the resin than other non-specific bound proteins (22). Thus, it is recommended to use the minimum amount of resin so that the tagged protein will fill most of the available binding sites in process reducing the binding of non-specific bound proteins (22). Proteins can be released from the complexes by reducing the affinity constant between the immobilized metal ion and the tagged

protein (34). There are different methods such as change in the salt concentration, pH and displacing by a competitive agent (imidazole) which resembles the histidine residues that can be employed (34). In our case, different concentrations of imidazole were used to displace the bound polyhistidine tagged protein from the complex. The figure 10 shows the elution of BFO 2294 using a step gradient of 50 mM, 75 mM, 100 mM, 125 mM, 150 mM, 175 mM, 200 mM, 225 mM, 250 mM and 260 mM imidazole. The predicted molecular weight of this protein is around 24 kDa. The gel looks poor, potentially because of the elution buffer not being fresh. Then the fractions containing the proteins were pooled (75 mM to 250 mM) and dialyzed in a buffer to exchange salts and imidazole. In Biochemistry, dialysis is a simple technique for separating molecules in solution by diffusion (35). Diffusion is a process by which atoms, ions or molecules move from an area where they are high in concentration to the area where they are low in concentration (35). A semipermeable membrane, in this case the dialysis tubing only allows particles smaller than a certain size to pass through it which are salts, imidazole and other contaminants. Thus, the protein of interest will be retained in the dialysis tubing. During dialysis in a buffer (dialysate), some of the salts and imidazole will move out from the area of high concentration (dialysis tubing) to the area of low concentration (dialysate) until an equilibrium is established. After 1 hour, concentrations of the salts and imidazole in the sample protein will be reduced to a significant amount. When we replace the old buffer with the fresh new buffer, another equilibrium will be established between the sample and dialysate in the process further reducing the concentrations of the salts and imidazole in the sample. The basic idea of dialysis is to reduce the concentrations of the salt and imidazole from the protein sample which can otherwise impede the process of crystallization. The next step was to further purify the protein sample using ion exchange chromatography (36). Ion exchange chromatography is a common

technique for separating charged molecules such as proteins, peptides, amino acids and nucleotides (36). Proteins are made up of amino acids that contain both positively charged and negatively charged chemical groups. Proteins may carry a net positive charge, a net negative charge or no charge. The pH at which proteins have no net charge is called its isoelectric point or pI. A protein will carry a net negative charge in a buffer with pH greater than pI and therefore an anion exchange resin (positively charged) will be used for binding the protein. A protein will carry a net positive charge in a buffer with pH lower than pI and therefore a cation exchange resin (negatively charged) will be used for binding the protein. Anion exchange chromatography was performed to remove any other protein contaminants that may impede crystallography. The figure 14 shows protein band for BFO 2291 at 74 from different tubes containing the protein fraction. We can see the protein looks very pure after anion exchange chromatography. Anion exchange instead of cation exchange was performed for BFO 2294 because the theoretical isoelectric point (pI) of this protein is 5.57 and the buffer used was 50 mM Tris-HCl at pH 8.0. This makes the overall charge of the protein (BFO 2294) negative. Anion exchange was also performed for BFO 2291 because the theoretical isoelectric point (pI) of this protein is 6.44 and the buffer used was 20 mM sodium phosphate buffer at pH 7.8. This makes the overall charge of the protein (BFO 2291) negative. Anion exchange chromatogram for BFO 2291 and BFO 2294 are shown in figure 12 and 13. The SDS gel is shown in figure 14 that shows purified protein (BFO 2291) after anion exchange chromatography.

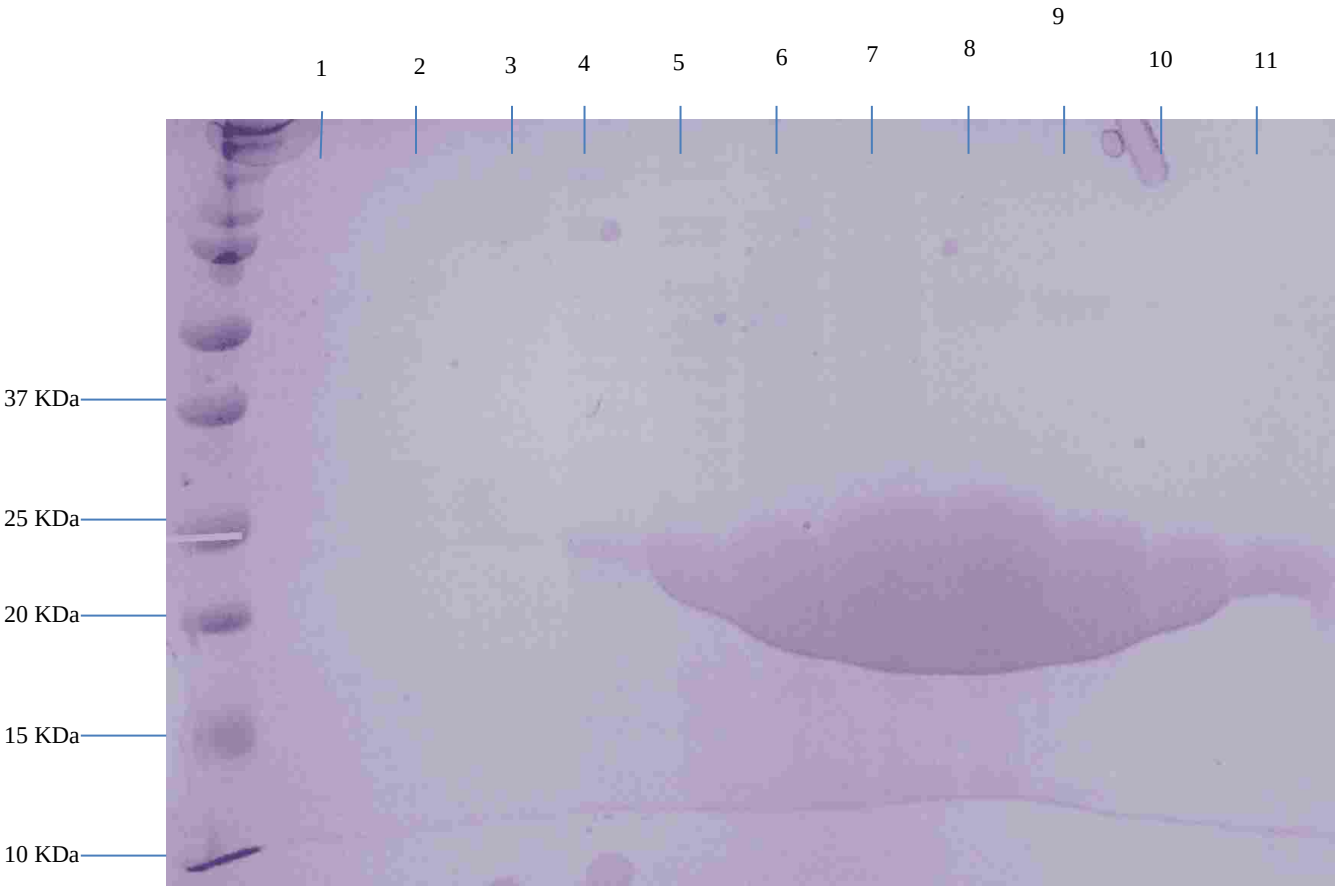



**Figure 10:** BFO 2294 IMAC Purification run at 200 V for 40 min. 12% (w/v) SDS PAGE analysis of protein fractions during Ni-NTA IMAC purification. Column 1 to 11 represent as follows: Flow through (FT), 50 mM imidazole, 75 mM imidazole, 100 mM imidazole, 125 mM imidazole, 150 mM imidazole, 175 mM imidazole, 200 mM imidazole, 225 mM imidazole, 250 mM imidazole and 260 mM imidazole. BFO 2294 which is predicted to be 24 kDa can be seen close to 25 kDa on the ladder band. Protein elutes strongly above 125 mM imidazole.

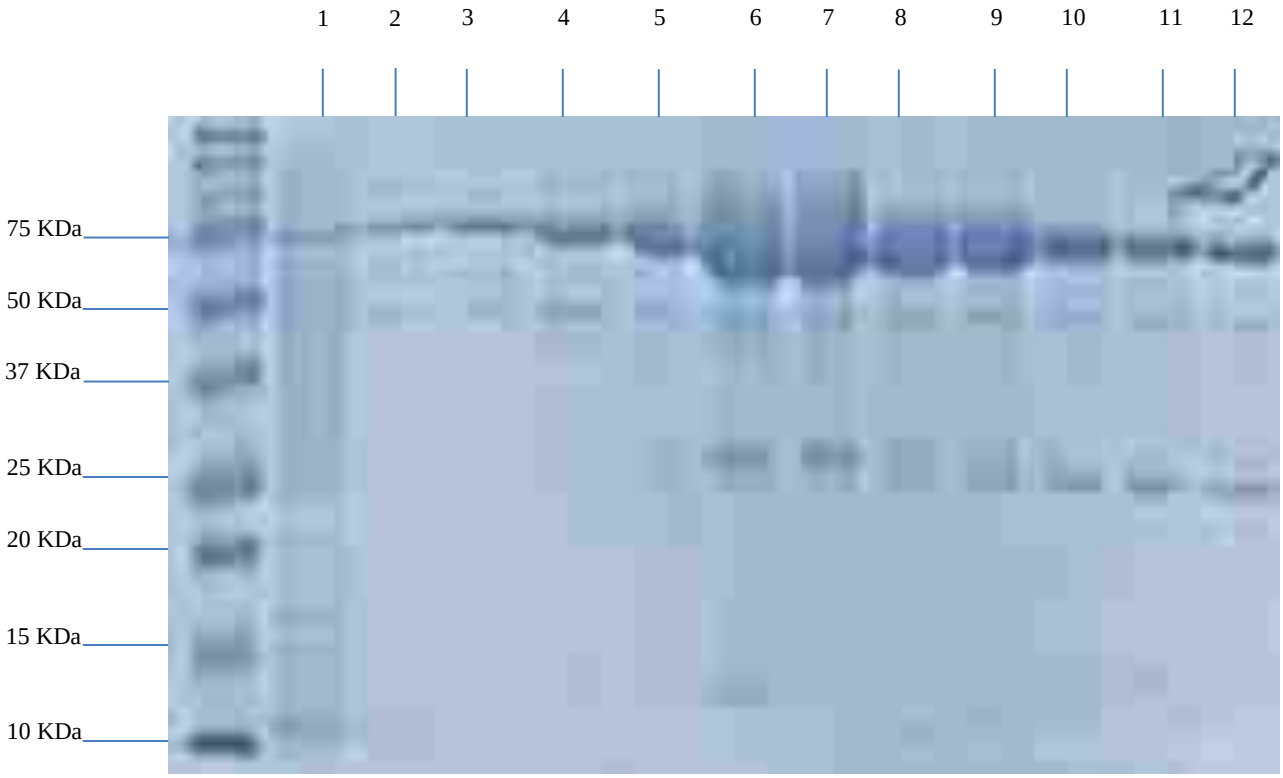



**Figure 11:** BFO 2291 IMAC Purification run at 200 V for 40 min. 12% (w/v) SDS PAGE analysis of protein fractions during Ni-NTA IMAC purification. Column 1 to 12 represent as follows: Flow through (FT), 5 mM imidazole, 25 mM imidazole, 50 mM imidazole, 75 mM imidazole, 100 mM imidazole, 125 mM imidazole, 150 mM imidazole, 175 mM imidazole, 200 mM imidazole. 225 mM imidazole and 250 mM imidazole. BFO 2291 which is predicted to be 74 kDa can be seen close to 75 kDa on the ladder band. Protein elutes strongly above 50 mM imidazole.

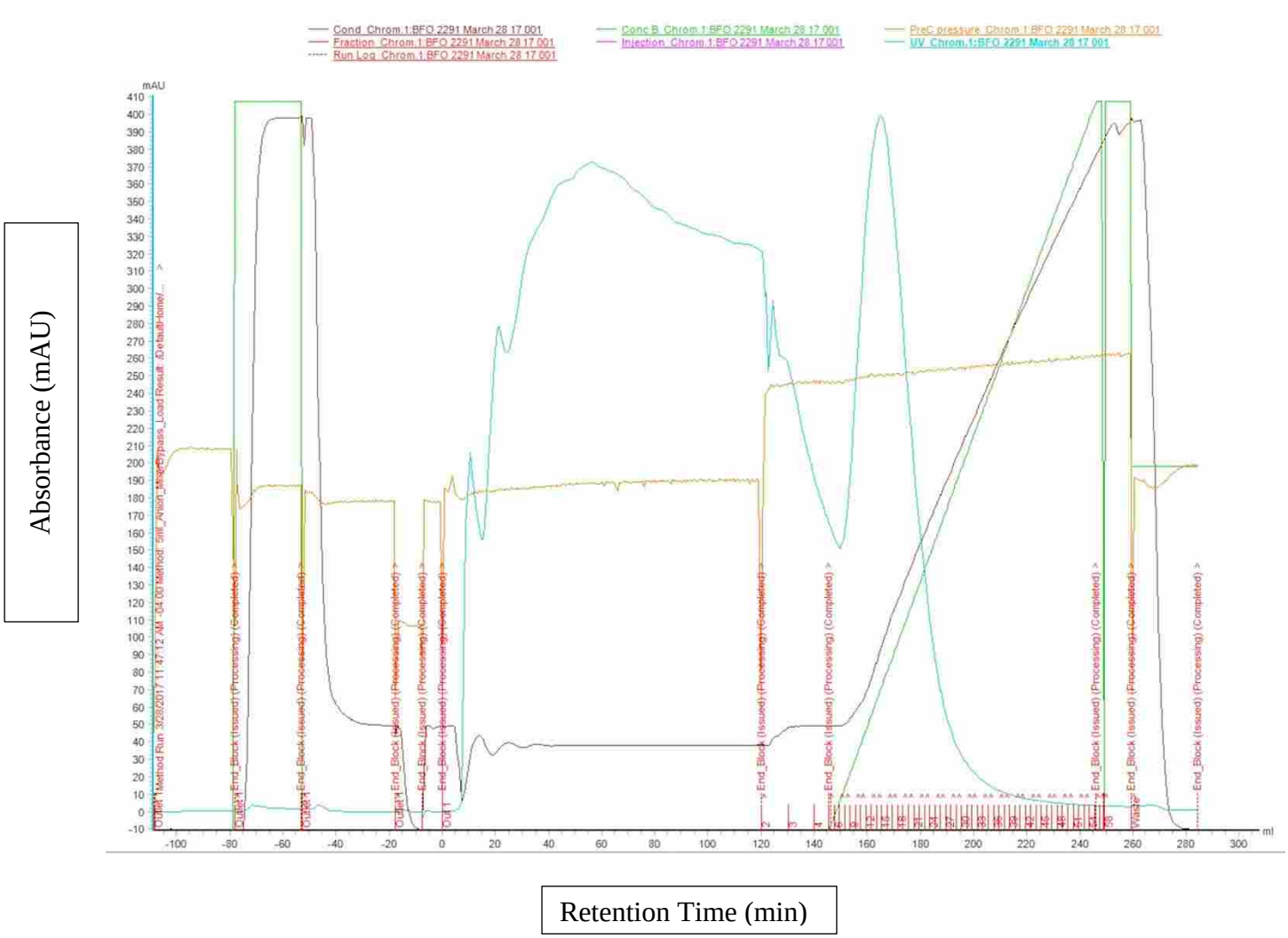
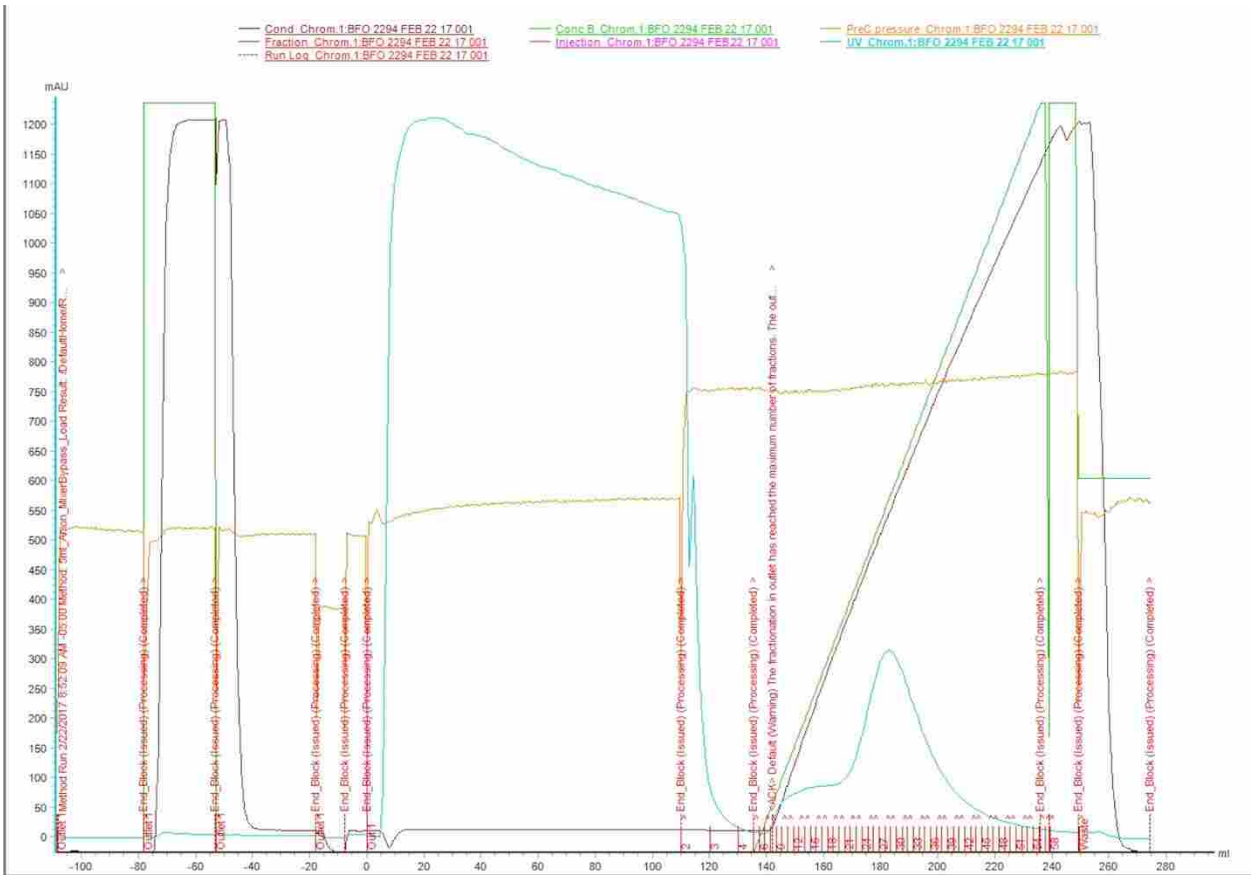



**Figure 12:** Anion exchange chromatogram for BFO 2291. This chromatogram shows a single large curve starting from about 150 mAU reaching the peak at about 400 mAU at 280 nm. The retention time of the peak is around 170 min. The protein starts to elute at about 375 mM of NaCl and is completely eluted by the time 200 min. The elution buffer used was 20 mM sodium phosphate, 1 M NaCl at pH 7.8

Absorbance (mAU)




Retention Time (min)

**Figure 13:** Anion exchange chromatogram for BFO 2294. This chromatogram shows a single curve starting from about 100 mAU reaching the peak at about 310 mAU at 280 nm. The retention time of the peak is about 183 min. The protein starts to elute at about 83 mM of NaCl and is completely eluted by the time 220 min. The elution buffer used was 50 mM Tris-HCl, 1 M NaCl at pH 8.0.

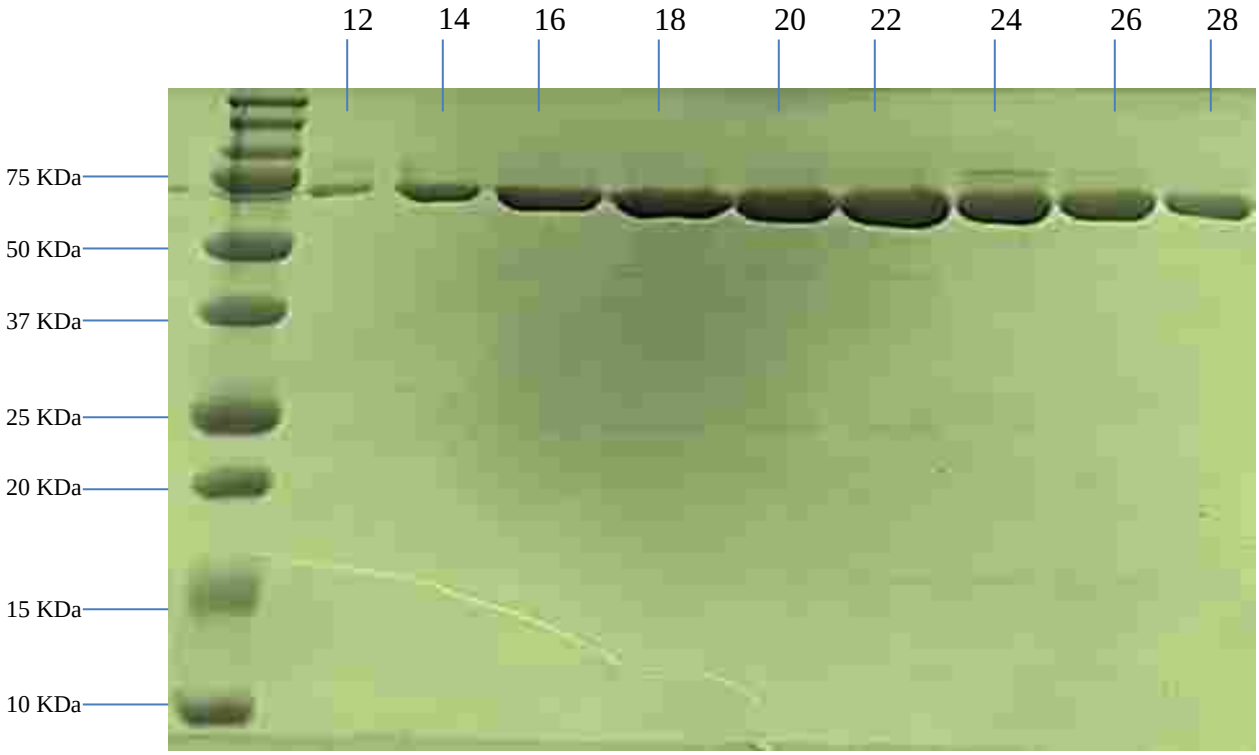



**Figure 14:** Anion exchange of BFO 2291. Purified protein collected from different tubes numbered 12 to 28 (during FPLC) and corresponds to the retention time between 150 min to 190 min. The protein starts to elute at about 375 mM of NaCl

## 4.3 Substrate specificity for BFO 2291

Based on bioinformatics analyses, BFO 2291 (GAG lyase) belongs to heparinase II/III like protein family (**see appendix II**). This protein was suggested to have a domain containing chondroitin AC/Alginate lyase activity. We eliminated alginate lyase based on the fact that alginate is not a glycosaminoglycans. Based on this information, we did some initial kinetics in order to establish the enzyme substrate relationship. If we look at figure 35 in appendix III, we can see that there is a small indication of increase in absorbance at 232 nm. By using the molar extinction coefficient 3800 $M^{-1}$ $cm^{-1}$, we calculated the amount of product formed to be 0.68 µM/min (26). For figure 36 in appendix III, there is a decrease in absorbance and that is not what

we should have seen if lyase activity was detected. It gave us a negative slope. Lastly, for figure 37 in appendix III, we observe some activity as well with shark chondroitin sulfate. In theory, it should show some activity because shark chondroitin sulfate contains all the mixture of chondroitin sulfate A, B and C. We calculated the amount of product formed to be 0.26 µM/min. The table 4 shows amount of product formed for different types of chondroitin sulfates. We can use these data as a pilot study since, the increase in absorbance for the substrate chondroitin sulfate A is not highly significant. Based on the analysis, we can summarize the results saying that chondroitin sulfate A is the best substrate for the enzyme BFO 2294 which is predicted to be chondroitin AC lyase based on bioinformatics. The table 5 shows the amount of product formed in the presence and absence of the enzyme. We can see that the products are increased in the presence of the enzyme showing the enzyme is required for the product formation.

**Table 4:** Amounts of products formed for different types of chondroitin sulfates by BFO 2291

| Compounds | Amount of Product Formed (µM/min) |
|---|---|
| Chondroitin Sulfate A | 0.68 |
| Chondroitin Sulfate B | -1.1 |
| Shark Chondroitin | 0.26 |

**Table 5:** Amounts of products formed for different types of chondroitin sulfates in the absence or presence of BFO 2291

| Compounds | Amount of Product Formed (µM/min) without BFO 2291 | Amount of Product Formed (µM/min) with BFO 2291 |
|---|---|---|
| Chondroitin Sulfate A | 0.13 | 6.5 |
| Shark Chondroitin | 0.16 | 5.0 |

# 4.4 pH profile for enzymatic activity for BFO 2291

The pH-optimum means the pH at which enzymatic activity is maximal (37). There are many macromolecular properties such as activity and protein stability that depend on pH. The pH is very important when it comes to enzymatic activity (38). Pepsin, one of the digestive enzymes

that breaks down protein molecules is highly active at pH 2.0-3.0 (39). Most of the dental bacteria live in an acidic condition (40). The pH of the oral cavity is around 7 but during periodontitis, the pH gets more acidic (40). It means that most of the enzymes produced by dental bacteria are more active in acidic pH. The extremely high or low pH can cause complete loss of the enzyme activity (37). The pH dependence of activity and stability usually resemble a typical bell-shaped curve with a single or several maxima and the maximum (maxima) is called pH-optimum (38). Such curves reflect the ionization of certain amino residues that must be in certain ionization state for enzymatic activity (41). Most enzymes are active within a narrow pH range from 5 to 9 (41). This phenomenon is an effect of pH on a combination of factors and they are binding of the substrate to an enzyme, the ionization of the substrate, the ionization states of the amino acid residues involved in the catalytic activity of the enzyme (41). In figure 15, we can see the pH-optimum is established. Based on the graph, we can conclude that sodium phosphate and pipes have the maximum activity at pH 6.5 but phosphate buffer was chosen as the buffer of choice because this specific protein was soluble in phosphate buffer. Before establishing the pH profile for this protein, we were doing the kinetics in phosphate buffer at pH 7.6 but after establishing the pH profile, we are able to study kinetics under optimal conditions. We also showed one of the rate curves (see figure 16) used to generate the figure 15.

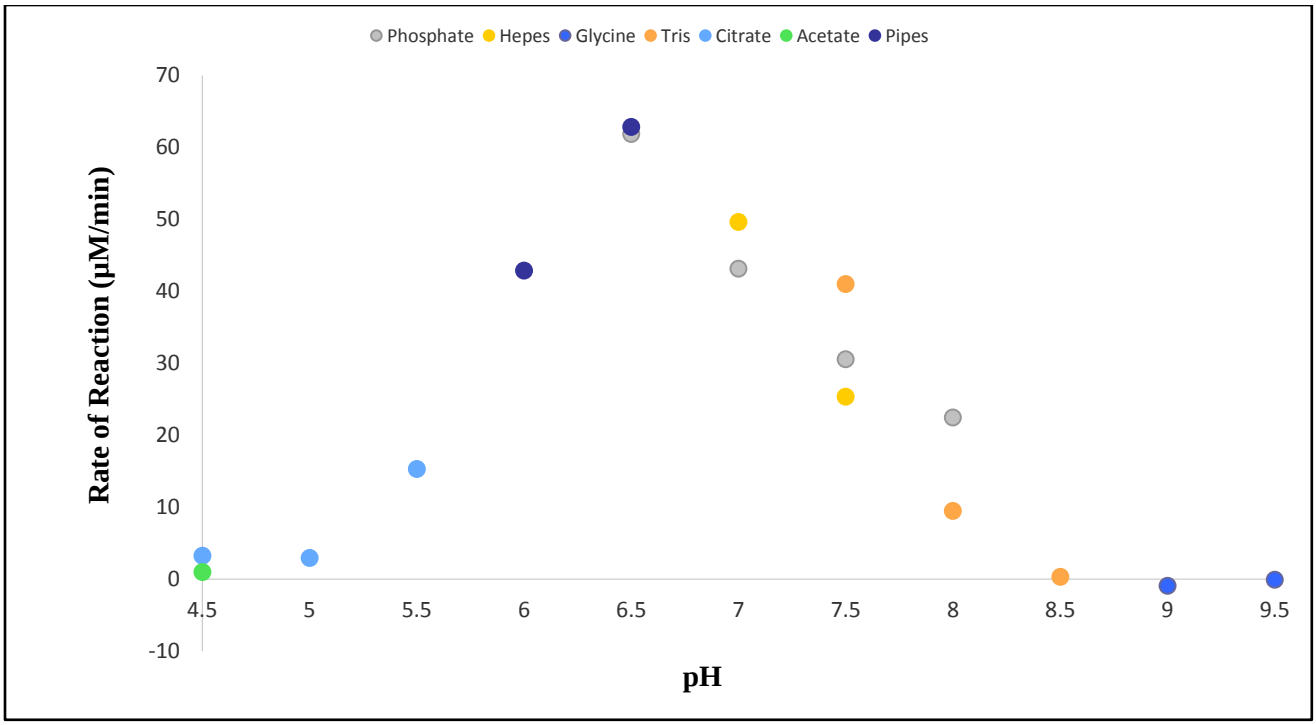



**Figure 15:** pH profile for BFO 2291. The curve shows the effects on reaction rate at different pH range. The enzyme concentration was 50 µg/mL, the substrate concentration of chondroitin sulfate A was 1 mg/mL and the concentrations of all the buffers were 50 mM.

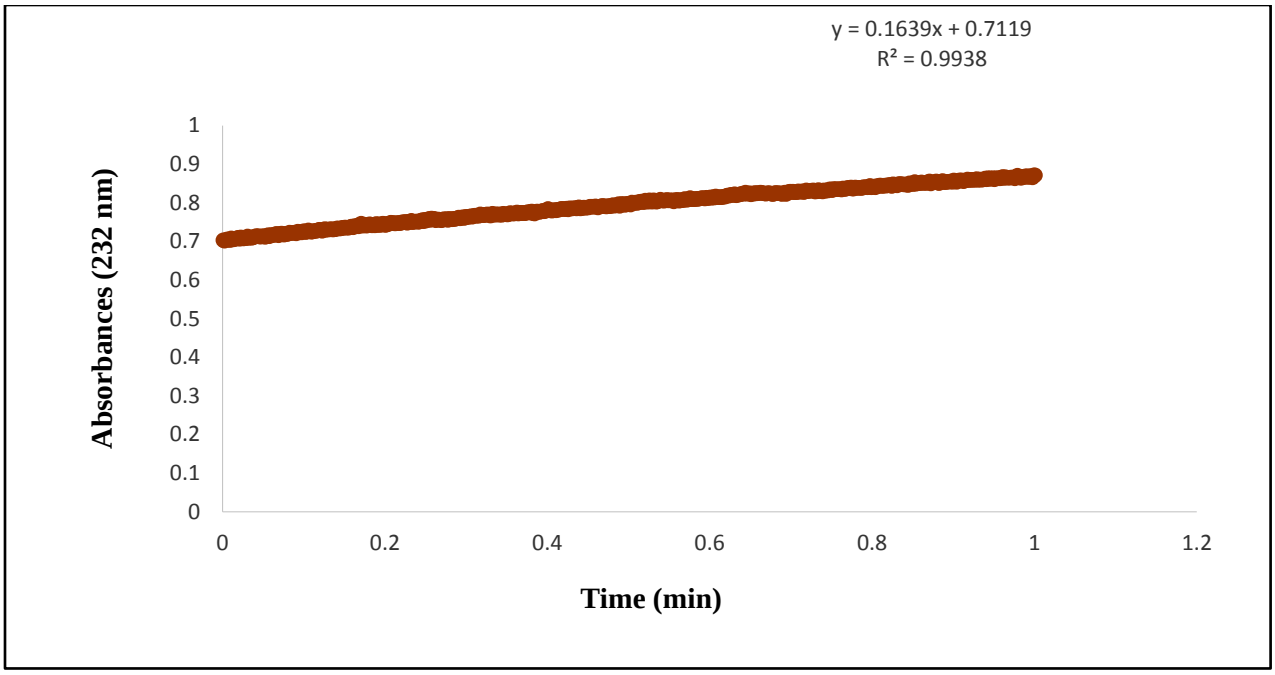



**Figure 16:** Rate curves for phosphate buffer at pH 7.0. One of the rate curves used to generate the figure 15. The enzyme concentration was 50 µg/mL, the substrate concentration of chondroitin sulfate A was 1 mg/mL and the concentrations of all the buffers were 50 mM.

# 4.5 Enzyme Kinetics for BFO 2291

The primary function of an enzyme is to increase the rate of reactions so they are compatible with the needs of the organism (42). In order to understand the function of an enzyme, we need a kinetic description of an enzyme activity (42). For most of the enzymes, initial reaction velocity ($V_0$) varies with substrate concentration [S]. The initial reaction velocity increases linearly as the substrate concentration increases and begins to reach a plateau at higher substrate concentration (42). The plateau occurs because all available enzymes are occupied by the substrate molecules and additional substrate molecules can only bind the enzyme when enzyme becomes available for binding. In Figure 17, the initial reaction velocity ($V_0$) is plotted against substrate

concentration [S] for BFO 2291. We are measuring the initial rate of reaction ($V_0$) right after adding the enzyme to the substrate at specific concentration because the reaction will slow down to zero as all the substrates are used up. The same concentration of enzyme was used for each reaction so that we keep the enzyme concentration constant and only changing around the substrate concentration. Each reaction was done in triplicates so that we have a reproducible data. After a certain substrate concentration, the graph levels off to a plateau as expected. The figure 18 shows smoothed curve with experimental values with error bars**.**

In order to analyze the kinetics data, we need to be familiarized with certain terms. The maximum velocity or $V_{max}$ is the theoretical maximum value for $V_0$ and Michaelis constant $K_M$ is the concentration of [S] producing a $V_0$ of $V_{max}/2$. The equation that accounts for the kinetic data shown in table 2 is called Michaelis-Menten equation. The equation is $V_0 = V_{max}$ [S]/$K_M$+[S]. When the substrate concentration is very low, [S] is much lower than $K_M$, $V_0 = (V_{max}/K_M)$ [S]. It means the rate is directly proportional the substrate concentration. When the substrate concentration is very high, [S] is much higher than $K_M$, $V_0 = V_{max.}$ It means the rate is at maximum, independent of the substrate concentration. If we look at the Michaelis-Menten equation, we can say that $V_0 = 1/2V_{max}$ when $K_M$ = [S]. Thus, we can conclude that $K_M$ = [S] at which reaction rate is half of it maximum value. For most of the enzymes, $K_M$ lies between $10^{-1}$ to $10^{-7}$ M. The question is what is defined by the $K_M$ of an enzyme? The Michaelis constant gives us the concentration at which half of the active site of an enzyme is filled by the substrate. By using the value of $K_{M,}$ we can calculate the fraction of active sites ($f_{ES}$) filled at any substrate concentration using the following equation: $f_{ES}$ = [S]/[S]+$K_M$. We can also make some assumptions about the strength of enzyme-substrate [ES] complex based on the $K_M$ values. The enzyme-substrate complex is the intermediate step when enzyme binds with specific substrate in

order to produce the product. If the $K_M$ value is large, then we can say the binding of enzyme-substrate complex is weak and if the $K_M$ value is small, then then we can say the binding of enzyme-substrate complex is strong. If that is true, then we can say a low $K_M$ value also means that the enzyme requires only small amount of substrate to become saturated. In this case, the $V_{max}$ will reach at low substrate concentration. If the $K_M$ value is large, then we can say the enzyme requires higher amount of substrate to reach the $V_{max.}$ We can also conclude saying that $V_{max}$ reveals the turnover number of an enzyme and that is the number of substrate molecules converted into product by an enzyme molecule in a unit time when the enzyme is fully saturated with substrate. It is also equal to kinetic constant $k_2$ or $k_{cat}$ can be calculated when the enzyme is fully saturated with the substrate molecule. The $V_{max}$ reveals the turnover number of the particular enzyme if the concentration of the active site $[E]_t$ is known. The equation can be written as $V_{max} = k_2[E]_t$ and can be rearranged as $k_2 = V_{max}/[E]_t$. For most of the enzymes, the values of $K_M$ lie between $10^{-1}$ to $10^{-7}$ M (42). All of the kinetic parameters were calculated via non- linear regression. By analyzing the $K_M$ value of BFO 2291, we can explain the behaviour of this enzyme. As we previously discussed, low $K_M$ value means that the enzyme requires only small amount of substrate to become saturated, while high $K_M$ value means the enzyme requires higher amount of substrate to become saturated. In the case of BFO 2291, the $K_M$ value of BFO 2291 is 396 μg/mL. The table 7 gives $K_M$ value for Chondroitin AC lyase from *B. stercoris* which is 388 μg/mL. If we compare the $K_M$ value of Chondroitin AC lyase from *Bacteroides stercoris* and the $K_M$ value of BFO 2291, we can conclude that the values of both enzymes are very close indicating nature of BFO 2291 is very similar to Chondroitin AC lyase. By examining the Table 4, we can see the value of $k_{cat}$ is 1.55/sec. The value of $k_{cat}$ reveals the turnover number which is defined as the maximum number of substrate molecules per unit time that get converted

by a single catalytic site at a given enzyme concentration. The turnover numbers of most enzymes with their corresponding substrates fall in the range from 1 to $10^4$ per second (42). When the concentration of the substrate is much greater than $K_M$, the rate of reaction is equal to $k_{cat}$ but most enzyme are not saturated with their corresponding substrates (42). When [S]<< $K_M$, the rate of reaction is much less than $k_{cat}$ because most of the active sites are unoccupied (42). The value of $k_{cat}/K_M$ is helpful in determining substrate specificity, catalytic efficiency and kinetic perfection (42). The rate constant for association of E+S is $k_1$ which sets an upper limit value for $k_{cat}/K_M$. How fast two molecules can bind (enzyme and substrate) is limited by how fast the two molecules can diffuse (bump into each other) so they can react (42). Thus, the maximum value of $k_{cat}/K_M$ for an enzyme is $10^8$-$10^9$ $M^{-1}Sec^{-1}$ (42). An enzyme with the $k_{cat}/K_M$ value in that range approaches the diffusion control limit and approaches kinetic perfection (42). It means the rate at which the enzyme's active site can convert substrate into product is limited only by the rate at which it encounters the substrate in solution (42). The higher the value of $k_{cat}/K_M$ for an enzyme, the more efficient an enzyme is. It also means that enzyme works better on that particular substrate. The $V_{max}$ of BFO 2291 is 62.0+/- 2.8 μM/min and the $V_{max}$ of Chondroitin AC lyase from *B. stercoris* is 76.7+/-1.5 μM/min. This again indicates that the values of $V_{max}$ are very close. By comparing the values of BFO 2291 with the Chondroitin AC lyase from *B. stercoris* tells us that the efficiency and nature of the two enzymes are very similar. The figure 18 shows the smoothed curve which is generated by applying the values in the equation $V_0 = V_{max}$ $[S]/K_M+[S]$.

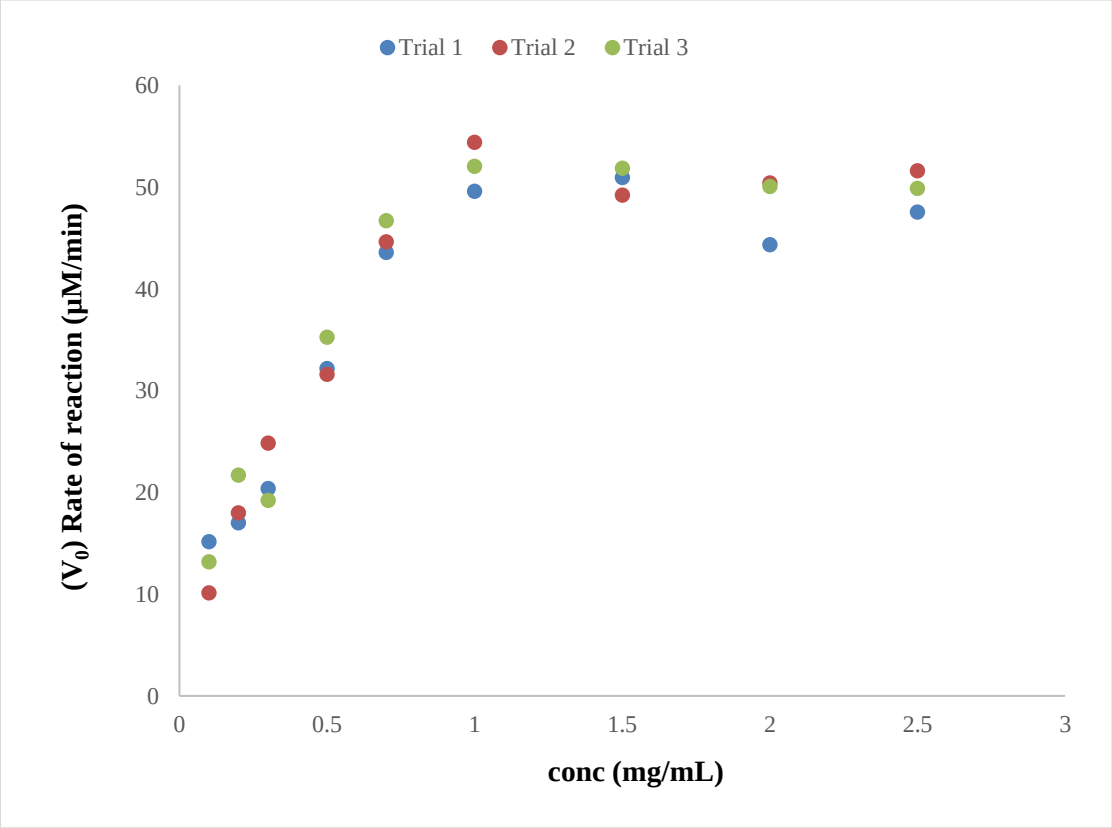



**Figure 17:** BFO 2291 Enzyme Kinetics. The figure shows the relationship between the increase in concentration of the chondroitin sulfate A[S] and the initial reaction velocity ($V_0$). The values for $K_M$, $k_{cat}$, and $V_{max}$ were found using the online software called $IC_{50}$ tool kit.

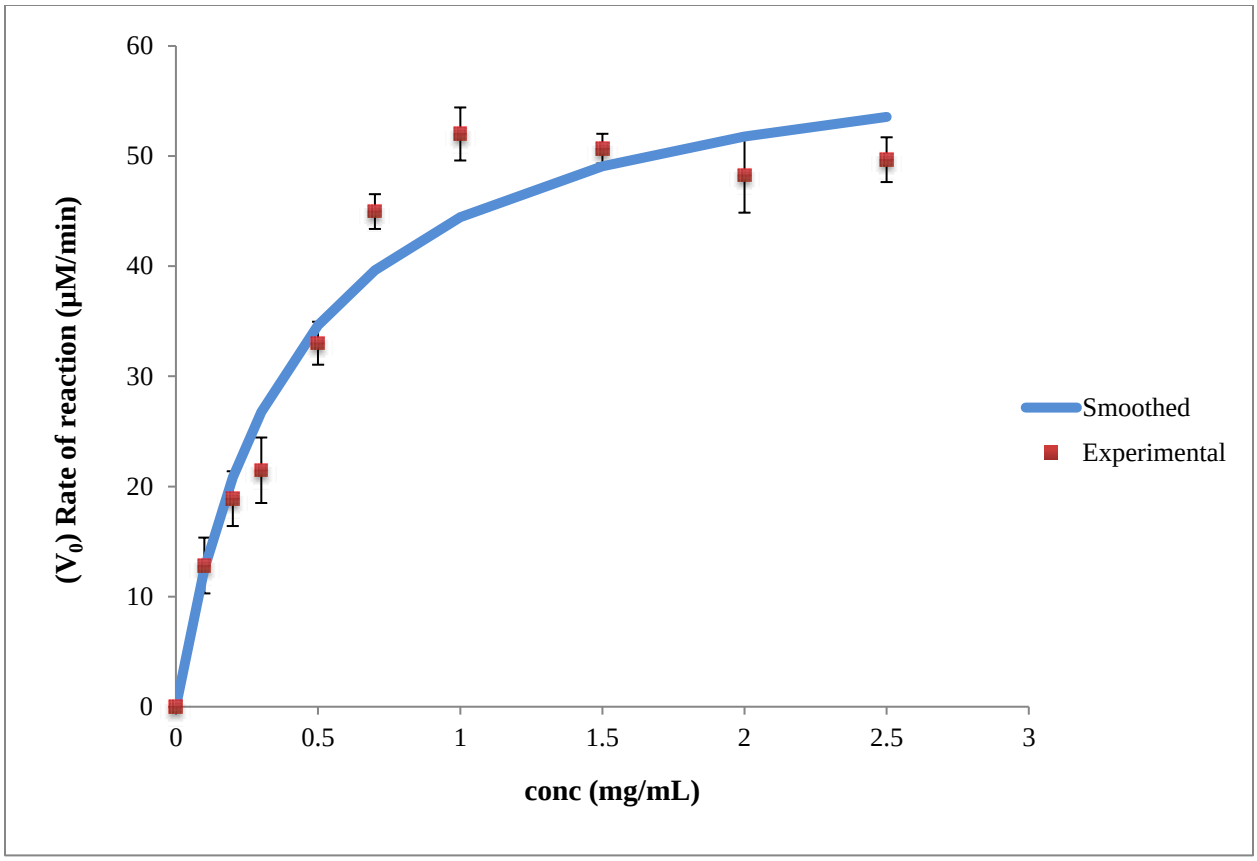



**Figure 18:** BFO 2291 Enzyme Kinetics. This figure shows smoothed curve with experimental values with error bars and the values for $K_M$, $k_{cat}$, and $V_{max}$ were found using the online software called $IC_{50}$ tool kit.

**Table 6:** Kinetic parameters of BFO 2291 and the substrate is chondroitin sulfate A

| $K_M$ (µg/mL) | $k_{cat}$ ($sec^{-1}$) | $V_{max}$ (µM/min) | $k_{cat}/K_M$ ($mg/ml^{-1}.sec^{-1}$) |
|---|---|---|---|
| 396 | 1.55 | 62.0+/- 2.8 | 3.92 |

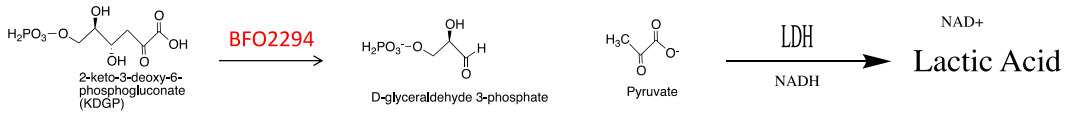
**Table 7:** Kinetic parameters of Chondroitin AC lyase from *Bacteroides stercoris* taken from (43)

| Substrate | $V_{max}$ (µM/min) | $K_M$ (µg/mL) |
|---|---|---|
| Chondroitin Sulfate A | 76.7+/-1.5 | 388 |

# 4.6 Enzyme Kinetics for BFO 2294

To measure the activity of BFO 2294 (Aldolase), we had to perform a coupled assay. According to figure 6, aldolase converts the substrate KDPG into pyruvate and D-glyceraldehyde-3-phosphate. The pyruvate is converted to lactic acid by lactic dehydrogenase using cofactor NADH which in turn gets oxidized to $NAD^+$. We are measuring the amount of NADH being converted to $NAD^+$ at 340 nm. The full reaction equation is shown in figure 19. The enzymatic activity was also tested by using different types of negative controls. This was done just to confirm whether any other factors are contributing to the activity that is seen in figure 23. The figure 20 shows there is no specific activity if the aldolase is not added in the mixture. In the figure 21, there is no activity without the aldolase and LDH and also in figure 22, there is no activity if the LDH is not added as well. In figure 23, there is a clear trend if both the aldolase and LDH are added. It proves that both the enzymes are needed in order to see the aldolase activity. If we look at the figure 37 in appendix IV, we can see that we only took single reading for each concentration because we just wanted to test whether the enzyme has aldolase activity or not. The data in figure 37 in appendix IV is inconclusive and needs further study to establish the kinetics. The table 8 shows amount of product being formed for different types of controls. Based on this table, we can say that maximum amount of product is being formed when KDPG, Aldolase and LDH are mixed together.

# Testing Aldolase Activity (BFO 2294)




16

**Figure 19:** KDPG aldolase is broken down by BFO 2294 into pyruvate. Pyruvate is converted into lactic acid using LDH and a cofactor NADH which in turn gets oxidized into $NAD^+$. The disappearance of NADH is measured at 340 nm.

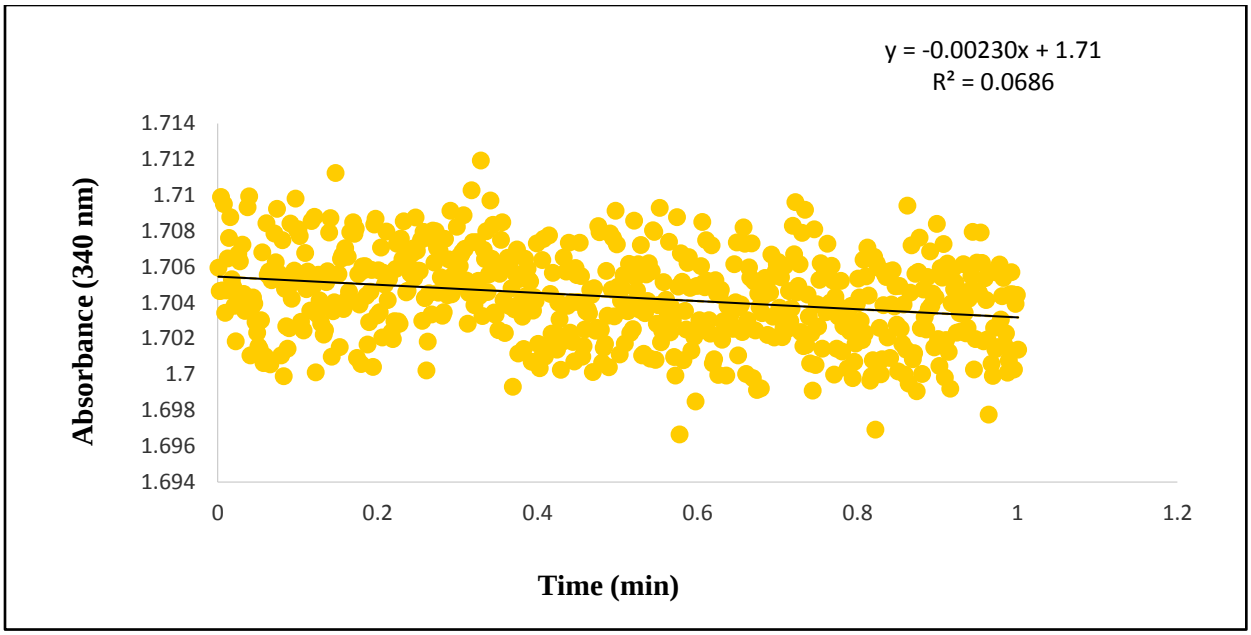



**Figure 20:** KDPG plus LDH. This figure shows there was no specific activity if the aldolase is not added in the mixture. The buffer used was 50 mM Hepes-NaOH at pH 7.0, 500 mM stock concentration of the substrate 2-keto-3-deoxy-6-phosphogluconate (KDPG), L-lactic dehydrogenase from rabbit muscle: Type II, ammonium sulfate suspension, 800-1,200 units/mg protein

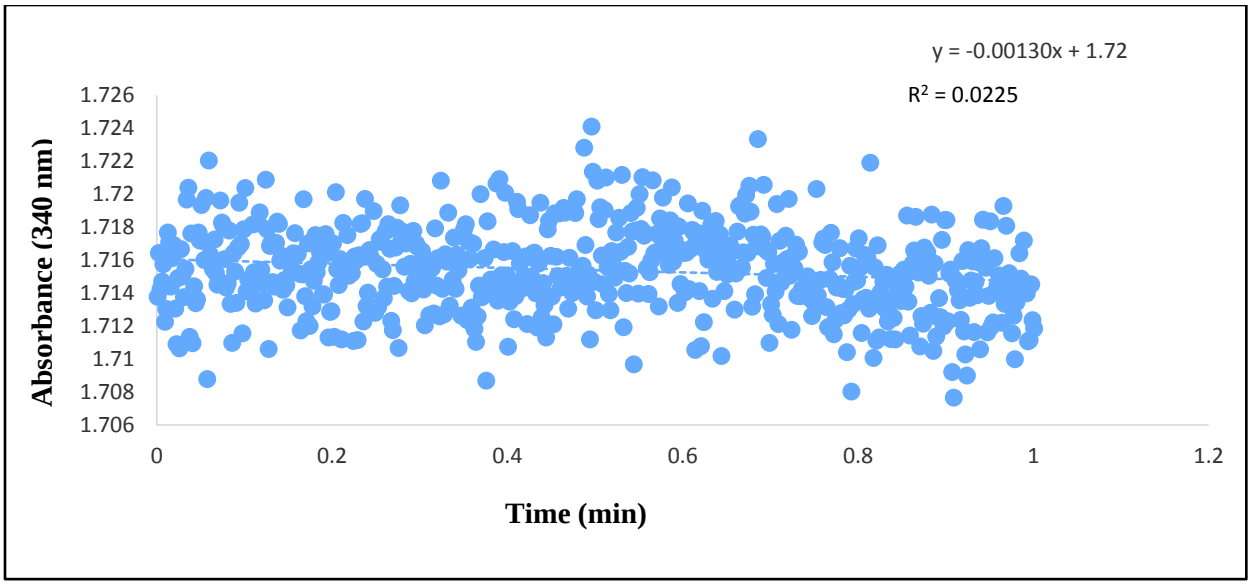



**Figure 21:** KDPG without LDH And Aldolase. This figure shows there was no specific activity if the LDH and aldolase are not added in the mixture. The buffer used was 50 mM Hepes at pH 7.0, and 500 mM stock concentration of the substrate 2-keto-3-deoxy-6-phosphogluconate (KDPG) was used.

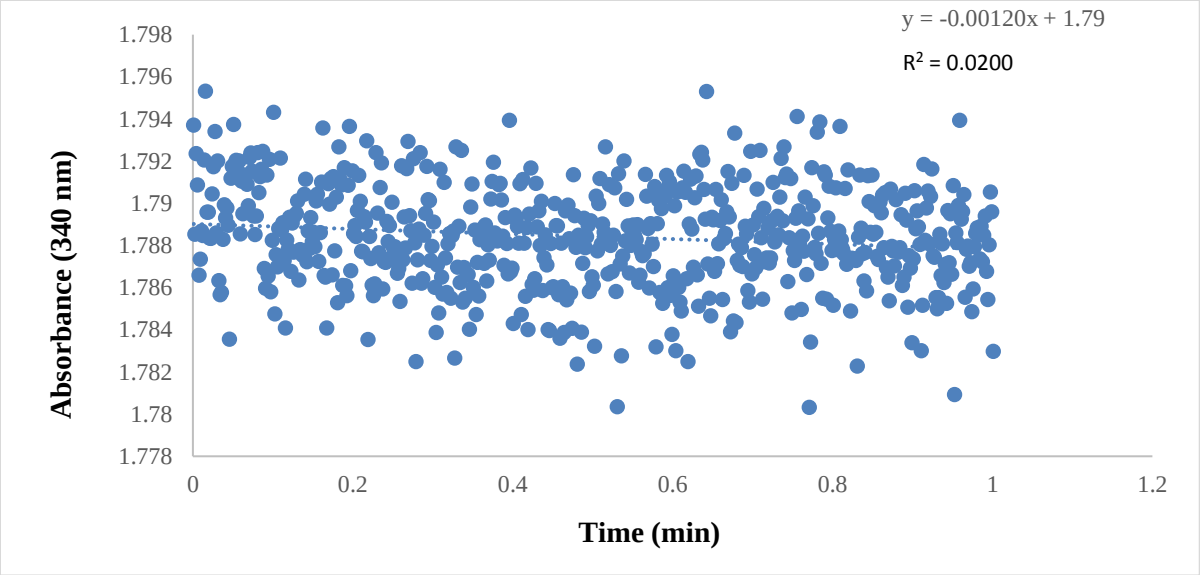



**Figure 22:** KDPG plus Aldolase. This figure shows there was no specific activity if the LDH is not added in the mixture. The buffer used was 50 mM Hepes at pH 7.0, 500 mM stock concentration of the substrate 2-keto-3-deoxy-6-phosphogluconate (KDPG), and the enzyme concentration was 50 μg/mL.

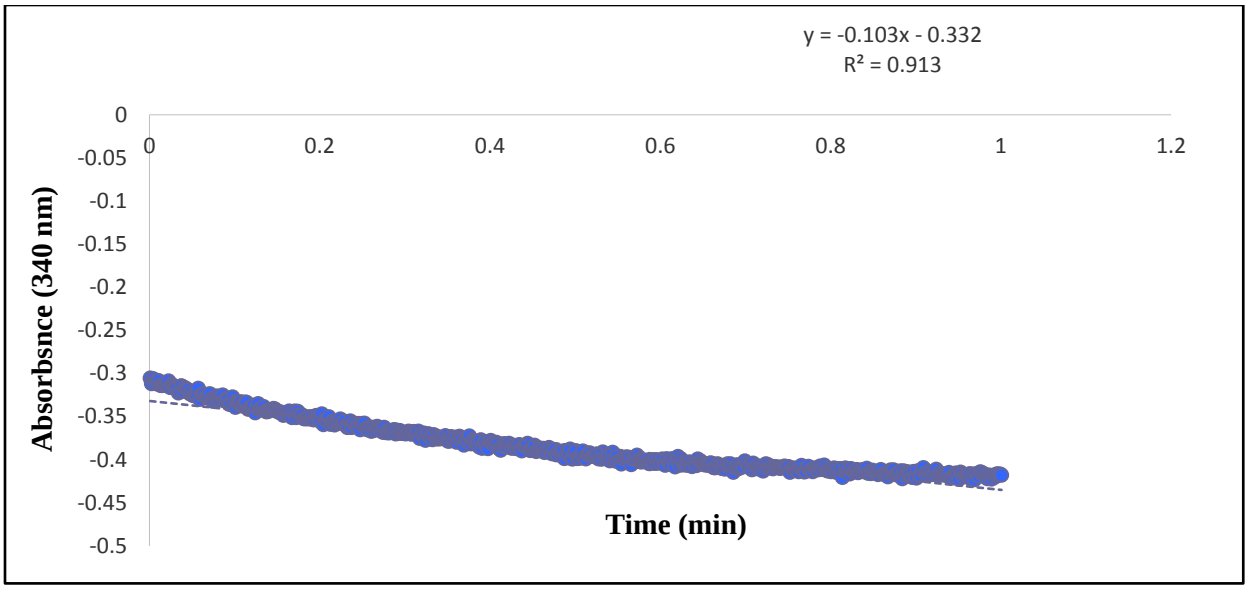



**Figure 23:** KDPG, Aldolase and LDH. This figure shows there was an activity if both the aldolase and LDH are added in the mixture. The buffer used was 50 mM Hepes at pH 7.0, 500 mM stock concentration of the substrate 2-keto-3-deoxy-6-phosphogluconate (KDPG), the enzyme concentration was 50 µg/mL, and L-lactic dehydrogenase from rabbit muscle: Type II, ammonium sulfate suspension, 800-1,200 units/mg protein

**Table 8:** Amount of product formed for different types of controls

| Controls | Amount of Product Formed (µM/min) |
|---|---|
| KDPG and LDH onl | 0.370 |
| KDPG without LDH and Aldolase | 0.209 |
| KDPG plus Aldolase | 0.193 |
| KDPG, Aldolase and LDH | 16.7 |

## 4.7 Principles of Protein Crystallization

X-ray crystallography is the most convenient method for determining the structure of a purified protein in crystal form (29). Protein crystallization was discovered by accident about 150 years ago and was developed as a purification tool in the late $19^{th}$ century (44). The aim of this technique is to obtain a three-dimensional molecular structure from a crystal (28). A crystal consists of atoms arranged in a pattern that repeats in three-dimensional structure (45). The pattern can consist of an atom or group of atoms, a molecule or group of molecules (45). The main feature of a crystal is having this pattern repeated for multiple times at regularly spaced interval with same orientation (45). The good quality of crystal for structure determination is a rate limiting step (28). Protein crystallization is mainly a trial and error procedure which requires protein solution to be slowly precipitated out of its solution (46). The presence of impurities, crystallization nuclei and the unknown factors play major roles in the crystallization of a protein (46). The general crystallization process is shown in the figure 24. The crystallization of the protein involves four steps and the first step is to check purity of the protein. If the purity of the protein is not sufficient based on SDS gel after running ion exchange chromatography, then further purification steps will be needed (46). The second step involves the protein to be dissolved in the suitable solvent from which it must be precipitated in crystalline form. The solvent could be the buffer in which the protein of interest is dialyzed or in extreme examples, a solvent called 2-methyl-2,4-pentanediol (46). The third step is to bring the solution in a state of supersaturation and in this step, small aggregates are formed which are known as nuclei that proceeds crystallization (46). The spontaneous formation of a nuclei requires surface tension energy and the crystal growth starts to begin when the energy barrier has been overcome (46). The spontaneous formation of a nuclei is best achieved at higher supersaturation because the

energy barrier can easily be overcome at higher supersaturation (46). Finally, the fourth step involved the formation of crystal after the nuclei is formed (46). The level of supersaturation must be reduced to a lower level to achieve a good crystal growth because high supersaturation would result in too many nuclei leading to many small crystals (46). The degree of supersaturation can be changed by changing the temperature and pH (46). The precipitation of the protein is achieved by increasing the concentration of the protein by adding a salt (salting out) or polyethyleneglycol (PEG) (46). Some proteins are poorly soluble in the solution but their solubility increases by the addition of salt (46). By slowly removing the salt, the protein can be precipitated (46).


**Figure 24:** Crystallization process. Protein solution is supersaturated through vapour diffusion which proceeds nucleation. From nucleation, crystal growth occurs and 3-D structure of the crystal is deduced.

## 4.8 Supersaturation, Nucleation And Crystal Growth

Crystallization can be divided into two processes and they are nucleation and crystal growth (47). In order to crystallize a macromolecule, it needs to be in a supersaturated state (44). Supersaturation is termed as a non-equilibrium condition in which the amount of macromolecule is in excess of the solubility limit (44). Then, as the protein associates into a regular lattice of a crystal, equilibrium is established again when the crystal is formed and matures (44). This phenomenon is explained by using the phase diagram in figure 25 below. This phase diagram is divided into different regions. The stable (undersaturated) region describes when the concentration of the macromolecules is lower than the solubility limit. In this region, the process of crystallization does not occur and the macromolecules are completely dissolved in the solution. In the metastable region, crystal growth is not visible because nuclei are not formed but crystal growth could occur if we put a nucleus in that solution. In this region, nuclei will form into crystals but no nucleation will occur. In the labile region, nuclei will form and crystals may grow from that nuclei. In the precipitation region, protein precipitates because of high concentration of both protein and precipitants. If the protein precipitates, then we need to either fix the concentration of the protein or precipitants or sometimes we need to design a whole new condition in order to avoid the precipitation region. Crystallization starts with nucleation and the nucleation of crystal determines many properties of crystalline phase (48). It is known that if nucleation occurs quickly then many crystals form simultaneously (48). These crystals deplete the medium of the solute and may end the process of nucleation at the later stages of crystallization (48). This leads to a formation of crystals of identical sizes (48). If the nucleation is slow then only a few crystals nucleate at one time (48). Thus, the supersaturation of the solution decreases slowly and the nucleation of the new crystals continues to happen (48). This

leads to formation of crystals of varying sizes and forms (48). In this way, we can control the size, size distribution, polymorphism and others properties of the crystals by controlling nucleation (48). The process of growth of crystals is well understood than the process of nucleation (44). Protein crystals grow by the mechanism of dislocation growth and growth by two-dimensional nucleation (44). It is important to note that both the nucleation and crystal growth depend on supersaturation (44).

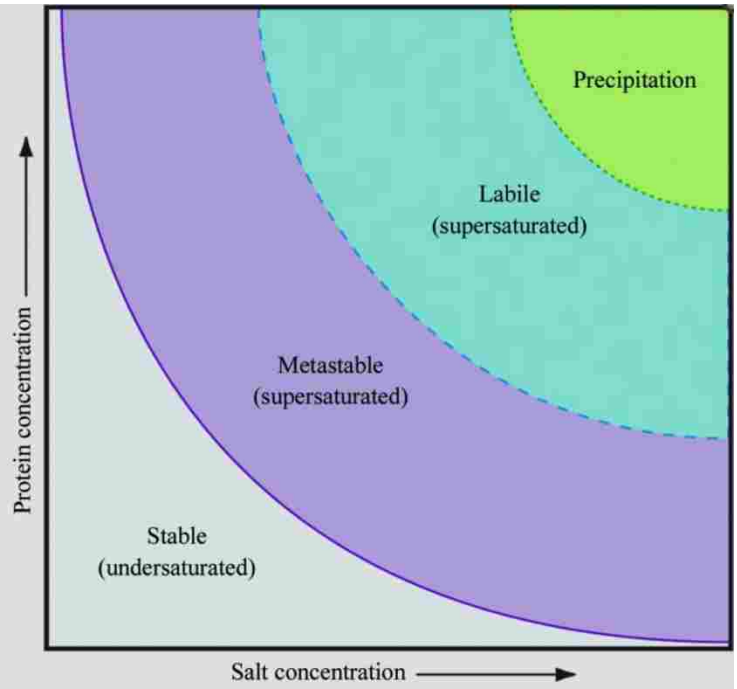



**Figure 25:** The process of crystallization goes through different region in the phase diagram. Stable region explains the situation when macromolecules are completely dissolved in the solution. In the metastable region, nuclei are not formed but crystal growth could occur if we put a nucleus in that solution. In the labile region, nuclei will form and crystals may grow from the nuclei. Finally, the precipitation region explains the condition when protein precipitates out of the solution.

## 4.9 Crystallization Techniques

There are many crystallization techniques can be used to grow crystals and they are Batch crystallization, Liquid-liquid diffusion and vapour diffusion. For our research purpose, we will concentrate on vapour diffusion technique. This technique is described in figure 26. This technique can be performed in either the sitting-drop or hanging-drop format and both are feasible techniques that can be prepared in 24 or 96 well plates (28). In the hanging drop vapour diffusion, the drops are prepared on a siliconized glass cover slip by mixing different ratios of protein solution to the different ratios of reservoir solution containing different precipitating agents, buffers and salts (28). In this method, the droplet contains protein, precipitating agents and buffer is equilibrated against the reservoir solution containing the same precipitating agent at higher concentration than the droplet (49). After a certain time period, equilibrium is established by the evaporation of the volatile component until the vapour pressure of the droplet equals the vapour pressure of the reservoir (49). The crystals will be formed in the droplet (49). Some suggested pre-made crystallization screens that may be used are from Microlytic Anatrace MCSG-1-4 crystal screen HT (1.7mL), Hampton Research Index Solution Set 1 reagents 1-48 HR2-144 and Solution Set 2 reagents 49-96 HR-144. The crystallization incubation period may range from days to months depending on the screens and conditions that viable crystals can be grown. In the figure 27, we can see growth of BFO 2294 crystals in a pre-made crystallization screens and in the figure 28, we can see the growth of BFO 2294 crystals in an expansion plate grown over a period of three weeks. Once crystals are obtained, they are sent to a facility to apply X-ray crystallography and data is collected for computational analysis using logarithmic construction of the 3D structure (29).

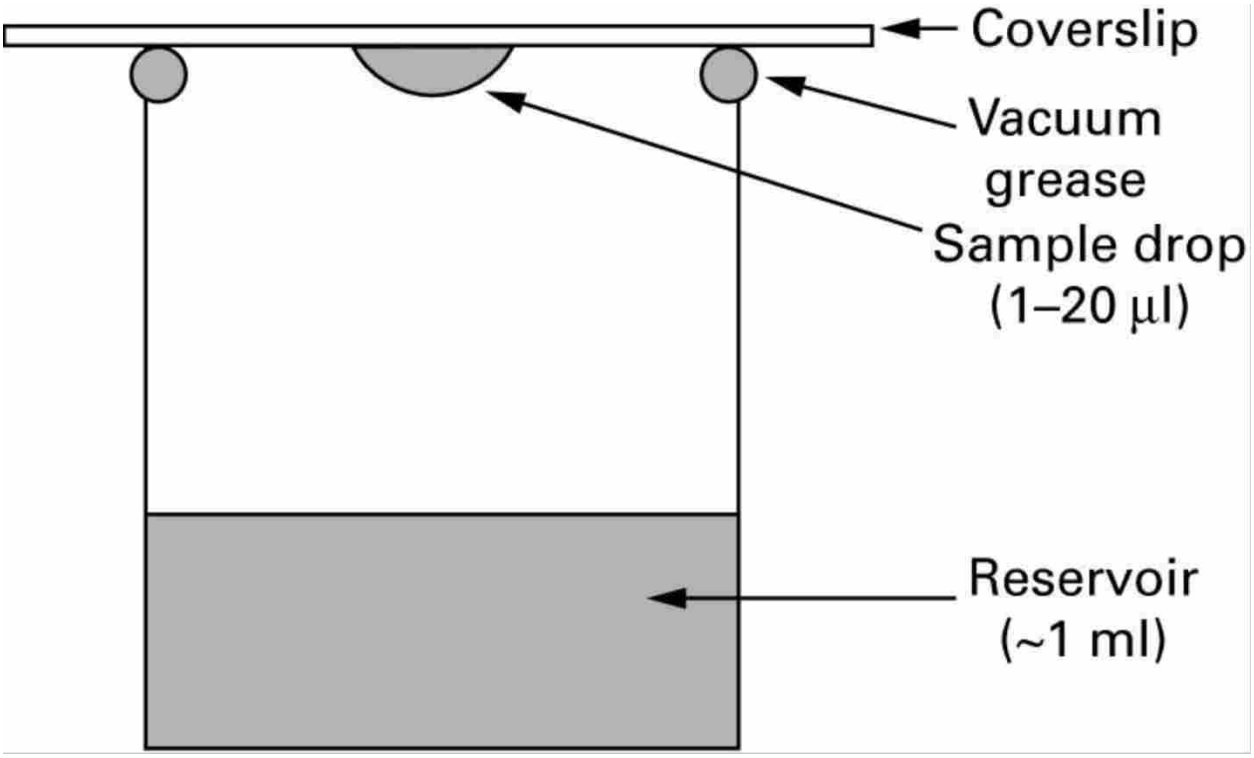



**Figure 26:** Vapour diffusion. Hanging drop vapour diffusion setup shows coverslip usually made of glass or plastic. The solution is placed on the coverslip and suspended over reservoir solution. Figure is taken from the reference (28).

**Figure 27:** Crystal growth of BFO 2294 in MCSG 2 over seven days. The condition was B4: 1.8 M $NaH_2PO_4/K_2HPO_4$ pH 8.2

**Figure 28:** Crystal growth of BFO 2294 in the expansion plate over three weeks. The condition was 6.25 M of glycerol, 400 μl or 40 M of Tacsimate 7.0 and 106 μl or 11 M of Tacsimate 9.0 at pH 8.2

## 4.10 Seeding

Sometimes poor crystal growth pattern results because of the limited formation of nuclei or spontaneous nucleation occurring at intense level of supersaturation (44). In that case, we can seed metastable, supersaturated protein solution with the crystals from earlier trials (44). The seeding techniques can involve two types of seeds and they are seeding with micro crystals and macro crystals (44). The method of seeding with micro crystals involves the danger of introducing too many nuclei (44). To overcome this problem, microcrystals could be diluted over a wide range (44). The solution containing higher concentration of seeds will produce microcrystals and the solution containing very low concentration of seeds will likely to produce nothing (44). Another method of seeding is with macrocrystals and the aim should be not to introduce too many nuclei (44). The technique of seeding is useful for inducing the nucleation, initiating the growth of crystals and growth at lower level of supersaturation (44).

# 4.11 X-Ray Data Collection and Refinement of BFO 2294

After harvesting single crystals, infusing them with cryoprotectants, and freezing them in liquid nitrogen data may be collected in oscillation mode wherein images are produced by rotating the crystals and taking regular images of the diffraction pattern through different angles (50). The resolution of the data starts to increase from the center of the image (50). We need both the amplitudes of the diffracted X-ray waves and their phase angles in order to calculate the electron density map of the protein of interest (50). The amplitudes can be measured as the intensities and positions of the diffraction spots but the phase information is sometimes lost (50). The lost phase information can be obtained by using the structure of a homologous, also known as molecular replacement (MR) (50). Another way of obtaining the phase information is to incorporate the heavy atoms such as Hg, Se into the crystals (50). Molecular replacement (MR) is a technique used to solve a crystal structure of a known molecular model to solve the unknown crystal structure of a related molecule (51). The main principle of the MR is very simple. We use a known model related to the unknown structure we are trying to deduce and also a set of measured diffraction intensities form the unknown structure using X-ray crystallography (51). We try all the possible orientations and positions of the unknown crystal model and try to match predicted diffraction with the observed diffraction (51). Another way of obtaining the phase information is to incorporate the heavy atoms such as mercury or selenium into the crystals by soaking (50). Once the correct phase information has been obtained, the new model is built through the electron density map in order to obtain the structure of the protein (51). The structure of the crystal BFO 2294 was solved by molecular replacement using the template PDB ID 1WA3. According to *Phyre*$^2$results , the confidence is100.0 between the template and BFO2294 .showing that the match between them was suggested to have high homologyThe sequence

identity is 33% between the template and BFO2294 .As we can see from the table 9, the resolution of this model is about 2.20 Å. The higher resolution data can provide details about electron density maps (50). At lower resolution, the electron density map is not well defined and the amino acids side chains cannot be well-modelled (50). A model with a resolution of 2.0 Å can provide us with active site residues, the binding pattern of an inhibitor, and the information about hydrogen bonding network (50). One can also measure the unit cell dimensions (50). The unit cell is the smallest repeating unit making up the crystals (50). The dimensions of the unit cell are written as three lengths: a, b, and c and three angles: α, β, and γ (50). The crystals are classified into seven crystal systems according to the shape (50). For the crystal of BFO 2294, we can see that α= ß = γ (°) = $90^0$ for the unit cell dimensions. This tells us that this crystal resembles the cubic crystal system $P2_3$ (50). The more diffracted beams (reflections) can be observed if the unit cell is larger. The intensities of reflections are influenced by the position of each atom in the crystal. A diffraction analysis requires measuring a large number of reflection intensities, but some reflections have identical intensity because of the crystals having certain symmetry. Redundancy is calculated by the number of measured reflections divided by the number of unique reflections. The average reflection intensity is more accurate when the redundancy is higher.

**Table 9:** Data collection and refinement statistics (molecular replacement) of BFO 2294

| **Space group** | $P2_3$ |
|---|---|
| **Cell dimensions** | |
| **a = b = c (Å)** | 98.15 |
| **α= ß = γ (°)** | $90^0$ |
| **Resolution (Å)** | 2.20 |
| **Completeness (%)** | 100 |
| **Redundancy** | 11.0 (7.4) |
| **No. reflections** | 179533 |

## 4.12 Multiple Sequence Alignment

Multiple sequence alignment (MSA) of DNA or protein sequence is one of the most commonly used techniques in sequence analysis (52). Multiple alignments can give us information regarding the phylogeny, sequence homology and the structure prediction (52). For this particular MSA, M-Coffee was used. It is a meta method for assembling MSA by joining the output of several individual methods into one single MSA (52). M-Coffee is a part of T-Coffee that uses consistency to estimate a consensus alignment (52). M-Coffee depends on consistency values and assumes that incorrect alignments are less likely to be consistent than the correct alignments (52). In figure 29, we showed the MSA of the protein sequence BFO 2294. The KDPG sequence of other organisms were aligned with the target sequence BFO 2294. The top portion is colour coded with different colours such as blue, green, yellow and pink. The sequences coloured in pink correspond to alignment portions with a strong support in the primary library used by T-Coffee (53). The sequences coloured in other colours such as blue, green and yellow do not correspond to aligned portions with a strong support in the primary library (53). Thus, the graphic colored output represent the level of consistency between the final alignment and the library used by T-Coffee (53). The main score represents the total consistency value and a value of 100 means strong relationship between the considered alignment and its associated primary library (53). The higher value of consistency value are proven to show high accuracy (53). As we can see from the figure 29, each of the sequences were given a consistency value and the consistency value of each sequence lies in the good region. The residue colour scheme follows the primary library support for the alignment of the considered residue on a scale indicating 0 (blue, poorly supported) and 9 (pink, strongly supported) (52). The study suggests that the residues with a consistency score higher than 5 (yellow, pink) are considered to be

correctly aligned (52). Thus, this scoring scheme is considered to be the most used method to identify the highly unreliable residues in a sequence that are aligned in a biologically meaningful way (52). As we can see, most of the residues of the sequences are coloured in pink and yellow meaning they are considered to be correctly aligned. If we look at the figure 19, we can see some of the residues are marked with asterisk signs meaning those residues are exactly the same or highly conserved in that position. We can see the amino acids glycine, glutamic acid, proline, phenylalanine and lysine are highly conserved meaning they could be important residues for the active sites. The residues with semicolons can signify that the residues in that position are very similar, while dots indicate the residues are less similar. The residues with no sign indicate there is no similarity between the residues.


**Figure 29:** MSA generated by T-Coffee. The sequences coloured in pink correspond to alignment portions with a strong support in the primary library used by T-Coffee but the sequences coloured in other colours such as blue, green and yellow do not correspond to aligned portions with a strong support in the primary library.

## 4.13 Structure of KDPG Aldolase

The enzyme 2-keto-3-deoxy-6 phosphogluconate (KDPG) aldolase is an important enzyme in the Entner-Doudoroff pathway which breaks down KDPG through class I Schiff-base mechanism (54). This enzyme plays an important role in the Entner-Doudoroff pathway of bacteria where it catalyzes reversible cleavage of KDPG to pyruvate and glyceraldehyde-3-phosphate (55). The structure of this enzyme has been determined to be a monomer containing 226 amino acids

residues according to research paper. The research paper shows that this enzyme has tertiary folding and the subunit contains eight α-helices and eight β-strands (56) (57). This is also known as α/β-barrel structure or TIM barrel structure. The α/β-barrel structure is situated on the N-terminal side and the carboxylic side contains the active site (54). The subunit contains a phosphate ion which is bound to the aldolase binding site (58). KDPG aldolase is a class I aldolase because KDPG aldolase involves the Schiff base formation between carbonyl substrate and lysine residue in the active site of this enzyme (55). The active site of BFO 2294 contains zwitterionic pair of Glu-49 and Lys-141 according to the research paper (55). The residue lysine is involved in the formation of Schiff base which is coordinated by phosphate ion and two water molecules (54) (58). The first water molecule acts as a shuttle between the glutamic acid and the substrate and the second water molecule is the product of the dehydration of the of the carbinolamine leading to the formation of the Schiff base (58). It also functions as a nucleophile during the hydrolysis of the enzyme product complex which leads to the release of the pyruvate molecule (58). The structure of BFO 2294 is shown in figure 30.

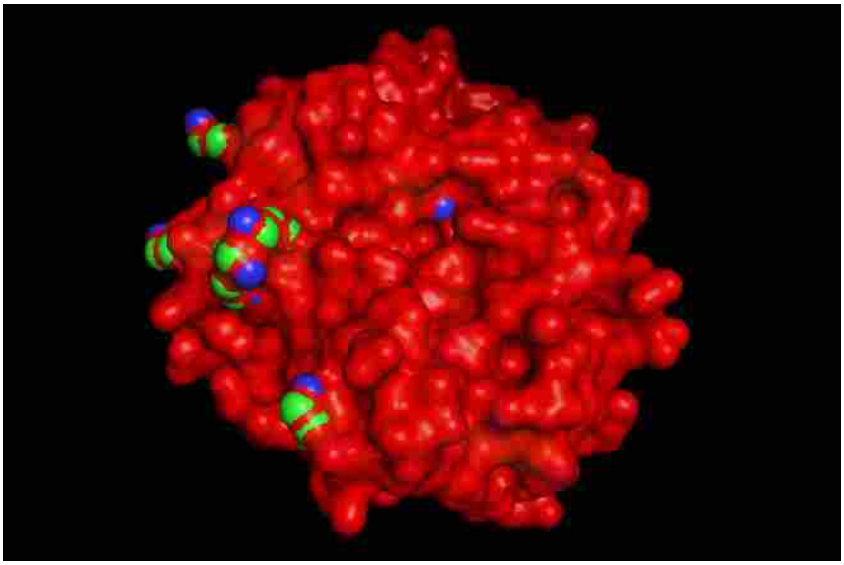


**Figure 30:** Structure of BFO 2294. The structure showing active residues lysine (blue) and glutamate (green).

## 4.14 Mechanism of KDPG Aldolase

Based on the different research paper on KDPG aldolase structure and mechanism, we came up with mechanism of BFO 2294 which is shown in figure 31. The enzyme BFO 2294 catalyzes the reversible cleavage of 2-keto-3deoxy-6-phosphogluconate (KDPG) into pyruvate and D-glyceraldehyde-3-phosphate by aldol cleavage (55) (58). In the first step, a proton transfer occurs between the zwitterionic pair Glu-49/Lys-141 in the active site and it activates lysine to participate as a nucleophile in the reaction (55). In the second step, glutamate is involved in the base catalysis in the carbon-carbon cleavage (55). The residue lysine-141 acts as a nucleophile and attacks the carbonyl group of 2-keto-3deoxy-6-phosphogluconate that forms a protonated carbinolamine intermediate or Schiff base (55) (58). This intermediate is stabilized by the hydrogen bonding with the residues (lysine and glutamate) in the active site (55). Then, the three carbon molecule glyceraldehyde 3-phosphate is cleaved off by base catalysis using water molecule and Glu-49 (58) (55). The other product pyruvate is generated by the nucleophilic attack of water on the Schiff base to form a ketone and also aromatic interaction with Phe-143 makes sure that the stereospecific addition is involved in the reverse process (55).

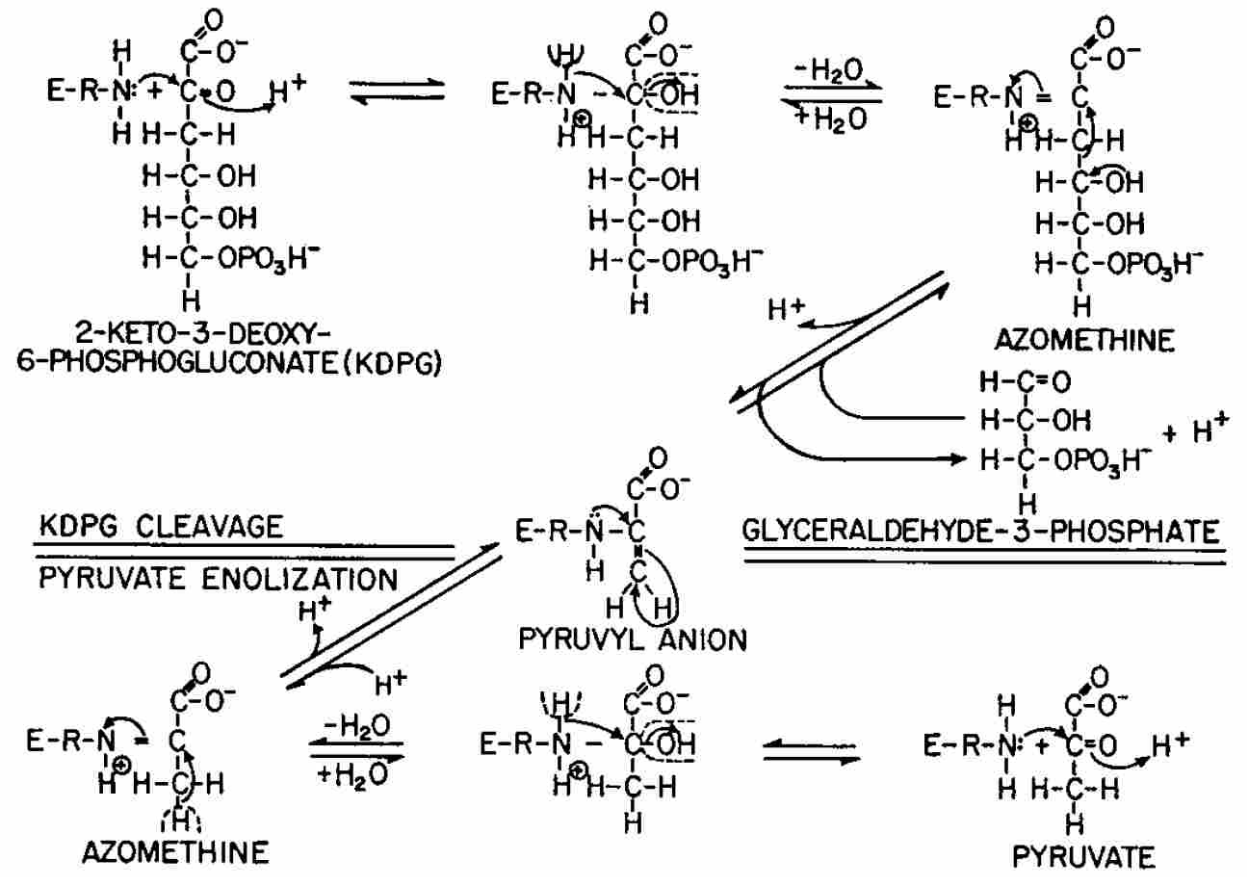



**Figure 31:** Mechanism of action of KDPG aldolase. In the first step, proton transfer occurs between the zwitterionic pair Glu-49/Lys-141 in the active site and it activates lysine to act as nucleophile. In the second step, glutamate is involved in the base catalysis in the carbon-carbon cleavage. Figure is taken from the reference (59).

# 5. CONCLUSION

The research presented herein was centred on glycosaminoglycans (GAGs) degradation by the dental pathogen *T. forsythia*. Using bioinformatics, GAGs degradations pathway was established which explains how different enzymes secreted from the bacterium *Tannerella forsythia* degrades one of the GAGs called chondroitin sulfate A. The two putative enzymes: BFO 2291 (chondroitin AC lyase) and BFO 2294 (KDPG aldolase) were the main focus of the research. Based on the bioinformatics and enzyme kinetics, it was found that the chondroitin sulfate A was the substrate for the enzyme BFO 2291. It was shown that there is a noticeable activity of this enzyme when acted upon chondroitin sulfate A but it requires further testing using better reproducibility and correct temperature condition. The research also showed the optimum pH at which BFO 2291 had the maximum activity. The pH at which the enzyme had the maximum activity was phosphate buffer at pH 6.5. The other enzyme BFO 2294 had noticeable activity towards the substrate KDPG using coupled enzymatic assay. The enzyme BFO 2294 was successfully crystallized and its structure and mechanism of action was found using various research journals. Doing research into characterizing the following gene products BFO 2291 and BFO 2294 can guide us in understanding the degradation of periodontal tissues during periodontitis. In the future, crystallization BFO 2291 could lead to a better understanding of the mechanism of this enzyme during depolymerisation of chondroitin sulfate A. Another part of this research was to identify the products of enzyme-substrate reaction and it could be another part of this research where the products can be identified. If one looks at the GAGs degradations pathway, one can see that there are other different enzymes involved in degradations of chondroitin sulfate A. In future, other researchers can work towards solving the whole pathway and can have a better understanding on how these enzymes work together during pathogenesis.

This may lead to the development of some drugs that can halt the progression of periodontitis by deactivating certain enzymes in the GAGs degradations pathway.

# REFERENCES


1. 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*. **11**, 1–20
2. Shimotahira, N., Oogai, Y., Kawada-Matsuo, M., Yamada, S., Fukutsuji, K., Nagano, K., Yoshimura, F., Noguchi, K., and Komatsuzawa, H. (2013) The surface layer of Tannerella forsythia contributes to serum resistance and oral bacterial coaggregation. *Infect. Immun.* **81**, 1198–1206
3. Sanghavi, T., Shah, N., Shah, R., and Sanghavi, A. (2014) Investigate the correlation between clinical sign and symptoms and the presence of P. gingivalis, T. denticola, and T. forsythia individually or as a "Red complex" by a multiplex PCR method. *J. Conserv. Dent.* **17**, 555
4. Socransky, S., Haffajee, A., Cugini, M., Smith, C., and Kent, R. J. (1998) Microbial complex in subgingival plaque. *J Clin Periodontol*. **25**, 134–44
5. Younis, L., Hassan, M., Anuar, S., Yunus, F., and Yusof, N. (2015) The Role of Reactive Oxygen Species in Initiation and Progression of Periodontal Diseases. *Br. J. Appl. Sci. Technol.* **8**, 541–549
6. Younis, L., Hassan, M., Anuar, S., Yunus, F., and Yusof, N. (2015) The Role of Reactive Oxygen Species in Initiation and Progression of Periodontal Diseases. *Br. J. Appl. Sci. Technol.* **8**, 541–549
7. Suzuki, N., Yoneda, M., and Hirofuji, T. (2013) Mixed red-complex bacterial infection in periodontitis. *Int. J. Dent.* **2013**, 587279
8. How, K. Y., Song, K. P., and Chan, K. G. (2016) Porphyromonas gingivalis: An overview

of periodontopathic pathogen below the gum line. *Front. Microbiol.* **7**, 1–14

9. Kornman, K. S., Page, R. C., Tonetti, M. S., Embery, G., Waddington, R. J., Hall, R. C., Last, K. S., Tobergte, D. R., and Curtis, S. (1997) The host response to the microbial challenge in periodontitis: assembling the players. *Periodontol. 2000*. **24**, 193–214
10. Juan, J. (2009) Structure and functions of heparan sulfate/heparin – Importance of glucuronyl C5-epimerase and heparanase
11. Smith, a J., Addy, M., and Embery, G. (1995) Gingival crevicular fluid glycosaminoglycan levels in patients with chronic adult periodontitis. *J. Clin. Periodontol.* **22**, 355–361
12. Linhardt, R. J., Avci, F. Y., Toida, T., Kim, Y. S., and Cygler, M. (2006) CS lyases: structure, activity, and applications in analysis and the treatment of diseases. *Adv. Pharmacol.* **53**, 187–215
13. Price, M. N., Arkin, A. P., and Alm, E. J. (2006) The life-cycle of operons. *PLoS Genet.* **2**, 0859–0873
14. Marcenes, W., Kassebaum, N. J., Bernabé, E., Flaxman, A., Naghavi, M., Lopez, A., and Murray, C. J. L. (2013) Global burden of oral conditions in 1990-2010: A systematic analysis. *J. Dent. Res.* **92**, 592–597
15. Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., De Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A. F., Rivoire, C., Sangrador-Vegas, A.,

Selengut, J. D., Sigrist, C. J. A., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P. D., Wu, C. H., Yeats, C., and Yong, S. Y. (2012) InterPro in 2011: New developments in the family and domain prediction database. *Nucleic Acids Res.* **40**, 306–312

16. Seeliger, D., and De Groot, B. L. (2010) Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided. Mol. Des.* **24**, 417–422
17. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) Europe PMC Funders Group The Phyre2 web portal for protein modelling , prediction and analysis. *Nat. Protoc.* **10**, 845–858
18. 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–D452
19. Gonzalez, R., Jennings, L. L., Knuth, M., Orth, A. P., Klock, H. E., Ou, W., Feuerhelm, J., Hull, M. V., Koesema, E., Wang, Y., Zhang, J., Wu, C., Cho, C. Y., Su, A. I., Batalov, S., Chen, H., Johnson, K., Laffitte, B., Nguyen, D. G., Snyder, E. Y., Schultz, P. G., Harris, J. L., and Lesley, S. A. (2010) Screening the mammalian extracellular proteome for regulators of embryonic human stem cell pluripotency. *Proc. Natl. Acad. Sci.* **107**, 3552–3557
20. Benov, L., and Al-Ibraheem, J. (2002) Disrupting Escherichia coli: A Comparison of Methods. *J. Biochem. Mol. Biol.* **35**, 428–431
21. Pethica, B. A. (1958) Bacterial Lysis. *Journals Gen. Microbiol.* **18**, 473–480
22. Bornhorst, J., and Falke, J. (2010) Purification of Proteins Using Polyhistidine Affinity

Tags. *Methods Enzymol.* **326**, 245–254

23. Purification of Polyhistidine-Containing Recombinant Proteins with Ni-NTA Purification System (2011) [online] https://www.thermofisher.com/ca/en/home/references/protocols/proteins-expression-isolation-and-analysis/protein-purification-protocol/purification-of-polyhistidine-containing-recombinant-proteins-with-ni-nta-purification-system.html#prot1 (Accessed October 1, 2017)
24. Bahadir, O. (2013) Ion-Exchange Chromatography and Its Applications. in *Column Chromatography* (Martin, D. ed), InTech
25. Andrew, S. M., Titus, J. A., and Zumstein, L. (2001) Dialysis and concentration of protein solutions. *Curr. Protoc. Immunol.* **Appendix 3**, Appendix 3H
26. Shaya, D., Hahn, B. S., Nam, Y. P., Sim, J. S., Yeong, S. K., and Cygler, M. (2008) Characterization of chondroitin sulfate lyase ABC from Bacteroides thetaiotaomicron WAL2926. *Biochemistry*. **47**, 6650–6661
27. Walters, M. J., Srikannathasan, V., Mcewan, A. R., and James, H. (2012) Characterization and crystal structure of Escherichia coli KDPGal aldolase. **14**, 3002–3010
28. Smyth, M. S., and Martin, J. H. J. (2000) x Ray crystallography. *J. Clin. Pathol. - Mol. Pathol.* **53**, 8–14
29. Woolfson, M. M. (1997) *An Introduction to X-ray crystallography*, 2nd Ed., Cambridge University Press
30. Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., and von Mering, C. (2009) STRING 8 - A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids*

*Res.* **37**, 412–416

31. Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S. Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J. A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D., and Finn, R. D. (2015) The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221
32. Kelley, L., and Sternberg, M. (2009) Protein structure prediction on the web : a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371
33. Peterson, T. N., Brunak, S., von Heijne, G., and Nielson, H. (2011) SignalP 4.0 : discriminating signal peptides from transmembrane regions. *Nat. Methods*. **8**, 785–786
34. Wong, J. W., Albright, R. L., and Wang, N. H. L. (1991) Immobilized metal ion affinity chromatography (IMAC) chemistry and bioseparation applications. *Sep. Purif. Rev.* **20**, 49–106
35. Ward, R. A., and Wathen, R. L. (1982) Principles of Dialysis. *Int. J. Dermatol.* **21**, 154–158
36. Ion Exchange Chromatography | Applications & Technologies | Bio-Rad [online] http://www.bio-rad.com/en-ca/applications-technologies/liquid-chromatography-principles/ion-exchange-chromatography (Accessed October 27, 2017)
37. Effects of pH (Introduction to Enzymes) (2016) *Worthingt. Biochem. Corp.* [online] http://www.worthington-biochem.com/introbiochem/effectsph.html (Accessed June 12, 2017)

38. Talley, K., and Alexov, E. (2010) On the pH-optimum of activity and stability of proteins. *Proteins Struct. Funct. Bioinforma.* **78**, 2699–2706

39. Hall, J. E. (2016) *Guyton and Hall Textbook of Medical Physiology*, 13th editi, Elsevier, Philadelphia, PA

40. Baliga, S., Muglikar, S., and Kale, R. (2013) Salivary pH: A diagnostic biomarker. *J. Indian Soc. Periodontol.* **17**, 461–5

41. Voet, D., Voet, J. G., and Pratt, C. W. (2016) *Fundamentals of biochemistry: life at the molecular level*, 5th Editio, John Wiley & Sons, Hoboken, NJ

42. Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry*, 5th Editio, W.H. Freeman, New York

43. Hong, S. W., Kim, B. T., Shin, H. Y., Kim, W. S., Lee, K. S., Kim, Y. S., and Kim, D. H. (2002) Purification and characterization of novel chondroitin ABC and AC lyases from Bacteroides stercoris HJ-15, a human intestinal anaerobic bacterium. *Eur. J. Biochem.* **269**, 2934–2940

44. McPherson, A., and Gavira, J. A. (2014) Introduction to protein crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70**, 2–20

45. Sands, D. E. (2012) *Introduction to Crystallography*, Dover, New York

46. Drenth, J. (1999) *Principles of Protein X-ray Crystallography*, 2nd Ed., Springer New York, New York, NY

47. Yamazaki, T., Kimura, Y., Vekilov, P. G., Furukawa, E., Shirai, M., Matsumoto, H., Van Driessche, A. E. S., and Tsukamoto, K. (2017) Two types of amorphous protein particles facilitate crystal nucleation. *Proc. Natl. Acad. Sci.* **114**, 2154–2159

48. Vekilov, P. G. (2010) Nucleation. *Cryst. Growth Des.* **10**, 5007–5019

49. Ladd, M., and Palmer, R. (2014) *Structure Determination by X-ray Crystallography: Analysis by X-rays and Neutrons Paperback – May 1 2013*, 5th Ed., Springer US

50. Lamb, A. L., Kappock, T. J., and Silvaggi, N. R. You are lost without a map: Navigating the sea of protein structures. *Biochim. Biophys. Acta - Proteins Proteomics*. **1854**, 258–268

51. Evans, P., and McCoy, A. (2007) An introduction to molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 1–10

52. Wallace, I. M., O'Sullivan, O., Higgins, D. G., and Notredame, C. (2006) M-Coffee: Combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699

53. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., Taly, J. F., and Notredame, C. (2011) T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**, 13–17

54. Bell, B. J., Watanabe, L., Rios-Steiner, J. L., Tulinsky, A., Lebioda, L., and Arni, R. K. (2003) Structure of 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase from Pseudomonas putida. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 1454–1458

55. Allard, J., Grochulski, P., and Sygusch, J. (2001) Covalent intermediate trapped in 2-keto-3-deoxy-6- phosphogluconate ( KDPG ) aldolase structure at 1 . 95-Å resolution. **98**, 3679–3684

56. Richardson, J. (1979) The singly-wound parallel β barrel: A proposed structure for 2-keto-3-deoxy-6-phosphogluconate aldolase. *Biochem. Biophys. Res. Commun.* **90**, 285–290

57. Mavridis, I., Hatada, M., Tulinsky, A., and Lebioda, L. (1982) Structure of 2-keto-3-

deoxy-6-phosphogluconate aldolase at 2.8 Å resolution. *J. Mol. Biol.* **162**, 419–444

58. Fullerton, S. W. B., Griffiths, J. S., Merkel, A. B., Cheriyan, M., Wymer, N. J., Hutchins, M. J., Fierke, C. A., Toone, E. J., and Naismith, J. H. (2006) Mechanism of the Class I KDPG aldolase. *Bioorg. Med. Chem.* **14**, 3002–3010

59. Meloche, H., and Wood, W. (1964) The mechanism of 2-keto-3deoxy-6-phosphogluconic Aldolase. *J. Biol. Chem.* **239**, 3511–3514

# APPENDICES

## Appendix I

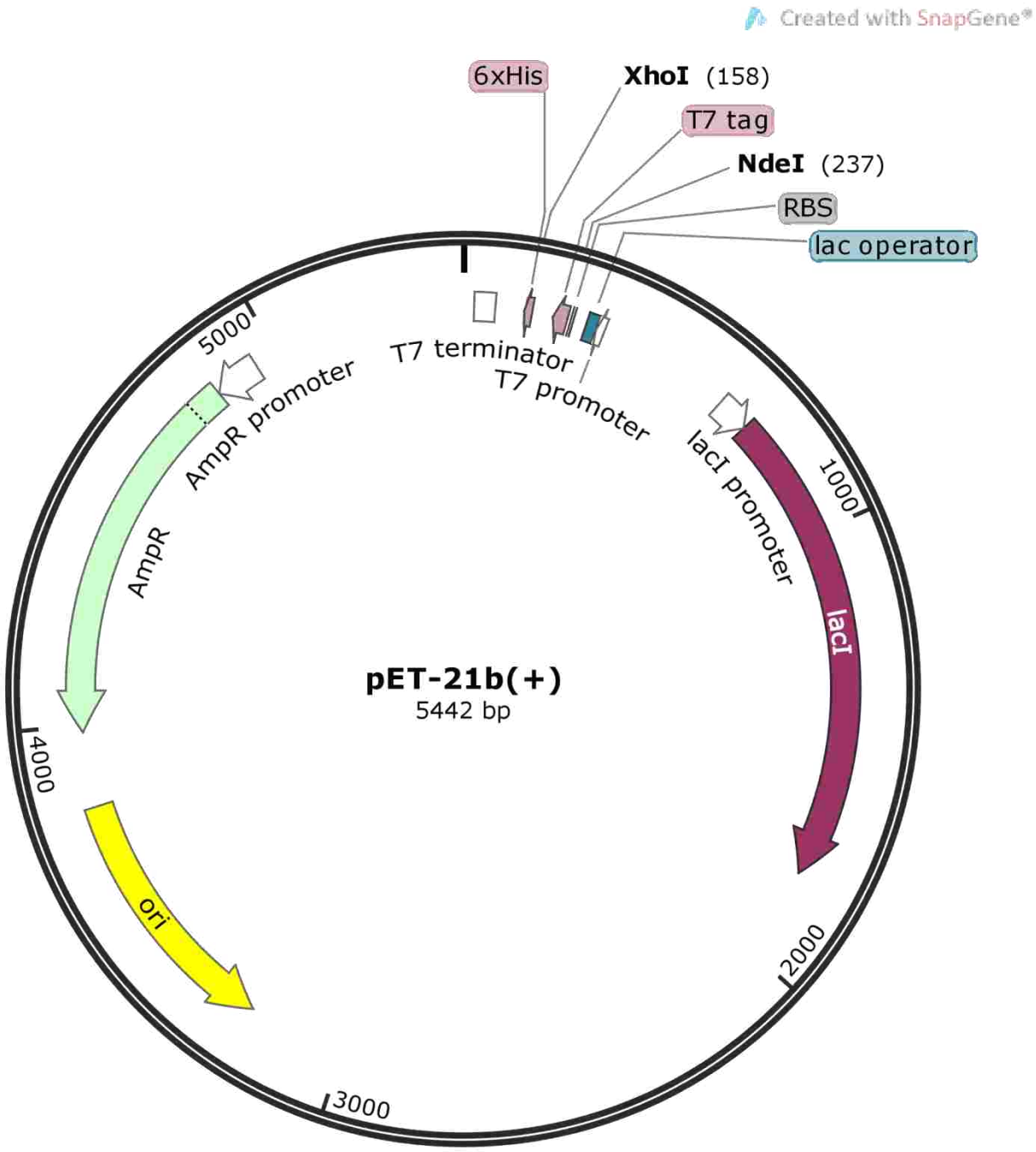



**Figure 32:** pET 21b+ plasmid. The figure shows different components of the plasmid. It also shows different restriction enzyme cut sites.

# Appendix II


**Figure 33:** Interpro results of BFO 2291. The figure tells us the protein family BFO 2291 belongs to and the major domains of BFO 2291.

# Appendix III

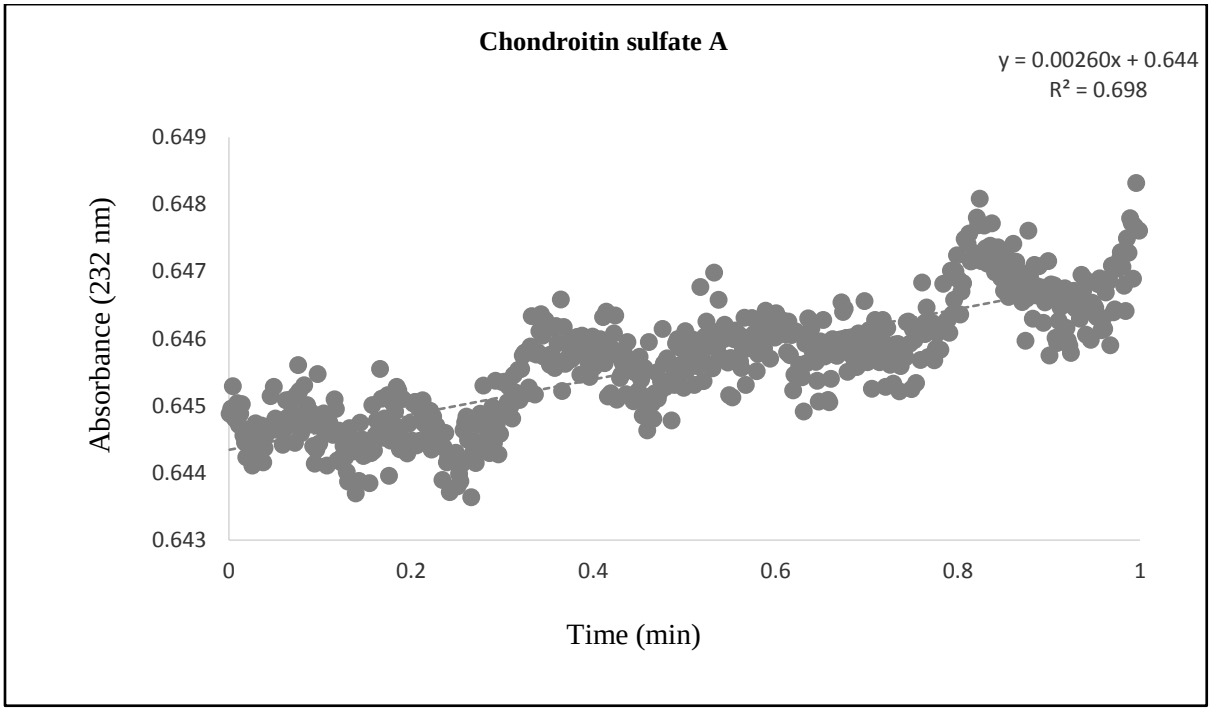



**Figure 34:** Substrate specificity. Chondroitin sulfate A shows a small increase in absorbance. The concentration of the substrate was 1 mg/mL, enzyme concentration (BFO 2291) was 50 µg/mL, the buffer was 50 mM of sodium phosphate at pH 6.5.

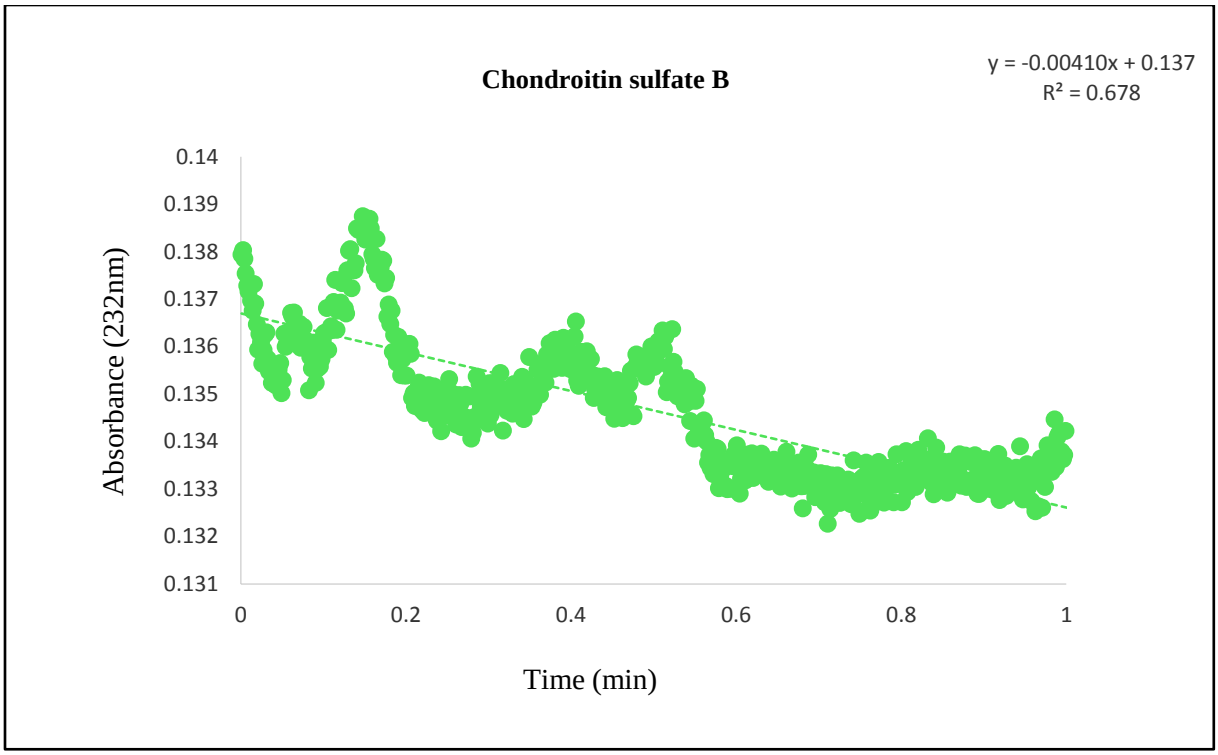



**Figure 35:** Substrate specificity. Chondroitin sulfate B shows decrease in absorbance. The concentration of the substrate was 1 mg/mL, enzyme concentration (BFO 2291) was 50 µg/mL, the buffer was 50 mM of sodium phosphate at pH 6.5.

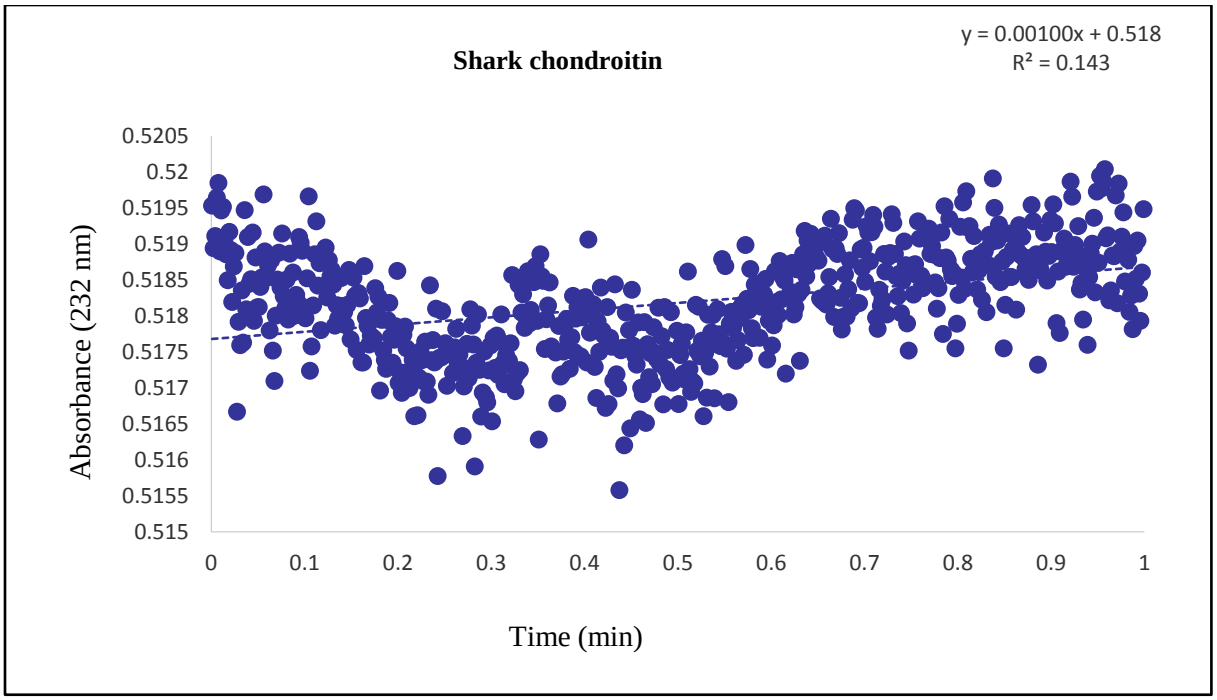



**Figure 36:** Substrate specificity. Shark chondroitin sulfate also shows small increase in absorbance as well due to the presence of chondroitin sulfate A. The concentration of the substrates was 1 mg/mL, enzyme concentration (BFO 2291) was 50 µg/mL, the buffer was 50 mM of sodium phosphate at pH 6.5.

# Appendix IV

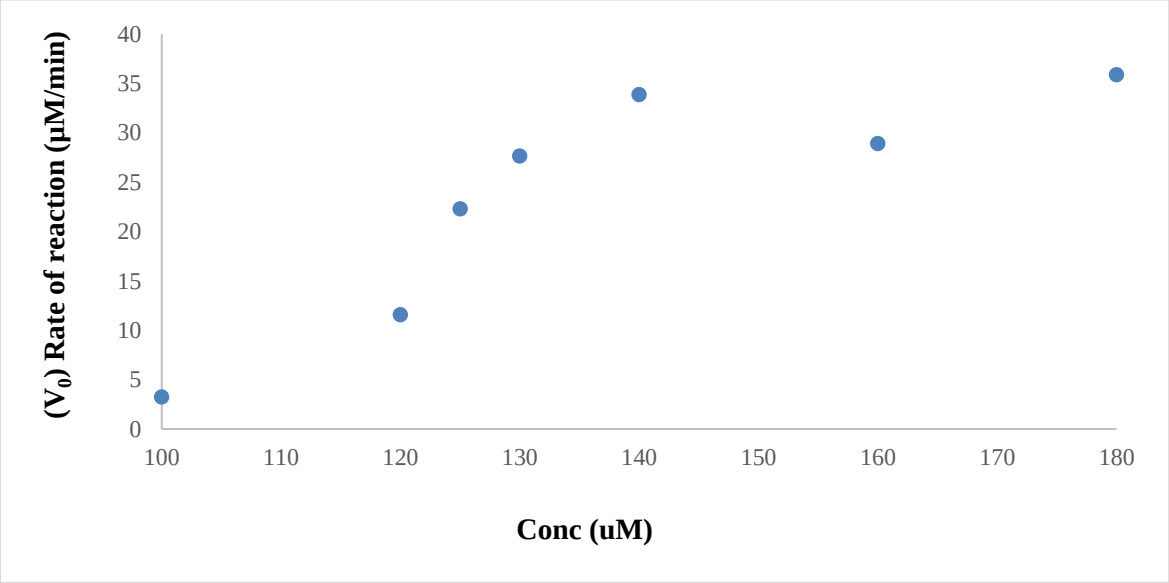



**Figure 37:** Enzyme kinetics for BFO 2294. This shows the relationship between the increase in concentration of the KDPG and the amount of product being formed. The buffer used was 50 mM Hepes at pH 7.0, 500 mM stock concentration of the substrate 2-keto-3-deoxy-6-phosphogluconate (KDPG), the enzyme concentration (BFO 2294) was 50 µg/mL, and L-lactic dehydrogenase from rabbit muscle: Type II, ammonium sulfate suspension, 800-1,200 units/mg protein