# STRUCTURAL STUDIES OF ACETYL ESTERASE AND MALT FROM *ESCHERICHIA COLI*

By

MAMIKO NISHIDA

Bachelor of Science in Biology and Chemistry

Pittsburg State University

Pittsburg, Kansas

2005

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2010

STRUCTURAL STUDIES OF ACETYL ESTERASE

AND MALT FROM *ESCHERICHIA COLI*

Dissertation Approved:

Dr. Stacy D. Benson
Dissertation Adviser

Dr. Neil Purdie
Committee Member

Dr. Nicholas F. Materer
Committee Member

Dr. Junpeng Deng
Committee Member

Dr. Mark E. Payton
Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who supported me during my time here in graduate school. I also would like to thank many people who offered help and advice in various matters. There are some people who deserve special acknowledgement.

I would like to express my sincere gratitude to my advisor, Dr. Stacy D. Benson, who is understanding and intelligent. Because of his advice for research, seminars, and writing papers, I could overcome all of the severities in the completion of the doctorate. Thank you very much for allowing me to learn and investigate new areas on my own. I also would like to thank you for listening to me and giving me advice about many personal matters.

I am grateful to Dr. Neil Purdie, Dr. Nicholas F. Materer, Dr. John Gelder, Dr. Junpeng Deng, and Dr. Asfaha Iob for numerous suggestions and constructive comments for research, writing, and also the laboratory teaching to get through these years. I also would like to thank Dr. Steve Hartson, Janet Rogers and Lisa Whitworth from the DNA/Protein resource facility in the Biochemistry and Molecular Biology department for their technical assistance and advice toward my research. I would like to thank all the professors who gave me great lectures that helped my research. I am also grateful to Dr. Masato Kumauchi and Dr. Furuta Kenji who gave me much advice for my personal

.

matters. Because of your encouragements, I could survive the life here and get to this point.

I also would like to express my gratitude to my colleagues and friends who supported me in my research work and also personal matters. I wish to thank Takeo Mohri for his suggestions, help, and encouragement and for giving me the opportunities of enjoying life outside the laboratory. I am also grateful to Jessica, Radika, and Xinyi for being such great friends and for their help and suggestions. Especially I would like to thank Danny for being such a big brother both inside and outside the laboratory and for listening and advising me on numerous matters.

Finally, I would like to dedicate my work to my parents for their encouragement, support and selfless love through my entire life and I have and will continue to make you proud.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

# ABBREVIATIONS

| | |
|---|---|
| Amp | ampicillin |
| bp | base pairs |
| cAMP | cyclic adenosine monophosphate |
| CAP | catabolite activator protein |
| Cα RMSD | root mean square deviation at C-alpha atoms |
| CCR | carbon catabolite repression |
| CTAB | hexadecyltrimethyl ammonium bromide |
| EB | ethidium bromide |
| EDTA | ethlenediamine tetraacetic acid |
| EI | Enzyme I |
| EIIA | component A of Enzyme II |
| EIIBC | component BC of Enzyme II |
| EST | carboxylesterases |
| E-value | Expect value |
| GS1 | gel stabilization buffer |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| His | histidine |
| HPr | histidine-containing phosphocarrier protein |
| HSL | hormone sensitive lipase |
| IPTG | isopropyl-β-D-1-thiogalactopyanoside |
| kDa | kilodaltons |
| kbp | kilobase pairs |
| LB | Luria Bertani |

| | |
|---|---|
| LPL | lipolytic lipase |
| MALDI-TOF | matrix-assisted laser desorption/ionization time-of-flight mass spectrometer |
| Mlc | make large colonies |
| MOWSE | molecular weight search |
| MSA | multiple sequence alignment |
| NCS | non-crystallographic symmetry |
| OD | optimal density |
| PCR | polymerase chain reaction |
| PDBePISA | protein interfaces, surfaces, and assemblies service PISA at European Bioinformatics Institute |
| PEG | polyethylene glycol |
| PEP | phosphoenolpyruvate |
| Phyre | protein homology/analogy recognition engine |
| PTS | phosphotransferase system |
| RCSB PDB | Research Collaboratory for Structural Bioinformatics Protein Data Bank |
| r.m.s | root mean square |
| rpm | revolutions per minute |
| RT | room temperature |
| SDS | sodium dodecyl sulfate |
| SDS-PAGE | sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| S. O. C. | super optimal broth with catabolite repression |
| TAE | Tris-acetate-EDTA |
| TLS | translation/libration/screw |
| TLSMD | TLS motion determination |
| W9 | wash buffer |

# CHAPTER 1

# BACKGROUND

## 1.1 Metabolism in *Escherichia coli*

The major food sources for bacteria to live are carbohydrates, especially more common monosaccharides and disaccharides such as glucose, maltose and lactose. Depending on the availability of the substrates, specific genes that express proteins needed for uptake and consequent metabolism are encoded. Therefore, if a particular substrate is not available, genes involved in the metabolism of that substrate are repressed [1]. In *Escherichia coli,* glucose and fructose are preferred carbon sources and other sugars such as maltose, lactose and arabinose are less preferred. Therefore, if preferred carbon sources are available, proteins required for the metabolism of less favored carbon sources are not synthesized, which is known as carbon catabolite repression (CCR). After the availability of the favored carbon sources is exhausted, proteins needed for the transport and metabolism of more complex sugars are synthesized depending on their presence in the environment [1].

The major role in bacterial CCR is performed by the phosphoenolpyruvate (PEP)-dependent sugar phosphotransferase system (PTS) (Figure 1-1) [1-3]. The PTS is an energy conserving transport system, that uses PEP to phosphorylate carbohydrates during transport [1-2]. The energy required for the phosphorylation and the translocation of carbohydrates is generated by PEP [4]. Many sugars are transported by PTS, for example, glucose, fructose, and mannose, while other sugars, such as lactose, melibiose, and maltose are not [3]. The former is called PTS sugars and the latter is called non-PTS sugars [5].

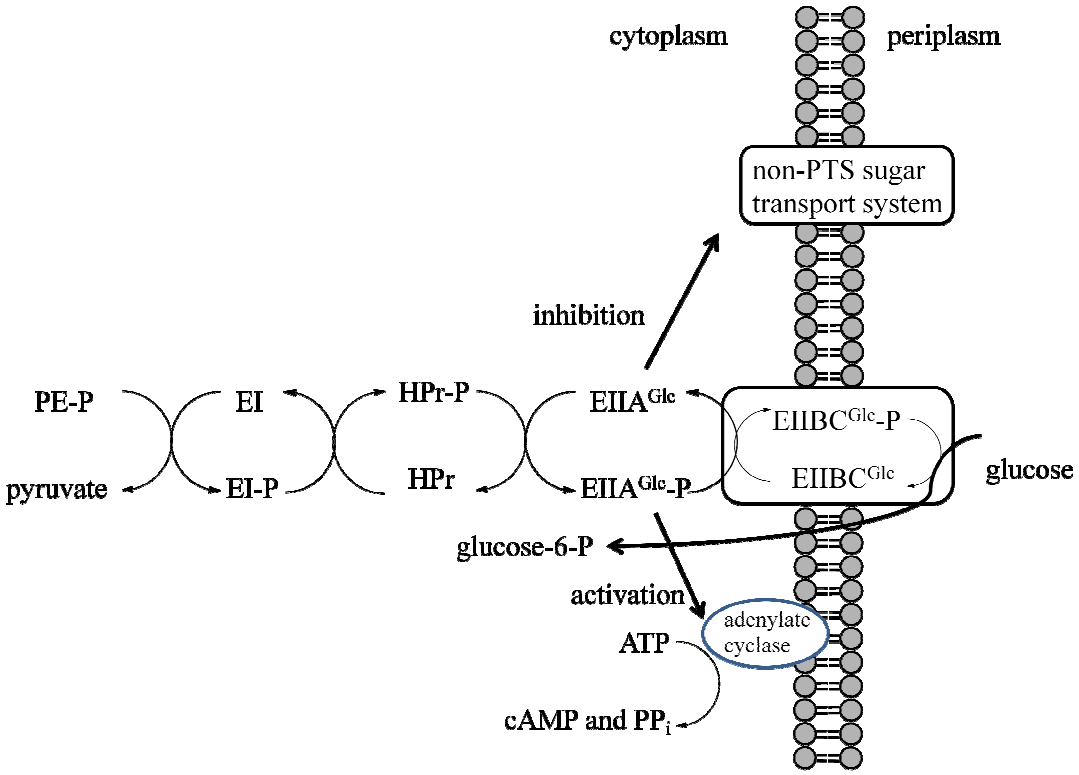



**Figure 1-1. The phosphoenolpyruvate-dependent phosphotransferase system (PEP-PTS) for glucose.** PEP phosphorylates Enzyme I (EI) followed by the phosphoryl transfer to histidine-containing phosphocarrier protein (HPr), which participates in the transport of all sugars. The phosphoryl transfer to component A of Enzyme II (EIIA) occurs, subsequently, to component BC of EII (EIIBC). The Enzyme II components are specific to individual sugars with $EIIA^{Glc}$ and $EIIBC^{Glc}$ being specific for glucose. Phosphorylated EIIA activates adenylate cyclase while non-phosphorylated EIIA inhibits non-PTS transport proteins such as the lactose permease [3, 5].

One of the essential molecules in PTS is Enzyme II (EII) consisting of at least three functional components: EIIA, EIIB, and EIIC. EIIA is a separate subunit in *E. coli* but EIIB and EII C are linked. Each component is specific for a particular sugar such as $EIIA^{Glc}$ for glucose. $EIIA^{Glc}$ controls the synthesis of cyclic adenosine monophosphate (cAMP) and the activity of many membrane-bound transport systems [2]. The mechanism of catabolite repression is based on the level of cAMP, which is a product of adenylate cyclase that is regulated by $EIIA^{Glc}$ of the PEP-PTS system [6]. The concentration of cAMP is lowered if glucose is present [7]. More specifically, the phosphorylated $EIIA^{Glc}$ activates the adenylate cyclase, leading to the regulation of cAMP concentration in the cell, while non-phosphorylated $EIIA^{Glc}$ hinders some proteins that are necessary in the metabolism of non-PTS carbohydrates, such as the lactose permease and MalK, a protein which restricts the activity of the transcriptional activator for the maltose regulon (see section 1.3) [2, 4, 8].

### 1.2. Transcriptional Regulation in Bacteria

It takes only minutes for prokaryotes to respond to sudden environmental changes while eukaryotic cells need hours to days. The reason for such a quick response to the environment is because transcription and translation are closely coupled in prokaryotes [7]. The expression level of proteins depends on the physiological conditions of bacterial cells and the protein expression is controlled by the level of transcription [2]. The level of transcription is under the control of transcriptional regulation, which turn on and off the expression of genes depending on the needs of the cells [2].

An operon is a cluster of consecutive genes under the control of a single promoter (a site that precedes the transcribed portion of the gene) and an operator (a control site) (see Figure 1-2) [2, 9]. The operator acts as a regulator and controls the expression of the genes in the operon as a group. The regulon, on the other hand, is a collection of genes or operons whose expression is controlled by the same regulatory protein or proteins [10]. The genes of the operon are located closely together whereas those of the regulon can be spread over the entire genome.

Unique regulatory sequences are present in the genome so that particular enzymes or regulatory proteins are able to bind and regulate the transcription of their target genes. In prokaryotes, regulatory proteins such as activators and repressors bind to a specific site on the DNA molecule for the control of transcriptional initiation. An activator protein binds to a specific control site to stimulate the initiation of transcription while a repressor protein binds to an operator or a promoter, leading to the prevention of the transcription initiation [9]. For instance, the CAP-cAMP (catabolite activator protein with cyclic adenosine monophosphate) complex stimulates the transcription of catabolite repressed operons and thus CAP is a positive regulator by initiating transcription [7]. Another example of a positive regulator is MalT, which is a transcriptional activator for the maltose regulon that binds to operator sites to control the expression of *mal* genes (see section 3.1). The sites for positive regulators are found upstream of the promoter and the activity of the RNA polymerase is enhanced by binding of the activator, leading to transcription initiation [2].

There are several well studied operons and one of them is the *lac* operon in *E. coli* (Figure 1-2). The *lac* operon consists of consecutive genes, Z (β-galactosidase), Y (galactoside permease), and A (thiogalactoside transacetylase), that encode proteins used in the metabolism of lactose. The *lac*I gene, which is located in the *lac* operon, encodes the *lac* repressor that binds to the *lac* promoter in the absence of allolactose (an inducer), leading to the prevention of transcription of the structural genes (genes which specifically encode a polypeptide) [7, 9]. A small amount of lactose will be modified to allolactose if lactose is available, and it will bind to the repressor. As a result, RNA polymerase is able to initiate the transcription of the structural genes (Z, Y, and A) [9]. Since glucose is the most favorable food source for *E. coli*, the expression of more than 100 genes involved in the metabolism of other food sources such as lactose is prevented if enough glucose is available, which is the basis of the CCR. Since the promoter of the *lac* operon is a CAP-dependent promoter, the transcription of structural genes is further stimulated when the CAP-cAMP complex binds to the CAP-cAMP binding site of the *lac* operon in the absence of adequate amounts of glucose (Figure 1-2) [7].

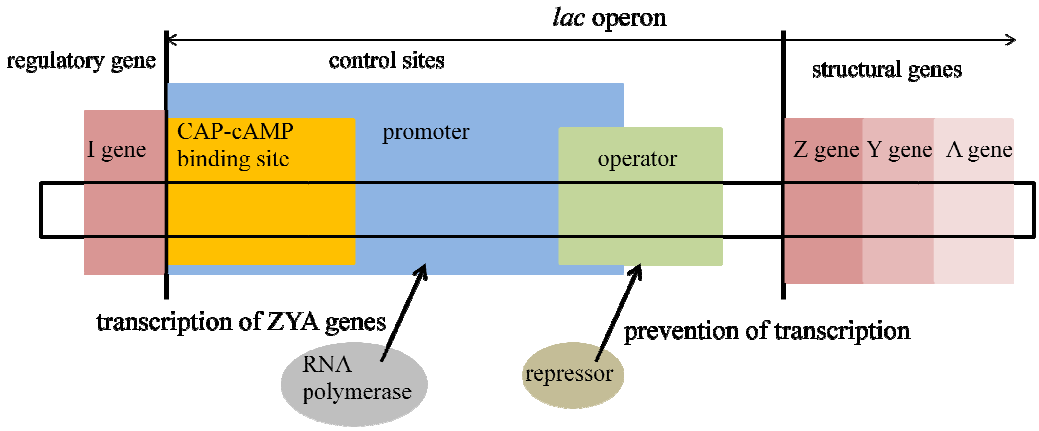



**Figure 1-2. The *lac* operon in *E. coli*.** The open black box represents the DNA sequence. The I gene encodes the repressor that binds to the operator site, leading to the prevention of transcription of the *lac* operon in the absence of the inducer. When lactose is available, an inducer generated from lactose binds to the repressor, thus

releasing it from the DNA, thereby RNA polymerase can bind to the promoter and transcribe the structural genes Z, Y, and A. [9].

### 1.3. The Maltose System

The maltose system is another notable example of transcriptional regulation in *E. coli* [2]. The maltose system is responsible for the catabolism of maltodextrins and glucose polymers containing up to 7 to 8 glucose units [6]. The maltose regulon is composed of five operons (*malEFG*, *malK-lamB-malM*, *malPQ*, *malS*, and *malZ*), and in *E. coli* there are ten *mal* genes that are regulated by MalT, the transcription regulator [6, 11]. There are many genes that affect maltose metabolism and regulation of *mal* gene expression and not all of them are part of the maltose regulon [6]. In this dissertation, the regulator MalT is the focus. The product of each gene of the maltose operon is a periplasmic, cytoplasmic or membrane protein. The operons *malEFG* and *malK-lamB* encode the binding protein dependent ABC transporter, *malP* and *malQ* encode essential enzymes, and *malS* and *malZ* encode nonessential enzymes for maltose and maltodextrin metabolism [11]. *malEFG* and *malK-lamB-malM* rely on the binding of CAP-cAMP in addition to MalT while *malPQ* and *malZ* do not [6, 12]. However, the function of *malM* , another *mal* gene regulated by MalT, is still unclear [6].

MalT, which is a positive regulator, is the transcriptional activator for the *mal* genes and is encoded by *malT* [2, 6, 12-13]. Since the expression of *malT* is under the regulation of the CCR, it is controlled by the CAP-cAMP complex and repressed by Mlc (make large colonies), which is a regulatory protein of the PTS and curbs the utilization of glucose and is

inactivated by $EIIB^{Glc}$ (Figure 1-3) [2, 6, 11-12]. MalT is activated by multimerizing, which is induced by maltotriose and ATP and repressed by MalK, MalY, and Aes (see section 3.1) [14]. MalY and Aes repress MalT by binding the N-terminus of MalT, which is also the ATP binding site, and competing with maltotriose for MalT binding [2]. It seems that the mechanism for these proteins to inhibit the MalT activity is similar but their binding sites are not identical, although they do seem to overlap [15].

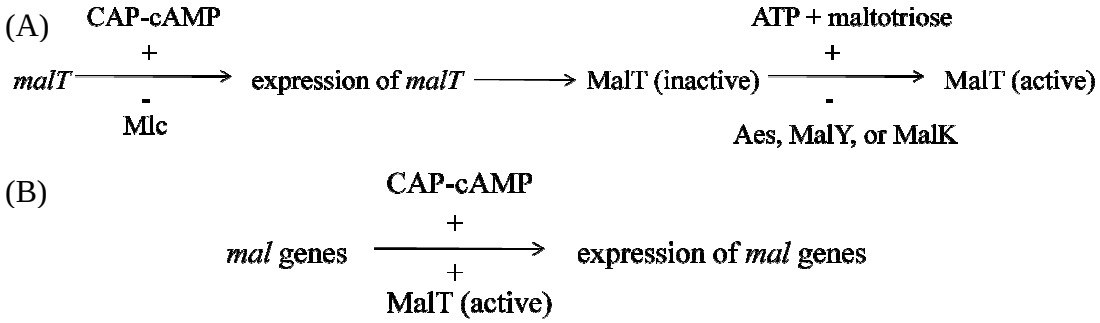



**Figure 1-3. The regulation of the maltose system in *E. coli*.** (A) The expression of *malT* is activated (+) by CAP-cAMP and repressed (-) by Mlc. MalT is activated by ATP and maltotriose and repressed by Aes, MalY, and MalK [2, 16]. (B) The expression of some of the *mal* genes is activated by the active form of MalT (active) and the CAP-cAMP complex, whereas the control of some of the other *mal* operons only depend on active MalT.

The maltose system can be induced by other sugars besides maltodextrins, such as galactose, lactose, and trehalose. Furthermore, maltotriose can be generated endogenously without maltodextrins [11]. It is suggested that there is a free internal unphosphorylated glucose generated in *E. coli* and it is involved in the pathway of endogenous maltotriose biosynthesis, leading to the induction of the maltose system. However, the origin of free glucose is still unclear [11]. One of the inducers for MalT, maltotriose, is formed when

maltodextrin or non-maltodextrin is available to the cells and, thereby, *mal* genes are also induced with the degradation of galactose or lactose [11].

There are still many functions that are unclear in the maltose system. To better understand this remarkable *E. coli* metabolic system, the crystal structure of Aes, one of the down-regulators for MalT, has been determined in this study since the crystal structure determination of the protein gives insight into functions and assemblies of the protein. Moreover, additional studies on the structures of MalT have begun to better understand the maltose system.

CHAPTER 2

ACETYL ESTERASE

## 2.1 Introduction

### 2.1.1 Literature Review for Aes

An acetyl esterase, also known as Aes from *Escherichia coli*, is a soluble protein consisting of 319 amino acids. Aes, whose molecular weight is 36 kDa, is involved in catabolic carbohydrate metabolism [17-18]. Aes belongs to the hormone sensitive lipase (HSL) family [19-21] and hydrolyzes substrates whose acyl chain length is less than eight carbons, and an example reaction is shown in Figure 2-1. Aes does not effectively hydrolyze large molecules such as trioleoyl glycerol and cholesterol oleate, thus showing no lipase activity [11, 19].

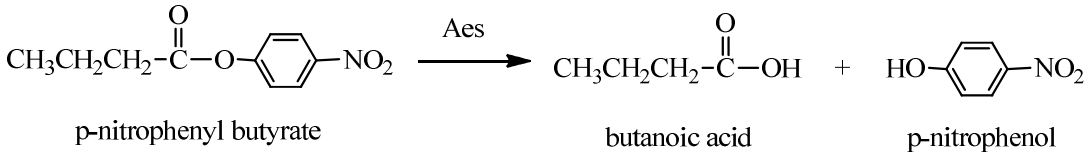



**Figure 2-1. Acetyl esterase activity of Aes.** The reaction above is carried out as the colormetric enzyme assay in order to know whether purified Aes is active. Aes catalyzes the hydrolysis of p-nitrophenyl esters of fatty acids whose acyl chain length is less than eight carbons.

Besides the catalytic activity, Aes down regulates MalT, which is the transcriptional activator of the maltose regulon [15, 22-23] (see section 3.1). Moreover, a recent study suggests that an interaction forms between Aes and α-galactosidase [24]. The *E. coli* α-galactosidase catalyzes the hydrolysis of α-linked-galactosides such as those found in the more complex sugars raffinose and melibiose [25]. Since Aes interacts with several important proteins in *E. coli*, it should play critical roles in the control of carbohydrate metabolism.

Aes is evolutionary related to the mammalian HSL, which is a member of the GDXG (X can be any amino acid) family, suggesting histidine or serine will be in the active site in either the residues 87 to 103 (LFYLHGGGFILGNLDTH) and/or 158 to 170 (IGFAGDSAGAMLA) [32]. The Gly-X-Ser-X-Gly pattern is found to be the consensus sequence around the active serine of the HSL family and is also found in esterases, serine proteases and lipases. Even though the GXSXG pattern is found in lipases, esterases and serine proteases, the tertiary structure of the serine proteases is different from the others and aspartate is sometimes substituted with glutamate to form the catalytic triad in some proteins [26-29]. Besides the GXSXG pattern, HGGG or HG for the lipolytic lipase (LPL) family (LPL, hepatic and pancreatic lipases) is also a conserved pattern in these: HSL, EST (carboxylesterases), and lipases families [20].

The enzymatic catalysis is necessary for living system and the active site (catalytic site) is the pocket where the enzyme-catalyzed reaction takes place. The

substrate is bound to the active site, which encloses and sequesters it from the solution [10]. The catalytic triad, consisting of Ser, His, and Asp, is found in the active site of certain enzymes, such as proteases, and plays the critical role in the enzymatic activity. Therefore, the recognition of the active site of the enzyme will lead to a better understanding of the function of the enzyme. According to recent studies, the catalytic triad of Aes is suggested to be Ser165, Asp262 and His292, and this triad is conserved in distantly related prokaryotes as well as in the mammalian hormone sensitive lipase (HSL) [20, 30].

Aes is classified as a lipolytic enzyme (PROSITE: PDOC00903) [31-32] and there are different classes of lipolytic enzyme produced in bacteria: carboxylesterases, lipases and phospholipases. Caryboxylesterases hydrolyze small esters and lipases hydrolyze the ester bond in long triglycerides. The substrates of carboxylesterases are soluble while lipases act efficiently on the insoluble part of the substrate. Such enzymes, which come from fungus or bacteria, are used in biotechnology as food ingredients, biocatalysts, and pitch control in the pulp and paper industry [30]. Therefore, a better understanding of such proteins will help in the discovery and development of novel enzymes used in these fields.

### 2.1.2 Classification and HSL

Cholinesterases, carboxylesterases, esterases, and lipases are the members of the α/β hydrolase fold family that shares the similar structure of a central β-sheet surrounded

by α-helices. This family is separated into three subfamilies, C-family, L-family and H-family. More specifically, acetylcholinesterases, cholesterolesterases, carboxylesterases, neuroactins and lipases from fungi belong to the C-family, lipases such as pancreatic lipase and lipoprotein lipase belong to the L-family, and hormone-sensitive lipase (HSL) and lipases from bacteria fit in the H-family [33]. HSL shows high relative activity at low temperature, below 15 °C, however, interestingly, the sequence similarity is observed in thermophilic (high temperature loving), mesophilic (moderate temperature loving) and psychrophilic (low temperature loving) organisms [34]. Therefore, temperature stability and activity preferences among them are not from the sequence conservation [30]. HSL is involved in many different metabolic processes and are present in heart, skeletal muscle, testes, and brown adipose tissue but not present in kidney and liver. Since HSL plays a role in energy metabolism, it may be involved in diabetes and obesity. HSL releases energy by hydrolyzing triacylglycerol, which is the major energy source for the heart, to glycerol and fatty acids under hormonal and neural control [35]. HSL is also in charge of hydrolysis of the neutral cholesterol ester in macrophages [36]. Therefore, it is possible that the accumulation of cholesterol ester in atherosclerosis would be reversed by the activation of HSL [35]. Moreover, cold adaptability could be controlled by HSL in poikilotherms whose body temperature varies with the environment and hibernators that depend on fat stores as their energy source during winter [37]. Therefore, it should be necessary for HSL to function at lower temperatures. HSL is a unique enzyme whose activity is controlled by cAMP-dependent phosphorylation and its structure obtained from sequence information and modeling is not homologous to other lipases, suggesting a unique function [38]. Therefore, it is interesting to understand the structure and function

of HSL that is involved in the energy homeostasis [38-39]. Moreover, since Aes from *E. coli* is homologous to the catalytic domain of HSL even though the whole HSL structure may be different to Aes, the structure determination of Aes should lead to understanding the functional and evolutional relationship of proteins belonging to the H-family including HSL [19-20, 40].

## 2.2 Materials and Methods

### 2.2.1 Genomic DNA Isolation from *Escherichia coli*

One Shot TOP10 Chemically Competent *Escherichia coli* cells (*E. coli* K12 strain, Invitrogen) were used to isolate genomic DNA [41]. First, 5 mL Luria-Bertani (LB) liquid media with 50 µg/mL ampicillin ($LB/Amp^{50}$) was inoculated with *E. coli* cells and incubated at 37 °C overnight with shaking (200 rpm, MaxQ 500 shaking incubator, Barnstead Lab-Line). An aliquot (1.5 mL) of the culture was centrifuged (Marathon microcentrifuge, Fisher Scientific) in a microcentrifuge tube and the supernatant was discarded. The pellet was then resuspended in 567 µL TE (50 mM Tris-HCl pH 8.0 and 10 mM EDTA (ethylenediamine tetraacetic acid)) buffer by pipetting and 30 µL of 10% SDS (sodium dodecyl sulfate) and 3.0 µL of 20 mg/mL proteinase K were added. The solution was mixed well and incubated for 1 hour at 37 °C. Sodium chloride (100 µL of 5 M) was added to the solution and mixed well, followed by the addition of 80 µL of CTAB/NaCl solution (prepared by dissolving 4.1 g NaCl in 80 mL water and adding 10 g hexadecyltrimethyl ammonium bromide (CTAB) while heating at 65 °C, adjusting the final volume to 100 mL). Afterwards, an equal volume of 24:1

chloroform/isoamyl alcohol was added, mixed thoroughly, and spun for 5 minutes in a microcentrifuge. The aqueous supernatant was transferred to a fresh microcentrifuge tube, an equal volume of 25:24:1 phenol/chloroform/isoamyl alcohol was added to the supernatant and the mixture was spun in a microcentrifuge for 5 minutes. The supernatant was transferred to a fresh microcentrifuge tube and 0.9 mL of isopropanol was added to the supernatant to precipitate the genomic DNA. The mixture was shaken well and spun in a centrifuge to harvest the DNA. The DNA was washed with 70% ethanol in water and centrifuged for 5 minutes to repellet. The supernatant was removed carefully and the pellet was dried by blotting on a paper towel. Afterwards, the pellet of DNA was redissolved in 100 µL TE buffer and stored at -4 °C.

### 2.2.2 Polymerase Chain Reaction and Gel Extraction of PCR Product

The aes gene (*aes*) was amplified by polymerase chain reaction (PCR) using a PTC-150 Minicycler (MJ Research). The forward primer 5'-CACCATGAAGCCGGAAAACAA ACTAC-3' and reverse primer 5'-CTAAAGCTGAGCGGTAAAGAACTG-3' were purchased from Invitrogen and these primers were diluted with sterile deionized water to 20 µM to be used in PCR. To perform PCR, 5.0 µL of 10X AccuPrime PCR Buffer II (Invitrogen), 1.0 - 4.0 µL of both the forward and reverse primers (20 µM), 1.0 - 4.0 µL of DNA isolated from *E. coli* (section 2.2.1), and 1.0 µL AccuPrime Taq Polymerase (Invitrogen) and sterile distilled water to adjust the final volume to 50 µL, were mixed in sterile PCR tubes. In the 1$^{st}$ cycle, the DNA was denatured for 6 minutes at 94 °C, annealed for 30 seconds at 55 °C, and extended for 1 minute at 68 °C. From the 2$^{nd}$ to the 30$^{th}$ cycle, denaturation for 45

seconds at 94 °C, annealing for 30 seconds at 55 °C, and extension for 1 minute at 68 °C were performed. As the final cycle, extension was done for 5 minutes at 68 °C (Table 2-1).

**Table 2-1. Program used in minicycler in order to amplify *aes* by PCR.** The minicycler was programmed in this way in order to amplify *aes*.

| | 1st Cycle | 2nd -30th Cycle | Final Cycle |
|---|---|---|---|
| Denaturation 94 ˚C | 6 min. | 45 sec. | |
| Annealing 55 ˚C | 30 sec. | 30 sec. | |
| Extension 68 ˚C | 1 min. | 1 min. | 5 min. |

Most of the 957 bp PCR product (40 µL) was mixed with 8.0 µL of loading buffer (30% glycerol, 20 mM EDTA, and 100 µg/mL crystal violet) and loaded onto a 1.0 % agarose gel (1.5 g agarose in 150 mL of TAE (Tris-acetate-EDTA: 40 mM Tris pH 7.6, 20 mM acetic acid, and 1 mM EDTA buffer)) with 750 µL of crystal violet solution (2 mg/mL) in deionized water for the detection of the PCR product. The PCR product was verified through agarose gel electrophoresis (80 V for 3 hours, Sub-Cell GT agarose gel electrophoresis, BioRad).

The Invitrogen PureLink Quick Gel Extraction Kit was used in order to purify the PCR product from the agarose gel so that any impurities were removed from the PCR reaction mixture. The bands containing the correct size of the PCR product were cut from the agarose gel with a sterile razor blade and up to 400 mg of gel were placed into 1.5 mL microcentrifuge tubes. The Gel Stabilization Buffer (GS1) was added to the tube with 30 µL for every 10 mg of the gel. The tube was incubated for 15 minutes at 50 °C

with mixing every 3 minutes. After the gel slice was visibly dissolved, the tube was incubated for an additional 5 minutes to ensure the complete dissolution of the gel.

The mixture was transferred into the Quick Gel Extraction Column that was placed onto the Wash Tube and spun for 1 minute in the microcentrifuge at the maximum speed. The flow through was discarded, 500 µL of GS1 was added to the column, and the tube was incubated at room temperature (RT) for 1 minute. The flow through was discarded and 700 µL of Wash Buffer (W9) with 70% ethanol was added to the column, followed by incubation at RT for 5 minutes. Afterwards, it was centrifuged for 1 minute and the flow through was discarded. The tube was centrifuged for 1 more minute to ensure that any buffer was removed. The column was placed onto the Recovery Tube and 50 µL of pre-warmed Tris-HCl (10 mM, pH 8.0) was added to the center of the cartridge. The tube was incubated at RT for 1 minute and centrifuged for 2 minutes at the maximum speed. The purified DNA was stored at -20 °C.

The agarose gel electrophoresis was performed to confirm the size of the purified PCR product in the same manner as written above except for using ethidium bromide (EB, 0.5 µg/mL of agarose gel) for the detection of the DNA. The DNA was visualized with UV light and the concentration of the purified PCR product was determined as 4-10 µg/ µL according to the DNA mass ladder (Invitrogen). Afterwards, the purified PCR product was used to construct the plasmid.

### 2.2.3 The TOPO Cloning Reaction and Transformation

The TOPO cloning reaction was performed using the Champion$^{TM}$ pET Directional TOPO Expression Kits (Invitrogen) [42]. Reaction mixtures were prepared in sterile microcentrifuge tubes ((A), (B), and (C) in Table 2-2), followed by the incubation at RT for 5 minutes. Two negative controls (tubes (A) and (B)) and one positive control (tube (D)) were performed in order to check the background of vector and PCR product, and the competency of the *E. coli* strain.

**Table 2-2. Reaction mixtures set up for the TOPO cloning reaction.** (A) and (B) are negative controls to check the background of the PCR product and the vector, respectively. (C) is the cloning reaction mixture. (D) is the positive control using control pUC 19 that is a part of the TOPO expression kit. Values are in µL.

| | (A) | (B) | (C) | (D) |
|---|---|---|---|---|
| Gel extracted PCR product | 1.5 | - | 1.5 | - |
| Salt solution | 1 | 1 | 1 | - |
| Sterile water | 3.5 | 4 | 2.5 | - |
| TOPO vector | - | 1 | 1 | - |
| pUC 19 | - | - | - | 1 |
| Total volume | 6 | 6 | 6 | 1 |

After the incubation, the tube was transferred onto ice and 3.0 µL of the reaction mixture was mixed into a vial of One Shot TOP10 Chemically Competent *E. coli* (Invitrogen) on ice without pipetting up and down. For (D), pUC19 (1 µL) was added directly to a vial of competent cells on ice. The vials were incubated on ice for 15 minutes and then the bacterial cells were heat-shocked for 30 seconds at 42 °C in a water bath. The vials were transferred back to the ice bath and 250 µL of room temperature S.O.C. medium was added to each of the vials. The mixtures were then incubated horizontally at 37 °C with shaking at 200 rpm for 1 hour. An aliquot (200 µL) of each of the transformation mixtures were spread using a sterile bent glass rod on pre-warmed (37

˚C) LB agar plates containing 100 µg/mL of ampicillin (LB/Amp$^{100}$). The plates were inverted and incubated overnight at 37 °C.

### 2.2.4 Analysis of the Transformants

Well isolated colonies (about 10 colonies) were picked using sterile toothpicks from the LB/Amp$^{100}$ agar plate and each colony was used to inoculate 10 mL of LB/Amp$^{100}$ medium. They were cultured for 12 hours at 37 °C shaking at 200 rpm. For propagation and maintenance of the plasmid, 0.9 mL of the culture was mixed with 0.3 mL of 60% sterile glycerol solution in a 2 mL cryo-vial and stored at -80 °C.

The cells were harvested with centrifugation at 4500 rpm for 5 minutes at 4 °C (Allegra X-15R Benchtop Centrifuge, Beckman Coulter) in order to isolate the plasmid from the cells with the use of the Wizard Plus SV Miniprep DNA Purification System (Promega) [43]. The supernatant was discarded and the tube was inverted onto a paper towel so that the supernatant was removed from the tube. The pellet was resuspended in 250 µL of the Cell Resuspension Solution (50 mM Tris-HCl pH 7.5, 10 mM EDTA, and 100 g/mL RNase A) by vortexing. The resuspended cells were transferred to a sterile 1.5 mL microcentrifuge tube and mixed with 250 µL of the Cell Lysis Solution (0.2 M NaOH and 1% SDS) by inverting the tube 4 times. The mixture was incubated at RT for 4 minutes and 10 µL of Alkaline Protease Solution was added to the mixture, followed by inverting the tube 4 times. The mixture was incubated at RT for 5 minutes. This process inactivates the endonuclease and other proteins released during the lysis of bacterial cells.

A 350 µL volume of the Neutralization Solution (4.09 M guanidine hydrochloride, 0.759 M potassium acetate, and 2.12 M glacial acetic acid) was then added to the mixture and immediately mixed by inverting the tube 4 times. The lysate was spun at maximum speed with the Marathon Microcentrifuge (Fisher Scientific) for 20 minutes at RT. The cleared lysate was then transferred to the Spin Column, which was inserted into a 2.0 mL Collection Tube. It was centrifuged at maximum speed for 1 minute at RT. The flow through was discarded from the tube and 750 µL of the Column Wash Solution (60 mM potassium acetate, 8.3 mM Tris-HCl, 60 % ethanol, and 0.04 mM EDTA) was added to the column. The tube was spun for 1 minute at RT and the flow through was discarded. The column was washed with 250 µL of Column Wash Solution and spun for 2 minutes at RT. The Spin Column was placed into a sterile microcentrifuge tube and the plasmid was eluted by the addition of 100 µL of Nuclease-Free water, followed by centrifugation for 2 minutes. Afterwards, the isolated plasmid was stored at -20 °C.

To confirm the presence of the insert of the aes gene (*aes*), the plasmid was subjected to restriction enzyme digestion and the fragments (2 kbp and 4 kbp fragments were expected to be seen on the agarose gel if *aes* was inserted) were run on a 1% agarose gel. As shown in Table 2-3, 6.5 µL of the sterile deionized water, 8 µL of plasmid DNA (10 µg/µL), 4 µL of MultiCore Buffer (Promega), and 0.5 µL acetylated BSA (10 µg/µL) were mixed in sterile microcentrifuge tubes and centrifuged for 30 seconds. The restriction enzymes SacI, which cuts the vector at position 416, and EcoRV, which cuts the vector at 556 and 4167, (both are from Promega), were added to

the tubes gently by pipetting and the tubes were spun for 30 seconds. The tubes were then incubated in a water bath at 37 °C for 2 hours [44].

| **Table 2-3. Restriction enzyme digestion.** The following components were mixed in sterile microcentrifuge tubes to verify the plasmids by use of restriction enzymes. | |
|---|---|
| Sterile deionized water | 6.5 µL |
| 10x restriction enzyme buffer (multicore) | 4.0 µL |
| Acetylated BSA (10 µg/ µL) | 0.5 µL |
| Plasmid DNA (ng/µL) | 8.0 µL |
| Sac I (10 units/µl) | 0.5 µL |
| Eco RV (10 units/µl) | 0.5 µL |
| Total volume | 20.0 µL |

A 1.0 % agarose gel was prepared in the same way as stated in section 2.2.2 using ethidium bromide as a DNA detection tool. After the incubation of the plasmid with the restriction enzymes, the digested samples (5 µL) were electrophoresed at 80 V for 2 hours. The plasmids that were confirmed to carry the insert by the restriction enzyme analysis were further verified with DNA sequencing at the Recombinant DNA/Protein Resource Facility of the Biochemistry and Molecular Biology Department (Oklahoma State University). The T7 promoter and terminator primers were used to verify the insert of *aes* into the pET100/D-TOPO vector with the correct size and orientation.

### 2.2.5 Over-Expression of Aes

The recombinant plasmids were transformed into BL21 Star™ (DE3) One Shot cells (Invitrogen) which were designed for the expression of genes regulated by the T7 promoter. To transform the construct, 2 µL of the plasmid DNA (10-15 µg/µL) was added into the vial containing BL21 Star cells on ice and stirred gently with a pipette tip.

The vials were incubated on ice for 30 minutes and then heat-shocked for 30 seconds at 42 °C in a water bath. The vials were immediately transferred back onto the ice, followed by the addition of S.O.C. medium (250 µL, pre-warmed to RT). Afterwards, the vials were incubated at 37 °C for 30 minutes with shaking at 200 rpm. The transformation mixture (one vial per transformation) was added to 10 mL of LB/Amp$^{100}$ liquid medium (a seed culture) and grown overnight at 37 °C with shaking at 200 rpm.

Since each recombinant protein has different characteristics for its expression, a time course study of the expression of Aes was performed in order to investigate the best condition for the protein expression. Two aliquots of seed culture (750 µL) were used to inoculate two tubes, each with 10 mL of LB/Amp$^{100}$, one tube was for the induction of the protein and the other was not induced. The bacteria were grown for about 2 hour at 37 °C with shaking until an $OD_{600}$ reached 0.6. The cells in the first tube were induced for over-expression of the recombinant protein with 1 mM of isopropyl-β-D-1-thiogalactopyranoside (IPTG, Sigma) and an aliquot (500 µL) from each tube was collected in a microcentrifuge tube and centrifuged for 30 seconds every hour, starting at the time of induction. After discarding the supernatant, cell pellets were frozen at -20 °C for later sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. The incubation of the remaining culture was continued at 37 °C for 6 hours.

To analyze the protein expression, pellets were thawed at RT and resuspended in 80 µL of SDS-PAGE sample buffer by vortexing. After incubation at 100 °C for 5

minutes, the microcentrifuge tubes were centrifuged briefly and 5-10 µL of sample from each tube was loaded onto a 15% SDS-PAGE gel, followed by electrophoresis at 200 V for about 30 minutes using a Mini-Protein 3 Cell electrophoresis apparatus (BioRad) [45]. The solubility of the protein was also studied by sonicating (Sonifier 150 Liquid Processor, Branson) the pellet from 4 mL LB culture, which was resuspended in 500 µL of ice-cold Buffer A (20 mM TrisHCl pH 7.9, 500 mM NaCl, and 20 mM imidazole), for 30 seconds and 1 minute rest on the ice bath. This process was repeated three times so that the majority of the cells could be disrupted completely. The sample was centrifuged at 4 °C for 1 minute to pellet insoluble proteins.

The supernatant was transferred to a microcentrifuge tube and 500 µL of 2x SDS-PAGE sample buffer was added to the tube, followed by boiling at 100 °C for 5 minutes. The pellet was resuspended with 450 µL of 1x SDS-PAGE sample buffer, boiled at 100 °C for 5 minutes, and spun for 1 minute. The samples from the supernatant (10 µL) and the pellet (5 µL) were loaded onto a 15% SDS-PAGE gel and electrophoresed. Since Aes (confirmed by westernblotting) was found more in the pellet (insoluble phase) than in the supernatant (soluble phase), the optimal conditions for culturing to make Aes more soluble was investigated. After trials, it was found that culturing should be carried out at 30 °C after the induction and the addition of lysozyme allowed Aes to be more soluble.

Since a large amount of protein was necessary to obtain protein crystals, the expression of the protein was scaled-up. To scale-up the expression, 200 mL of

LB/Amp$^{100}$ was inoculated with 5 mL of the seed culture and incubated at 37 °C until the $OD_{600}$ reached 0.6. The over-expression of the Aes was induced by adding 1 mM IPTG and the temperature was lowered to 30 °C and cultured for 4 additional hours with shaking at 200 rpm. After the time course of the large scale expression study, Aes was over-expressed and found mostly in the soluble phase. Thus, cells were harvested by centrifugation at 4500 rpm, 4 °C and stored at -80 °C until needed.

### 2.2.6 Purification of Aes 1: Ni Chelate Affinity Chromatography

The cell pellet from 400 mL culture was resuspended in 10 mL of Buffer A and 0.005 g of lysozyme (Invitrogen) was added, followed by incubation with shaking (200 rpm) at RT for 15 minutes. The sample was then sonicated (initially Branson Sonifier 150 and then XL-2000 series Ultrasonic Liquid processors) on an ice bath for 30 seconds and 1 minute rest three times. The lysate was centrifuged at 4500 rpm for 15 minute in a sterile 50 mL falcon tube and then the supernatant was transferred to a 15 mL centrifuge tube. The tube was centrifuged at 4500 rpm for a few hours until the supernatant became clear.

The Biologic Duo-Flow Chromatography System (BioRad) was used to purify Aes. Since the recombinant Aes contained an N-terminal poly-histidine tag ($His_6$-Aes), immobilized metal affinity chromatography was used for the purification [46]. The supernatant containing Aes was loaded onto the HisTrap$^{TM}$ HP Column (GE Healthcare) and Aes was eluted by a linear gradient with increasing imidazole concentration (1%

increase of imizadole concentration per minute). During the purification process (flow rate 1.0 mL/min), the buffer blender was used to blend Buffer A1 (40 mM TrisHCl and 1.0 M NaCl) and Buffer A2 (40 mM Tris and 1.0 M NaCl) so as to adjust the pH to 7.9. This was further blended with Buffer B1 (deionized water) and Buffer B2 (1.0 M imidazole). The linear gradient for the imidazole concentration was created by adjusting the percentage of B2. Fractions were collected and select fractions were used for a 15 % SDS-PAGE to verify the presence of Aes. Since Aes was eluted with 240 mM imidazole, step wise purifications were performed for subsequent purifications, which eluted most of the impurities at 20 mM imidazole, more tightly bound impurities at 100 mM and Aes at 500 mM imidazole.

### 2.2.7 Purification of Aes 2: Desalting Column Purification

Since impurities were not observed in fractions containing Aes on the 15% SDS-PAGE gel after the Ni chelate affinity chromatography, samples were loaded onto the desalting column (HiPrep 26/10 Desalting GE Healthcare) in order to remove excess amount of imidazole and to exchange the buffer from Tris to Buffer B (25 mM ammonium acetate pH 7.5, 150 mM NaCl, and 150 mM imidazole). The fractions from the desalting column were analyzed with 15 % SDS-PAGE. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF) at the recombinant DNA/Protein Resource Facility of the Biochemistry and Molecular Biology Department was used to confirm that Aes was present. Because purified protein was confirmed as Aes, fractions were pooled and concentrated to 5 mg/mL with a 20 mL VIVA SPIN, 10000 molecular weight cut-off (MWCO) (ViVa Science) by centrifugation at 1500 rpm

and 4 °C. The concentration was determined by the Bradford protein assay using BIO-RAD Quick Star Bradford Protein Assay Kit [47]. Moreover, the enzyme assay using p-nitrophenyl butyrate was performed to see whether the purified Aes was active [19].

### 2.2.8 Crystallization

To determine the preliminary crystallization conditions for Aes, Crystal Screen™ I and II (Hampton Research) were used in the hanging drop vapor diffusion method [48]. The purified Aes sample in Buffer B (5 mg/mL, filtered with a 0.45 µm HT Tuffryn membrane filter) was used in the screening. For Screen I and II, 24 well crystallization plates (Hampton Research) were used and 750 µL of each reagent was pipetted into the wells, which had the rim greased with high vacuum grease (DOW CORNING). On a microscope cover slip (VWR), a droplet of 3.0 µL of Aes and 3.0 µL of the reservoir solution was prepared, and the microscope cover slip was inverted over the well and sealed with the grease. Conditions were duplicated for RT and 4 °C. For the Index Screen (Hampton Research), the sitting drop vapor diffusion method was carried out with a 96 well crystallization plate (Axygen). A 150 µL aliquot of screen reagent was pipetted into the wells and 1.0 µL of Aes was mixed with 1.0 µL of the reservoir solution in a smaller well formed in the plate for sitting drop experiments. Afterwards, the plate was sealed with clear non-reactive tape and stored at RT. All plates were checked periodically for crystal growth. In addition to these methods, a silica hydrogel (Hampton Research) sitting drop vapor diffusion method was performed since rapid nucleation for most of the conditions was observed, especially for plates at 4 ˚C.

Crystals were found in a few weeks to a month with the following conditions (see section 2.3.3): Index # 9, 12, 15, 44, 47, Screen I # 7, 20, and Screen II # 31. Since crystals formed in Screen I # 7 (0.1 M sodium cacodylate trihydrate pH 6.5 and 1.4 M sodium acetate trihydrate) were single and relatively big, crystals were reproduced with a hanging drop vapor diffusion method with 1.0 mL freshly prepared reservoir solution and the crystallization condition was optimized to obtain well defined and high quality single crystals. To optimize the crystallization conditions, the concentration of Aes, the pH of the buffer, and the concentration of each reagent were altered. After optimization, crystals of the size 0.45 x 0.35 x 0.15 $mm^3$ grew in a few weeks.

### 2.2.9 Data Collection

Once good single crystals were formed, preliminary data collections were performed. Before the data collection, single crystals were lassoed by a nylon loop (0.05 to 0.5 mm diameter, Hampton Research) from the drop, dunked in oil (Paraton-N, Hampton Research), and then immediately mounted onto the goniometer head of the X-ray diffractometer (Bruker Microstar) under a nitrogen cryo flow (100 K: Oxford Cryosystems COBRA). The X-ray generator was set with the power at 2.5 kW (45 kV and 60 mA). Once the crystal was mounted, it was aligned in the center of the X-ray beam. To test whether diffraction occurs, various parameters were set as follows with the system software (Proteum2) [49]. The distance from the goniometer to the detector was set at 70 cm and exposure time was 120 seconds. If no diffraction was observed on the

detection screen, the crystal was removed and other crystals would be mounted to test whether diffraction occurred. If they diffracted, the unit cell was determined by Proteum2 and exposure time was extended to 300 seconds with the distance being set at 120 cm for preliminary data collections [50]. More specifically, for the preliminary data collection, a data collection scheme was developed by Proteum2 to expose the crystal to the X-ray beam for several frames at different angles. Each frame was collected for 300 seconds followed by a 0.5 degree rotation of the crystal. A total of 724 frames were collected. To determine the structure of Aes, diffraction data were collected for about 24 hours from a single crystal of $His_6$-Aes at the wavelength 1.5418 Å.

After the data collection, raw data were processed with SAINT-NT and XPREP [51-52]. More specifically, raw crystallographic data frames were converted by SAINT-NT into integrated intensity sets with error analysis, background subtraction, etc. The processed data were then analyzed by XPREP so as to determine the space group and evaluate data statistically. The number of molecules per asymmetric unit was also determined with Matthew's coefficient ($V_M$) using the following equation [53-54]:

$$V_M = V/ (n*M*X)$$

where V is the volume of the unit cell, n is the number of asymmetric units in the unit cell, M is the molecular weight and X is the number of molecules in the asymmetric unit.

When $V_M$ falls within the range from 1.68 to 3.53 $Å^3$/Dalton, the value of X is considered to be most probable. The Matthew's coefficient also gives the expected solvent content (obtained by $1\text{-}1.23/V_M$) that should be in the range of 27 to 65%. After the determination of the space group and the number of molecules per asymmetric unit, data were transferred to a Linux based workstation (OS: Ubuntu) in order to convert .hkl data to .mtz data that is hkl data in binary format used in CCP4 [55].

### 2.2.10 Molecular Replacement

In order to determine the initial phases for the data collected, molecular replacement was carried out with PHENIX [56-57] using the homologous proteins: acetyl esterase from *Salmonella typhimurium* (PDB code: 3GA7, 70% identity), hyper-thermophilic carboxylesterase from *Archaeon archaeoglobus* (PDB code: 1JJI, 29 % identity) [58], carboxylesterase from a metagenomic library (PDB code: 2C7B, 27% identity) [59], thermophilic carboxylesterase EST2 from *Alicyclobacillus acidocaldarius* (PDB code: 1EVQ, 26% identity) [60], and mutant M211S/R215L of carboxylesterase EST2 complexed with hexadecanesulfonate (PDB code: 1QZ3, 26% identity) [57, 59-63]. The percent identity for each homologous protein was obtained from RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) [64]. Since initial phases were determined using a homologous protein, the acetyl esterase from *Salmonella typhimurium* (3GA7), the amino acid sequence was based on this sequence. Therefore, after phase determination, the amino acids were altered with the molecular modeling software COOT [65] so that it would possess the same sequence as Aes from *E. coli* K-12.

First, molecular replacement was performed as follows with dataset $His_6$-Aes1, which was collected on a single crystal and refined in the rhombohedral *R*32 space group. The PDB file of 3GA7 was downloaded from the PDB and water molecules and a phosphate were removed from the file so that only the protein coordinates of 3GA7 were obtained. On a Linux workstation (OS:Ubuntu), *phenix.automr data.mtz* (reflection data collected in this study) *3ga7.pdb rms=1.0* (pdb file of 3ga7), italicized letters being the syntax to start the program, was carried out on the terminal. As a result, MR.1.pdb and MR.1.mtz were generated as protein coordinates and phased structure factors to generate the electron density map, respectively. In order to know whether the molecular replacement worked for the initial phase determination, *phenix.refine MR.1.pdb data.mtz* were performed. Since $R_{free}$ = 0.46663 dropped to $R_{free}$ = 0.4415 after the refinement, initial phase determination was considered successful and then further investigations were carried out as follows. The mtz file was opened in COOT to check whether protein coordinates fit to the electron density map (rendered at $\sigma$ = 1.0) and alternative conformations were removed from the protein coordinates (saved as start.pdb). Then, *phenix.refine start.pdb data.mtz* was performed and $R_{free}$ = 0.4213 before refinement and $R_{free}$ = 0.4052 after the refinement were observed. After the refinement, start_refine_002.def (show the result of refinement) was generated and then it was used in further refinement. Some parameters were modified on start_refine_002.def as b-factor = 20.0 and incorporated simulated annealing = True and then *phenix.refine start_refine_002.def* was performed. After repeating these refinement steps, $R_{work}$ lowered but $R_{free}$ increased even though both R values should lower if molecular

replacement succeeded in phase determination. Therefore, this phase determination was considered not good enough for further structure determination. After trying the same procedures for another dataset, $His_6$-Aes3 that was from a different crystal and also in the rhombohedral space group (possibly *R*3), the molecular replacement was not really successful.

The same procedure was then performed with the dataset $His_6$-Aes2, which was from a single crystal that crystallized in the tetragonal $P4_1$ space group. Since it showed better R values after refinements, protein coordinates generated from this data led to the structure determination of Aes. According to the Matthew's coefficient analysis and molecular replacement, six molecules were found in the asymmetric unit. Therefore, after initial phase determination, relationships among these molecules were investigated with the molecular visualization program Rasmol in order to see the non-crystallographic symmetry [66]. According to the symmetry investigation, there were three dimers in the asymmetric unit in which chain A forms a dimer with chain E, chain C forms a dimer with chain F, and chain B forms a dimer with chain D.

Afterward, the model of Aes in $P4_1$ ($His_6$-Aes2) was manually re-built with COOT by adjusting the amino acid coordinates of the model used in the molecular replacement to fit the electron density map obtained from the crystal data and the model phases [65]. At the early stage of this process, model building was performed on chain A and the same modification was applied to all other chains using "SSM superposed" on

COOT, which simply copied the modified chain A to the location of the other five molecules. To achieve this, set reference structure = pdb file containing all chains and set moving structure = chain A. Moreover, change chain ID from A to B so that the modification on chain A is applied to chain B. Performing the same translation to the other chains, six molecules in the asymmetric unit have the same structure as chain A. Therefore, non-crystallographic symmetry (NCS) was applied while refining the model with PHENIX. The model was alternatively refined with PHENIX (see section 2.2.11) and built manually with COOT until the agreement between the data and model converged, in other words, $R_{work}$ and $R_{free}$ had reached to less than 30% and no more improvement could be obtained [57, 67].

### 2.2.11 Refinements

After all the amino acids were altered to the ones corresponding to Aes from *E. coli*, further investigations for each amino acid in chain A were performed with COOT in order to obtain a better model of Aes. Initially, side chains of amino acids in chain A were truncated into alanine if poor electron density was observed and the same changes were applied to other chains as mentioned in 2.2.10. Moreover, since the electron density was missing around residues 26 to 35, these amino acids were removed at this point. Refinements, using PHENIX, were then performed with simulated annealing in order to obtain a better model that fit to the electron density map [68]. After each refinement, protein coordinates were observed in COOT to investigate whether the electron density map had improved so that the residues truncated to alanine could be altered to the proper amino acid. Depending on the density obtained from the refinement, some amino acid

residues were altered or built to fit to the density map. The same modifications were also performed on the other chains followed by refinement with PHENIX.

After several rounds of refinement, translation/libration/screw (TLS) [69-70] was applied in the refinement. TLS parameterizes a model in a physically sensible form (group) that will be refined by translation, libration, and screw components [71]. Initially, a TLS parameter file was prepared (tls_group_selections.params) where each chain was treated as a group. Since there were 6 molecules in the asymmetric unit, chain A was firstly refined and then copies of chain A were produced to generate all 6 molecules of the asymmetric unit along with simulated annealing and TLS to improve the model and electron density map.

At the beginning, the electron density of an N terminal loop (residues 26-36) was noisy and not contiguous as mentioned above. However, after several rounds of refinement, part of this density appeared. Therefore, a few amino acid residues were built step by step to complete the loop region around the N-terminus. Since this loop region did not obey the NCS well with each chain showing a different conformation, this region was excluded from the NCS definition and the loop of each molecule was manually modeled based on the Ramachandran plot, torsion angle and geometry around each amino acid residue. After the fifth round of refinement, water molecules were added automatically by PHENIX. Moreover, investigations for some positive density around the plausible active site for Aes were performed to determine what types of small

molecules fit and were chemically and physically reasonable. However, these were not successful. Therefore, the model without water molecules or small molecules were retained and refined to determine the structure of Aes.

When almost all of the amino acid coordinates were built, Molprobity [72] was performed to validate the model (see Table 2-7). According to the comprehensive validation, there were still a few residues that were out of the preferred region in the Ramachandran plot (see Table 2-8). Thus, these residues were manually refined and also refined with TLS, and moderate NCS excluding residues that were unique among each chain.

Afterwards, using the model, TLS Motion Determination (TLSMD) was performed in order to analyze the flexibility of the protein crystal structure [73-74], leading to the generation of a new TLS parameter file that better represents the motion of the six molecules in the asymmetric unit of the crystal. After trying several different TLS partitioning for each molecule, it worked better that chain A and F were partitioned into 5 parts, chain B and C into 4, and chain D and E into 3. Therefore, the TLS parameter file for PHENIX was generated in the TLSMD server with this partitioning and used in further refinements. In addition to this new TLS parameter, rotamer and geometry validation were performed in COOT (see Table 2-9). Based on rotamer validation and the Ramachandran plot, some residues that were unique and outliers were modified so that they obtained preferred rotamers and geometries and fell into the favored region on

the Ramachandran plot. Moreover, these residues were excluded from NCS in refinement since they showed different orientations in the different chains. After modifications to the rotamers and outliers, refinement with PHENIX was performed followed by validation using Molprobity. These processes were repeated several times to improve the model. Moreover, twining analysis was re-performed using updated software with *phenix.xtriage* [56-57].

Once the structure was determined, proteins with similar fold were found using the DaLi server [75] to find proteins that shared the fold with Aes. Moreover, the multiple sequence alignment (MSA) was performed by ClustalW [76] so as to understand the relationship between the sequence similarity and structure, using the following homologous proteins: an acetyl esterase from *Salmonella typhimurium* (PDB ID: 3GA7), a hyper-thermophilic carboxylesterase from *Archaeoglobus fulgidus* (1JJI), a new thoermophilic and thermostable carboxylesterase cloned from a metagenomic library (2C7B), a thermophilic carboxylesterase Est2 from *Alicyclobacillus acidocaldarius* (1EVQ), a mutant M211S/R215L of a carboxylesterase Est2 complexed with hexadecanesulfonate (1QZ3), a hormone-sensitive lipase like Este5 from a metagenomic library (3FAK), 3DNM is a hormone-sensitive lipase from a metagenomic library, a brefeldin A esterase that is a bacterial homologue of human hormone-sensitive lipase (1JKM), HSL from *Homo sapiens*, and HSL from *Rattus norvegicus*.

## 2.3 Results and Discussion

### 2.3.1 PCR

The set of reactions that were set up to lead to the production of the *aes* gene in PCR is summarized in Table 2-4. Initially, the reaction volume was 25 µL but it was not enough for sequencing and further research. Therefore, 50 µL of reaction mixture was prepared for further research.

| **Table 2-4. Reactions that led to the production of *aes*.** (Volumes in µL). | | | | | |
|---|---|---|---|---|---|
| 10X Accu Prime PCR Buffer II | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| *aes* forward (20 µM) | 0.5 | 1.5 | 1.0 | 2.0 | 0.5 |
| *aes* reverse (20 µM) | 0.5 | 1.5 | 1.0 | 2.0 | 0.5 |
| DNA template | 1.0 | 0.5 | 1.0 | 0.5 | 0.5 |
| Accu Prime Taq Polymerase | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| Sterile Distilled Water | 20.0 | 18.5 | 19.0 | 17.5 | 20.5 |
| Total Volume | 25 | 25 | 25 | 25 | 25 |

Agarose gel electrophoresis was carried out to check the size of the PCR product that was purified from the agarose gel (Figure 2-2). The size of the PCR product was found to be about the same size as the *aes* gene and the concentration of the purified PCR product was found to be about 10 ng/µL. After the validation of the PCR product, purified PCR product was used in the plasmid construction.

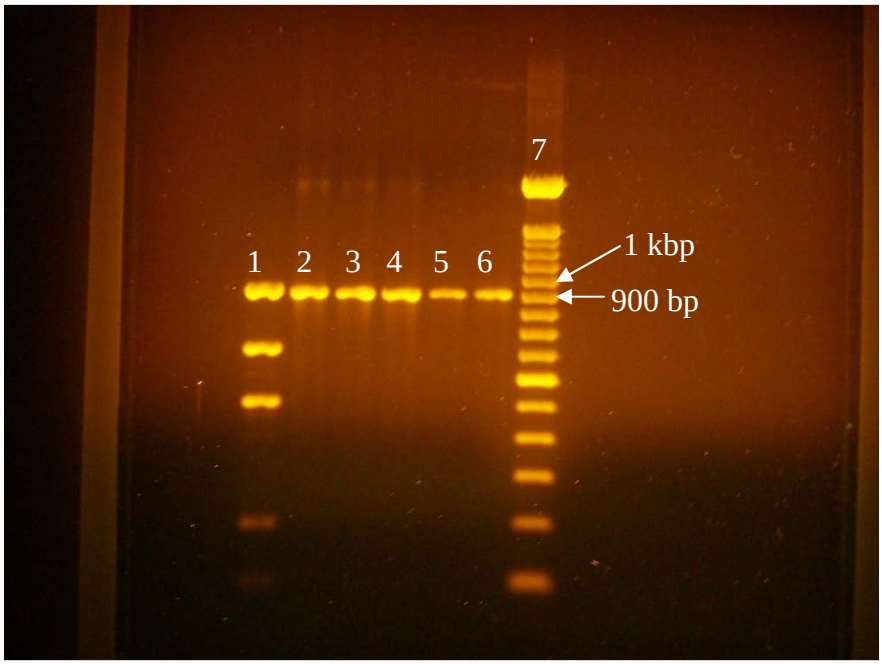



**Figure 2-2. Size analysis of the PCR product by agarose gel electrophoresis.** 1% agarose gel was stained with ethidium bromide. Lane 1: density marker, lanes 2-6: PCR products extracted from the agarose gel, and lane 7: 100 bp size marker.

### 2.3.2 Over-Expression and Purification of Aes

The plasmids carrying the *aes* gene were constructed and verified by restriction enzyme analyses and DNA sequencing as described in Materials and Methods. According to sequencing, there were no mutations found in the recombinant *aes*. After a series of investigations about the optimal condition for *aes*'s over-expression, it was found that *aes* needed to be incubated for 4 hours at 30 °C after induction by the addition of 1.0 mM IPTG. At the beginning, Aes was found more in the insoluble phase after sonication but the solubility of Aes was improved with more Aes in the soluble fraction by lowering the temperature during culturing (Figure 2-3). As a result, a large quantity of Aes was purified through Ni (II) chelate affinity chromatography (Figure 2-4).

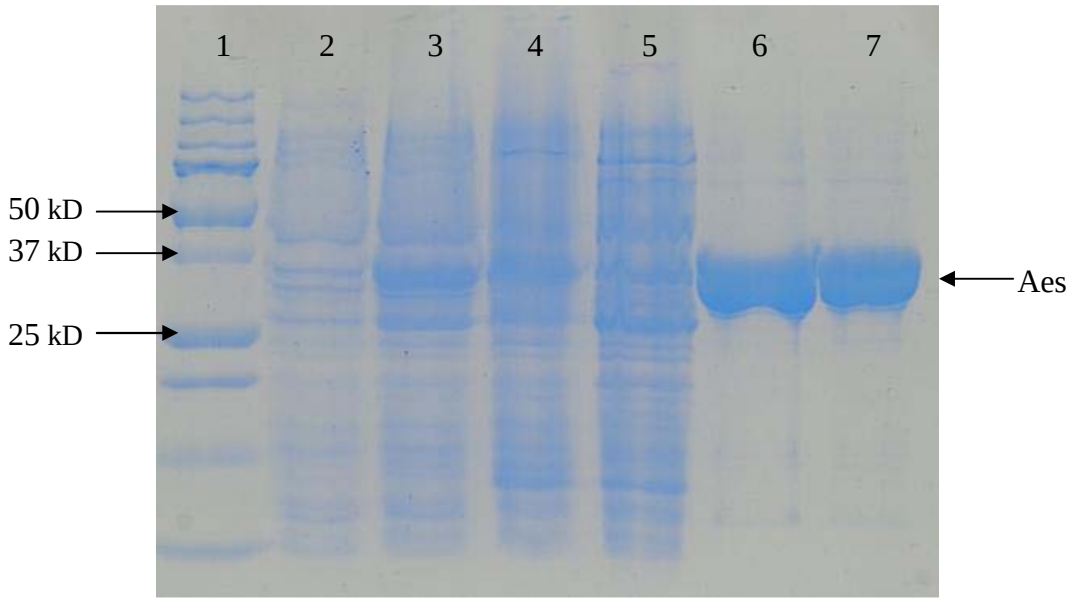



**Figure 2-3. SDS-PAGE gel analysis showing solubility and purity of Aes.** 15% SDS-PAGE gel stained with Coomassie blue. Lane 1: size marker (Precision Plus Protein™ Standards, Bio-Rad), lane 2: crude extract from un-induced culture, lane 3: crude extract from induced culture, lane 4: sample from insoluble phase after sonication, lane 5: sample from soluble phase after sonication, lane 6: sample purified by Ni (II) chelate affinity column, and lane 7: sample purified with desalting column.

According to the chromatogram from Ni chelate affinity chromatography, Aes tightly bound to the column, being eluted with 500 mM imidazole. The other bacterial proteins were eluted as either flow through or with 100 mM imidazole (Figure 2-4). Therefore, Aes obtained from the Ni affinity column was mostly pure as shown on a 15% SDS-PAGE gel (Figure 2-5). The buffer was then exchanged with a desalting column so as to lower the salt concentration. The chromatogram from the desalting column chromatography showed Aes was eluted faster than the salt since the size of the protein is larger (Figure 2-6). There were no other bands observed on SDS-PAGE gel after purifications as shown in Figures 2-5 and 2-7.

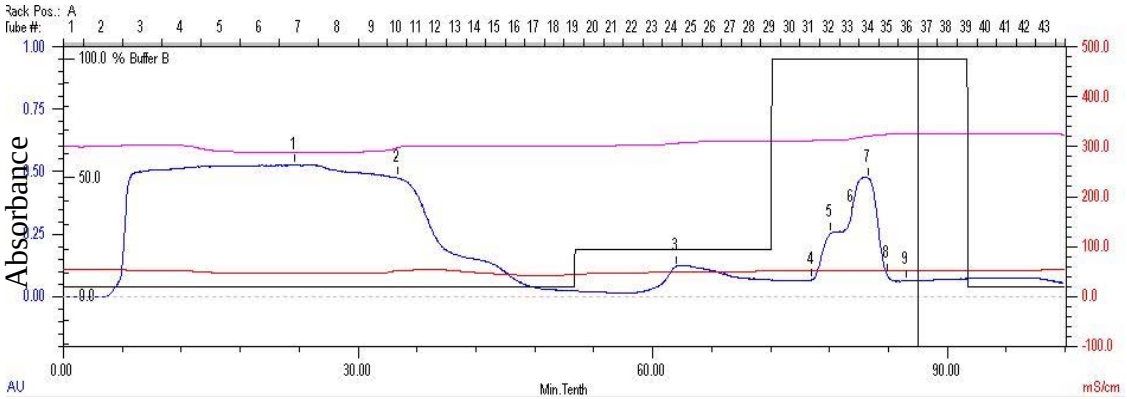



**Figure 2-4. Chromatogram generated by Bio-Rad DuoFlow as a result of Ni chelate affinity column purification of Aes.** The black line shows the percentage of 500 mM imidazole used to elute the proteins, pink is pH of the solution, red is the conductivity and blue is the absorbance. Impurities are mostly eluted at the beginning as flow through, indicated by the elevation of the blue line from test tube #3 to #13. Aes is eluted at 500 mM imidazole into test tube #33 to #35. Samples labeled as #1 to #9 were checked with SDS-PAGE.

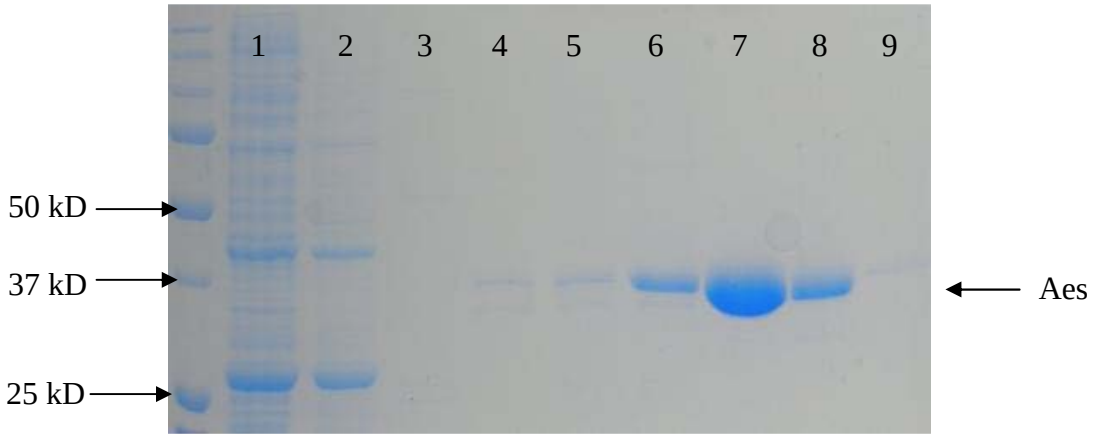



**Figure 2-5. SDS-PAGE analysis of Aes purified by Ni chelate affinity column chromatography.** 10% SDS-PAGE gel stained with Coomassie blue. First lane from the left contains size marker and the numbers for each lane correspond to the sample number on the chromatogram. Aes was eluted with 500 mM imidazole and the amount of impurities was negligible. Samples # 6, 7 and 8 were pooled and used for further purification by a desalting column.

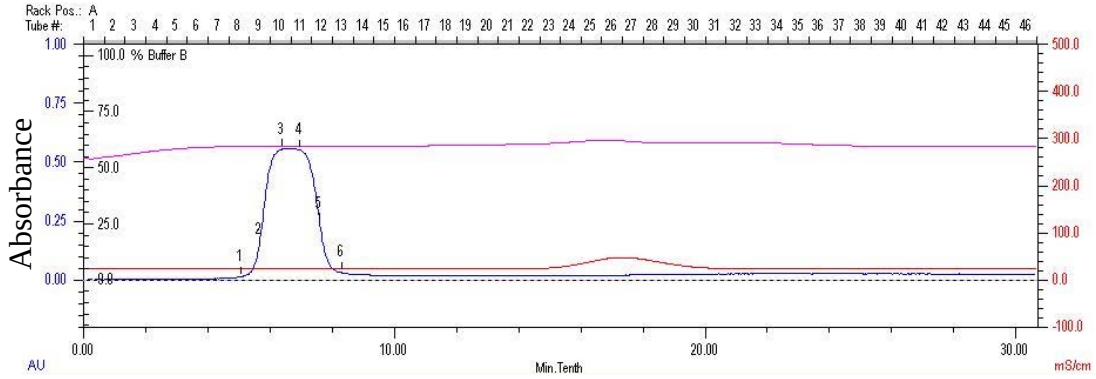



**Figure 2-6. Chromatogram generated by Bio-Rad DuoFlow as a result of desalting column purification of Aes.** Aes is eluted into test tubes #9 to #12. Samples labeled as #1 to #6 were checked with SDS-PAGE.

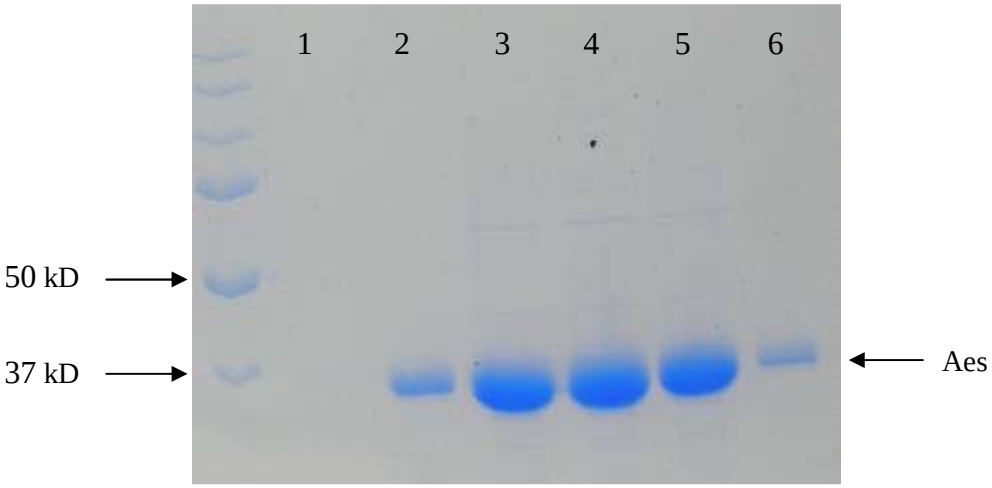



**Figure 2-7. SDS-PAGE analysis of Aes purified by desalting column chromatography.** 10% SDS-PAGE gel stained with Coomassie blue. First lane from the left contains size marker and the numbers correspond to the sample number on the chromatogram. Sample #2 to 5 were pooled and used in crystallization.

According to the MALDI-TOF analysis (matrix-assisted laser desorption/ionization time-of-flight mass spectrometer), purified protein was confirmed as carboxylesterase (*ybaC*, synonym of Aes) from *E. coli* K-12 strain (Figure 2-8) using molecular weight search (MOWSE) and Mascot as scoring and search engines [77-79]. About 15 mg of Aes (the concentration of the final pooled sample was about 1.5 mg/mL), determined by Bradford Assay, was obtained from 400 mL bacterial culture after purification and Aes was concentrated to 5 mg/mL in buffer B to be used in crystallization. Moreover, according to the enzyme assay, purified Aes was active.

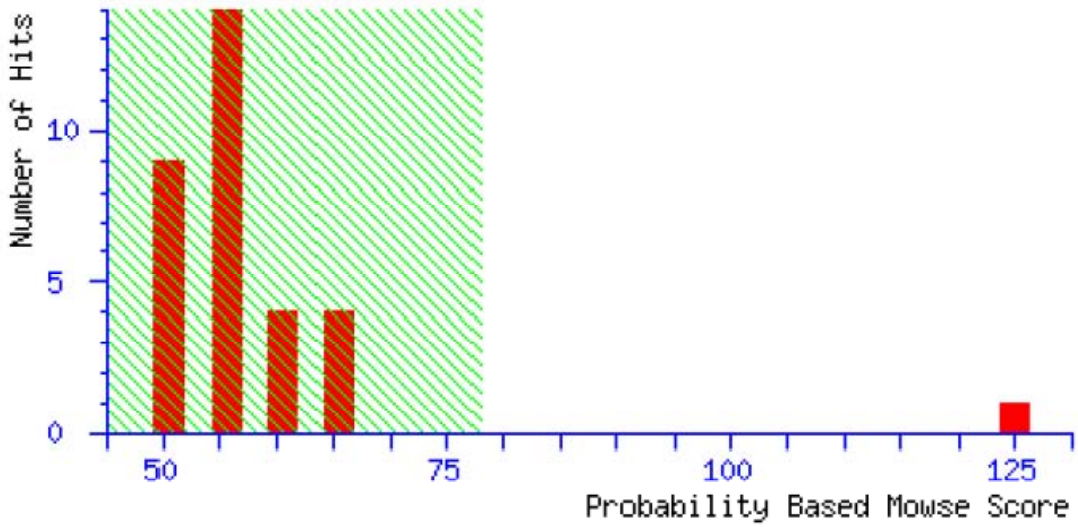



**Figure 2-8. MALDI-TOF result of band after purification.** Proteins whose probability based MOWSE score is greater than 78 are considered as significant and the top score was 125, confirming it as probably carboxylesterase, *ybaC* (gene synonym of *aes*), from *E. coli* (strain K-12).

### 2.3.3 Crystallization

Crystal screening was performed by the hanging drop vapor diffusion method to investigate crystallization conditions of Aes using Crystal Screen kits from Hampton Research and the result is summarized in Table 2-5.

| **Table 2-5. Crystallization conditions that led to the formation of crystalline material or small crystals for Aes.** | | |
|---|---|---|
| Protein Name | Preliminary Crystallization Condition | Appearance |
| Aes | Screen I #7 (0.1 M sodium cacodylate trihydrate pH 6.5 and 1.4 M sodium acetate trihydrate) | Crystals |
| | Screen I #20 (0.2 M ammonium sulfate, 0.1 M sodium acetate trihydrate pH 4.6, and 25% w/v PEG 4000) | Small plate |
| | Screen II # 31 (0.1 M HEPES pH 7.5 and 20% v/v Jeffamine M600) | Square crystal |
| | Index #44 (0.1 M HEPES pH7.5 and 25% w/v PEG 3350) | Small rod |

Since crystals obtained with Crystal Screen condition #7 (0.1 M sodium cacodylate trihydrate pH 6.5, 1.4 M sodium acetate trihydrate) were single, well-shaped, and larger than others, suggesting better ordered crystals, this condition was chosen for optimization so as to obtain better crystals of Aes for data collection (Figure 2-9). In order to achieve this, the concentration of Aes, temperature, pH range and the amount of reagents were varied. As a result of these trials, crystals with better quality, which were able to be used for X-ray crystallography, were produced and reproduced as shown in Figure 2-10. According to trials related to temperature for nucleation and crystal growth, rapid nucleation occurred and crystals grew faster at low temperature. Aes crystals appeared in about 5 days and grew to their final size (usually 0.45 x 0.35 x 0.15 $mm^3$) in a few weeks. Typical crystals of Aes formed in a few weeks are shown in Figure 2-10.

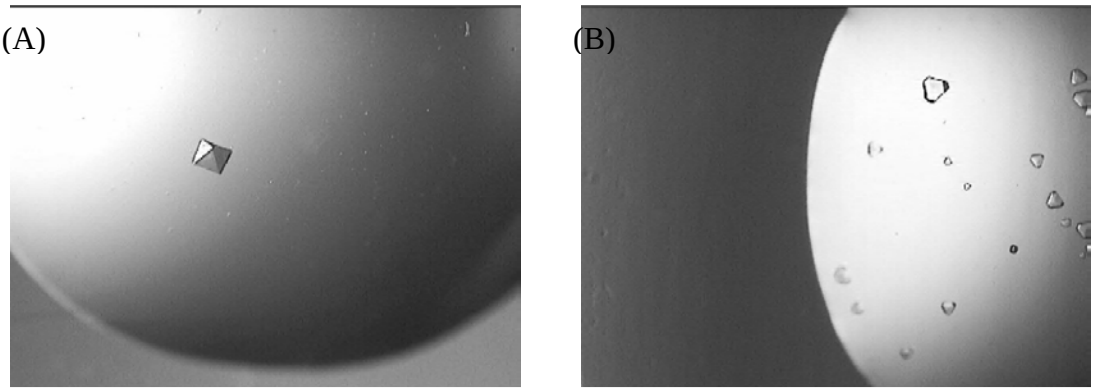



**Figure 2-9. The result of crystal screening for Aes.** (A) Crystal of Aes (5 mg/mL) formed in Crystal Screen (Hampton Research) condition #7 (0.1 M sodium cacodylate trihydrate pH 6.5 and 1.4 M sodium acetate trihydrate) at RT. (B) Crystals of Aes (5 mg/mL) formed in the same reagent at 4 °C. It took less time for nucleation and more crystals formed at 4 °C than at RT.

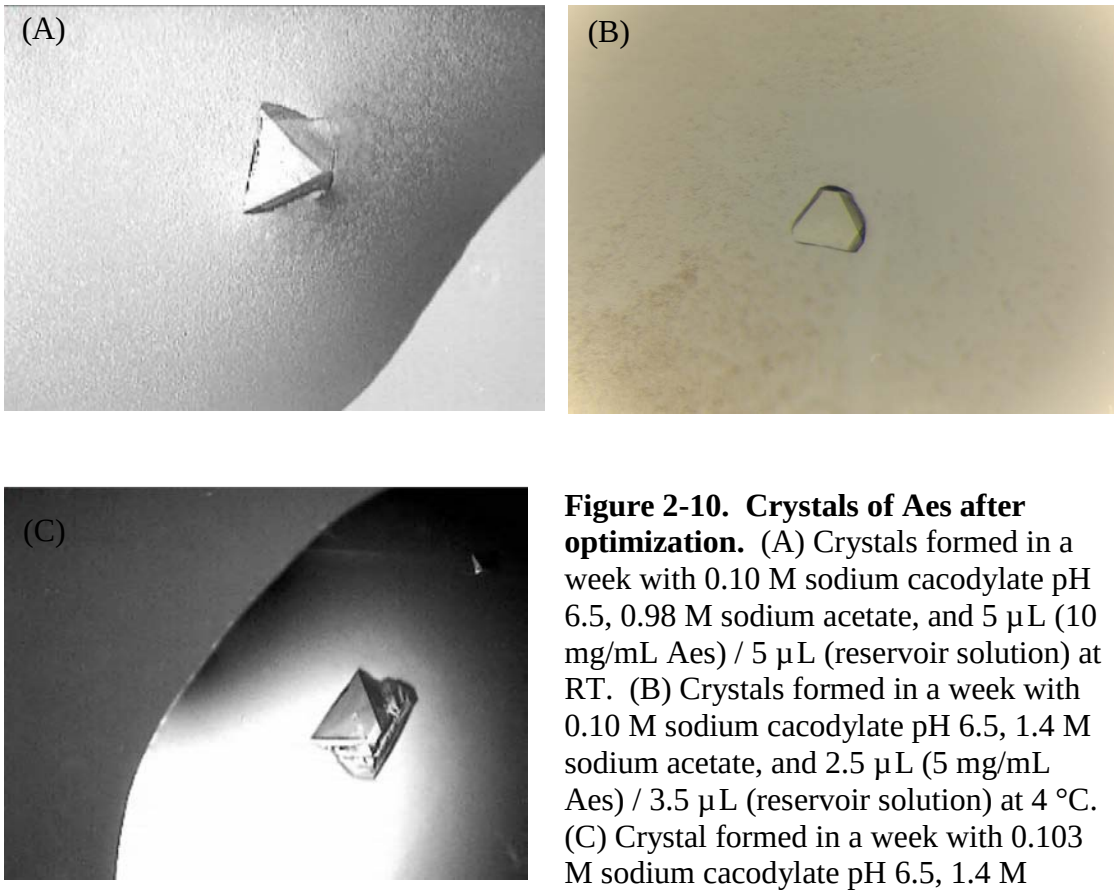



**Figure 2-10. Crystals of Aes after optimization.** (A) Crystals formed in a week with 0.10 M sodium cacodylate pH 6.5, 0.98 M sodium acetate, and 5 µL (10 mg/mL Aes) / 5 µL (reservoir solution) at RT. (B) Crystals formed in a week with 0.10 M sodium cacodylate pH 6.5, 1.4 M sodium acetate, and 2.5 µL (5 mg/mL Aes) / 3.5 µL (reservoir solution) at 4 °C. (C) Crystal formed in a week with 0.103 M sodium cacodylate pH 6.5, 1.4 M sodium acetate, and 3 µL (10 mg/mL Aes) / 3 µL (reservoir solution) at 4 °C. Crystals seemed to be twinned or a cluster.

The number of crystals obtained at RT was less than that of crystals at 4 °C but were better shaped. However, crystals formed at RT did not diffract with high resolution. The concentration of Aes used in crystallization was also varied from 1.5 mg/mL to 10 mg/mL to investigate the effect of concentration on nucleation and crystal growth. As the concentration increases, both nucleation and precipitation occur rapidly especially at 4 °C. As shown in Figure 2-10, many clusters rather than single crystals were observed and crystals also seemed to be cracked when 10 mg/mL of Aes was used. Trials suggested that protein concentration and temperature had a large effect on nucleation and crystal growth of Aes with these reagents. Crystals with different shapes from the same reagents were obtained; the shapes of the crystals were triangular, octahedral, or rod but all the crystals were colorless (Figure 2-10). Crystals of Aes were stable and could bear complete data collection.

### 2.3.4 X-Ray Crystallography

A couple of the many crystals examined diffracted moderately and three of them diffracted to a resolution of 2.7 – 2.8 Å that was considered decent for further data collections. Data were collected from three crystals that consisted of the full-length Aes protein with an N-terminal tag containing six histidines in a row: $His_6$-Aes1, $His_6$-Aes2, and $His_6$-Aes3. X-ray data collection statistics are summarized in Table 2-6. According to the data analysis, space groups were determined to be rhombohedral (*R*3 or *R*32) and tetragonal ($P4_1$). The space group of $His_6$-Aes1 was determined to be *R*3 with two molecules in the asymmetric unit, that of $His_6$-Aes2 was $P4_1$ with six molecules in the asymmetric unit (Table 2-6), and that of $His_6$-Aes3 was *R*32 with one molecule per

asymmetric unit (data not shown due to uncertainty). For $His_6$-Aes2, the Matthew's coefficient suggested a range from 6 to 8 as being optimal and molecular replacement using acetyl esterase from *Salmonella typhimurium* (PDB code: 3ga7, 70% identity) in PHENIX [56-57] found only 6 molecules.

**Table 2-6. Statistics of X-ray data collections for $His_6$-Aes.**

| | $His_6$-Aes1 | $His_6$-Aes2 |
|---|---|---|
| Protein concentration (mg/mL) | 10 | 10 |
| Crystallization condition | 0.1 M sodium cacodylate pH 6.8, 1.4 M sodium acetate, 5 µL/3 µL, at 4 °C | 0.1 M sodium cacodylate pH 6.6, 1.4 M sodium acetate, 5 µL/3 µL, at 4 °C for three months and left at RT for a day |
| Resolution limit (Å) | 2.7[1] | 2.8[2] |
| Wavelength (Å) | 1.5418 | 1.5418 |
| Unit cell parameters (Å, degree) | a=113.7, b=113.7, c=151.8, α=β=90, γ=120 | a=113.6, b=113.6, c=284.7, α=β=γ=90 |
| Space group | *R*3 | $P4_1$ |
| Reflections | 180082 | 335684 |
| Unique reflections | 20063 | 83938 |
| Completion (%) | 99.6 | 96.8 |
| Redundancy | 8.9 | 3.9 |
| $R_{int}$ | 0.1240 (0.4395)[1] | 0.1446 (0.3034)[2] |
| I/σI | 10.93 (2.07)[1] | 7.75 (1.68)[2] |
| Number of molecules in asymmetric unit | 2 | 6 |

[1] Value in parentheses was for highest resolution bin 2.8-2.7.
[2] Value in parentheses was for highest resolution bin 2.9-2.8.

At the beginning, molecular replacements were performed to determine the initial phases by using the homologous proteins 1JJI (29% identity), 1EVQ (26%), 1QZ3 (26%) and 2C7B (27%) that were available from the PDB, but the trials were not successful. However, 3GA7, the Aes from *Salmonella typhimurium*, had been recently determined and with the percent identity to Aes being 70%, was successful in providing the initial

phases of Aes. The data from $His_6$-Aes2, which crystallized in the tetragonal space group $P4_1$, led to the structure determination of Aes. Moreover, the initial phases of *R*3 were found but the refinement did not go well and stuck with high R-factors probably because of twining. Therefore, the model was built using the data from $His_6$-Aes2.

After the molecular replacement solution was found, some of the amino acids for the search molecule were converted to correspond to the ones in Aes from *Escherichia coli* since the initial model was based on the homologous protein used in molecular replacement. Afterward, the side chain of 32 residues (Leu-7, Asp-11, Leu-12, Lys-18, Leu-24, Gln-25, Thr-45, Leu-46, Glu-57, Gln-71, Glu-73, Asp-82, Gln-115, Ile-138, His-146, Gln-148, Glu-150,Asp-151, Gln-153, Ile-154, Met-156, Arg-158, Asp-184, Lys-187, Arg-201, Val-212, Trp-213, Gln-218, Gln-219, Glu-226, Glu-250, Lys-302, Glu-306, Arg-309, and Gln-313) in chain A was initially truncated into alanine due to poor density around the side chain of these residues. Moreover, residues of a loop (Pro-26, Asp-27, Leu-28, Pro-29, Pro-30, Trp-31, Pro-32, Ala-33, Thr-34, and Gly-35) in chain A were removed due to the poor electron density since residues 26-35 in 3GA7 were missing.

Almost all of the residues were built in five rounds of refinement except for the loop region of the N-terminal cap, residues 26 to 35, which showed poor density. In order to build this region, each amino acid was manually built with COOT to fit to the electron density map, obeying proper geometry. After the manual refinement, all six molecules were refined with PHENIX with simulated annealing, TLS and modified NCS

that excluded the loop region mentioned above and amino acids 217-219 and 247-250 that showed relatively weak electron density among the molecules in the asymmetric unit and slightly different orientation for the amino acids. Molprobity was performed to validate the structure of Aes (Table 2-7). According to the comprehensive validation, there were still a few residues that were out of the preferred region in the Ramachandran plot (see Table 2-8) and unique rotamers for each chain (see Table 2-9).

| **Table 2-7. Model validation by Molprobity in PHENIX.** | | |
|---|---|---|
| Geometry outliers | | 1 residue |
| Ramachandran | Favored | 94.7% |
| | Outliers | 0.7% |
| Rotamer outliers | | 5.9% |
| C-β outliers | | 0 |
| Clashscore | | 24.68% |

| **Table 2-8. Ramachandran outliers for each chain.** The number is the residue number of Aes. | |
|---|---|
| Chain A | 30, 32 |
| Chain B | 26 |
| Chain C | 32 |
| Chain D | 34, 35 |
| Chain E | 29, 33, 35, 36 |
| Chain F | 30, 31, 36 |

| **Table 2-9. Unique rotamers for each chain.** The number is the residue number of Aes. | |
|---|---|
| Chain A | 17, 23,42, 45, 60, 110, 217, 232, 265, 269, 298, 299, 304, 316, 318 |
| Chain B | 17, 37, 39, 42, 46, 60, 68, 110, 218, 219, 232, 248, 249, 265, 269, 298, 299, 303, 316, 318 |
| Chain C | 17, 23, 24, 39, 42, 46, 60, 68, 110, 232, 265, 269, 298, 299, 303, 316, 318 |
| Chain D | 17, 23, 24, 28, 36, 39, 42, 46, 60, 68, 110, 232, 265, 269, 298, 299, 303, 316, 318 |
| Chain E | 17, 23, 24, 28, 31, 34, 36, 39, 42, 46, 60, 64, 68, 110, 232, 265, 269, 298, 299, 303, 316, 318 |
| Chain F | 17, 23, 24, 28, 34, 39, 42, 46, 60, 64, 68, 110, 232, 265, 298, 299, 303, 316, 318 |

The crystal structure of Aes is shown in Figure 2-11 and the model statistics are in Table 2-10. X-ray crystallography revealed that Aes contained a typical α/β hydrolase fold with the central β-strands surrounded by α-helices. There are eight β-strands (β1: Thr-60 to Val-65, β2: Val-72 to Phe-77, β3: Thr-86 to Leu-90, β4: Thr-117 to Ile-121, β5: Arg-158 to Asp-164, β6: Gly-190 to Trp-194, β7: Cys-254 to Ala-259, β8: Cys-282 to Tyr-287) and thirteen α-helices (α1: Val-9 to Leu-12, α2: Ala-15 to Asn-22, α3: Ile-37 to Asn-52, α4: Leu-100 to Ser-114, α5: Gln-133 to Tyr-152, α6: Ala-166 to Lys-181, α7: Val-204 to Leu-208, α8: Gln-218 to Tyr-228, α9: Asp-232 to Glu-236, α10: Leu-242 to Asp-244, α11: Leu-265 to His-278, α12: Phe-294 to Tyr-297, α13: Lys-302 to Gln-318). A portion (residues 1-4 and 26-35) of the N-terminal region of Aes from *Salmonella typhimurium*, which was used to determine the initial phases, was not built. However, in this study, residues 26-35 of the Aes from *E. coli* were built even though residue 1 and 2 were unable to be built. This new finding should help in better understanding the structure and function of homologous proteins. The current model consists of six molecules in the asymmetric unit with 1902 amino acid residues. The residuals are $R_{work}$ = 19.79% and $R_{free}$ = 26.46% (Table 2-10). The N-terminal residues were disordered and not modeled, including the $His_6$ tag (36 amino acids including six histidines) and some N-terminal residues of Aes (residues 1 and 2). However, the C-terminal residues were ordered enough to be modeled. The N-terminus of Aes was highly flexible, suggesting they act as a lid which has been reported for lipases (Figure 2-11) [63].

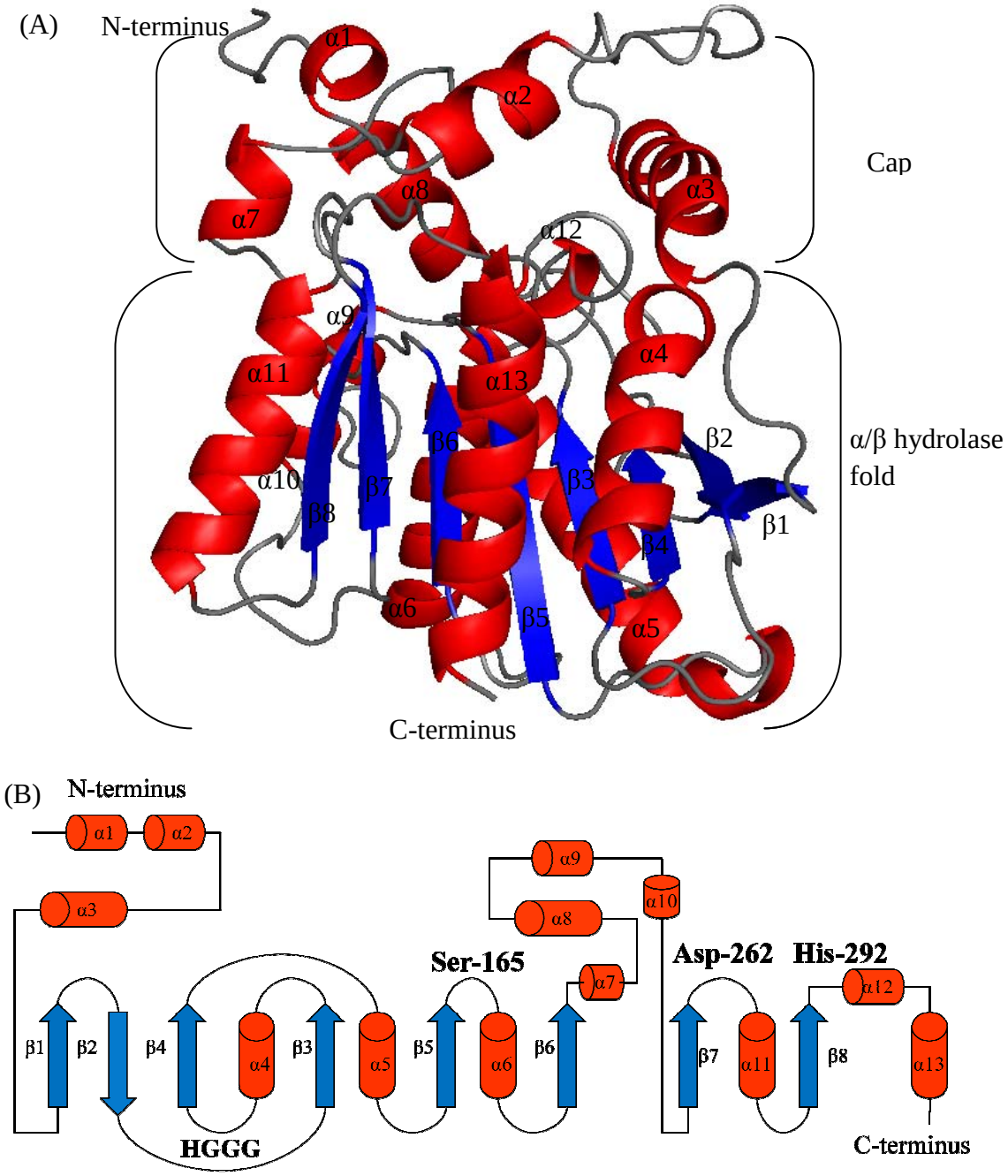



**Figure 2-11. X-ray crystal structure of Aes generated by PyMOL and topology diagram of the Aes structure.** (A) A view of Aes molecule generated by PyMOL [80]. Eight β-strands (blue) are surrounded by α-helices (red), showing the α/β hydrolase fold. (B) Helices are red, β-strands are blue, and the connecting regions (random coil and loops) are black. The second β-strand is anti-parallel to the others, leading to the formation of the α/β hydrolase fold. HGGG represents the histidine (H) - glycine (G) motif of Aes and Ser-165, Asp-262 and His-292 are the residues of the catalytic triad.

| **Table 2-10. Statistics for the model of $His_6$-Aes2.** | |
|---|---|
| Resolution | 2.8 |
| $R_{work}$ (%) | 19.79 |
| $R_{free}$ (%) | 26.46 |
| r.m.s bond deviation (Å) | 0.008 |
| r.m.s angle deviation (°) | 1.072 |
| Number of amino acid residues | 1902 |
| Number of water molecules | 0 |
| B-factor for all atoms (Å) | 60.7 |
| | |
| $R_{work} = \Sigma\|F_{obs} - F_{calc}\|/\Sigma\|F_{obs}\|$, where summation is over data and $R_{free}$ contains only 1.1% of data excluded from all refinement. | |

According to the three dimensional structure of Aes, a catalytic triad is present consisting of Ser-165, Asp-262 and His-292 (Figure 2-11). The catalytic triad is located near the C-terminus in three different loop regions between an α-helix and a β-strand (Figure 2-11). More specifically, Ser-165 is located between β5 and α6, Asp-262 is located between β7 and α11, and His-292 is located between β8 and α12. His-292 interacts with Ser-165 and Asp-262 with a distance of 3.2 Å and 2.9 – 3.3 Å respectively (see Figures 2-12 and 2-14). Another motif of Aes, HGGG (histidine-glycine motif), is found between α4 and β4 (see Figures 2-12 and 2-14). The N-terminus is composed of a long loop consisting of three relatively short α-helices (α1, α2, and α3) (Figure 2-12). The second β-strand (β2) is anti-parallel to the others, which leads to the formation of the α/β hydrolase fold where the central eight β-strands (β1, β2, β3, β4, β5, β6, β7, and β8) are surrounded by α-helices (α4, α5, α6, α11, α12, and α13). Moreover, there is another loop with α-helices (α7, α8, α9, and α10) between β6 and β7, which may provide the flexibility and selectivity for substrates of Aes. A total of six molecules were found in COOT arranged as three dimers in the asymmetric unit (Figure 2-12).

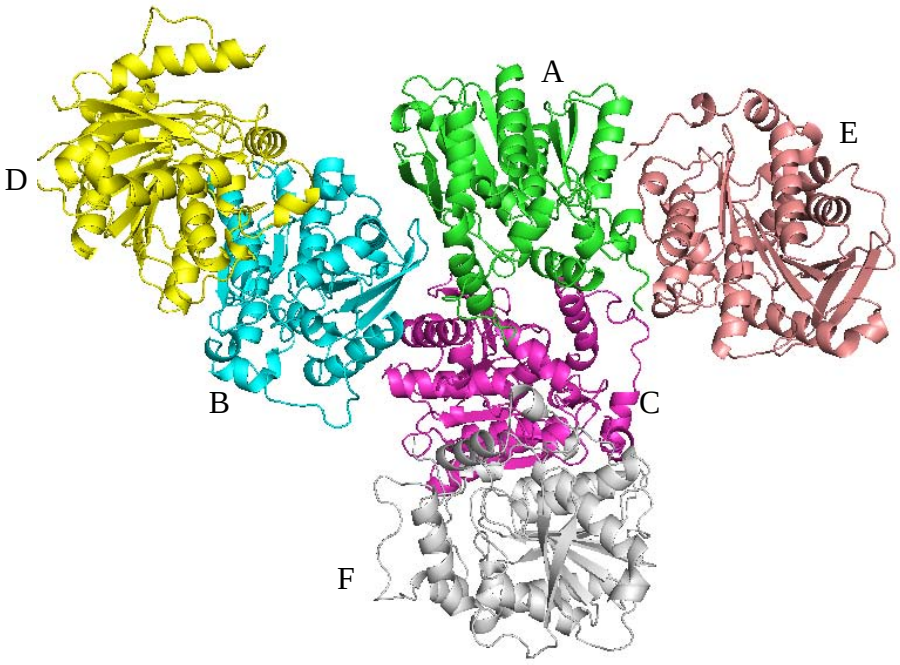



**Figure 2-12. A view of the six molecules in the asymmetric unit.** Three dimers are shown. Figure was produced by PyMOL [80]. Chain A forms a dimer with chain E, chain B forms a dimer with chain D, and chain C forms a dimer with chain F.

A dimer is shown perpendicular to the 2-fold axis that relates the two molecules by a 180˚ rotation (Figure 2-13 (A)). Moreover, according to PDBePISA (Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute) that predicts the quaternary structures of proteins [81], chain A and E (AE), chain B and D (BD), and chain C and F (CF) form stable dimers with the surface area 24850 $Å^2$, 24560 $Å^2$, and 25000 $Å^2$ and buried area 2140 $Å^2$, 2240 $Å^2$, and 2100 $Å^2$ respectively (Table 2-11). The buried area of BD is slightly larger than the others, showing tight contact than the other dimers (AE and CF). According to structural studies on homologous proteins used in MSA, they all contain the α/β hydrolase fold but the crystal structures of 3GA7, 1EVQ,

1QZ3, and 3FAK are monomers, 2C7B and 1JKM contain a dimer, and 1JJI and 3DNM contain two dimers in the asymmetric unit (Figure 2-13).

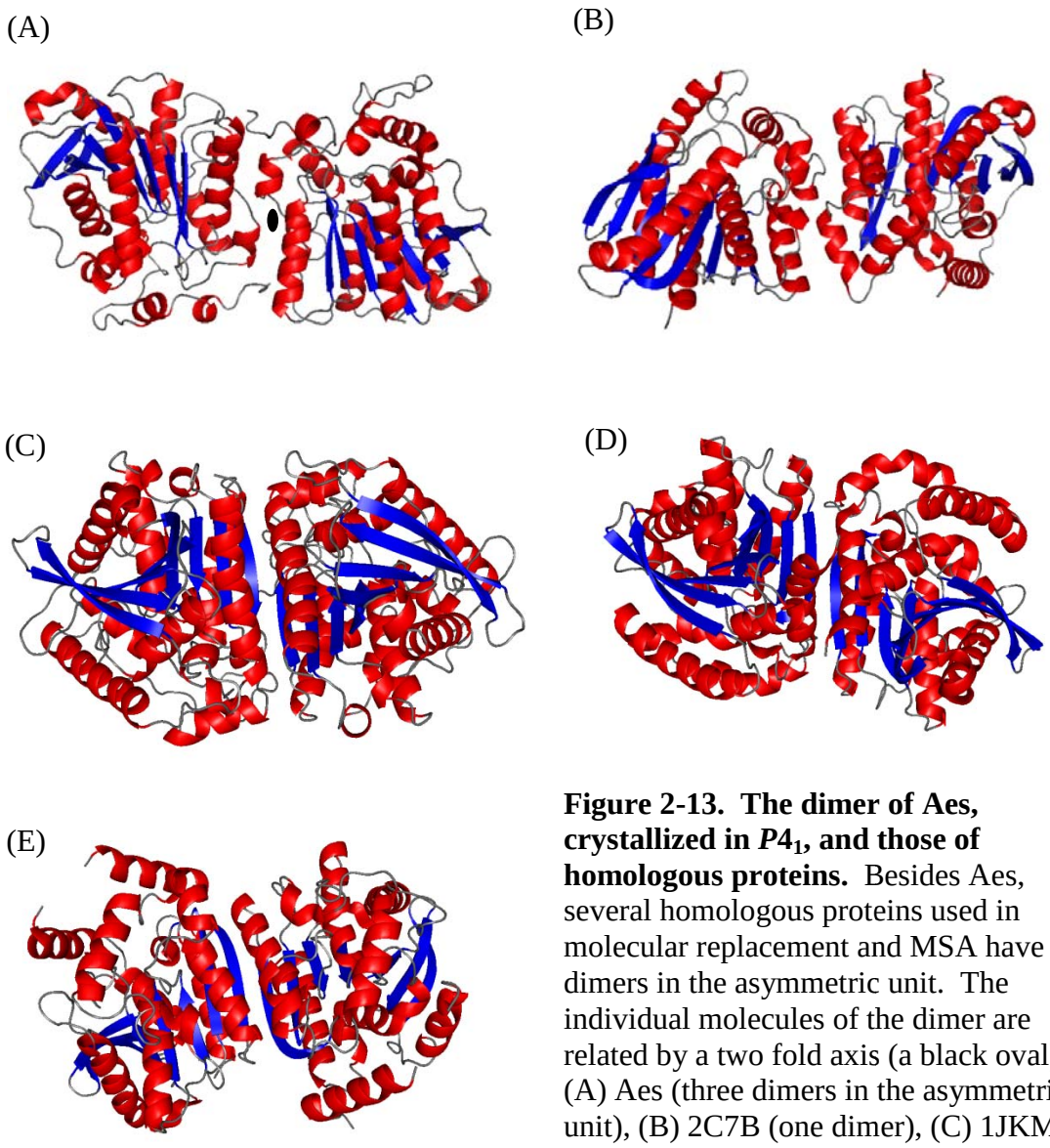



**Figure 2-13. The dimer of Aes, crystallized in $P4_1$, and those of homologous proteins.** Besides Aes, several homologous proteins used in molecular replacement and MSA have dimers in the asymmetric unit. The individual molecules of the dimer are related by a two fold axis (a black oval). (A) Aes (three dimers in the asymmetric unit), (B) 2C7B (one dimer), (C) 1JKM (one dimer), (D) 1JJI (two dimers), and (E) 3DNM (two dimers). They have an α/β hydrolase fold.

**Table 2-11. Analysis of protein interfaces.** Chain A (A) forms a dimer with chain E (E), chain B (B) forms a dimer with chain D (D), and chain C (C) forms a dimer with chain F (F), which are consistent to the finding from the crystal structure of Aes in the asymmetric unit (Figure 2-12). According to the PISA analysis, these dimers are stable in solution.

| Composition | Surface area, $Å^2$. | Buried area, $Å^2$. |
|---|---|---|
| AE | 24850 | 2140 |
| BD | 24560 | 2240 |
| CF | 25000 | 2100 |

Since Aes was crystallized at pH 6.6, Asp (pKa=3.65) is deprotonated and Ser should be protonated and a hydrogen atom on His-292 binds to the oxygen atom on Asp-262 and a nitrogen atom ($\varepsilon^2$N) of His-292 binds to the hydrogen atom on Ser by hydrogen bonding (see Figures 2-14 and 2-15). X-ray crystallography also revealed that the catalytic triad of Aes is stabilized by weak hydrogen bonds formed between Ser-165 and the backbone residues of the HGGG motif (see Figures 2-14 and 2-15). In addition to these residues, there are several His and Asp residues around the active site of Aes. They probably are involved in hydrogen bonding to stabilize the active site (Figure 2-15). Moreover, there are several positive densities around the active site but identifications of them are still undetermined due to low resolution (Figure 2-14).

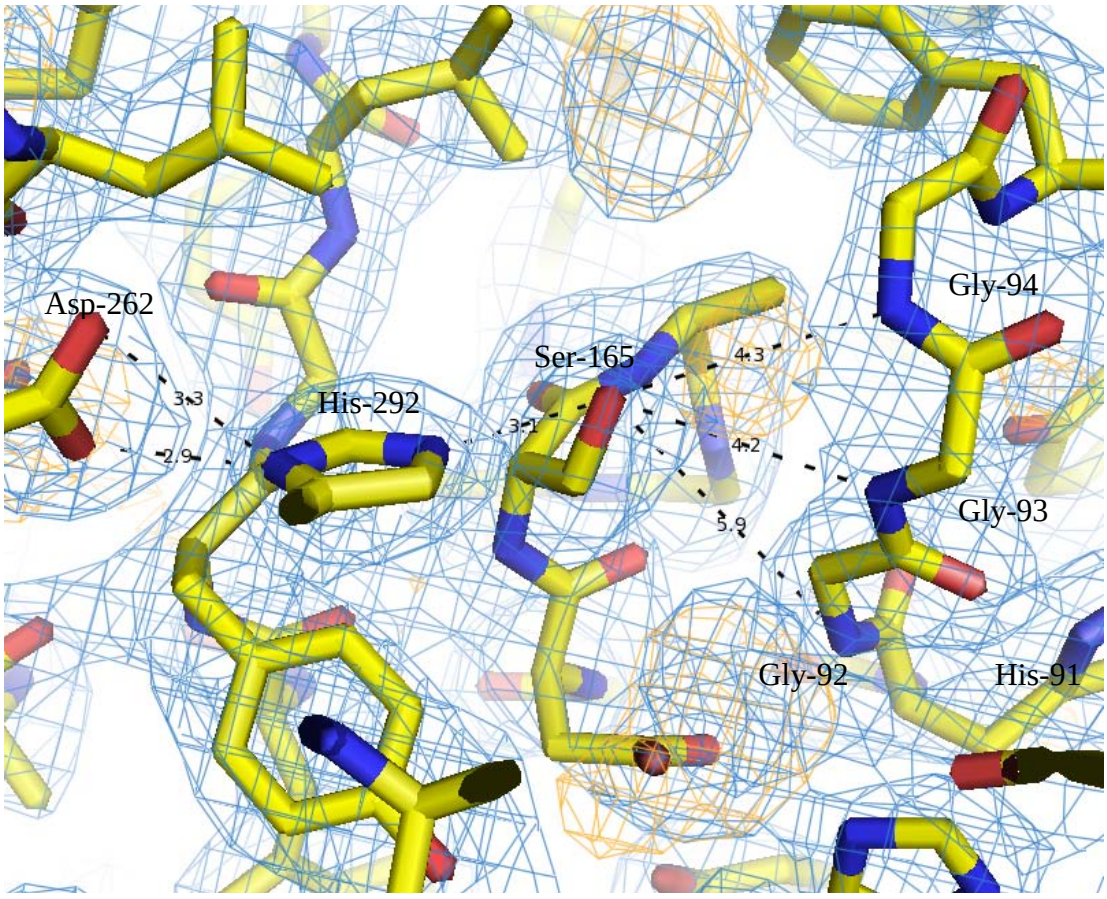



**Figure 2-14. The view around the catalytic triad and HGGG motif with electron density map.** The protein coordinate is shown as yellow sticks, electron density map (rendered 1σ) is shown as blue grids, and positive density is shown as orange grids. His-292 binds to Asp-262 and Ser-165 by hydrogen bonding (shown in dashed line) and Ser-165 is stabilized by hydrogen bonds that forms between the oxygen atom (γO) of Ser-165 and the nitrogen atom (blue of sticks) of glycines from HGGG, the hydrogen atom of Ser-165 and the nitrogen atom ($\varepsilon^2$N) of histidine, and the hydrogen atom of histidine and the oxygen atom of aspartic acid.

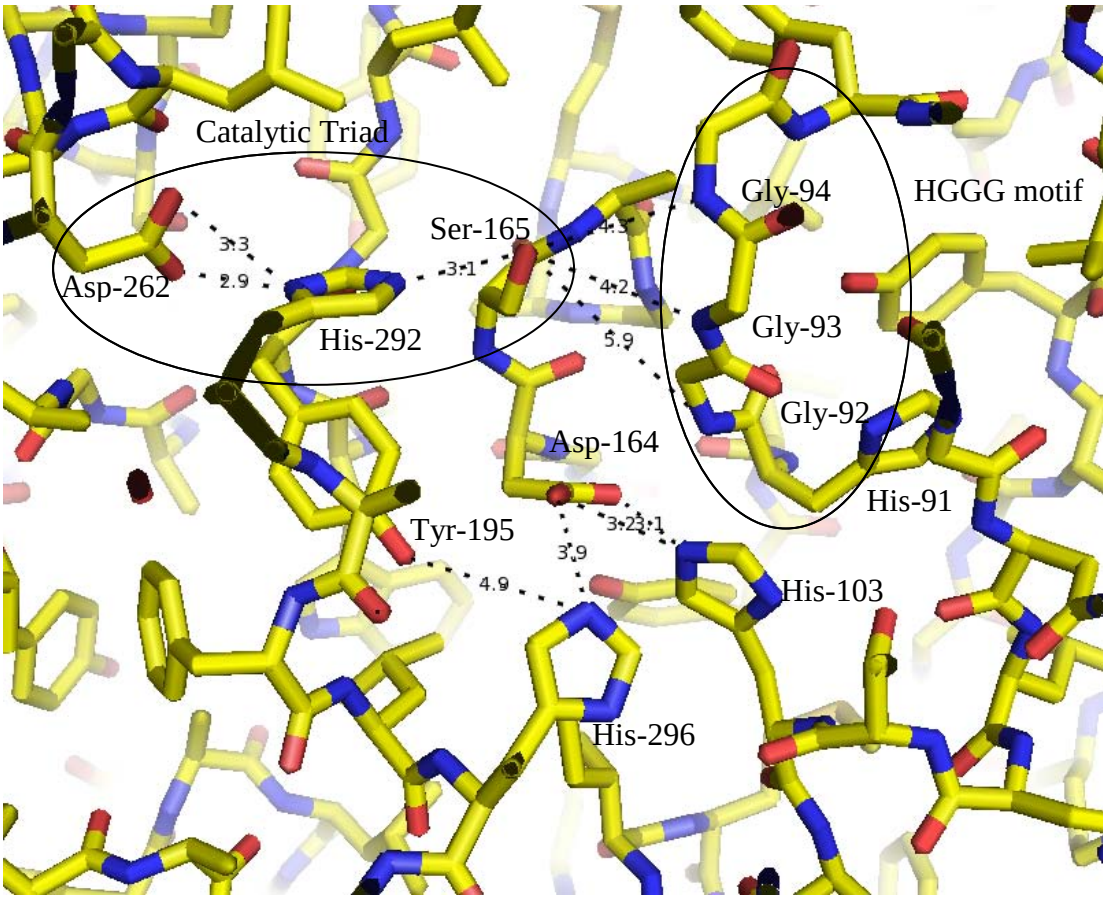



**Figure 2-15. The view around the active site.** The protein coordinate is shown as yellow sticks. There are several residues that are involved in the stabilization of the active site through hydrogen bonds (dashed line). Asp-164 is stabilized by His-103 and His-296 through relatively strong hydrogen bonds and also His-296 is stabilized by Tyr-195 through a weak hydrogen bond. Because of these hydrogen bonds, the catalytic triad is stabilized.

Moreover, according to the multiple sequence alignment (MSA) performed by ClustalW [76], the HGGG motif (residue 91, 92, 93, and 94 in Aes) and residues of the catalytic triad (Ser-165 and His-292) are both highly conserved among bacterial homologous proteins and these are also conserved in HSL from human and rat [30, 33]. Even though Glu is found instead of Asp in some homologous proteins such as hormone-sensitive lipase like Este5 from a metagenome library (PDB ID: 3FAK) and hormone-sensitive lipase from a metagenome library (PDB ID: 3DNM), which are from uncultured

bacteria, the nucleophile such as Asp or Glu should be necessary for such hydrolases to function. Moreover, there are some highly conserved residues; His-103, Asp-122, Pro-132, and His-252. Asp-122, Pro-132, and His-252 are found in the loop region, while His-103 is found on an α-helix. According to the structural studies using PyMOL, His-103 stabilizes Gly-92 (4.1 Å) and Asp-122 is also involved in the stabilization of His-91 (4.1 Å) by hydrogen bonds.

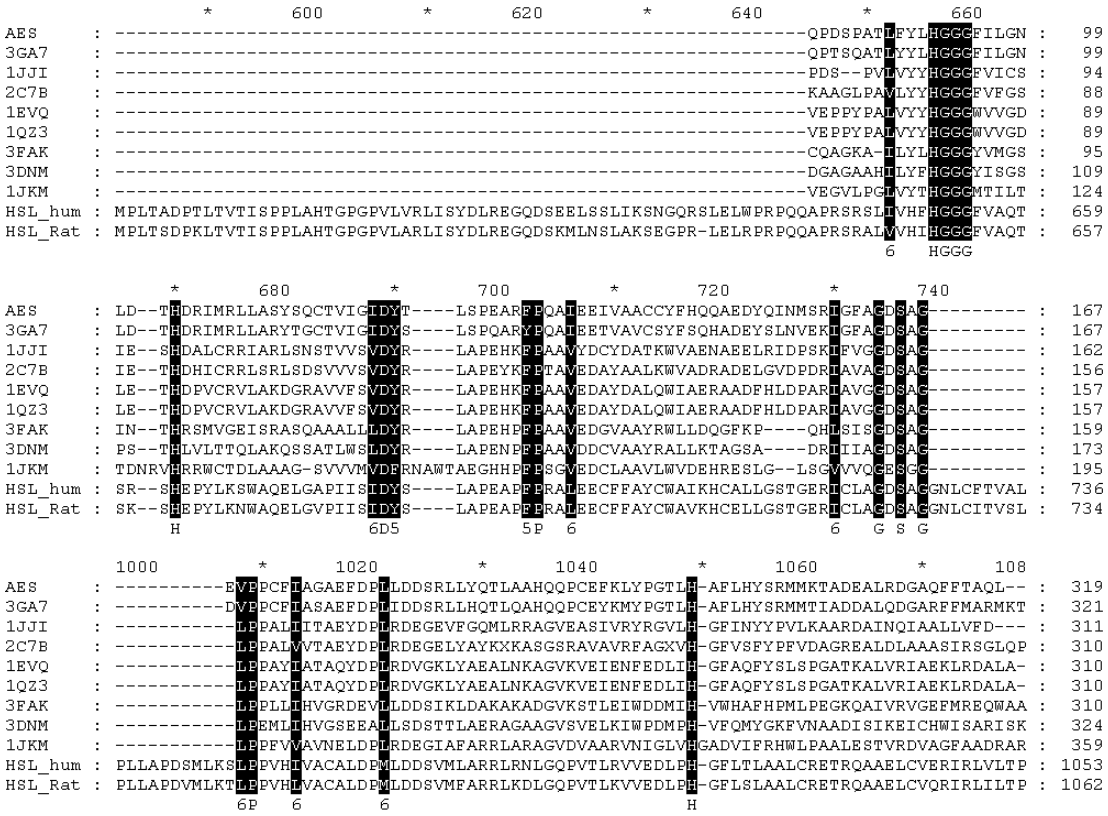


**Figure 2-16. Multiple sequence alignment of Aes to homologous proteins.** 3GA7 is an acetyl esterase from *Salmonella typhimurium*, 1JJI is a hyper-thermophilic carboxylesterase from *Archaeoglobus fulgidus*, 2C7B is a new thermophilic and thermostable carboxylesterase cloned from a metagenomic library, 1EVQ is a thermophilic carboxylesterase Est2 from *Alicyclobacillus acidocaldarius*, 1QZ3 is a mutant M211S/R215L of a carboxylesterase Est2 complexed with hexadecanesulfonate, 3FAK is a hormone-sensitive lipase like Este5 from a metagenomic library, 3DNM is a hormone-sensitive lipase from a metagenomic library, and 1JKM is a brefeldin A esterase, a bacterial homologue of human hormone-sensitive lipase. Some homologous proteins are used in molecular replacement to find the search model in molecular replacement. The name

corresponds to the name used in PDB except for HSL_hum and HSL_Rat. HSL_hum refers to HSL from *Homo sapiens* and HSL_Rat refers to HSL from *Rattus norvegicus*.

The catalytic triad and HGGG motif are found in a cleft, of the molecule surrounded by hydrophobic residues, where a flexible loop of the N-terminus covers them by hydrophobic interactions. However, it seems that there is partial accessibility to the triad (Figure 2-17). Therefore, the N-terminal region does not appear to act as a perfect lid as reported for other homologous proteins belonging to the H-group [63]. If a crystal structure is determined with a ligand, it is possible that the loop could move and cover the active site.

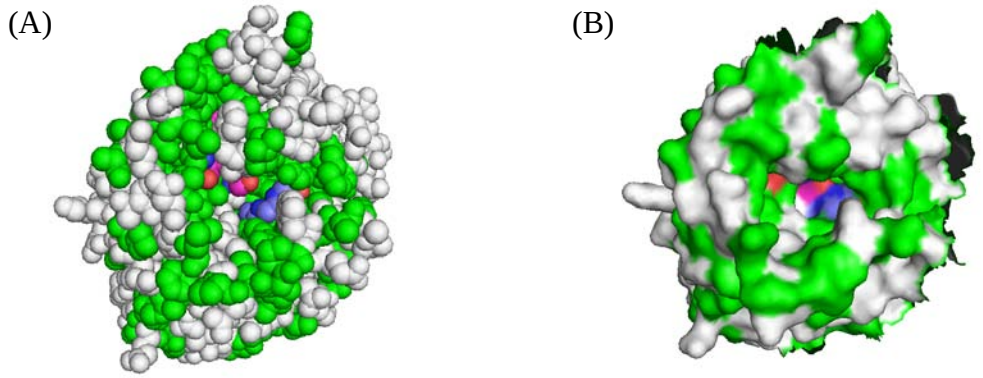



**Figure 2-17. Hydrophobicity plot of Aes.** Hydrophobic residues are green, the catalytic triad is magenta, HGGG motif is blue, and others are white. Sphere representation (A) and surface representation (B) of Aes viewed from the N-terminus. The catalytic triad and HGGG are hidden in the shallow cleft and there is a partial accessibility from the surface.

Based on X-ray crystallography, chemical properties around the active site of Aes and the mechanism of serine protease and α/β hydrolase [10], a plausible mechanism can

be proposed as shown in Figure 2-18. Since the actual substrate for Aes in *E. coli* is still unknown, p-nitrophenyl butyrate is used to explain the proposed mechanism. (1) The hydroxyl group (OH group) of Ser-165 attacks the carbonyl carbon of the substrate. The lone pair on the nitrogen atom ($\varepsilon^2$N) of His accepts the hydrogen (H) from the OH group and a pair of electrons from the double bond of the carbonyl oxygen moves to the oxygen atom (O), forming an enzyme-substrate intermediate (2). (2) The electrons between C-O of the substrate attack H of His, leading to the bond formation of O-H on the substrate. The electrons on $O^-$ move back to recreate the C=O and His is regenerated, releasing p-nitrophenol (3). Now, water comes in, attacking the carbonyl carbon, and electrons from the C=O double bond move back to O. Afterward, the bond between O of water and C of the substrate is formed and $\varepsilon^2$N of His accepts H from water (4). (5) The electrons forming the bond between C-O from Ser attack H on His to reform the OH group on Ser and then the double bond is reformed between C=O. Finally, the carboxylic acid is released and the catalytic triad goes back to the initial state (6).

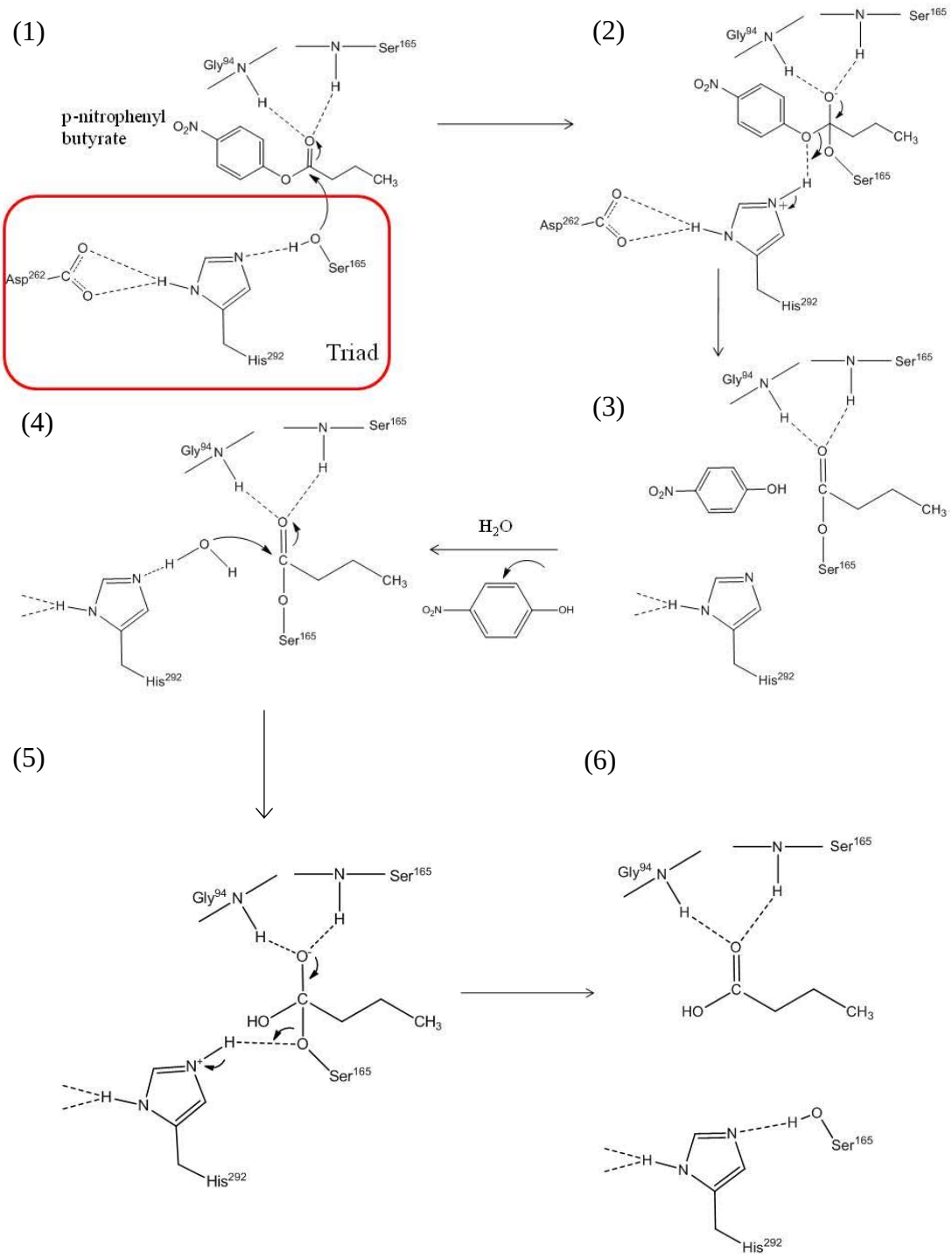



**Figure 2-18. Proposed mechanism of Aes.**

As mentioned before, only 3GA7 could work as a successful search model in molecular replacement to determine the initial phases. To understand the reasons why the other search models did not work, Aes was superimposed with the homologous proteins used in the molecular replacements with the program PyMOL (Figure 2-19). Compared to the crystal structures for homologous proteins, one of the reasons why the other structures failed in molecular replacement was that Aes has a long loop around the N-terminus, which was found to be completely different to the other proteins. The other reason is probably because α11 and β8 in Aes have slightly different conformations compared to the others. Since a flexible loop in the N-terminus of 3GA7 was not built and also the identity of 3GA7 was high to Aes, it worked as a good model in molecular replacement. The overall structures of all the homologous proteins are similar, showing the α/β hydrolase fold; however, there are slight shifts between Aes and the other homologous proteins except for 3GA7. Moreover, since some structures of homologous proteins have a ligand, the catalytic triad proposed in this study is consistent with their findings. Therefore, also in Aes, the cleft mentioned should be the binding pocket for the substrate in *E. coli* and the substrate should be relatively small due to the size of the shallow cleft.

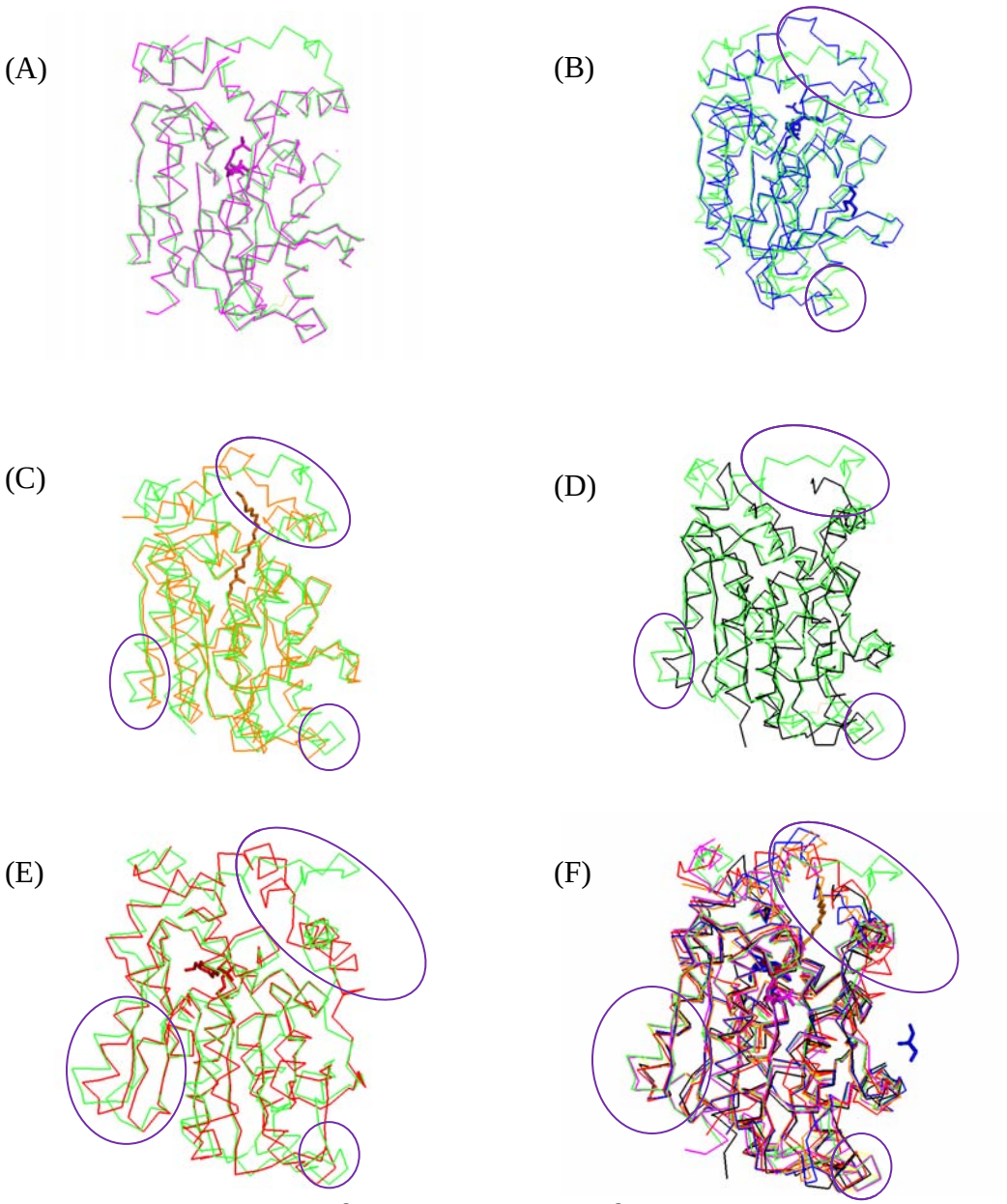



**Figure 2-19. Comparison of crystal structures of homologous proteins to Aes.** Cα RMSD value for 3GA7 was determined by PyMoL with syntax PyMOL> align chain ChA_aes (pdb file of chain A of Aes) and name ca, 3GA7 (pdb file of 3GA7) and name ca. The same procedure was performed for each homologous protein in order to obtain Cα RMSD values. (A) Aes (green) is aligned with 3GA7 (magenta), Cα RMSD = 0.364 Å. (B) Aes is aligned with 1EVQ (blue), Cα RMSD = 1.375 Å. (C) Aes is aligned with 1QZ3 (orange), Cα RMSD = 1.351. (D) Aes is aligned with 2C7B (black), Cα RMSD = 1.043 Å. (E) Aes is aligned with 1JJI (red), Cα RMSD = 1.061 Å. (F) 3D alignment of Aes (green) with homologous proteins: 3GA7 (magenta), 1EVQ (blue), 1QZ3 (orange), 2C7B (black), and 1JJI (red). The ligands are shown as sticks with corresponding color to the protein. The ligand of 1EVQ is 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid and 2-amino-2-hydroxymethyl-

propane-1, 3-diol, that of 1QZ3 is 1-hexadecanosufonic acid, and that of 1JJI is 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid. The phosphoserine, which is also shown as sticks, is found in 3GA7. The greatest differences are circled (purple).

## 2.4 Summary

X-ray crystallography revealed that Aes contained an α/β hydrolase fold, the central β-strands being surrounded by α-helices. Moreover, according to the structural analyses and literature review, the catalytic triad of Aes consists of Ser-165, Asp-262 and His-292, Ser-165 being stabilized by His-292 through hydrogen bonding. His-292 is also stabilized by Asp-262 by hydrogen bonding. Aes has a shallow cleft consisting of hydrophobic residues, which leading to the substrate binding pocket and the catalytic triad. Aes also has a long loop with short α-helices around the N-terminus, suggesting that it acts as a flexible lid that protects a substrate from the environment during the catalytic activity.

As future work, crystallization with a substrate should be performed in order to investigate whether there is a conformational change of the structure of Aes with the presence of the substrate. Moreover, the structure determination of Aes using the data of $His_6$-Aes1 (*R*3) should be carried out to see whether there are any structural differences compared to the structure from $His_6$-Aes2 ($P4_1$). The co-crystallization with MalT is also challenging but worth investigating in order to know how Aes and MalT interact to better understand the remarkable maltose system in *E. coli*.

# CHAPTER 3

# MALT

## 3.1 Introduction

The maltose system is a member of the periplasmic binding protein-dependent ATP-Biding Cassette (ABC) high affinity transport systems and the maltose regulon consists of five operons: *malPQ, malS, malEFG, malK-lamB-malM,* and *malZ* [2, 6, 16]. The *malPQ, malS*, and *malZ* encode proteins that play roles in maltose and maltodextrin metabolism while the *malEFG* and *malK* encode proteins involved in the uptake of maltose and maltodextrin (see section 1.3) [14, 16]. The expression of these genes are controlled by MalT and since the expression of *malEFG* and *malK-lamB-malM* is under catabolite repression, these *mal* genes are controlled by CAP [2, 16].

MalT from *E. coli* is a 103 kDa protein and consists of 901 amino acids. It is a transcriptional activator of the maltose regulon, showing a weak ATPase activity [82]. The activity of MalT is regulated by many different regulatory signals. MalT is activated by ATP and maltotriose [2, 16, 82] while it is down regulated by three proteins, MalK, MalY, and Aes [14]. There are three regulators for the expression of MalT: CAP, Mlc acting as a glucose inducible repressor, and H-NS (nucleoide-associated protein), that adjust the expression of MalT [83-87].

MalK is the ATP biding cassette (ABC) protein that transports maltose and maltodextrin and is stimulated by the periplasmic maltose binding protein (MBP) [88]. MalK consists of two domains, the nucleotide binding domain (NBD) and the residue regulatory domain (RD). NBD shows an α/β motif and binds to ATP while RD displays β folds and binds to MalT [88]. MalK sequesters MalT in the inactive form and it releases MalT by hydrolyzing ATP when substrates are transported [14]. MalK interacts with MalT directly without maltotriose and inhibits binding of maltotriose to MalT [2, 89].

The other regulator MalY belongs to the gene cluster of *malIXY* and is regulated by MalI, in other words, MalY is not controlled by MalT. The gene cluster, *malIXY*, is conserved among some bacteria but MalY from other bacteria does not function as a down-regulator in *E. coli* [87]. MalY acts as a cystathionase but this enzyme activity is not necessary for the down regulation of MalT [14]. MalY and Aes (see Chapter 2) act in a similar manner by stabilizing the inactive form of MalT, the monomeric conformation. Crystal structures of MalK and MalY have been determined [87-88] and the crystal structure of Aes has been determined in this study (see Chapter 2) (see Figures 2-11 and 3-1). Interestingly, parts of MalY and MalK show a similar fold to Aes: β-strands surrounded by α-helices.

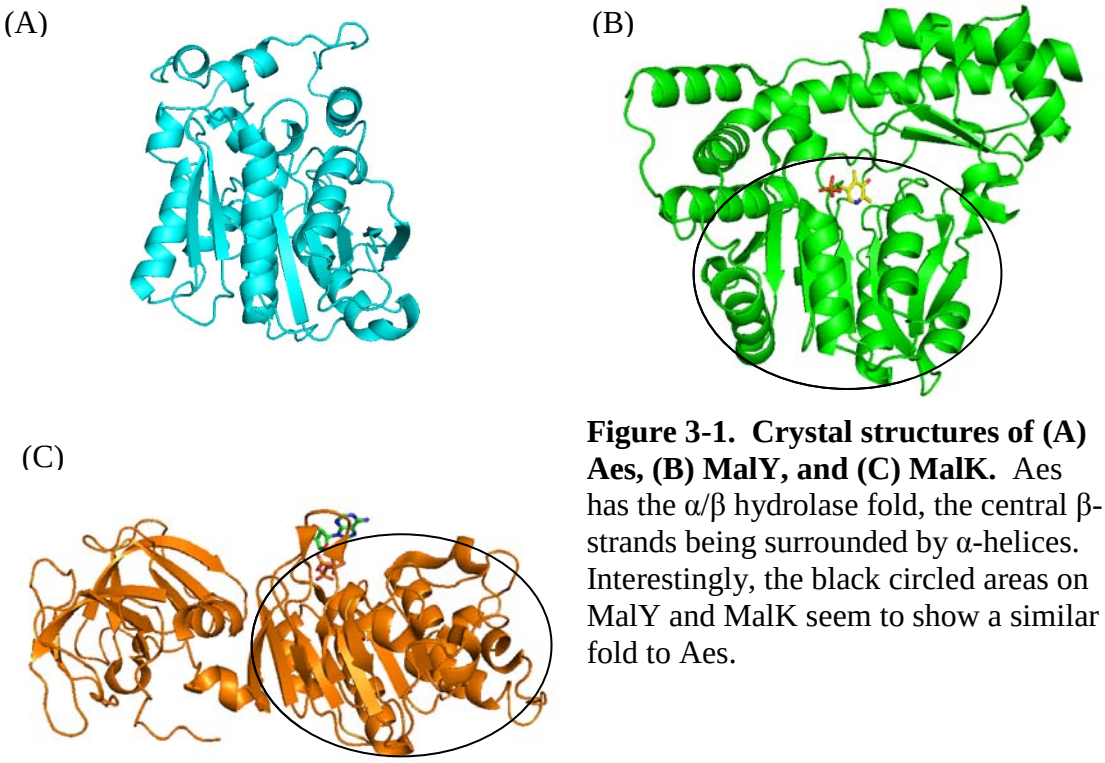



**Figure 3-1. Crystal structures of (A) Aes, (B) MalY, and (C) MalK.** Aes has the α/β hydrolase fold, the central β-strands being surrounded by α-helices. Interestingly, the black circled areas on MalY and MalK seem to show a similar fold to Aes.

MalT is a relatively large protein consisting of four domains, domain I (DT1: residues 1-241), domain II (DT2: residues 242-436), domain III (DT3: residues 437-806) and domain IV (DT4: residues 807-901) [22, 90-91]. Its function is complicated as it can interact with several ligands to be either activated or repressed [82]. According to studies on MalT, DT1 is a binding site for ATP which is one of the inducers for MalT. ATP or ADP, besides maltotriose, is necessary for MalT to be active (to be multimer). It is suggested that ATPase activity of MalT, even though it is slow, plays a role in the competition between the positive and negative effectors [15, 22]. However, the hydrolysis of ATP is not required for the activation of open complex formation (the unwinding of DNA with RNA polymerase and transcription factors to initiate the

transcription) by MalT [15]. The binding site of MalY is DT1 and that of Aes includes both DT1 and DT2 [14-15]. DT3, whose structure has been determined [92], is the binding site for maltotriose, the second inducer, and DT4 is the DNA binding site [14, 22]. Since a recent study suggested Aes bound to DT1 or both DT1 and DT2, structural studies for DT1, the first two domains from the N-terminus (DT1-DT2), and the whole MalT protein have also been conducted in this research.

### 3.2 Materials and Methods

Basically the same methods described in section 2.2 were used for PCR, plasmid construction, over-expression, purification, and crystallization with a few changes as described below. PCR was performed in order to amplify DT1, DT1-DT2, and *malT* with primers listed in Table 3-1. The stop codon was added on each reverse primer shown as bold in Table 3-1. Using these primers, genes of domain I (DT1), and the first two domains from the N-terminus (DT1-DT2) were amplified by PCR with the same program as used in the PCR amplification of the *aes* gene (see section 2.2.2). Since amplification did not work for the *malT* gene with this program, probably because *malT* is a much larger gene compared to the others, a longer extension step was needed to obtain the whole gene. Thus, different PCR cycles were tried (Table 3-2) and the reaction set up for each gene is listed in Tables 3-3, 3-4, and 3-5.

| **Table 3-1. Forward and reverse primers for DT1, DT1-DT2, and *malT*.** DT1 consists just of domain I of MalT, DT1-DT2 consists of domains I and II, and *malT* consists of the whole *malT* gene from *E. coli*. | |
|---|---|
| *malT* forward | ***CAC***CATGCTGATTCCGTCAAAACTAAG |
| *malT* reverse | **TTA**CACGCCGTACCCCAT |
| DT1 forward | Same as *malT* forward |
| DT1 reverse | **CTA**GCGTGCCGACTTATGG |
| DT1-DT2 forward | Same as *malT* forward |
| DT1-DT2 reverse | **CTA**TTCAGCACGGGCTAGCA |

| **Table 3-2. Program used in minicycler in order to amplify *malT* by PCR.** | | | |
|---|---|---|---|
| | 1st Cycle | 2nd - 30th Cycles | Final Cycle |
| Denaturation 94 ˚C | 6 min. | 45 sec. | |
| Annealing 55 ˚C | 30 sec. | 30 sec. | |
| Extension 72 ˚C | 2 min. | 1 min. | 10 min. |

| **Table 3-3. Reactions that led to the production of DT1.** (Volumes are in µL) | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10X Accu Prime PCR Buffer II | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| *malT* forward (20 µM) | 2 | 3 | 3 | 4 | 4 | 4 | 1 |
| DT1 reverse (20 µM) | 2 | 3 | 3 | 4 | 4 | 4 | 1 |
| DNA template | 1 | 1 | 2 | 1 | 2 | 4 | 1 |
| Accu Prime Taq Polymerase | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Sterile Distilled Water | 39 | 37 | 36 | 35 | 34 | 32 | 41 |
| Total Volume | 50 | 50 | 50 | 50 | 50 | 50 | 50 |

| **Table 3-4. Reactions that led to the production of DT1-DT2.** (Volumes are in µL) | | | | | |
|---|---|---|---|---|---|
| 10X Accu Prime PCR Buffer II | 5 | 5 | 5 | 5 | 5 |
| *malT* forward (20 µM) | 2 | 2 | 3 | 4 | 4 |
| DT1-DT2 reverse (20 µM) | 2 | 2 | 3 | 4 | 4 |
| DNA template | 1 | 4 | 1 | 1 | 2 |
| Accu Prime Taq Polymerase | 1 | 1 | 1 | 1 | 1 |
| Sterile Distilled Water | 39 | 36 | 37 | 35 | 34 |
| Total Volume | 50 | 50 | 50 | 50 | 50 |

| **Table 3-5. Reactions that led to the production of *malT*.** (Volumes are in µL) | | | |
|---|---|---|---|
| 10X Accu Prime PCR Buffer II | 5 | 5 | 5 |
| *malT* forward (20 µM) | 3 | 1 | 1 |
| *malT* reverse (20 µM) | 3 | 1 | 1 |
| DNA template | 1 | 1 | 2 |
| Accu Prime Taq Polymerase | 1 | 1 | 1 |
| Sterile Distilled Water | 37 | 41 | 40 |
| Total Volume | 50 | 50 | 50 |

After PCR, agarose gel electrophoreses were performed followed by gel extraction in order to purify the PCR product from the gel (see section 2.2.2). Purified PCR products were then used in TOPO cloning reactions and transformations, followed by analyses of transformants (20 to 30 transformants were analyzed per gene product due to high background) using restriction enzymes: SacI and AccI for DT1 and PstI instead of AccI for DT1-DT2 and *malT* (see sections 2.2.3 and 2.2.4). Using 1.0% or 0.7% agarose gel, agarose gel electrophoreses were performed to analyze plasmids digested by the restriction enzymes, followed by DNA sequencing. After the confirmation of the genes being inserted in the plasmids with the correct orientation, each gene (DT1, DT1-DT2, or *malT*) was over-expressed in *E. coli* BL21 Star$^{TM}$ (DE3) in the same manner as Aes (see section 2.2.5). Optimal growth conditions were discovered by conducting time course analyses for expression, adding different types of sugars (such as 0.5% maltose and 0.5% arabinose) in the LB media and varying the temperature for culturing (18 ˚C, 28 ˚C, 30 ˚C, 37 ˚C) after induction for each transformant. Moreover, optimization of the solubility of each protein was conducted by adding 0.1 mM, 0.2 mM, 0.5 mM, 1.0 mM, and 1.2 mM ATP into the cellular resuspension prior to the disruption of the bacterial cells. Since the yield of bacterial cells producing MalT was low compared to the others,

the pLysS *E. coli* strain, which reduced the basal expression of MalT, was also tried for the over-expression of MalT in order to slow the expression of this apparent toxic gene.

Bacterial cells were harvested in the same manner as for Aes with pellets made from 400 mL of bacterial culture that were resuspended in 10 mL of Buffer C (50 mM TrisHCl pH 7.8, 500 mM KCl, and 10% sucrose). After the addition of 0.2 mg/mL or 10 mg/mL lysozyme in the resuspension, they were incubated at 27 ˚C for 30 minutes. Afterwards, centrifuge tubes containing the cell resuspension were placed on ice for at least 10 minutes followed by the addition of 0.1 mM - 1.2 mM ATP. Bacterial cells were then disrupted by the sonicator in the same manner as Aes (see section 2.2.6).

Nickel chelate affinity chromatography was performed for DT1, DT1-DT2, and MalT with Buffer C. During purifications, the buffer blender was used in the similar manner as Aes (section 2.2.6.) with the Duo Flow System so as to adjust the pH to 7.8. After Ni affinity chromatography, either a desalting column or a gel filtration column was used to remove the excess amount of imidazole depending on the result of Ni affinity chromatography, more specifically, the desalting column was used if no impurities were observed on SDS-PAGE while the gel filtration column was used if impurities were observed after Ni affinity chromatography. For either case, 0.1 – 1.2 mM ATP was added to the sample prior to the chromatography using Buffer D (50 mM TrisHCl pH 7.8 and 100 mM KCl).

Since DT1 and MalT were successfully purified, crystallization screening was performed using Crystal Screen™ I and II, and Index Screen (Hampton Research) with the hanging drop vapor diffusion method using 24 well plates (see section 2.2.8). The concentration of DT1 and MalT used in crystal screening was 5 mg/mL according to the Bradford assay, and the protein solution was filtered with a 0.20 µm filter. Reagent (700 µL) was used for the reservoir and the rest of the procedure was similar to the crystallization methods used for Aes, such as the exploration of the size of the droplets and temperature (RT and 4 ˚C).

### 3.3 Results and Discussion

The plasmids carrying DT1, DT1-DT2 and the *malT* genes were successfully constructed even though only a few of the about 30 transformants tested carried the proper recombinant plasmid. According to DNA sequencing, a point mutation was found in DT1 at amino acid residue 182 compared to SwissProt P06993, resulting in Gln (codon CAG) to Arg (codon CGG) while no mutations were found in DT1-DT2 and *malT*. Time course analyses were performed to optimize culturing condition for each protein and several results of SDS-PAGE analyses of the optimization are shown in Figure 3-2. The optimal conditions for culturing are summarized in Table 3-6. Moreover, as a result of screening, some crystallization conditions for these proteins were found (see Table 3-7).

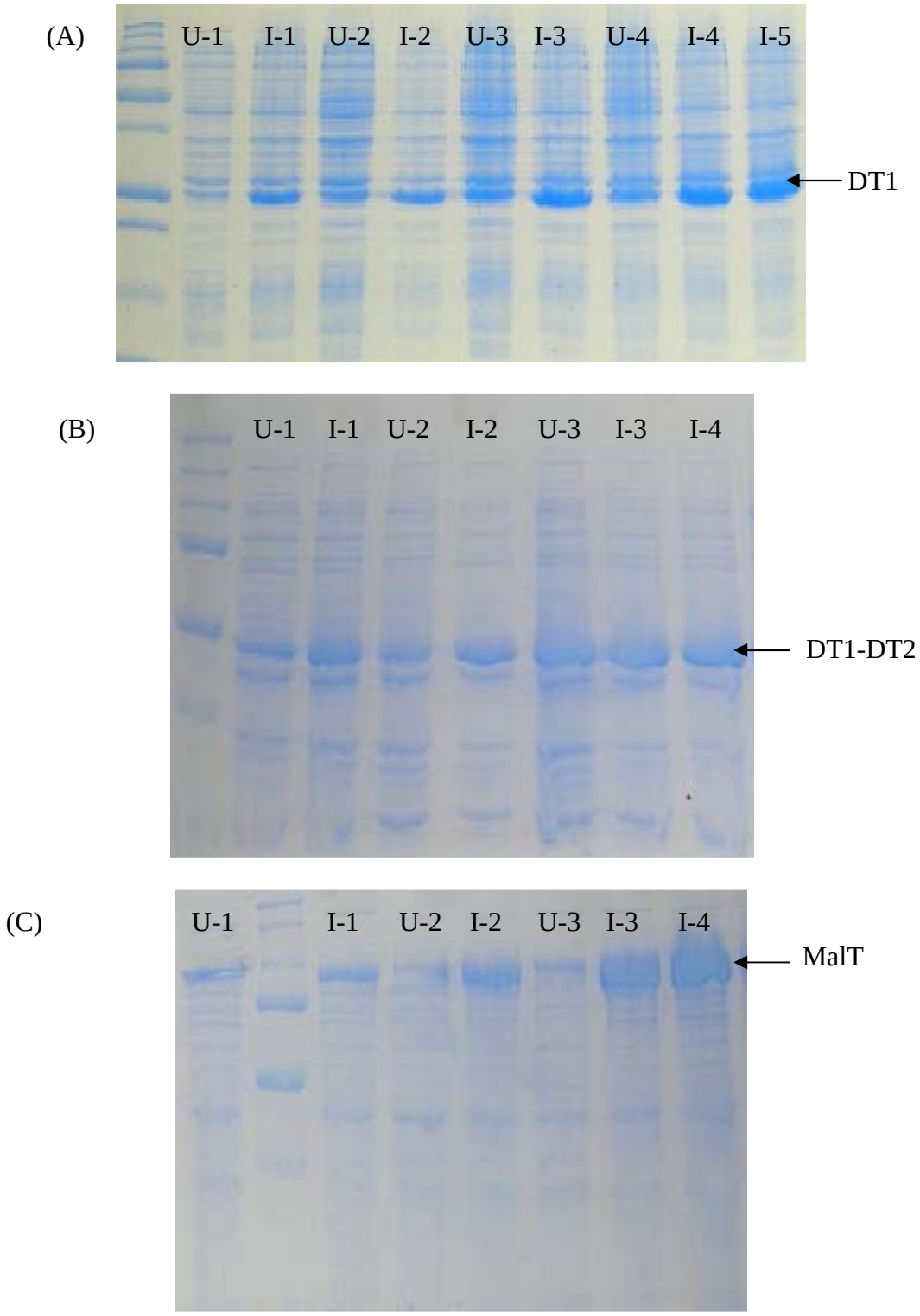



**Figure 3-2. SDS-PAGE analyses of over-expression of DT1, DT1-DT2, and MalT.** An aliquot was obtained from induced and un-induced culture every one hour for four to five hours after the induction. Unlabeled lane is a size marker. (A) Time course study of DT1 at 37 ˚C without any sugars, comparing the induced (I) and the un-induced (U) whole cell. DT1 was cultured for 5 hours (number shows

the hour after the induction) and over-expressed in 5 hours after the induction. (B) Time course analysis of DT1-DT2 at 37 ˚C, comparing the induced and the un-induced whole cell. DT1-DT2 was over-expressed in 4 hours after the induction. (C) Time course analysis of MalT at 37 ˚C, comparing the induced and the un-induced whole cell. MalT was over-expressed in 4 hours after the induction.

DT1 and MalT were successfully purified by chromatography as shown in Figures 3-3, 3-4, 3-5, and 3-6. Since there were impurities observed on a SDS-PAGE gel after Ni affinity column chromatography of DT1, a gel filtration column was used to obtain pure DT1. On the other hand, impurities were negligible on a SDS-PAGE gel after Ni affinity column chromatography of MalT, a desalting column was used to exchange buffer. DT1-DT2 was unable to be purified and the reason might be because six-histidine tag is hidden in the protein molecule due to the disorder and so unable to bind to the Ni affinity column during purification since DT1-DT2 flows through without any imidazole.

**Table 3-6. Optimal conditions for culturing in order to obtain high yield of each protein.**

| Protein Name | Growth condition after the induction by 1.0 mM IPTG at $OD_{600}$ = 0.6. |
|---|---|
| DT1 | 18-20 hours of incubation at 28 ˚C. |
| DT1-DT2 | 5 hours of incubation at 28 ˚C. |
| MalT | 5 hours of incubation at 18 ˚C. |

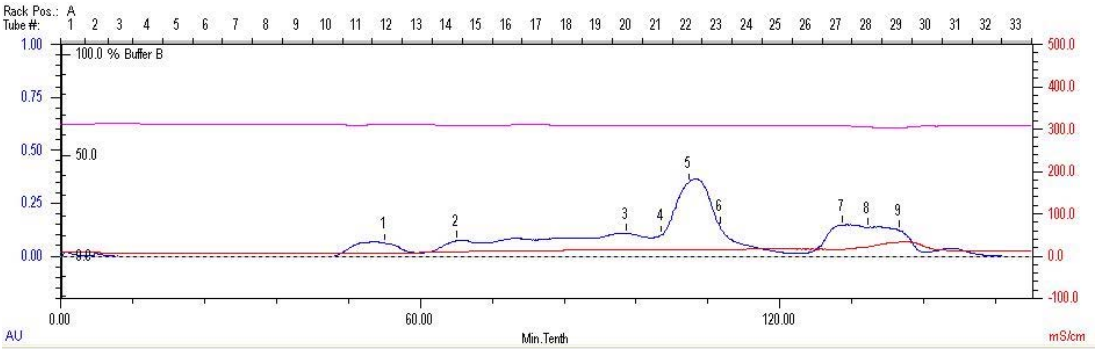



**Figure 3-3. Chromatogram generated by Bio-Rad DuoFlow as a result of gel filtration column purification of DT1.** DT1 was eluted into test tubes # 22 and 23 and samples #1 to 9 were used in SDS-PAGE analysis.

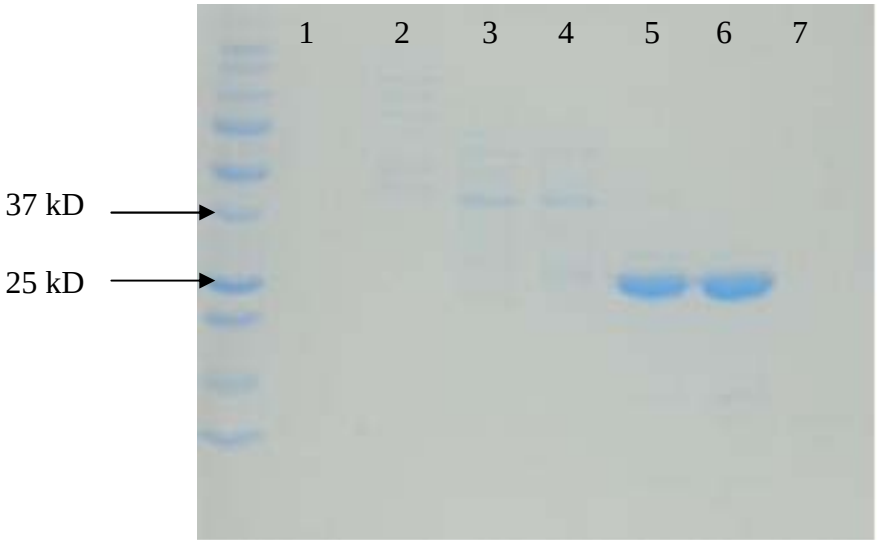



**Figure 3-4. SDS-PAGE analysis of DT1 purified by gel filtration column chromatography after Ni affinity column chromatography.** DT1 was eluted with 500 mM imidazole followed by the gel filtration column chromatography since impurities were observed. First lane from the left contains size marker and the numbers for each lane correspond to the sample number on the chromatogram. Samples # 5 and 6 were used in crystallization.

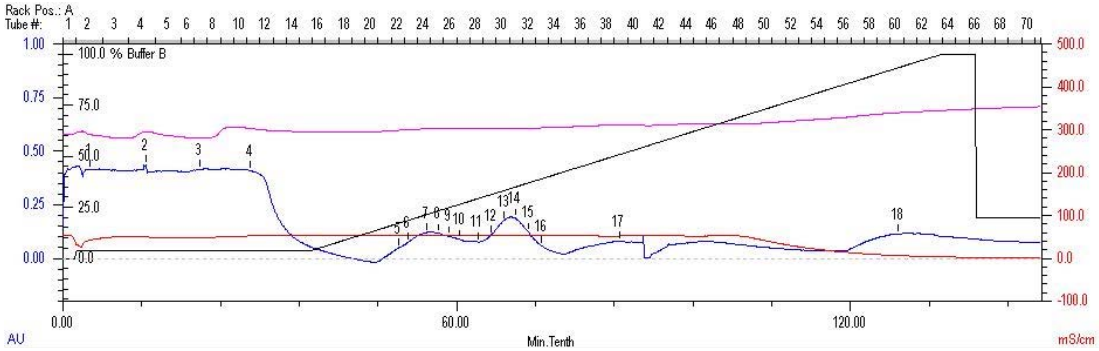



**Figure 3-5. Chromatogram generated by Bio-Rad DuoFlow as a result of Ni affinity column purification of MalT.** Impurities are mostly eluted at the beginning as flow through, indicated by the elevation of the blue line from test tube #1 to #13. MalT is eluted around 150 mM imidazole into test tube #29 to #33. Samples labeled as #1 to #18 were used in SDS-PAGE analysis.

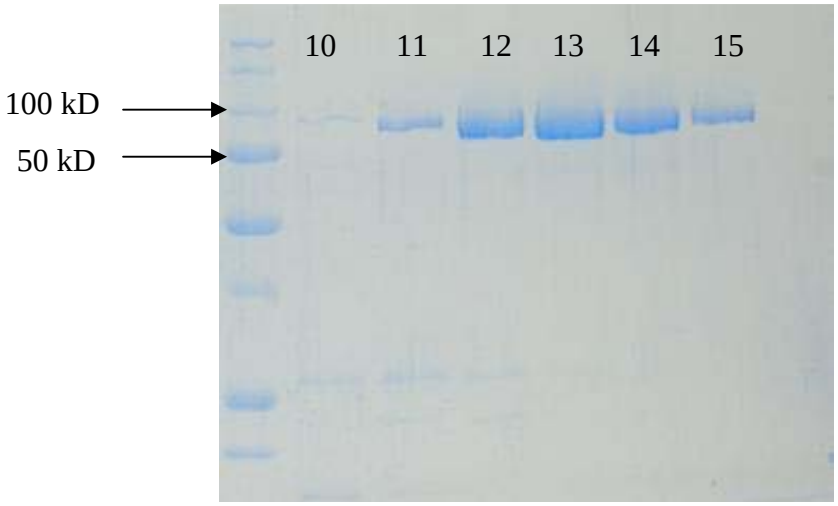
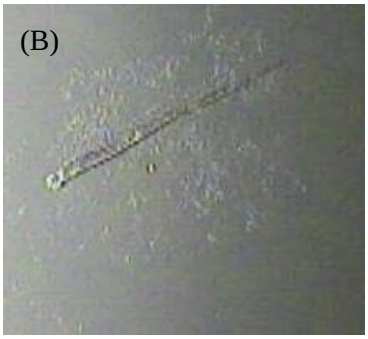



**Figure 3-6. SDS-PAGE analysis of MalT purified by Ni affinity column chromatography.** 15% SDS-PAGE gel stained with Coomassie blue. First lane from the left contains size marker and the numbers for each lane correspond to the sample number on the chromatogram. MalT was eluted with about 150 mM imidazole and the amount of impurities was negligible. Samples # 12 - 14 were pooled and used in further purification by a desalting column.

Based on the screening (Table 3-7 and Figure 3-7), the optimization of crystallization condition for DT1 and MalT was conducted as follows. Initially, for DT1, since crystals were found in Index #68 were relatively larger than the others and all reagents in Index #68 (Jaffamine M-600 consists of O-(2-aminopropyl)-O'-(2-

methoxyethyl)polypropylene glycol 500 and polypropylene glycol 500 mono-2-aminoethyl mono-2-methoxyethyl ether) were available, optimization was tried by varying the concentration of protein sample in a droplet (5-10 mg/mL of protein sample was used to prepare various sizes of droplets), the pH ranged from 7.2 to 7.7, and the trays were tried at temperatures 4 °C and RT. Moreover, the pH (7.8 or 8.8) during purification was varied in order to see whether it affected crystallization of DT1. As a result, larger needles were obtained with the protein purified at pH 8.8 and they were more separated in reservoirs with 7.5, 7.6 and 7.7 compared to the original conditions. However, they were not large enough for data collection. Therefore, optimization using Index #38 or #76 were also tried in order to obtain single crystals for data collection of DT1. So far, large single crystals have not been obtained even though the size of the crystals has been improved after optimization. In order to obtain better crystals of DT1, some trials should be performed such as changing the type of buffer for purification, the amount of ATP, and also using different crystallization buffers that led to the crystal formation of DT1 in the screening (Table 3-7).

| **Table 3-7. Crystallization conditions that led to the formation of crystalline material or small crystals for DT1 and MalT within a month.** | | |
|---|---|---|
| Protein Name | Preliminary Crystallization Condition | Appearance |
| DT1 | Screen I #40 (0.1 M sodium citrate tribasic dehydrate pH 5.6 and 25% v/v 2-propanol) | Needles |
| | Index #37 (25% w/v PEG1500) | Crystalline |
| | Index #38 (0.1 M HEPES pH 7.0 and 30 % v/v Jeffamine M-600 pH 7.0) | Crystalline |
| | Index #66 (0.2 M ammonium sulfate, 0.1 M BIS-Tris pH5.5 and 25% w/v PEG) | Needles |
| | Index #67 (0.2 M ammonium sulfate, 0.1M BIS-Tris pH 6.5, and 25% w/v PEG 3350) | Needles |
| | Index #68 (0.2 M ammonium sulfate, 0.1 M HEPES pH 7.5, and w/v 25% PEG 3350) | Needles |
| | Index #74 (0.2 M lithium sulfate monohydrate, 0.1 M BIS-Tris pH 5.5, and w/v 25% PEG3350) | Needles |
| | Index #75 (0.2 M lithium sulfate monohydrate, 0.1 M BIS-Tris pH 6.5, and w/v 25% PEG3350) | Needles and plates |
| | Index #76 (0.2 M lithium sulfate monohydrate, 0.1 M HEPES pH 7.5, and w/v 25% PEG3350) | Needles and plates |
| | | |
| MalT | Screen II #1 (2.0 M sodium chloride and 10% w/v PEG6000) | A tiny plate |
| | Screen II # 38 (0.1 M HEPES pH 7.5 and 20% w/v PEG 10000) | A tiny crystal |
| | Index #29 (60% v/v Tacsimate pH 7.0) | Crystalline |
| | Index #33 (1.1 M sodium malonate pH 7.0, 0.1 M HEPES pH 7.0, and 0.5% v/v Jeffamine ED-2001 pH 7.0) | A needle and tiny crystal |

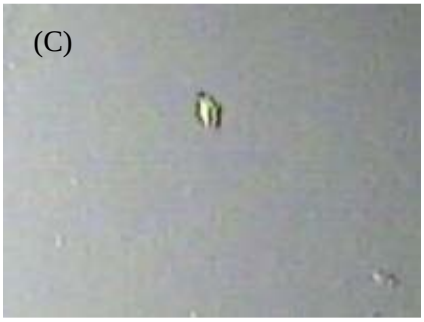




**Figure 3-7. Crystals or crystalline like objects found in screening for DT1 and MalT.** (A) DT1 in Index #38. (B) MalT in Index#33. (C) MalT in ScreenII #38.

Based on the screening (Figure 3-7), the optimization of the crystallization condition for DT1 and MalT was conducted as follows. Initially, for DT1, since crystals were found in Index #68 were relatively larger than the others and all reagents in Index #68 (Jaffamine M-600 consists of O-(2-aminopropyl)-O'-(2-methoxyethyl)polypropylene glycol 500 and polypropylene glycol 500 mono-2-aminoethyl mono-2-methoxyethyl ether) were available, optimization was tried by varying the concentration of protein sample in a droplet (5-10 mg/mL of protein sample was used to prepare various sizes of droplets), the pH range from 7.2 to 7.7, and the trays were tried at temperatures 4 °C and RT. Moreover, the pH (7.8 or 8.8) during purification was varied in order to see whether it affected crystallization of DT1. As a result, larger needles were obtained with the protein purified at pH 8.8 and they were more separated in reservoirs with 7.5, 7.6 and 7.7 compared to the original conditions. However, they were not large enough for data collection. Therefore, optimization using Index #38 and #76 were also tried in order to

obtain single crystals for data collection of DT1. So far, large single crystals have not been obtained even though the size of the crystals has been improved after optimization. In order to obtain better crystals of DT1, some trials should be performed such as changing the type of buffer for purification, the amount of ATP, and also using different crystallization buffers that led to the crystal formation of DT1 in screening (Table 3-7).

On the other hand, for MalT, even though crystals in Index #38 looked better than the others, PEG10000 was not available. Therefore, PEG8000 was used instead of PEG 10000 to optimize the crystallization condition for MalT by varying the concentration of protein sample in the droplet (about 5mg/mL was used to prepare various sizes of droplets), the pH ranged from 6.8 to 7.8, the percentage of PEG8000 (20%-32%), and the temperature (4 °C and RT). Moreover, using Tacsimate (Index #29, the mixture of 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), the optimization was performed at RT by altering the percentage of Tacsimate (from 50% to 70%) in the reservoir solution. Crystals were formed in a few weeks at 4 °C and in a month to a few months at RT (Figure 3-8). Initially, good precipitates and small crystalline like objects were observed in droplets that led to the formation of these crystals. However, it took a few weeks to months for these crystals to grow in the droplets. As a result, reservoir solutions whose droplets contained crystals were almost dried out when they were examined by X-ray diffraction. Crystals in Figure 3-8 were mounted on the X-ray diffractometer but did not diffract (simple scan by Dr. Leonard M. Thomas at the University of Oklahoma Macromolecular

Crystallography Laboratory). It is possible that crystals shown in Figure 3-8 are just salt or proteins that had just lost the order of the protein.

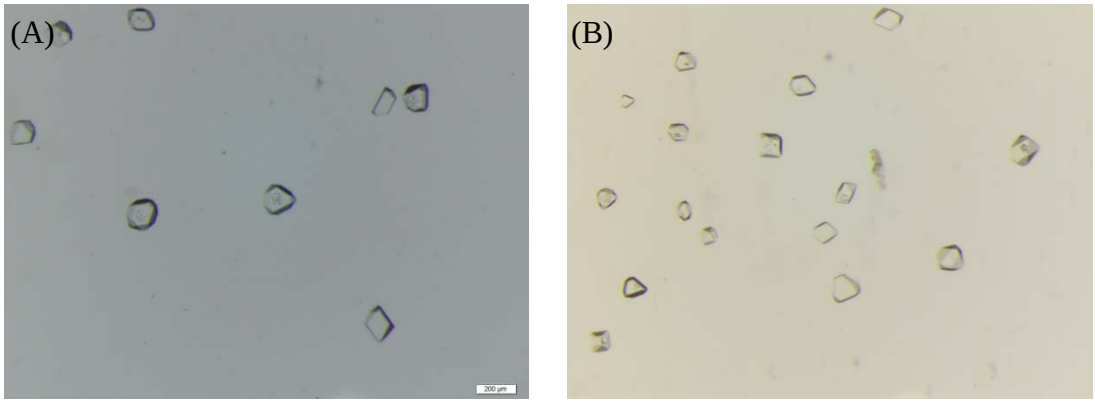



**Figure 3-8. Crystals found in optimizing the MalT crystallization.** (A) Crystals in a droplet containing 3 µL MalT (1 mg/mL) and 3 µL reservoir solution (60% Tacsimate) at RT. (B) Crystals in a droplet containing 3 µL MalT (1mg/mL) and 3 µL reservoir solution (65% Tacsimate) at RT.

Further investigations with these conditions are necessary to examine whether crystals obtained in this condition were protein or salt crystals. Moreover, buffers used in the purification should be investigated since in this research only Tris buffers were used based on the literature that successfully purified each domain of MalT along with whole MalT. Besides Tris buffer, HEPES, citrates and others may work better to purify these proteins and affect their crystallization. In addition to buffer conditions, crystallization conditions should be further investigated based on the screening since some reagents were not available at that time and thus some conditions listed in Table 3-7 were not investigated.

Since crystals of DT1 or MalT have not been obtained, preliminary structural studies have been performed by modeling DT1 based on the sequence using web servers, SWISS-MODEL (an automated comparative protein modeling server) [93-95] and Phyre (protein homology/analogy recognition engine) [96] (Figure 3-9). SWISS-MODEL generated a segment, a model 1 (residues 32 to 68), based on the crystal structure of the putative gluconate kinase from *Bacillus halodurans* (PDB ID: 2BDT and sequence identity = 34.375%) with Expect value (E-value) = $2.90e^{-5}$ and a model 2 (residues 9 to 225) based on the crystal structure of RuvB complexed with RuvA (Holliday junction DNA helicases) domain III from *Thermus thermophilus* (1IXS and identity = 14.041%) [97] with E-value = $8.9e^{-12}$. On the other hand, Phyre generated ten models with better E-value that is a factor showing the number of hits one can expect to see by chance [98]. A model 3, d2fnaa2 (residues 10 to 237), was generated based on P-loop containing nucleoside triphosphate hydrolases (identity = 12%) with E-value = $1.2e^{-18}$, which belong to the Family AAA-ATPase domain. A model 4, c2fnaB (residues 10 to 237), was generated based on the conserved hypothetical protein from *Sulfolobus solfataricus* (identity = 12%) with E-value = $9.9e^{-18}$. A model 5, c2a5yB (residues 8 to 236), was generated based on the crystal structure of apoptosome from the *Caenorhabditis elegans* (CED4, identity = 12%) with E-value = $2.2e^{-16}$ According to these models, DT1 may consist of α-helices with a few β-strands. However, since residue 60 and 61 are not connected in model 3 and model 4 while a β-strand is present around these residues of model 4, there are differences among these models even though they look similar. Therefore, further investigation is necessary to understand the structure of DT1 and how DT1 interacts with Aes. Moreover, Phyre and SWISS-MODEL were also used for

modeling MalT and DT1-DT2. However, only a part of them, which were almost the same as model 1 to 5, were generated.

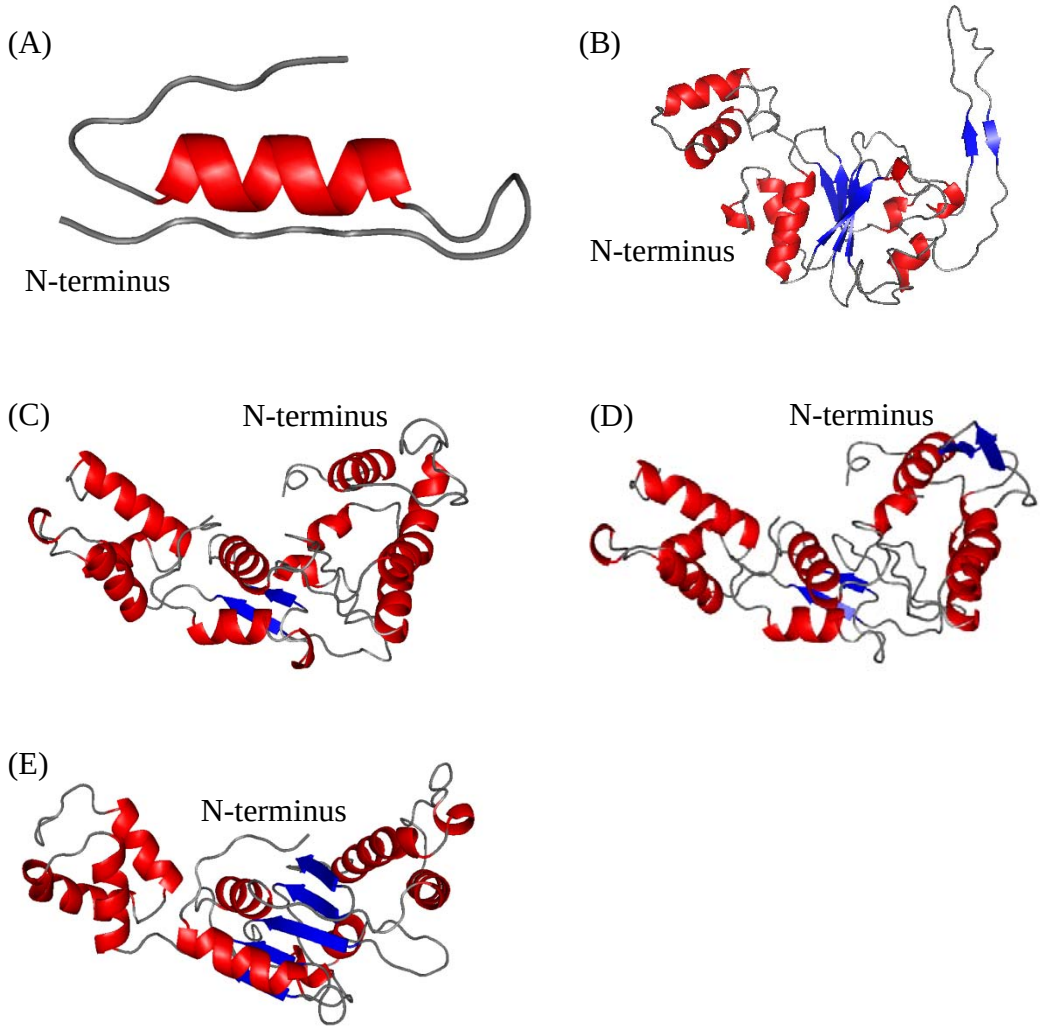



**Figure 3-9 Models of DT1.** These models were generated based on the sequence of DT1. (A) and (B) were generated by SWISS-MODEL and (C), (D), and (E) were generated by Phyre. (A) Model 1 was based on the crystal structure of the putative gluconate kinase from *Bacillus halodurans* (PDB ID: 2BDT) with Expect value (E-value) = $2.90e^{-5}$. (B) Model 2 was based on the crystal structure of RuvB complexed with RuvA (Holliday junction DNA helicases) domain III from *Thermus thermophilus* (1IXS). (C) Model 3 was generated by Phyre, based on the P-loop fold containing nucleoside triphosphate hydrolases. (D) Model 4 was generated by Phyre based on a hypothetical protein from *Sulfolobus solfataricus*. (E) Model 5 was generated by Phyre based on the crystal structure of apoptosome from the *Caenorhabditis elegans* (CED4).

### 3.4 Summary

The plasmids carrying DT1, DT1-DT2, and *malT* were constructed and DT1, DT1-DT2, and MalT were successfully over-expressed in *E. coli* in this study. Moreover, DT1 and MalT were also successfully purified and as a result of crystallization screening, there were some conditions that led to the formation of small crystals or crystalline like objects, which were necessary to be optimized so as to obtain single crystals that diffract and lead to the crystal structure determination of these proteins. Current work with some conditions is quite promising. Therefore, continued work in optimizing conditions listed in section 3.3 will lead to the structure determination of these proteins. The purification of DT1-DT2 should be also investigated so as to purify DT1-DT2 since this region is a critical part in the binding of Aes and ATP. The crystal structure determination of this part should help understanding the relationship between Aes and this region of MalT. Moreover, since the crystal structure of DT3 has been determined [92], crystal structure determination of DT1-DT2 will lead to understanding of the structure of MalT and subsequently the interaction between Aes and MalT. Purification condition for them can also be optimized using different types of buffers such as HEPES, which may lead to higher yield of these proteins, especially MalT since the yield of MaT is quite low compared to other proteins studied in this research. Co-crystallization of these proteins and Aes is challenging but worth trying in order to obtain co-crystals to understand how they interact. Moreover, according to preliminary models of DT1 obtained using Phyre, DT1 may contain α-helices with a few β-strands and model 3 might be close to the crystal structure of DT1 since DT1 shows week ATPase activity as its template. However, since there are differences among these models further investigation is necessary to conclude

the plausible model of DT1 and know how DT1 interacts with Aes. Moreover, structural studies of MalT and DT1-DT2 based on the sequence using SWISS-MODEL [93-95] with MSA and docking between Aes and these proteins can be investigated further in order to obtain better models so as to understand how Aes and MalT interact.

# CHAPTER 4

## CONCLUSION

This dissertation reports the cloning, over-expression, purification, crystallization and crystal structure analysis of Aes that belongs to the HSL family and is a down regulator for MalT which is a transcriptional activator of the maltose regulon (Chapters 2 and 3). Aes acts as an acetyl esterase and a recent study suggests an interaction occurs between Aes and α-galactosidase, which is involved in the hydrolysis of α-linked-galactosides such as those found in the more complex sugars raffinose and melibiose. Since Aes plays critical roles in carbohydrate metabolism, especially in the notable maltose system in *E. coli*, the crystal structure determination of Aes will help understanding the structures of homologous proteins and also lead to better understanding of the maltose system in the future. In this study, a recombinant Aes in a six histidine tag was crystallized (the space groups were *R*3, and $P4_1$) and the data set was obtained with the range from 2.7 to 2.8 Å. The crystal structure of Aes was determined using PHENIX and COOT (see section 2.3.4). X-ray crystallography revealed that Aes contained an α/β hydrolase fold similar to homologous proteins whose structures have been determined. Moreover, according to the structural analysis and literature review, the catalytic triad of Aes consists of Ser-165, Asp-262 and His-292 with Ser-165 being stabilized by His-292, which, in turn, is also stabilized by Asp-262 through hydrogen bonding. Furthermore,

Ser-165 is stabilized by weak hydrogen bonding from the HGGG motif. Aes also has a long loop with a few α-helices around the N-terminus and contains a shallow cleft consisting of hydrophobic residues, which lead to the substrate binding pocket and the catalytic triad. As mentioned in sections 2.2.10 and 2.3.4, several homologous proteins have been used to find the initial phases. However, acetyl esterase from *Salmonella typhimurium* (PDB code: 3GA7) has been the only structure successful in determination of initial phases. According to the comparison of crystal structures, there are slight shifts between Aes from *E. coli* and the search models except for 3GA7 even though the overall structures of all contain the α/β hydrolase fold. In addition, the N-terminus of all of them is highly flexible, leading to the difficulty in finding phases. The reasons why 3GA7 worked in molecular replacement is that the area of greatest difference, the loop of the N-terminus, is not modeled in the structure of 3GA7 and the overall fold of the rest of the molecule shows little structural difference with Aes from *E. coli* (see section 2.3.4). In this study, the loop region around the N-terminus of Aes was built even though some residues were still outliers in the Ramachandran plot. Since the crystal structure of Aes was determined in this study including the N-terminal region which should be an important part of Aes since it may act as a lid protecting the active site, the findings should help further crystal structure determination of homologous proteins, leading to the new finding about them. The ultimate goal is to know how MalT and Aes interact in *E. coli* so as to understand the remarkable maltose system. Therefore, the crystal structure of Aes will lead to this ultimate goal in the future since the knowledge of the structure will give insight into the function of the protein.

MalT is the transcriptional regulator for the maltose regulon that is a cluster of operons and genes involved in maltose metabolism (see section 3.1). MalT is activated by maltotriose and ATP and inactivated by three proteins; MalK, MalY, and Aes. Four operons involved in the maltose system in *E. coli* are mostly under the control of MalT. Since the crystal structure of DT3 has been determined, besides the crystal structure determination of Aes, another aim of this study has been to determine the crystal structures of DT1, DT1-DT2, and MalT in order to understand how MalT and Aes interact, which play a role in the remarkable maltose system. To achieve this, plasmids carrying each portion of the gene of interest have been constructed and the proteins have been over-expressed. In this study, DT1 and MalT were successfully purified and crystallization has also been carried out in this study (section 3.2). However, high quality crystals that diffract well enough for data collection have not been obtained yet. In order to obtain high quality crystals, the following should be investigated for both proteins: purification with different types of buffers and further optimization in crystallization based on the results of crystal screening. Moreover, purification of DT1-DT2 will have to be investigated with different buffers or methods. According to preliminary models of DT1, DT1 may contain α-helices with a few β-strands. However, the results were not conclusive. Therefore, further investigation of these proteins in modeling will have to be carried out in order to know how MalT and Aes interact in *E. coli*.

# REFERENCES


1. **Deutscher, J., C. Francke, and P. W. Postma**, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.*, 2006. **70**: p. 939-1031.
2. **Gerber, K.**, X-ray crystallographic studies on Mlc and Aes, two transcriptional modulators from *Escherichia coli*, in *Biology*. 2005, University Konstanz: Konstanz. p. 115.
3. **Voet, D. and J. G. Voet**, Biochemistry. Third ed. Chapter 20 Trabsport through Membranes, ed. Harris, D., P. Fitzgerald, R. Smith, B. Heaney, S. Dumas, H. Newman, E. Rieder, M. Lesure, S. Ingrao, and S. Malinowski. 2004: John Wiley and Sons, Inc.
4. **Potsma, P. W., J. W. Lengeler, and G. R. Jacobson**, Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.*, 1993. **57**: p. 543-594.
5. **Kuroda, M., T. H. Wilson, and T. Tsuchiya**, Regulation of Galactoside Transport by the PTS. *J. Mol. Microbiol. Biotechnol.*, 2001. **3**: p. 381-384.
6. **Boos, W. and H. Shuman**, Maltose/Maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Mocrobiol. Mol. Biol. Rev.*, 1998. **62**: p. 204-229.
7. **Voet, D. and J. G. Voet**, Biochemistry. Third ed. Chapter 31 Transcription, ed. Harris, D., P. Fitzgerald, R. Smith, B. Heaney, S. Dumas, H. Newman, E. Rieder, M. Lesure, S. Ingrao, and S. Malinowski. 2004: John Wiley and Sons, Inc.
8. **Lengeler, J. W.**, ed. The phosphoenolpyruvate-dependent carbohydrate: phosphotransferase system (PTS) and control of carbon source utilization. Regulation of gene expression in *Escherichia coli*., ed. Lin, E. C. C. and A. S. Lynch. 1996, R. G. Landes Company: Austin, Texas, USA.
9. **Voet, D. and J. G. Voet**, Biochemistry. Third ed. Chapter 5 Nucleic Acids, Gene Expression, and Recombinant DNA Technology, ed. Harris, D., P. Fitzgerald, R. Smith, B. Heaney, S. Dumas, H. Newman, E. Rieder, M. Lesure, S. Ingrao, and S. Malinowski. 2004: John Wiley and Sons, Inc.
10. **Nelson, D. L. and M. M. Cox**, Lehninger Principles of Biochemistry. fourth ed. 2005, New York: W. H. Freeman and Company.
11. **Schlegel, A., A. Böhm, S.-J. Lee, R. Peist, K. Decker, and W. Boos**, Network regulation of the *Escherichia coli* maltose system. *J. Mol. Microbiol. Biotechnol.*, 2002. **4**: p. 301-307.
12. **Chapon, C.**, Role of the catabolite activator protein in the maltose regulon of *Escherichia coli. J. Bacteriol.*, 1982. **150**: p. 722-729.

13. **Danot, O., D. Vidal-Ingigliardi, and O. Raibaud**, Two amino acid residues from the DNA-binding domain of MalT play a crucial role in transcripitonal activation. *J. Mol. Biol.*, 1996. **262**: p. 1-11.
14. **Schlegel, A., O. Danot, E. Richet, T. Ferenci, and W. Boos**, The N terminus of the *Escherichia coli* transcription activator MalT Is the domain of interaction with MalY. *J. Bacteriol.*, 2002. **184**: p. 3069-3077.
15. **Joly, N., O. Danot, A. Schlegel, B. Windried, and E. Richet**, The Aes protein directly controls the activity of MalT, the central transcriptional cctivator of the *Escherichia coli* maltose regulon. *J. Biol. Chem.*, 2002. **277 No.19**: p. 26606-16613.
16. **Richet, E. and O. Raibaud**, Purification and properties of the MalT protein, the transcription activator of the *Escherichia coli* maltose regulon. *J. Biol. Chem.*, 1987. **262, No.26**: p. 12647-12653.
17. **EcoCyc**. *Escherichia coli* K-12 substr. MG1655. 1997 [cited 2009 Wed Oct 14]; Available from: http://ecocyc.org/ECOLI/new-image?type=GENE&object=EG11101.
18. **Keseler, I. M., C. Bonavides-Martines, J. Collado-Vides, S. Gama-Castro, R. P. Gunsalus, D. A. Johnson, M. Krummenacker, L. M. Nolan, S. Paley, I. T. Paulsen, M. Peralta-Gil, A. Santos-Zavaleta, A. G. Shearer, and P. D. Karp**, EcoCyc: A comprehensive view of *Escherichia coli* biology. *Nucleic Acids Research*, 2009. **37**: p. D464-D470.
19. **Kanaya, S., T. Koyanagi, and E. Kanaya**, An esterase from *Escherichia coli* with a sequence similarity to hormone-sensitive lipase. *Biochem. J.*, 1998. **332**: p. 75-80.
20. **Hemila, H., T. T. Koivula, and I. Palva**, Hormone-sensitive lipase is closely related to several bacterial proteins, and distantly related to acetylcholinesterase and lipoprotein lipase: identification of a superfamily of esterases and lipases. *Biochim. Biophys. Acta*
21. **Gerber, K., A. Schiefner, P. Seige, K. Diederichs, W. Boos, and W. Welte**, Crystallization and preliminary X-ray analysis of Aes, an acetyl-esterase from *Escherichia coli. Acta Crystallogr. D*, 2004. **60**: p. 531-533.
22. **Richet, E., N. Joly, and O. Danot**, Two domains of MalT, the cctivator of the *Escherichia coli* maltose regulon, bear determinants essential for anti-activation by MalK. *J. Mol. Biol.*, 2005. **347**: p. 1-10.
23. **Peist, R., A. Koch, P. Bolek, S. Sewitz, T. Kolbus, and W. Boos**, Characterization of the *aes* gene of *Escherichia coli* encoding an enzyme with esterase activity. *J. Bacteriol.*, 1997. **179. No.24**: p. 7679-7686.
24. **Burstein, C. and A. Kepes**, The α-galactosidase from Escheirichia coli K12. *Biochim. Biophys. Acta*, 1971. **230**: p. 52-63.
25. **Mandrich, L., E. Caputo, B. M. Martin, M. Rossi, and G. Manco**, The Aes protein and the monomeric alpha-galactosidase from *Escherichia coli* form a non-covalent complex. *J. Biol. Chem.*, 2002. **277 No.50**: p. 48241-48247.
26. **Brenner, S.**, The molecular evolution of genes and proteins: a tale of two serines. *Nature*, 1988. **334**: p. 528-530.
27. **Winkler, F. K., A. D'arcy, and W. Hunzuker**, Structure of human pancreatic lipase. *Nature*, 1990. **343**: p. 771-774.

28. **Sussman, J. L., M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, and I. Silman**, Atomic structure of acetylcholinesterase from Torpedo california: a prototypic acetylcholine-binding protein. *Science*, 1991. **253**: p. 872-879.
29. **Schrag, J. D., Y. LI, S. WU, and M. Cygler**, Ser-His-Glu triad forms the catalytic site of the lipase from *Geotrichum candidum. Nature*, 1991. **351**.
30. **Arpigny, J. L. and K.-E. Jaeger**, Bacterial lipolytic enzymes: classification and properties. *Biochem. J.*, 1999. **343**: p. 177-183.
31. **Hulo, N., A. Bairoch, V. Bulliard, L. Cerutti, E. D. Castro, P. S. Langendijk-Genevaux, M. Pagni, and C. J. A. Sigrist**, The PROSITE database. *Nucleic Acids Res.*, 2006. **34**: p. D227-D230.
32. **PROSITE**. 2006, Swiss Institute of Bioinformatics (SIB).
33. **Cousin, X., T. Hotelier, K. Giles, P. Lievin, J.-P. Toutant, and A. Chatonnet**, The α/β fold family of proteins database and the cholinesterase gene server ESTHER. *Nucleic Acids Res.*, 1997. **25, No. 1**: p. 143-146.
34. **Krintel, C., C. Klint, H. Lindvall, M. Mörgelin, and C. Holm**, Quarternary structure and enzymological properties of the different hormone-sensitive lipase (HSL) isoforms. *PLoS One*, 2010. **5**: p. e11193.
35. **Yeaman, S. J.**, Hormone-sensitive lipase - a multipurpose enzyme in lipid metabolism. *Biochem. Biophys. Acta*, 1990. **1052**: p. 128-132.
36. **Small, C. A., J. A. Goodacre, and S. J. Yeaman**, Hormone-sensitive lipase is responsible for theneutral cholesterol ester hydrolase activity in macrophages. *FEBS Lett.*, 1989. **247**: p. 205-208.
37. **Nedergaard, J. and B. Cannon**, Preferential utilization of brown adipose tissue lipids during arousal from hibernation in hamsters. *Am. J . Physiol.*, 1984. **247**: p. R506 - 512.
38. **Langin, D., H. Laurell, L. S. Holst, P. Belfrage, and C. Holm**, Gene organization and primary structure of human hormone-sensitive lipase: possible significance of a sequence homology with a lipase of *Moraxella* TA144, an antarctic bacterium. *Proc. Natl. Acad. Sci.*, 1993. **90**: p. 4897-4901.
39. **Holm, C., T. G. Kirchgessner, K. L. Svenson, G. Fredrikson, S. Milsson, C. G. Miller, J. E. Shively, C. Heinzmann, R. S. Sparkes, T. Mohandas, A. J. Lysis, P. Belfrage, and M. C. Schotz**, Hormone-sensitive lipase: sequence, expression, and chromosomal localization to 19 cent-q13.3. *Science*, 1988. **241**: p. 1503-1506.
40. **Osterlund, T., B. Danielsson, E. Degerman, J. A. Contreras, G. Edgren, R. C. Davis, M. C. Schotz, and C. Holm**, Domain-structure analysis of recombinant rat hormone-sensitive lipase. *Biochem. J.*, 1996. **319**: p. 411-420.
41. **Wilson, K.**, Current Protocols in Molecular Biology: Prepartion of Genomic DNA from Bacteria, ed. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 2003, Townsville, Australia: John Wiley&Sons, Inc.
42. **Invitrogen, C.**, Champion™ pET Directional TOPO Expression Kits. 2006, Carlsbad, CA: Invitrogen Corporation. 1-62.
43. **Promega, C.**, Technical Bulletin: Wizard *Plus* Minipreps DNA Purification System. 2005, Madison, WI: Promega Corp.
44. **Promega, C.**, Restriction Enzyme User Manual. 1996-2006.

45. **BIO-RAD**, Mini-PROTEIN 3 Cell Instruction Manual. 2005, Hercules, CA.
46. **BIO-RAD**, BioLogic DuoFlow™ Chromatography System Starter Kit Instruction Manual, Hercules, CA.
47. **BIO-RAD**, Protein Assay, BIO-RAD, Editor.
48. **Hampton**, Crystal Screen - User Guide. 2000-2005, Hampton Research: Aliso Viejo, CA.
49. **BRUKER, A.**, PROTEUM2 User Manual. 2005.
50. **BRUKER**, **A.**, PROTEUM crystallographic software suit. 2009.
51. **BRUKER, A.**, SAINT-NT. 2003: Madison, WI.
52. **Sheldrick, G. M.** (2001) XPREP.
53. **Campbell, R. L.** Protein function discovery and department of biochemistry moleculr modelling and crystallographic computing facility. 2005 [cited 2009 Nov. 16]; Available from: http://pldserver1.biochem.queensu.ca/~rlc/pfd/index.shtml.
54. **Matthews, B. W.**, Solvent content of protein crystals. *J. Mol. Biol.*, 1968. **33**: p. 491-497.
55. **Collaborative Computational Project, N.**, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, 1994. **50**: p. 760-763.
56. **Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart**, PHENIX: a comprehensive Python-based system for macromolecular structure solution. 2010.
57. **Adams, P. D., R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, and T. C. Terwilliger**, PHENIX: building new software for automated crystallographic structure determination. *Acta Cryst.*, 2002. **D58**: p. 1948-1954.
58. **Simone, G. D., V. Menchise, G. Manco, L. Mandrich, N. Sorrentino, D. Lang, M. Rossi, and C. Pedone**, The crystal structure of a hyper-thermophilic carboxylesterase from the archaeon Archaeoglobus fulgidus. *J. Mol. Biol.*, 2001. **314**: p. 507-518.
59. **Byun, J.-S., J.-K. Rhee, N. D. Kim, J. Yoon, D.-U. Kim, E. Koh, J.-W. Oh, and H.-S. Cho**, Crystali structure of hyperthermophilic esterase EstE1 and the relationship between its dimerization and thermostability properties. *BMC Struct. Biol.*, 2007. **7**.
60. **Simone, G. D., S. Galdiero, G. Manco, D. Lang, M. Rossi, and C. Pedone**, A Snapshot of a transition state analogue of a novel thermophilic esterase belonging to the subfamily of mammalian hormone-sensitive lipase. *J. Mol. Biol.*, 2000. **303**: p. 761-771.
61. **Simone, G. D., L. Mandrich, V. Menchise, V. Giordano, F. Febbraio, M. Rossi, C. Pedone, and G. Manco**, A Substrate-induced switch in the reaction mechanism of a thermophilic esterase. *J. Biol. Chem.*, 2004. **279** p. 6815-6823.
62. **Minasov, G., Z. Wawrzak, J. Brunzelle, O. Onopriyenko, T. Skarina, P. Scott, A. Savchenko, and W. F. Anderson**, 1.55 Angstrom crystal structure of an acetyl esterase from *Salmonella typhimurium*. 2009, Center for Structural Genomics Of Infectious Diseases (CSGID).

63. **Simone, G. D., V. Menchise, V. Alterio, L. Mandrich, M. rossi, G. Manco, and C. Pedone**, The crystal structure of an EST2 mutant unveils structural insights on the H-Group of the carboxylesterase/lipase family. *J. Mol. Biol.*, 2004. **343**: p. 137-146.
64. **Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne**, The Protein Data Bank. *Nucleic Acids Res.*, 2000. **28**.
65. **Emsley, P. and K. Cowtan**, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, 2004. **60**: p. 2126-2132.
66. **Sayle, R. and E. J. Milner-White**, RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.*, 1995. **20**: p. 374.
67. **Brünger, A. T.**, Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 1992. **355**: p. 472-475.
68. **Kirkpatrick, S., C. D. G. Jr., and M. P. Vecchi**, Optimazation by simulated annealing. *Science*, 1983. **220**: p. 671-680.
69. **Cruickshank, D. W.**, On the lattice vibrations of benzene, naphthalene, and anthracene. *Rev. Mod. Phys.*, 1958. **30**: p. 163-167.
70. **Schomaker, V. and K. N. Trueblood**, On the rigid-body motion of molecules in crystals. *Acta Crystallogr. B*, 1968. **24**: p. 63-76.
71. **Rupp, B.**, Biomolecular Crystallography: Principles, Practice, and Application to Structral Biology, ed. Scholl, S. 2009, New York: Garland Science, Taylor & Francis Group, LLC.
72. **Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. A. III, J. Snoeyink, J. S. Richardson, and D. C. Richardson**, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, 2007. **35**: p. W375-W383.
73. **Painter, J. and E. A. Merritt**, Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D*, 2006. **62**: p. 439-450.
74. **Painter, J. and E. A. Merritt**, TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.*, 2006. **39**: p. 109-111.
75. **Holm, L., S. Kaariainen, P. Rosenstrom, and A. Schenkel**, Searching protein structure database with DaliLite v.3. *Bioinformatics*, 2008. **24**: p. 2780-2781.
76. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P.A.McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins**, ClustalW and ClustalX version 2. *Bioinformatics*, 2007. **23**: p. 2947-2948.
77. **Koenig, T., B. H. Menze, M. Kirchner, F. Monigatti, K. C. Parker, T. Patterson, J. J. Steen, F. A. Hamprecht, and HannoSteen**, Robust prediction of the MASCOT score for an improved quality aeesssment in mass spectrometric proteomics. *J. Proteome Res.*, 2008. **7**: p. 3708-3717.
78. **Perkins, D. N., D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell**, Probability-based protein identification by searching sequence database using mass spectrometry data. *Electrophoresis*, 1999. **20**: p. 3551-3567.
79. **Pappin, D. J. C., P. Hojrup, and A. J. Bleasby**, Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.*, 1993. **3**: p. 327-332.

80. **DeLano, W. L.**, The PyMOL Molecular Graphics System. 2002, DeLano Scientific: San Carlos, CA.
81. **Krissinel, E. and K. Henrick**, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, 2007. **372**: p. 774-797.
82. **Richet, E. and O. Raibaud**, MalT, the regulatory protein of the *Escherichia coli* maltose system, is an ATP-dependent transcriptional activator. *EMBO J.*, 1989. **8**: p. 981-987.
83. **Tagami, H. and H. Aiba**, A common role of CRP in trascription activation: CRP acts transiently to stimulate events leading to open complex formation at a diverse set of promoters. *EMBO J.*, 1998. **17**: p. 1759-1767.
84. **Valerie, S. and E. Richet**, Self-association of the *Escherichia coli* transcription activator in the presence of maltotriose and ATP. *J. Biol. Chem.*, 1999. **274**: p. 33220-33226.
85. **Decker, K., J. Plumbridge, and W. Boos**, Negative transcriptional regulation of a positive regulator: the expression of malT, encoding the transcriptional activator of the maltose regulon of *Escherichia coli*, is negatively controlled by Mlc. *Mol. Microbiol.*, 1998. **27**: p. 381-390.
86. **Johansson, J., B. Dagberg, E. Richet, and B. E. Uhlin**, H-NS and StpA proteins stimulate expression of the maltose regulon in *Escherichia coli. J. Bacteriol.*, 1998. **180**: p. 6117-6125.
87. **Clausen, T., A. Schlegel, R. Peist, E. Schneider, C. Steegborn, Y.-S. Chang, A. Haase, G. P.Bourenkov, H. D.Bartunik, and W. Boos**, X-ray structure of MalY from *Escherichia coli*: a pyridoxal 5'-phosphate-dependent enzyme acting as a modulator in *mal* gene expression. *EMBO J.*, 2000. **19**: p. 831-842.
88. **Chen, J., J. Lin, A. L. Davidson, and F. A. Quiocho**, A Tweezers-like motion of the ATP-binding casette dimer in an ABC transport cycle. *Mol. Cell*, 2003. **12**: p. 651-661.
89. **Panagiotidis, C. H., W. Boos, and H. A. Shuman**, The ATP-binding cassette subunit of the maltose transporter MalK antagonizes MalT, the activator of the *Escherichia coli mal* regulon. *Mol. Microbiol*, 1998. **30**: p. 535-546.
90. **Jain, E., A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B. E. Suzek, M. J. Martin, P. MacGarvey, and E. Gasteiger**, Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, 2009. **10**: p. 10-136.
91. **Consorthium, T. U.**, The Universal Protein Resouce (UniProt) in 2010. *Nucleic Acids Res.*, 2010: p. D142-D148.
92. **Steegborn, C., O. Danot, R. Huber, and T. Clausen**, Crystal structure of transcription factor MalT domai III: a novel helix repest fold implicated in regulated oligomerization. *Structure*, 2001. **9**: p. 1051-1060.
93. **Peitsch, M. C.**, Protein modeling by E-mail. *Bio/Technology*, 1995. **13**: p. 658-660.
94. **Arnold, K., L. Bordoli, J. Kopp, and T. Schwede**, The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 2006. **22**: p. 195-201.

95. **Kiefer, F., K. Arnold, M. Künzli, L. Bordoli, and T. Schwede**, The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, 2009. **37**: p. D387-D392.
96. **Kelley, L. A. and M. J. E. Stemberg**, Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc*, 2009. **4**: p. 363-371.
97. **Yamada, K., T. Miyata, D. Tsuchiya, T. Oyama, Y. Fujiwara, T. Ohnishi, H. Iwasaki, H. Shinagawa, M. Ariyoshi, K. Mayanagi, and K. Morikawa**, Crystal structure of the RuvA-RuvB complex: a structural basis for the holliday junction migrating motor machinery. *Mol. Cell*, 2002. **10**: p. 671-681.
98. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhan, W. Miller, and D. J. Lipman**, Gapped BLAST an PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 1997. **25**: p. 3389-3402.

APPPENDICES

**6.1. DNA sequence of *aes* from *E. coli* str. K12 substr.**

**GENE ID: 947514 aes**

CTAAAGCTGAGCGGTAAAGAACTGAGCGCCGTCGCGAAGAGCCTCGTCGGC
GGTTTTCATCATCCGTGAATAATGCAAAAAGGCGTGCAGCGTGCCTGGGTAG
AGTTTGAACTCACAGGGCTGCTGATGCGCCGCTAACGTCTGGTAAAGCAGAC
GGCTGTCATCCAGCAGCGGATCGAACTCCGCCCCGGCAATAAAACAGGGCGG
AACTTCGCGAGTGAGATCATTATTAAACAGACAGTAATACGGCGACTCGCGG
TCCGCGTCGTTGCTTAAATATGCCTCTTCGTACATCTGCAAATCCTGTTGCGTT
AAGCCATCCCAGACACCGCCCAACAGACGACGAGTCACGGAATCCCGTAATC
CGTAAAGCCCATACCACAGCAAAACGCCCGCAACTTTACCGCAATCGATCTG
TTTATCACGCAACCACAACGCACTGGCGAGCGCCAGCATGGCACCTGCGGAA
TCACCGGCAAAGCCAATGCGGGACATATTGATTTGATAATCCTCCGCCTGCTG
GTGGAAATAACAACAAGCAGCCACAATTTCCTCTATCGCTTGCGGAAAACGC
GCTTCAGGTGAAAGGGTGTAATCAATACCAATCACCGTACATTGGCTGTAGC
TTGCCAGCAGGCGCATGATGCGATCGTGGGTATCGAGATTGCCGAGAATAAA
ACCGCCTCCATGCAAATAAAATAGCGTCGCTGGGCTATCTGGCTGCGGACAA
AAGAGACGTGTTTCCACCTGCCCATATTTTGTTGGAACCATGTAAGCTCTGGT
TGCCATTTCTGGAGCGCCCGCATTCCAGAATCGGCGCTCAAGCGTGTAATACT
GTCGTTGCTCAGCAATCGTTCCCGTTGCGGGCCAGGGCGGTAAATCCGGCTG
AAGAGTATTCACAACGGTCTTCATTTCAGCAGAAATAAGGTCCAGAACAGGT
AGTTTGTTTTCCGGCTTCAT

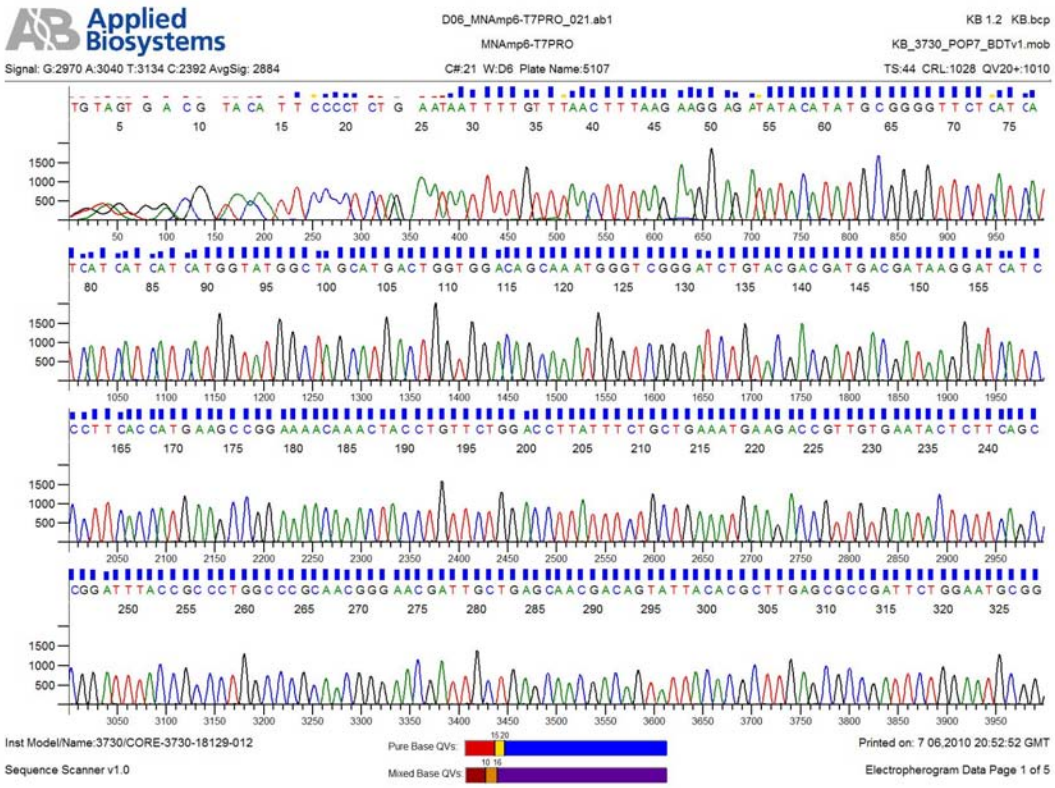
**6.2. Amino acid sequence of Aes from *E. coli*.**

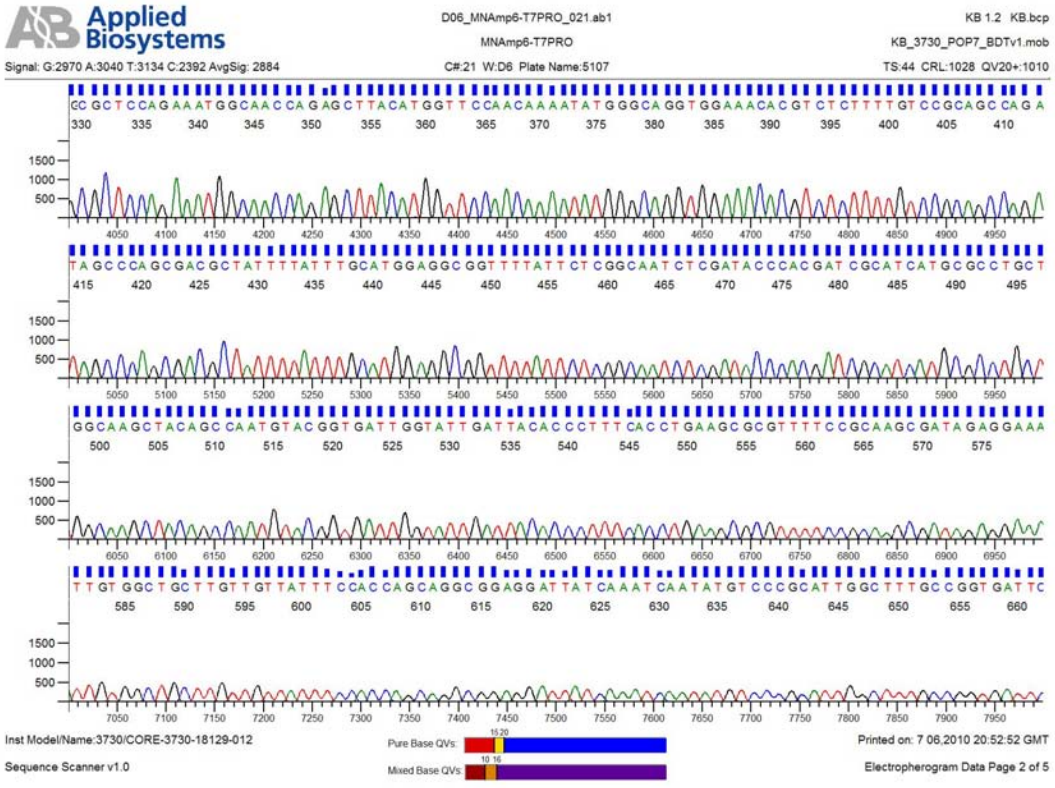
MKPENKLPVLDLISAEMKTVVNTLQPDLPPWPATGTIAEQRQYYTLERRFWNAG
APEMATRAYMVPTKYGQVETRLFCPQPDSPATLFYLHGGGFILGNLDTHDRIMR
LLASYSQCTVIGIDYTLSPEARFPQAIEEIVAACCYFHQQAEDYQINMSRIGFAGD
SAGAMLALASALWLRDKQIDCGKVAGVLLWYGLYGLRDSVTRRLLGGVWDGL
TQQDLQMYEEAYLSNDADRESPYYCLFNNDLTREVPPCFIAGAEFDPLLDDSRLL
YQTLAAHQQPCEFKLYPGTLHAFLHYSRMMKTADEALRDGAQFFTAQL

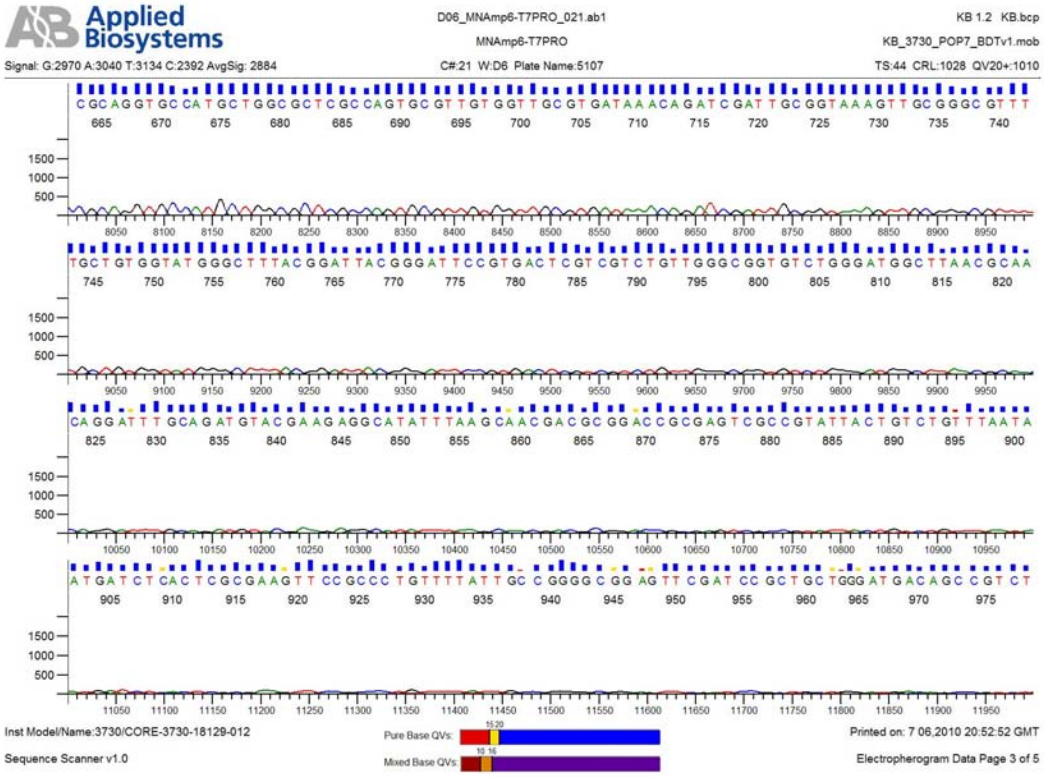
**6.3. The result of DNA sequencing for *aes*.**

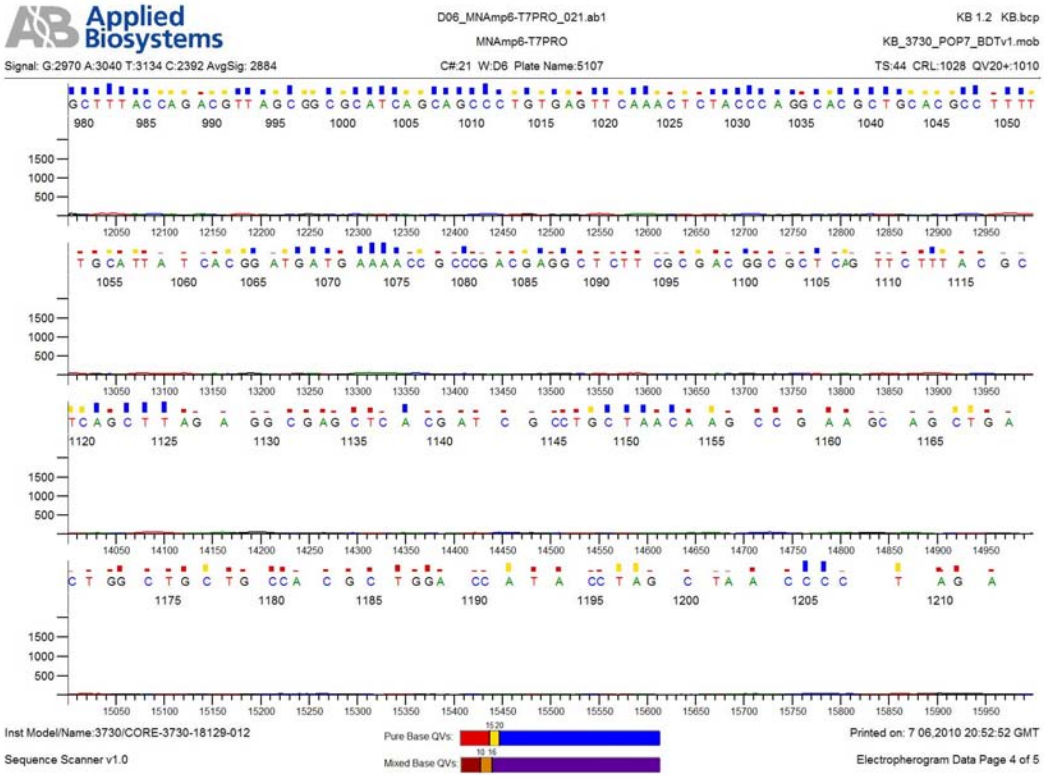
**T7promoter**

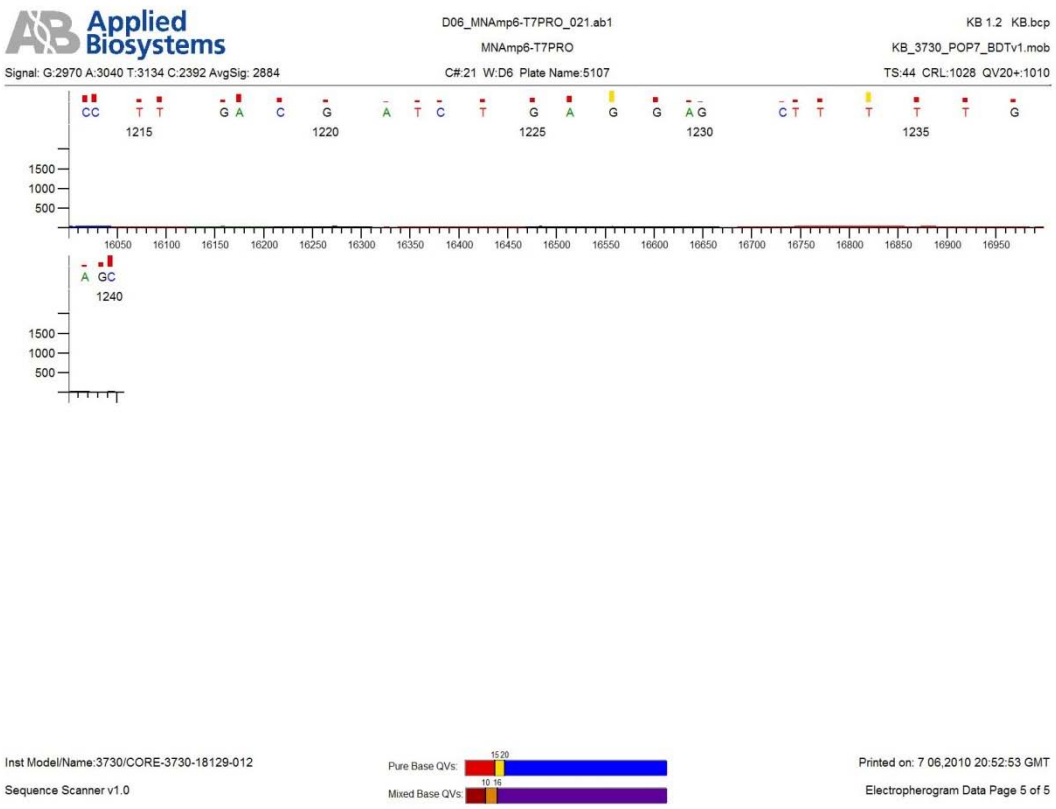
TGTAGTGACGTACATTCCCCTCTGAATAATTTTGTTTAACTTTAAGAAGGAGA
TATACATATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGA
CTGGTGGACAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCA
TCCCTTCACCATGAAGCCGGAAAACAAACTACCTGTTCTGGACCTTATTTCTG
CTGAAATGAAGACCGTTGTGAATACTCTTCAGCCGGATTTACCGCCCTGGCCC
GCAACGGGAACGATTGCTGAGCAACGACAGTATTACACGCTTGAGCGCCGAT
TCTGGAATGCGGGCGCTCCAGAAATGGCAACCAGAGCTTACATGGTTCCAAC
AAAATATGGGCAGGTGGAAACACGTCTCTTTTGTCCGCAGCCAGATAGCCCA
GCGACGCTATTTTATTTGCATGGAGGCGGTTTTATTCTCGGCAATCTCGATAC
CCACGATCGCATCATGCGCCTGCTGGCAAGCTACAGCCAATGTACGGTGATT
GGTATTGATTACACCCTTTCACCTGAAGCGCGTTTTCCGCAAGCGATAGAGGA
AATTGTGGCTGCTTGTTGTTATTTCCACCAGCAGGCGGAGGATTATCAAATCA
ATATGTCCCGCATTGGCTTTGCCGGTGATTCCGCAGGTGCCATGCTGGCGCTC
GCCAGTGCGTTGTGGTTGCGTGATAAACAGATCGATTGCGGTAAAGTTGCGG
GCGTTTTGCTGTGGTATGGGCTTTACGGATTACGGGATTCCGTGACTCGTCGT
CTGTTGGGCGGTGTCTGGGATGGCTTAACGCAACAGGATTTGCAGATGTACG
AAGAGGCATATTTAAGCAACGACGCGGACCGCGAGTCGCCGTATTACTGTCT
GTTTAATAATGATCTCACTCGCGAAGTTCCGCCCTGTTTTATTGCCGGGGCGG
AGTTCGATCCGCTGCTGGGATGACAGCCGTCTGCTTTACCAGACGTTAGCGGC
GCATCAGCAGCCCTGTGAGTTCAAACTCTACCCAGGCACGCTGCACGCCTTTT
TGCATTATCACGGATGATGAAAACCGCCCGACGAGGCTCTTCGCGACGGCGC
TCAGTTCTTTACGCTCAGCTTAGAGGCGAGCTCACGATCGCCTGCTAACAAGC
CGAAGCAGCTGACTGGCTGCTGCCACGCTGGACCATACCTAGCTAACCCCTG
AGACCTTGACGATCTGAGGAGCTTTTTGAGC


Applied Biosystems
D06_MNAmp6-T7PRO_021.ab1
MNAmp6-T7PRO
C#:21 W:D6 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:44 CRL:1028 QV20+:1010
Signal: G:2970 A:3040 T:3134 C:2392 AvgSig: 2884
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: 7 06,2010 20:52:52 GMT
Electropherogram Data Page 1 of 5


Applied Biosystems
D06_MNAmp6-T7PRO_021.ab1
MNAmp6-T7PRO
C#:21 W:D6 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:44 CRL:1028 QV20+:1010
Signal: G:2970 A:3040 T:3134 C:2392 AvgSig: 2884
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: 7 06,2010 20:52:52 GMT
Electropherogram Data Page 2 of 5


Applied Biosystems
D06_MNAmp6-T7PRO_021.ab1
MNAmp6-T7PRO
C#:21 W:D6 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:44 CRL:1028 QV20+:1010
Signal: G:2970 A:3040 T:3134 C:2392 AvgSig: 2884
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 20:52:52 GMT
Electropherogram Data Page 3 of 5

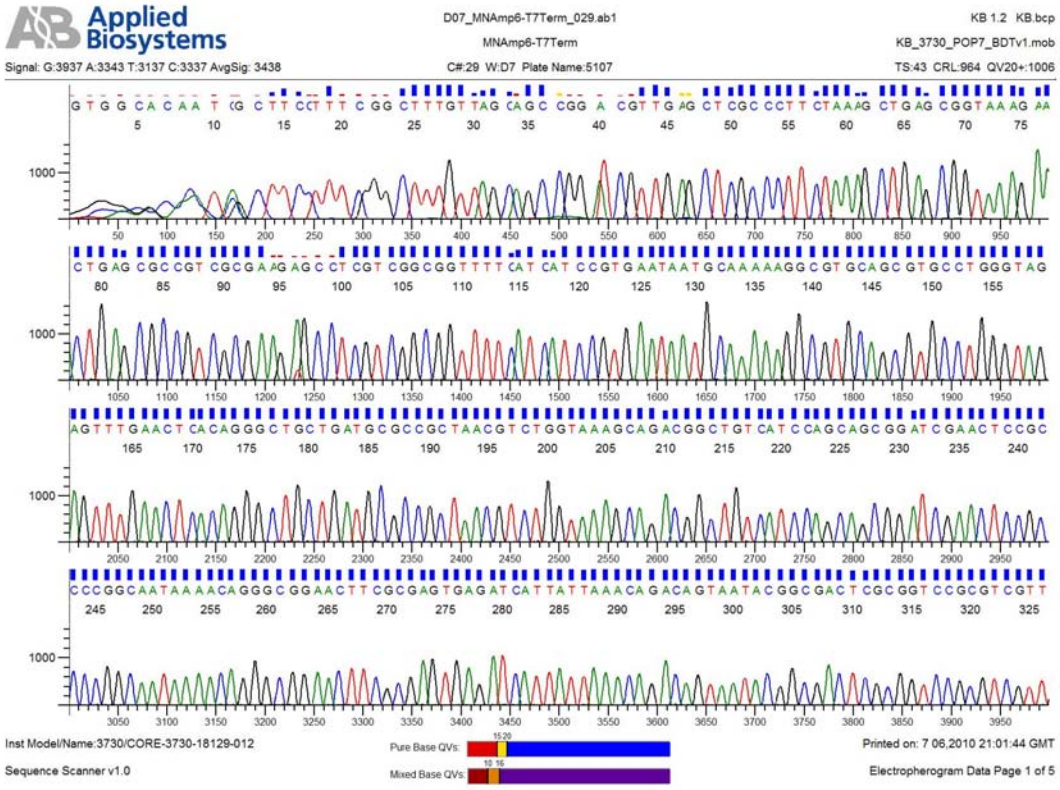

Applied Biosystems
D06_MNAmp6-T7PRO_021.ab1
MNAmp6-T7PRO
C#:21 W:D6 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:44 CRL:1028 QV20+:1010
Signal: G:2970 A:3040 T:3134 C:2392 AvgSig: 2884
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 20:52:52 GMT
Electropherogram Data Page 4 of 5

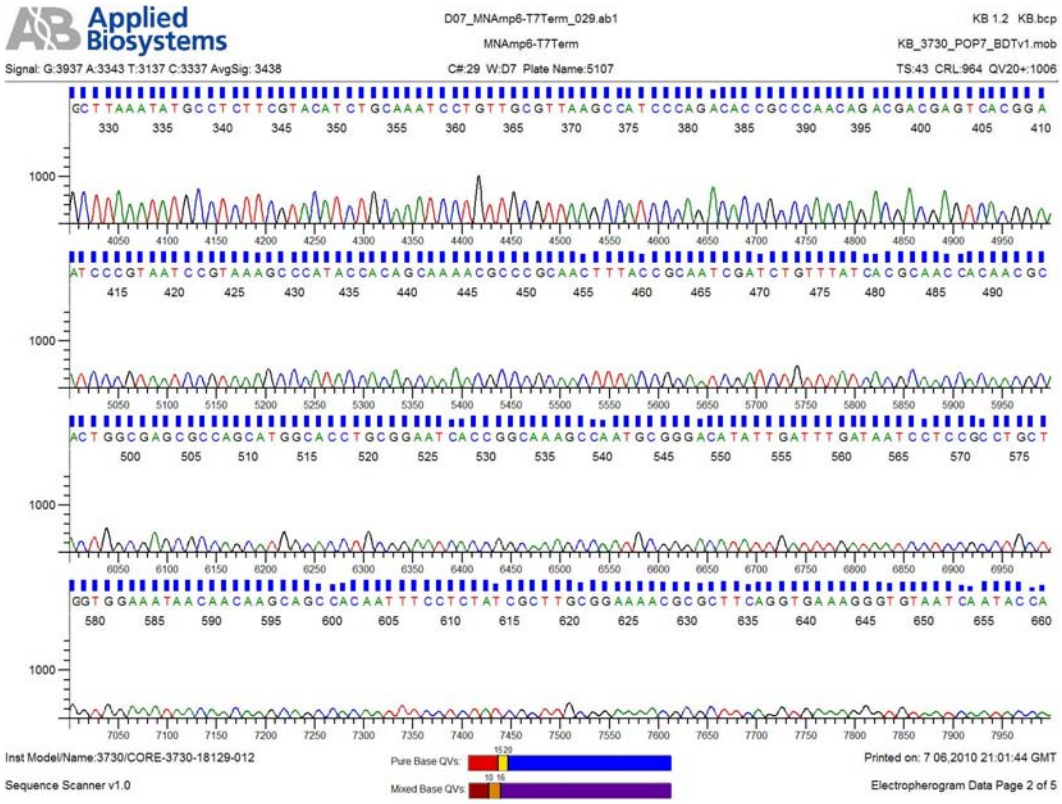


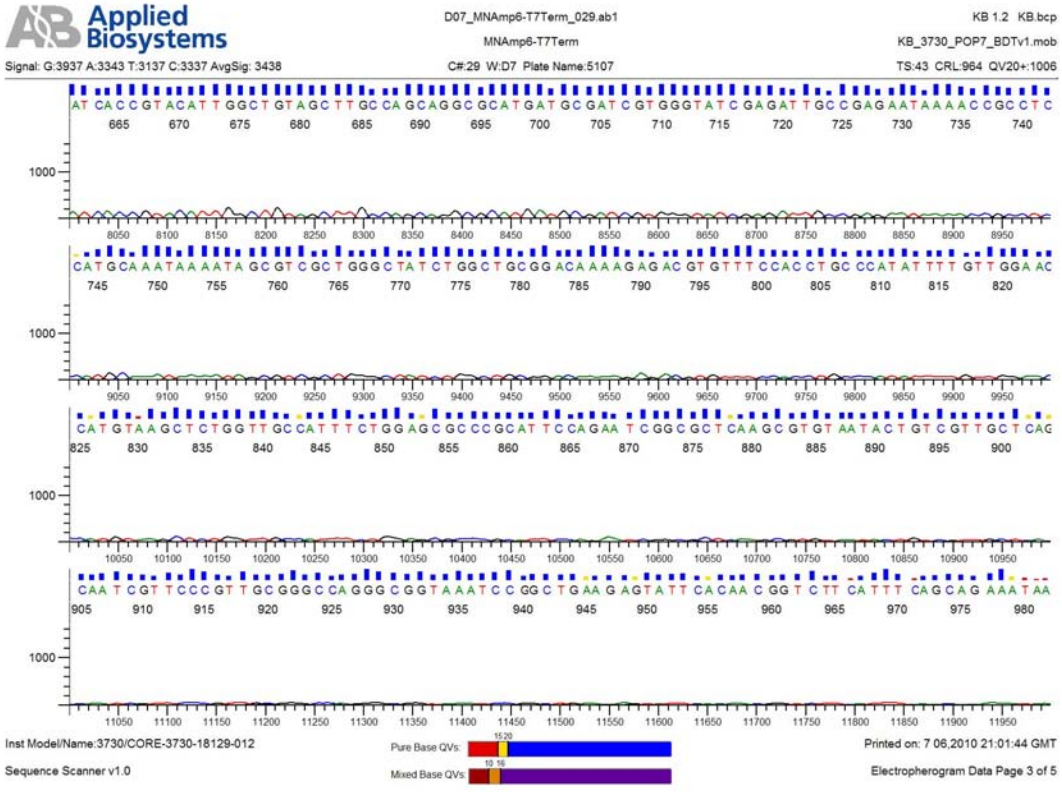

**T7terminator**

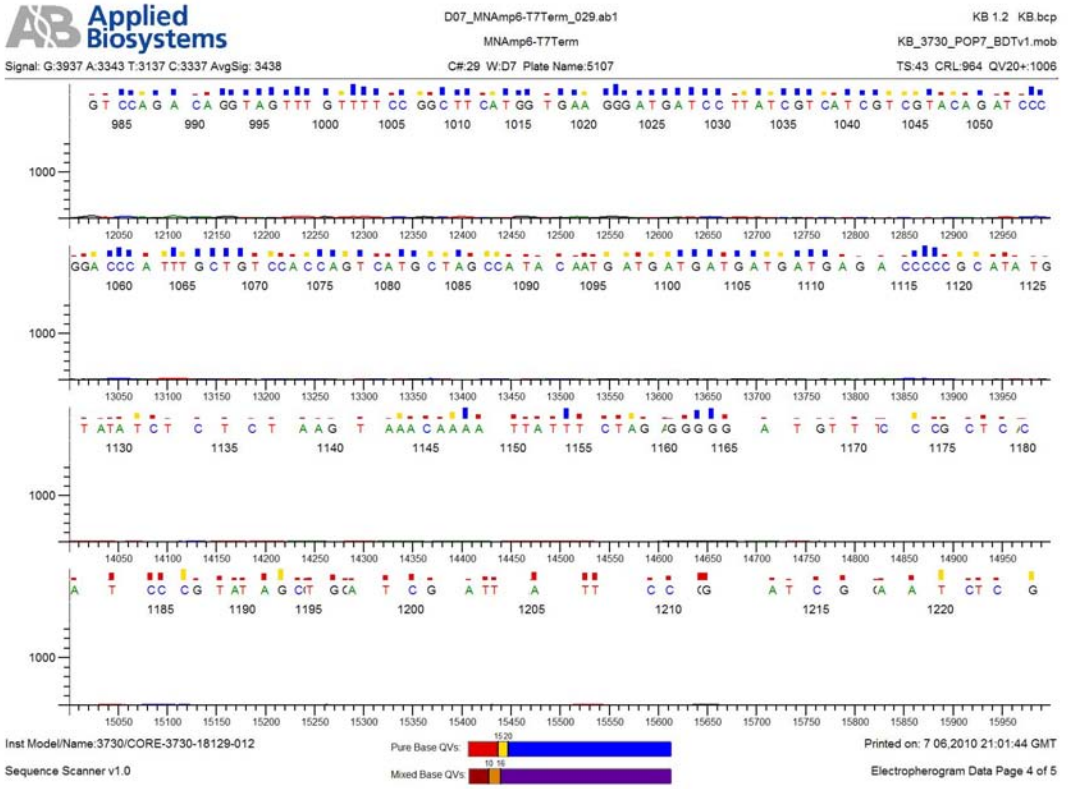
GTGGCACAATCGCTTCCTTTCGGCTTTGTTAGCAGCCGGACGTTGAGCTCGCC
CTTCTAAAGCTGAGCGGTAAAGAACTGAGCGCCGTCGCGAAGAGCCTCGTCG
GCGGTTTTCATCATCCGTGAATAATGCAAAAAGGCGTGCAGCGTGCCTGGGT
AGAGTTTGAACTCACAGGGCTGCTGATGCGCCGCTAACGTCTGGTAAAGCAG
ACGGCTGTCATCCAGCAGCGGATCGAACTCCGCCCCGGCAATAAAACAGGGC
GGAACTTCGCGAGTGAGATCATTATTAAACAGACAGTAATACGGCGACTCGC
GGTCCGCGTCGTTGCTTAAATATGCCTCTTCGTACATCTGCAAATCCTGTTGC
GTTAAGCCATCCCAGACACCGCCCAACAGACGACGAGTCACGGAATCCCGTA
ATCCGTAAAGCCCATACCACAGCAAAACGCCCGCAACTTTACCGCAATCGAT
CTGTTTATCACGCAACCACAACGCACTGGCGAGCGCCAGCATGGCACCTGCG
GAATCACCGGCAAAGCCAATGCGGGACATATTGATTTGATAATCCTCCGCCT
GCTGGTGGAAATAACAACAAGCAGCCACAATTTCCTCTATCGCTTGCGGAAA
ACGCGCTTCAGGTGAAAGGGTGTAATCAATACCAATCACCGTACATTGGCTG
TAGCTTGCCAGCAGGCGCATGATGCGATCGTGGGTATCGAGATTGCCGAGAA
TAAAACCGCCTCCATGCAAATAAAATAGCGTCGCTGGGCTATCTGGCTGCGG
ACAAAAGAGACGTGTTTCCACCTGCCCATATTTTGTTGGAACCATGTAAGCTC
TGGTTGCCATTTCTGGAGCGCCCGCATTCCAGAATCGGCGCTCAAGCGTGTAA
TACTGTCGTTGCTCAGCAATCGTTCCCGTTGCGGGCCAGGGCGGTAAATCCGG
CTGAAGAGTATTCACAACGGTCTTCATTTCAGCAGAAATAAGTCCAGACAGG
TAGTTTGTTTTCCGGCTTCATGGTGAAGGGATGATCCTTATCGTCATCGTCGT
ACAGATCCCGGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACAATGAT

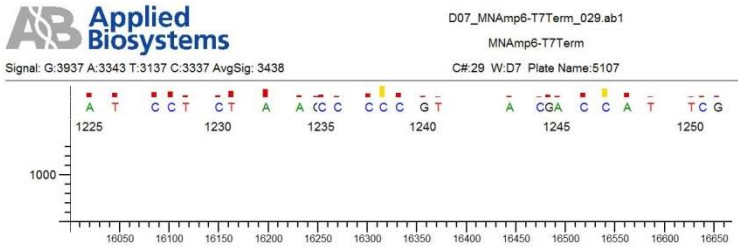
GATGATGATGATGAGACCCCCGCATATGTATATCTCTCTAAGTAAACAAAAT
TATTTCTAGAGGGGGATGTTTCCCGCTCACATTCCCGTATAGCGTGGATCGAT
TAATTCCGGATCGGAATCTCGATCCTCTAAGCCCCCGTACGACCATTCG




Applied Biosystems
D07_MNAmp6-T7Term_029.ab1
MNAmp6-T7Term
C#:29 W:D7 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:43 CRL:964 QV20+:1006
Signal: G:3937 A:3343 T:3137 C:3337 AvgSig: 3438
GCTTAAATATGCCTCTTCGTACATCTGCAAATCCTGTTGCGTTAAGCCATCCCAGACACCGCCCAACAGACGACGAGTCACGGA
ATCCCGTAATCCGTAAAGCCCATACCACAGCAAAACGCCCGCAACTTTACCGCAATCGATCTGTTTATCACGCAACCACAACGC
ACTGGCGAGCGCCAGCATGGCACCTGCGGAATCACCGGCAAAGCCAATGCGGGACATATTGATTTGATAATCCTCCGCCTGCT
GGTGGAAATAACAACAAGCAGCCACAATTTCCTCTATCGCTTGCGGAAAACGCGCTTCAGGTGAAAGGGTGTAATCAATACCA
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:01:44 GMT
Electropherogram Data Page 2 of 5


Applied Biosystems
D07_MNAmp6-T7Term_029.ab1
MNAmp6-T7Term
C#:29 W:D7 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:43 CRL:964 QV20+:1006
Signal: G:3937 A:3343 T:3137 C:3337 AvgSig: 3438
ATCACCGTACATTGGCTGTAGCTTGCCAGCAGGCGCATGATGCGATCGTGGGTATCGAGATTGCCGAGAATAAAACCGCCTC
CATGCAAATAAAATAGCGTCGCTGGGCTATCTGGCTGCGGACAAAAGAGACGTGTTTCCACCTGCCCATATTTTGTTGGAAC
CATGTAAGCTCTGGTTGCCATTTCTGGAGCGCCCGCATTCCAGAATCGGCGCTCAAGCGTGTAATACTGTCGTTGCTCAG
CAATCGTTCCCGTTGCGGGCCAGGGCGGTAAATCCGGCTGAAGAGTATTCACAACGGTCTTCATTTCAGCAGAAATAA
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:01:44 GMT
Electropherogram Data Page 3 of 5


Applied Biosystems
D07_MNAmp6-T7Term_029.ab1
MNAmp6-T7Term
C#:29 W:D7 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:43 CRL:964 QV20+:1006
Signal: G:3937 A:3343 T:3137 C:3337 AvgSig: 3438
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:01:44 GMT
Electropherogram Data Page 4 of 5


Applied Biosystems
D07_MNAmp6-T7Term_029.ab1
MNAmp6-T7Term
C#:29 W:D7 Plate Name:5107
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:43 CRL:964 QV20+:1006
Signal: G:3937 A:3343 T:3137 C:3337 AvgSig: 3438


Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:01:44 GMT
Electropherogram Data Page 5 of 5

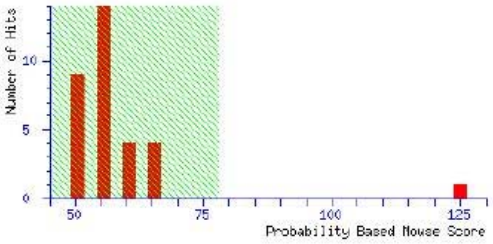
## 6.4. Mass spectometry

### Mascot Search Results

User : janet
Email : jrogers@biochem.okstate.edu
Search title : 571_2b_MN
Database : MSDB 20060831 (3239079 sequences; 1079594700 residues)
Timestamp : 19 Oct 2007 at 16:51:54 GMT
Top Score : 125 for C64778, probable carboxylesterase (EC 3.1.1.1) ybaC - Escherichia coli (strain K-12)

**Probability Based Mowse Score**

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 78 are significant (p<0.05).




**Concise Protein Summary Report**

Format As | Concise Protein Summary | Help

Significance threshold p< 0.05 Max. number of hits AUTO

Re-Search All | Search Unmatched

1. C64778 Mass: 36011 Score: 125 Expect: 1e-06 Queries matched: 14
probable carboxylesterase (EC 3.1.1.1) ybaC - Escherichia coli (strain K-12)
ABB65088 Mass: 36039 Score: 90 Expect: 0.0031 Queries matched: 12

CP000036 NID: - Shigella boydii Sb227
AAZ87239 **Mass:** 36015 **Score:** 90 **Expect:** 0.0031 **Queries matched:** 12
CP000038 NID: - Shigella sonnei Ss046
AES_SHIFL **Mass:** 36014 **Score:** 90 **Expect:** 0.0031 **Queries matched:** 12
Acetyl esterase (EC 3.1.1.-).- Shigella flexneri.
A90695 **Mass:** 36011 **Score:** 87 **Expect:** 0.0062 **Queries matched:** 10
probable lipase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
BAA14305 **Mass:** 31580 **Score:** 79 **Expect:** 0.038 **Queries matched:** 11
ECOADKVIS NID: - Escherichia coli
AAN79074 **Mass:** 36192 **Score:** 73 **Expect:** 0.17 **Queries matched:** 9
AE014075 NID: - Escherichia coli CFT073
Q1ZJ69_9GAMM **Mass:** 10119 **Score:** 52 **Expect:** 21 **Queries matched:** 4
Hypothetical protein.- Psychromonas sp. CNPT3.
Q3EAW1_ARATH **Mass:** 13692 **Score:** 51 **Expect:** 26 **Queries matched:** 5
Protein At3g30360.- Arabidopsis thaliana (Mouse-ear cress).
Q2DTY7_9CHLR **Mass:** 43036 **Score:** 50 **Expect:** 29 **Queries matched:** 6
Phosphoglycerate kinase (EC 2.7.2.3).- Dehalococcoides sp. BAV1.

---

2. Q60E08_ORYSA **Mass:** 199856 **Score:** 68 **Expect:** 0.54 **Queries matched:** 13
Putative polyprotein.- Oryza sativa (japonica cultivar-group).
Q2QPA9_ORYSA **Mass:** 198646 **Score:** 57 **Expect:** 6 **Queries matched:** 12
Retrotransposon protein, putative, Ty3-gypsy subclass.- Oryza sativa (japonica cultivar-group).
Q7XQM2_ORYSA **Mass:** 200379 **Score:** 57 **Expect:** 6.3 **Queries matched:** 12
OSJNBa0089E12.2 protein.- Oryza sativa (Rice).
Q8LLZ2_ORYSA **Mass:** 203060 **Score:** 57 **Expect:** 7.1 **Queries matched:** 12
Putative retroelement (Retrotransposon protein, putative, Ty3-gypsy sub-class).- Oryza sativa (japonica cultivar-group).
Q8LMC1_ORYSA **Mass:** 202939 **Score:** 57 **Expect:** 7.1 **Queries matched:** 12
Putative retroelement (Retrotransposon protein, putative, Ty3-gypsy sub-class).- Oryza sativa (japonica cultivar-group).
Q2QPK3_ORYSA **Mass:** 205152 **Score:** 56 **Expect:** 8.9 **Queries matched:** 12
Retrotransposon protein, putative, Ty3-gypsy subclass.- Oryza sativa (japonica cultivar-group).
Q4TBC7_TETNG **Mass:** 16150 **Score:** 52 **Expect:** 23 **Queries matched:** 5
Chromosome undetermined SCAF7151, whole genome shotgun sequence. (Fragment).- Tetraodon nigroviridis (Green puffer).

---

### Search Parameters

**Type of search** : Peptide Mass Fingerprint
**Enzyme** : Trypsin
**Variable modifications** : Gln->pyro-Glu (N-term Q),Oxidation (M),Propionamide (C)
**Mass values** : Monoisotopic
**Protein Mass** : Unrestricted
**Peptide Mass Tolerance** : ± 100 ppm

## MATRIX SCIENCE Mascot Search Results

### Protein View

Match to: C64778 Score: 125 Expect: 1e-06
**probable carboxylesterase (EC 3.1.1.1) ybaC - Escherichia coli (strain K-12)**

Nominal mass ($M_r$): 36011; Calculated pI value: 4.95
NCBI BLAST search of C64778 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Escherichia coli
Links to retrieve other entries containing this sequence from NCBI Entrez:
AAB40230 from Escherichia coli
AAC73578 from Escherichia coli K12
BAE76255 from Escherichia coli W3110
AES_ECOLI from Escherichia coli

Variable modifications: Gln->pyro-Glu (N-term Q),Oxidation (M),Propionamide (C)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 23
Number of mass values matched: 14
Sequence Coverage: 36%

Matched peptides shown in **Bold Red**

```
    1 MKPENKLPVL DLISAEMKTV VNTLQPDLPP WPATGTIAEQ RQYYTLERRF
   51 WNAGAPEMAT RAYMVPTKYG QVETRLPCPQ PDSPATLFYL HGGGFILGNL
  101 DTHDRIMRLL ASYSQCTVIG IDYTLSPEAR FPQAIEEIVA ACCYFHQQAE
  151 DYQINMSRIG FAGDSAGAML ALASALWLRD KQIDCGKVAG VLLWYGLYGL
  201 RDSVTRRLLG GVWDGLTQQD LQMYEEAYLS NDADRESPYY CLFNNDLTRE
  251 VPPCFIAGAE FDPLLDDSRL LYQTLAAHQQ PCEFKLYPGT LHAFLHYSRM
  301 MKTADEALRD GAQFFTAQL
```

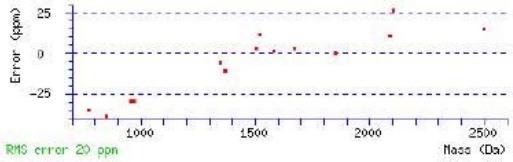
Show predicted peptides also

Sort Peptides By (•) Residue Number ( ) Increasing Mass ( ) Decreasing Mass

| Start - End | Observed | Mr(expt) | Mr(calc) | ppm | Miss | Sequence |
|---|---|---|---|---|---|---|
| 19 - 41 | 2504.3524 | 2503.3451 | 2503.3071 | 15 | 0 | K.TVVNTLQPDLPPWPATGTIAEQR.Q |
| 42 - 48 | 955.4240 | 954.4167 | 954.4447 | -29 | 0 | R.QYYTLER.R Gln->pyro-Glu (N-term Q) |
| 42 - 48 | 972.4500 | 971.4427 | 971.4712 | -29 | 0 | R.QYYTLER.R |
| 49 - 61 | 1506.7317 | 1505.7244 | 1505.7197 | 3 | 1 | R.RFWNAGAPEMATR.A |
| 49 - 61 | 1522.7400 | 1521.7327 | 1521.7147 | 12 | 1 | R.RFWNAGAPEMATR.A Oxidation (M) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 50 - 61 | 1350.6178 | 1349.6105 | 1349.6186 | -6 | 0 | R.FWHAGAPEMATR.A |
| 50 - 61 | 1366.6062 | 1365.5999 | 1365.6136 | -11 | 0 | R.FWHAGAPEMATR.A Oxidation (M) |
| 69 - 75 | 852.3876 | 851.3803 | 851.4137 | -39 | 0 | K.YGQVETR.L |
| 159 - 179 | 2091.1276 | 2090.1203 | 2090.0983 | 11 | 0 | R.IGFAGDSAGAMLALASALWLR.D |
| 159 - 179 | 2107.1570 | 2106.1497 | 2106.0932 | 27 | 0 | R.IGFAGDSAGAMLALASALWLR.D Oxidation (M) |
| 180 - 201 | 1579.9017 | 1578.8944 | 1578.8922 | 1 | 0 | K.VAGVLLWYGLYGLR.D |
| 286 - 299 | 1674.8794 | 1673.8721 | 1673.8678 | 3 | 0 | K.LYPGTLHAFLHYSR.M |
| 303 - 309 | 775.3670 | 774.3597 | 774.3872 | -35 | 0 | K.TADEALR.D |
| 303 - 319 | 1853.9030 | 1852.8957 | 1852.8955 | 0 | 1 | K.TADEALRDGAQFFTAQL.- |

**No match to:** 915.3870, 943.4690, 1207.4784, 1302.6248, 1305.7621, 1461.9479, 1611.9067, 2046.0906, 2520.3204




```
>P1;C64778
probable carboxylesterase (EC 3.1.1.1) ybaC - Escherichia coli (strain K-12)
C;Species C64778: Escherichia coli
C;Species AAB40230: Escherichia coli
C;Species AAC73578: Escherichia coli K12
C;Species BAE76255: Escherichia coli W3110
C;Species AES_ECOLI: Escherichia coli.
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: C64778; JU0313
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C64778
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-319
A;Cross-references: UNIPROT:P23872; UNIPARC:UPI000012564B; GB:AE000153; GB:U00096; NID:g1786671; PIDN:AAC73578.1; PID:g1786682;
A;Experimental source: strain K-12, substrain MG1655
R;Miyamoto, K.; Nakahigashi, K.; Nishimura, K.; Inokuchi, H.
submitted to JIPID, December 1990
A;Reference number: JU0312
A;Accession: JU0313
A;Molecule type: DNA

A;Residues: 1-166,'VPCWRSPVRCGCVINRSIAVKLRAFCCGMGFTDYGIP'
A;Cross-references: UNIPARC:UPI0000178D7C
A;Experimental source: strain K12, substrain CA274
C;Genetics:
A;Gene: ybaC
A;Map position: 11 min
C;Superfamily: probable lipolytic protein ybaC
C;Keywords: carboxylic ester hydrolase
F;91,165/Active site: His, Ser #status predicted
C;SRCDB PIR2
C;IDN_GENBANK AAB40230; AAC73578; BAE76255;
C;IDN_SWISSPROT AES_ECOLI;
```

Mascot: http://www.matrixscience.com/

## 6.5. DNA sequence of DT1 from *E. coli*.

ATGCTGATTCCGTCAAAACTAAGTCGTCCGGTTCGACTCGACCATACCGTGGT
TCGTGAGCGCCTGCTGGCTAAACTTTCCGGCGCGAACAACTTCCGGCTGGCG
CTGATCACGAGTCCTGCGGGCTACGGAAAGACCACCCTCATTTCCCAGTGGG
CGGCAGGCAAAAACGATATCGGCTGGTACTCGCTGGATGAAGGTGATAACCA
GCAAGAGCGTTTCGCCAGCTATCTCATTGCCGCCGTGCAGCAGGCAACCAAC
GGTCACTGTGCGATATGTGAGACGATGGCGCAAAAACGGCAATATGCCAGCC
TGACGTCACTCTTCGCCCAGCTTTTCATTGAGCTGGCGGAATGGCATAGCCCA
CTTTATCTGGTCATCGATGACTATCATCTGATCACTAATCCAGTGATCCACGA
GTCAATGCGCTTCTTTATTCGCCATCAACCAGAAAATCTCACCCTGGTGGTGT
TGTCACGCAACCTTCCGCAACTGGGCATTGCCAATCTGCGTGTTCGTGATCAA
CTGCTGGAAATTGGCAGTCAGCAACTGGCATTTACCCATCAGGAAGCGAAGC
AGTTTTTTGATTGCCGTCTGTCATCGCCGATTGAAGCCGCAGAAAGCAGTCGG

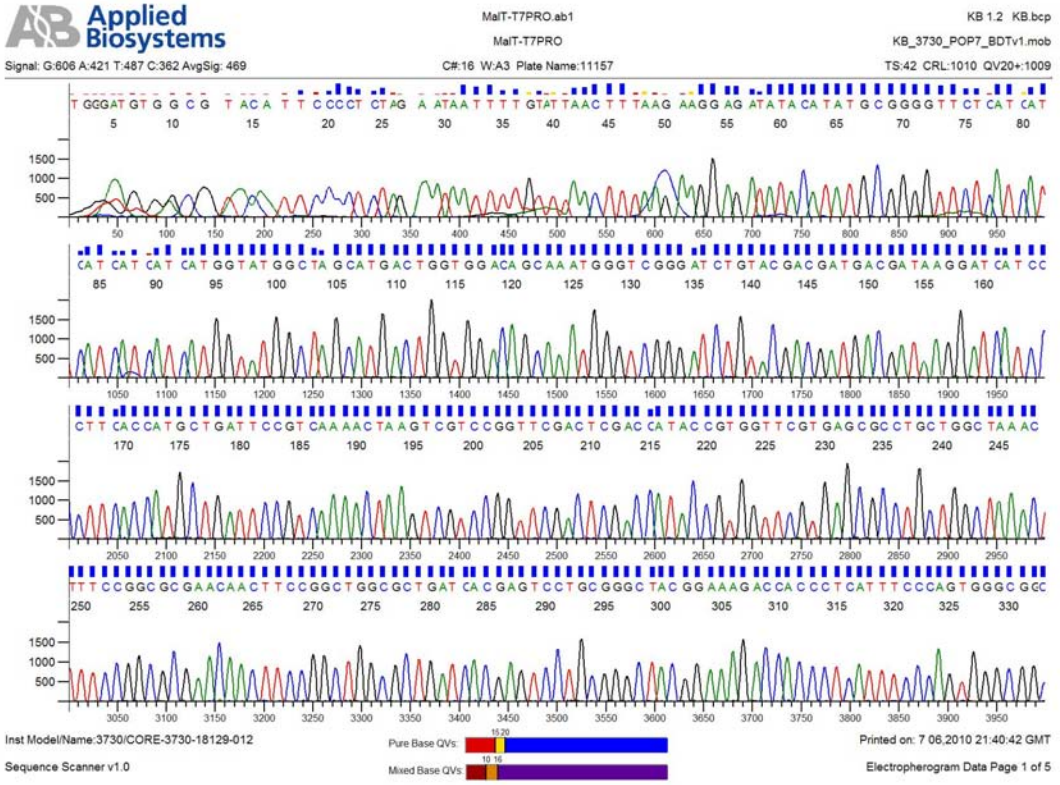
ATTTGCGATGACGTTTCCGGTTGGGCGACGGCACTACAGCTAATCGCCCTCTC
CGCCCGGCAGAATACCCACTCAGCCCATAAGTCGGCACGC

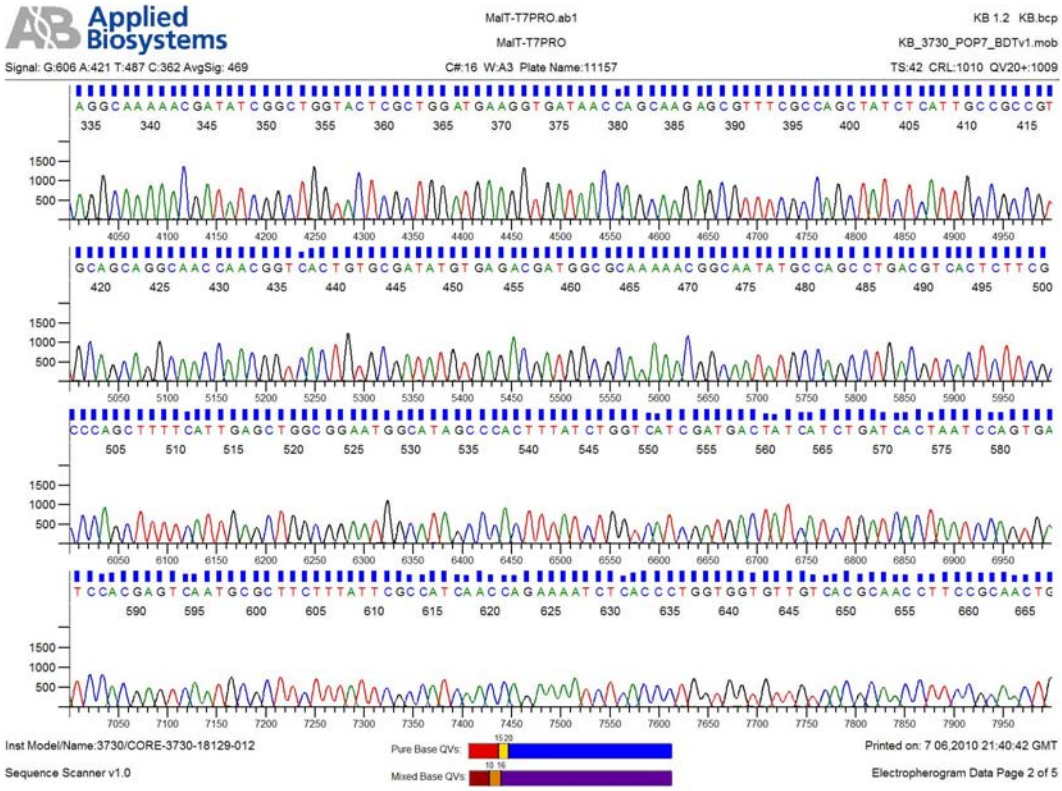
### 6.6. Amino acid sequence of DT1 from *E. coli*.

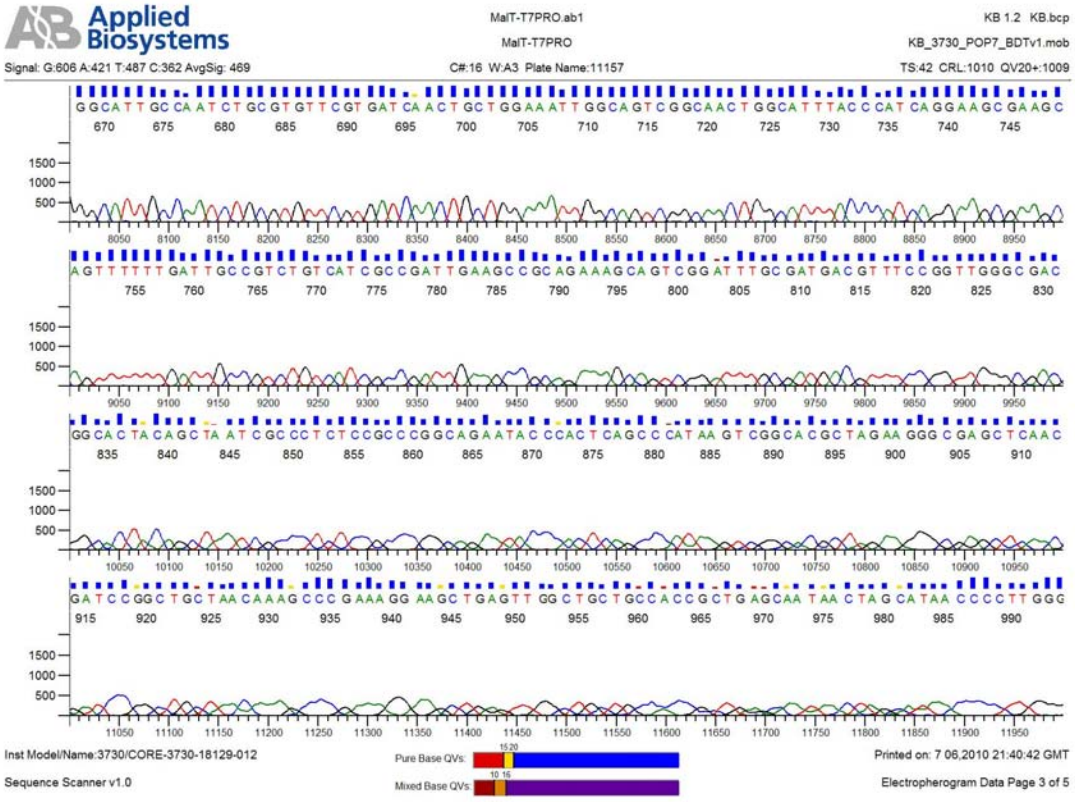
MLIPSKLSRPVRLDHTVVRERLLAKLSGANNFRLALITSPAGYGKTTLISQWAAG
KNDIGWYSLDEGDNQQERFASYLIAAVQQATNGHCAICETMAQKRQYASLTSLF
AQLFIELAEWHSPLYLVIDDYHLITNPVIHESMRFFIRHQPENLTLVVLSRNLPQLG
IANLRVRDQLLEIGSQQLAFTHQEAKQFFDCRLSSPIEAAESSRICDDVSGWATAL
QLIALSARQNTHSAHKSAR

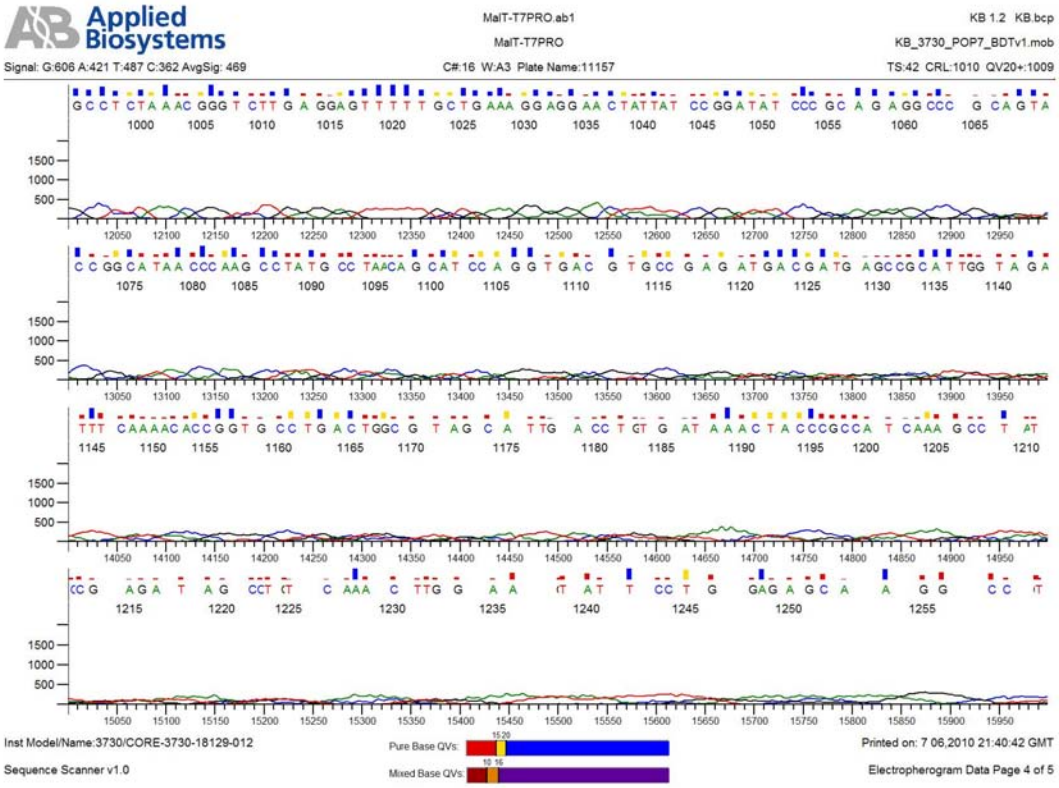
### 6.7. The result of DNA sequencing of DT1.

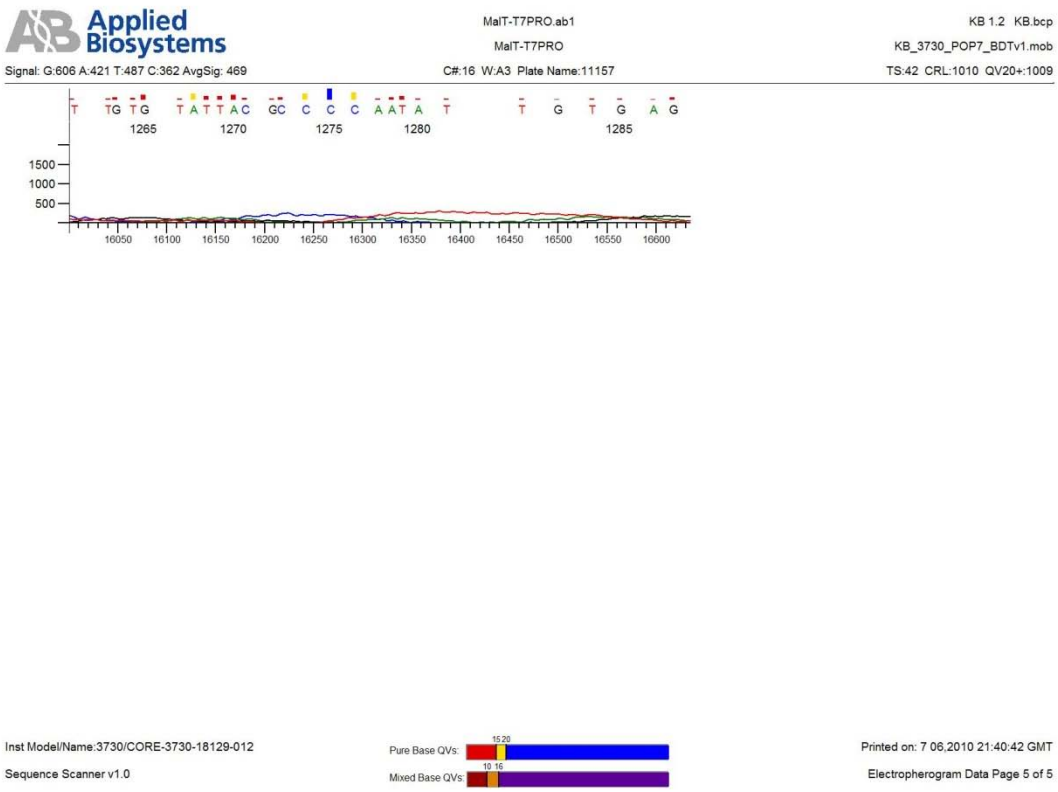
**T7 promoter**
TGGGATGTGGCGTACATTCCCCTCTAGAATAATTTTGTATTAACTTTAAGAAG
GAGATATACATATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGC
ATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGG
ATCATCCCTTCACCATGCTGATTCCGTCAAAACTAAGTCGTCCGGTTCGACTC
GACCATACCGTGGTTCGTGAGCGCCTGCTGGCTAAACTTTCCGGCGCGAACA
ACTTCCGGCTGGCGCTGATCACGAGTCCTGCGGGCTACGGAAAGACCACCCT
CATTTCCCAGTGGGCGGCAGGCAAAAACGATATCGGCTGGTACTCGCTGGAT
GAAGGTGATAACCAGCAAGAGCGTTTCGCCAGCTATCTCATTGCCGCCGTGC
AGCAGGCAACCAACGGTCACTGTGCGATATGTGAGACGATGGCGCAAAAAC
GGCAATATGCCAGCCTGACGTCACTCTTCGCCCAGCTTTTCATTGAGCTGGCG
GAATGGCATAGCCCACTTTATCTGGTCATCGATGACTATCATCTGATCACTAA
TCCAGTGATCCACGAGTCAATGCGCTTCTTTATTCGCCATCAACCAGAAAATC
TCACCCTGGTGGTGTTGTCACGCAACCTTCCGCAACTGGGCATTGCCAATCTG
CGTGTTCGTGATCAACTGCTGGAAATTGGCAGTCGGCAACTGGCATTTACCCA
TCAGGAAGCGAAGCAGTTTTTTGATTGCCGTCTGTCATCGCCGATTGAAGCCG
CAGAAAGCAGTCGGATTTGCGATGACGTTTCCGGTTGGGCGACGGCACTACA
GCTAATCGCCCTCTCCGCCCGGCAGAATACCCACTCAGCCCATAAGTCGGCA
CGCTAGAAGGGCGAGCTCAACGATCCGGCTGCTAACAAAGCCCGAAAGGAA
GCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG
CCTCTAAACGGGTCTTGAGGAGTTTTTGCTGAAAGGAGGAACTATTATCCGG
ATATCCCGCAGAGGCCCGCAGTACCGGCATAACCCAAGCCTATGCCTAACAG
CATCCAGGTGACGTGCCGAGATGACGATGAGCCGCATTGGTAGATTTCAAAA
CACCGGTGCCTGACTGGCGTAGCATTGACCTGTGATAAACTACCCGCCATCA
AAGCCTATCCGAGATAGCCTGTCAAACTTGGAAAGTATTCCTGGAGAGCAAG
GCCCTTTGTGTATTACGCCCCAATATTGTGAG


Applied Biosystems
MalT-T7PRO.ab1
MalT-T7PRO
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:606 A:421 T:487 C:362 AvgSig: 469
C#:16 W:A3 Plate Name:11157
TS:42 CRL:1010 QV20+:1009
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:40:42 GMT
Electropherogram Data Page 1 of 5


Applied Biosystems
MalT-T7PRO.ab1
MalT-T7PRO
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:606 A:421 T:487 C:362 AvgSig: 469
C#:16 W:A3 Plate Name:11157
TS:42 CRL:1010 QV20+:1009
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:40:42 GMT
Electropherogram Data Page 2 of 5


Applied Biosystems
MalT-T7PRO.ab1
MalT-T7PRO
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:606 A:421 T:487 C:362 AvgSig: 469
C#:16 W:A3 Plate Name:11157
TS:42 CRL:1010 QV20+:1009
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:40:42 GMT
Electropherogram Data Page 3 of 5

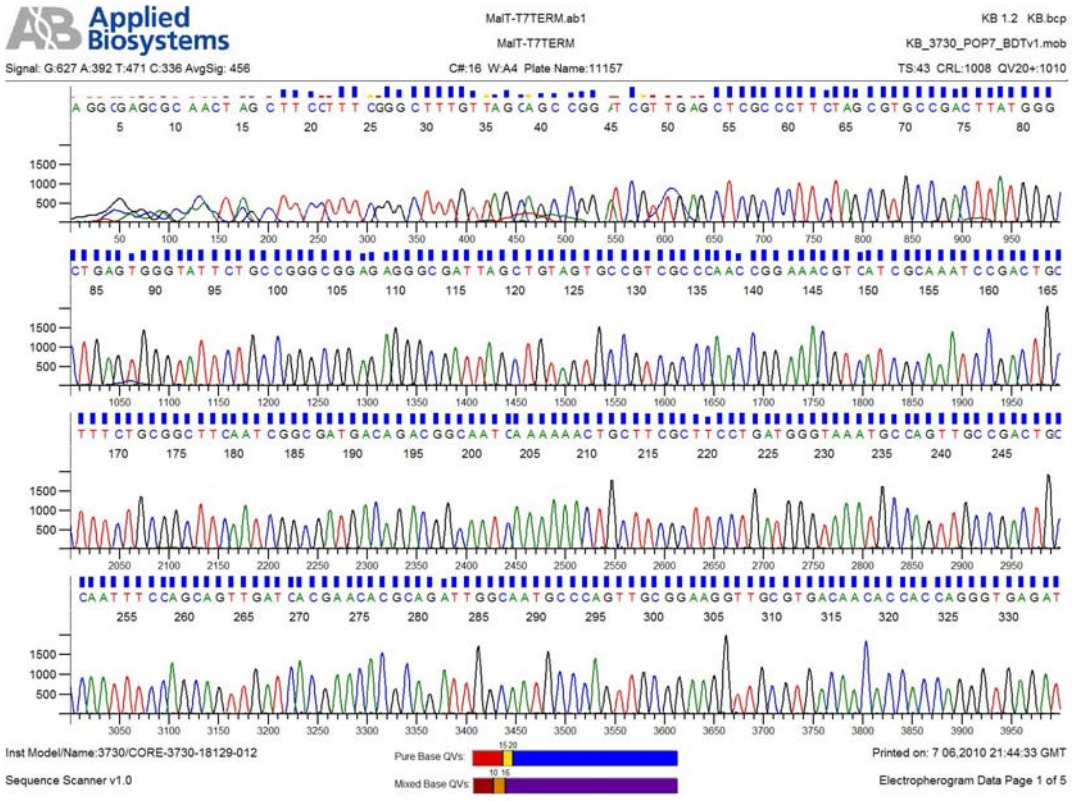

Applied Biosystems
MalT-T7PRO.ab1
MalT-T7PRO
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:606 A:421 T:487 C:362 AvgSig: 469
C#:16 W:A3 Plate Name:11157
TS:42 CRL:1010 QV20+:1009
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:40:42 GMT
Electropherogram Data Page 4 of 5

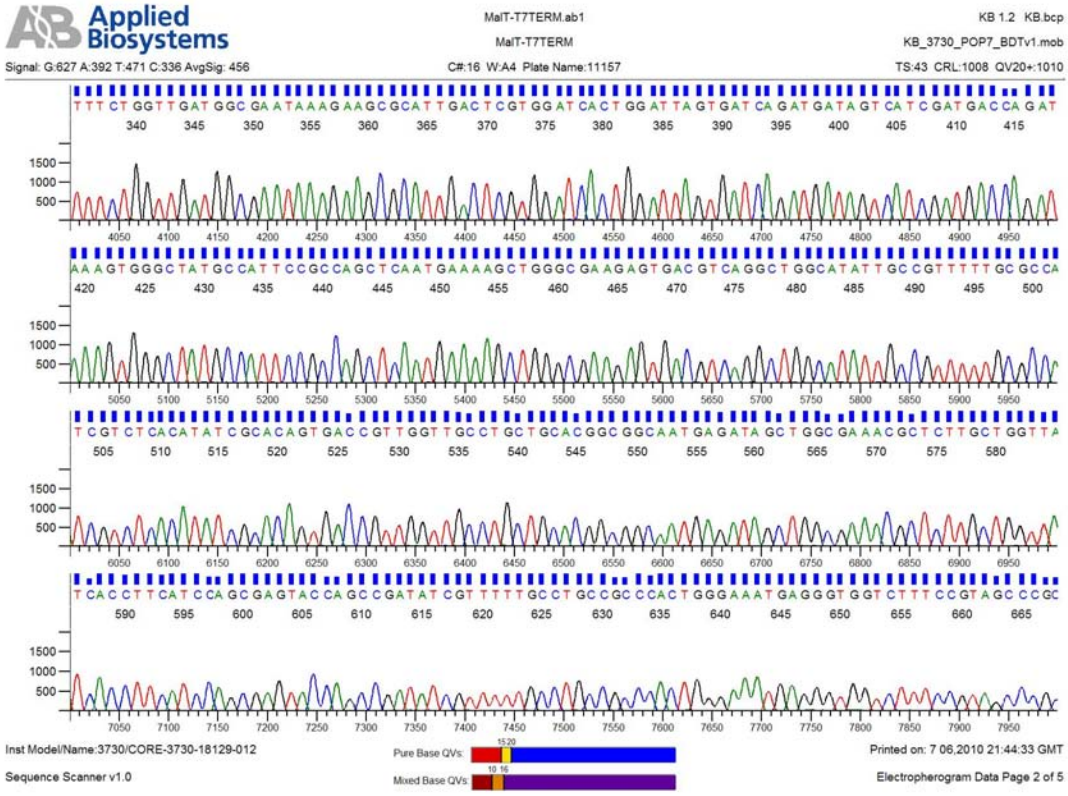


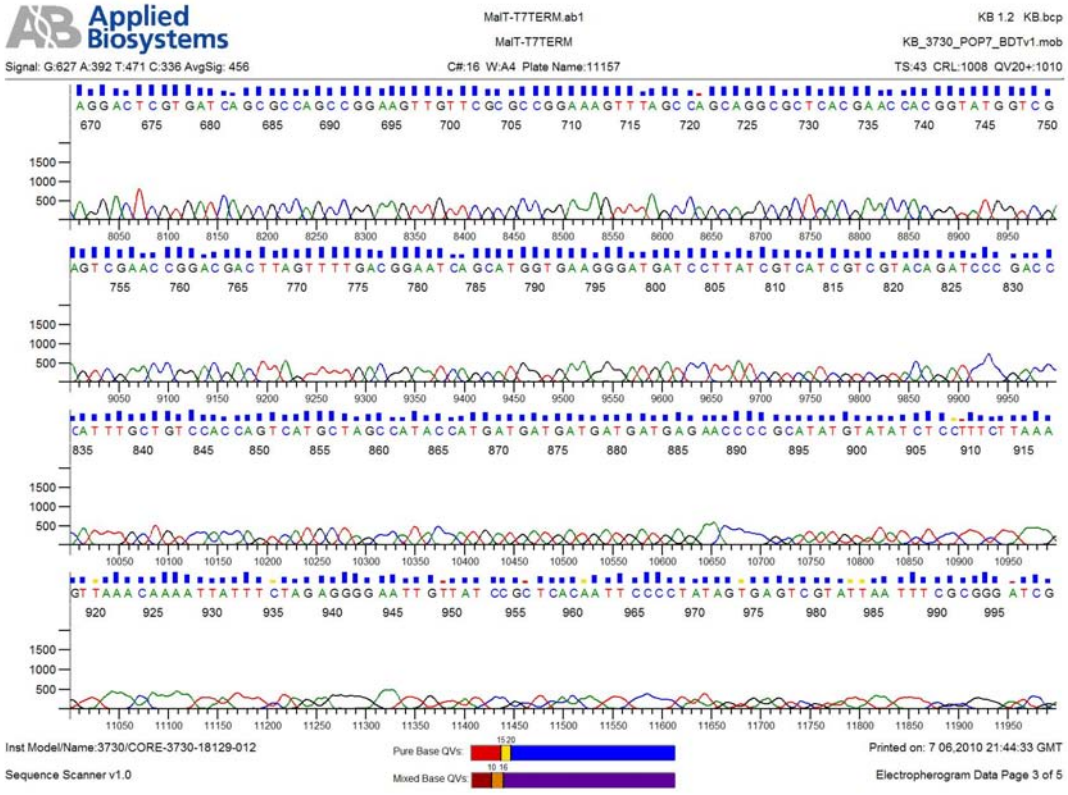

**T7 terminator**

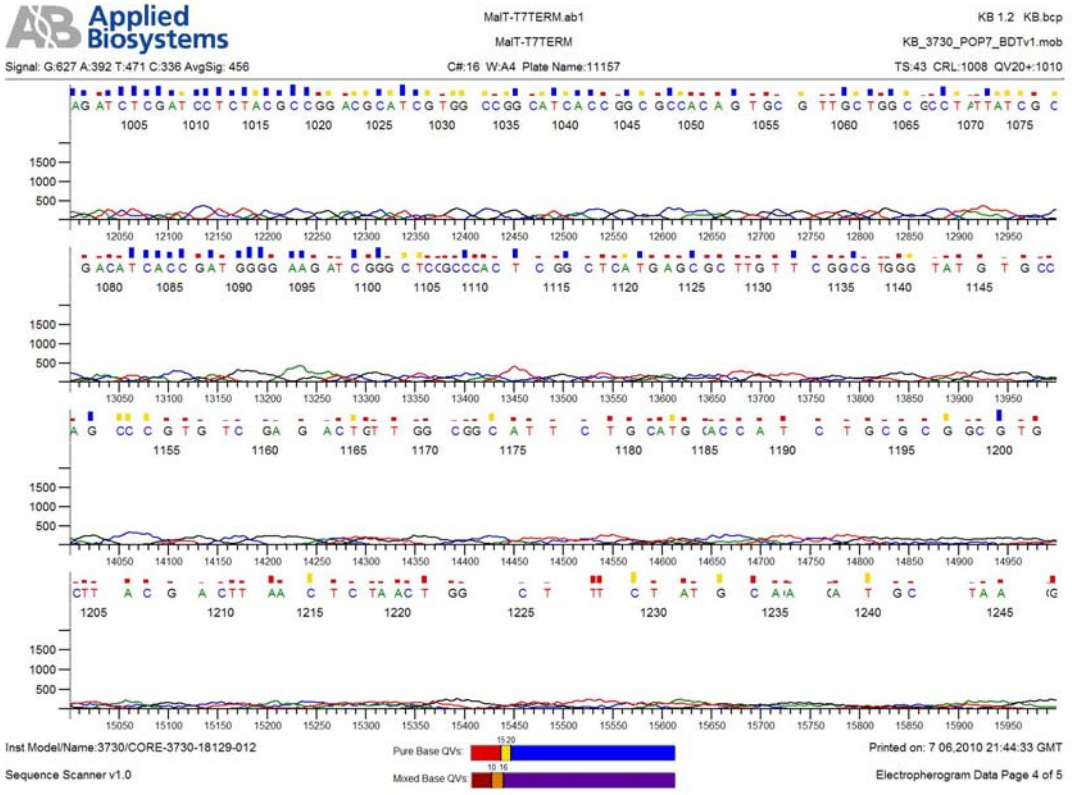
AGGCGAGCGCAACTAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCGTTGAG
CTCGCCCTTCTAGCGTGCCGACTTATGGGCTGAGTGGGTATTCTGCCGGGCGG
AGAGGGCGATTAGCTGTAGTGCCGTCGCCCAACCGGAAACGTCATCGCAAAT
CCGACTGCTTTCTGCGGCTTCAATCGGCGATGACAGACGGCAATCAAAAAAC
TGCTTCGCTTCCTGATGGGTAAATGCCAGTTGCCGACTGCCAATTTCCAGCAG
TTGATCACGAACACGCAGATTGGCAATGCCCAGTTGCGGAAGGTTGCGTGAC
AACACCACCAGGGTGAGATTTTCTGGTTGATGGCGAATAAAGAAGCGCATTG
ACTCGTGGATCACTGGATTAGTGATCAGATGATAGTCATCGATGACCAGATA
AAGTGGGCTATGCCATTCCGCCAGCTCAATGAAAAGCTGGGCGAAGAGTGAC
GTCAGGCTGGCATATTGCCGTTTTTGCGCCATCGTCTCACATATCGCACAGTG
ACCGTTGGTTGCCTGCTGCACGGCGGCAATGAGATAGCTGGCGAAACGCTCT
TGCTGGTTATCACCTTCATCCAGCGAGTACCAGCCGATATCGTTTTTGCCTGC
CGCCCACTGGGAAATGAGGGTGGTCTTTCCGTAGCCCGCAGGACTCGTGATC
AGCGCCAGCCGGAAGTTGTTCGCGCCGGAAAGTTTAGCCAGCAGGCGCTCAC
GAACCACGGTATGGTCGAGTCGAACCGGACGACTTAGTTTTGACGGAATCAG
CATGGTGAAGGGATGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTT
GCTGTCCACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACC
CCGCATATGTATATCTCCTTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGG

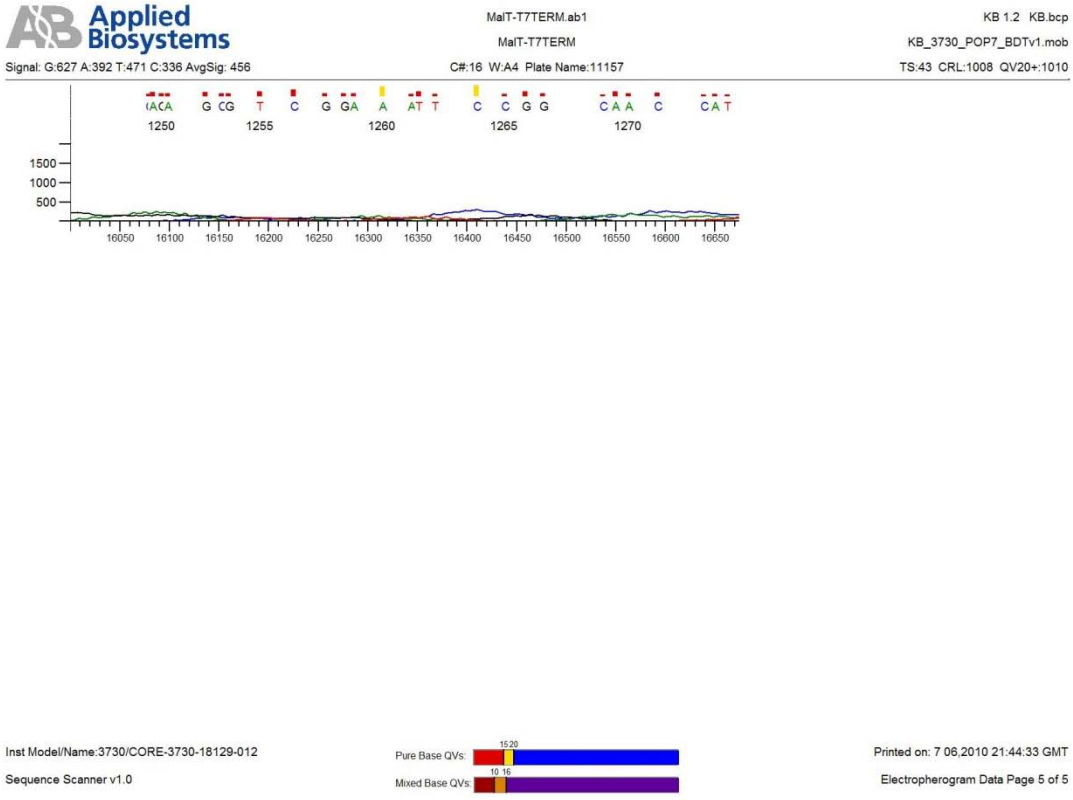
AATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGA
TCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGC
CACAGTGCGTTGCTGGCGCCTATTATCGCGACATCACCGATGGGGAAGATCG
GGCTCCGCCCACTCGGCTCATGAGCGCTTGTTCGGCGTGGGTATGTGCCAGCC
CGTGTCGAGACTGTTGGCGGCATTCTGCATGCACCATCTGCGCGGCGTGCTTA
CGACTTAACTCTAACTGGCTTTCTATGCAGAGATGCTAAGGCAGAGCGTCGG
AAATTCCGGCAACCAT




Applied Biosystems
MalT-T7TERM.ab1
MalT-T7TERM
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:627 A:392 T:471 C:336 AvgSig: 456
C#:16 W:A4 Plate Name:11157
TS:43 CRL:1008 QV20+:1010
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: 7 06,2010 21:44:33 GMT
Electropherogram Data Page 2 of 5


Applied Biosystems
MalT-T7TERM.ab1
MalT-T7TERM
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:627 A:392 T:471 C:336 AvgSig: 456
C#:16 W:A4 Plate Name:11157
TS:43 CRL:1008 QV20+:1010
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: 7 06,2010 21:44:33 GMT
Electropherogram Data Page 3 of 5


Applied Biosystems
MalT-T7TERM.ab1
MalT-T7TERM
C#:16 W:A4 Plate Name:11157
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:43 CRL:1008 QV20+:1010
Signal: G:627 A:392 T:471 C:336 AvgSig: 456
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:44:33 GMT
Electropherogram Data Page 4 of 5


Applied Biosystems
MalT-T7TERM.ab1
MalT-T7TERM
C#:16 W:A4 Plate Name:11157
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
TS:43 CRL:1008 QV20+:1010
Signal: G:627 A:392 T:471 C:336 AvgSig: 456
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:44:33 GMT
Electropherogram Data Page 5 of 5

**6.8. DNA sequence of DT1-DT2 from *E. coli*.**

ATGCTGATTCCGTCAAAACTAAGTCGTCCGGTTCGACTCGACCATACCGTGGT
TCGTGAGCGCCTGCTGGCTAAACTTTCCGGCGCGAACAACTTCCGGCTGGCG
CTGATCACGAGTCCTGCGGGCTACGGAAAGACCACCCTCATTTCCCAGTGGG
CGGCAGGCAAAAACGATATCGGCTGGTACTCGCTGGATGAAGGTGATAACCA
GCAAGAGCGTTTCGCCAGCTATCTCATTGCCGCCGTGCAGCAGGCAACCAAC
GGTCACTGTGCGATATGTGAGACGATGGCGCAAAAACGGCAATATGCCAGCC
TGACGTCACTCTTCGCCCAGCTTTTCATTGAGCTGGCGGAATGGCATAGCCCA
CTTTATCTGGTCATCGATGACTATCATCTGATCACTAATCCAGTGATCCACGA
GTCAATGCGCTTCTTTATTCGCCATCAACCAGAAAATCTCACCCTGGTGGTGT
TGTCACGCAACCTTCCGCAACTGGGCATTGCCAATCTGCGTGTTCGTGATCAA
CTGCTGGAAATTGGCAGTCAGCAACTGGCATTTACCCATCAGGAAGCGAAGC
AGTTTTTTGATTGCCGTCTGTCATCGCCGATTGAAGCCGCAGAAAGCAGTCGG
ATTTGCGATGACGTTTCCGGTTGGGCGACGGCACTACAGCTAATCGCCCTCTC
CGCCCGGCAGAATACCCACTCAGCCCATAAGTCGGCACGCCGCCTGGCGGGA
ATCAATGCCAGCCATCTTTCGGATTATCTGGTCGATGAGGTTTTGGATAACGT
CGATCTCGCAACGCGCCATTTTCTGTTGAAAAGCGCCATTTTGCGCTCAATGA
ACGATGCCCTCATCACCCGTGTGACCGGCGAAGAAAACGGGCAAATGCGCCT
CGAAGAGATTGAGCGTCAGGGGCTGTTTTTACAGCGGATGGATGATACCGGC
GAGTGGTTCTGCTATCACCCGCTGTTTGGTAACTTCCTGCGCCAGCGCTGCCA
GTGGGAACTGGCGGCGGAGCTGCCGGAAATCCACCGTGCCGCCGCAGAAAG
CTGGATGGCCCAGGGATTTCCCAGCGAAGCAATTCATCATGCGCTGGCGGCA
GGCGATGCGCTGATGCTGCGCGATATTCTGCTTAATCACGCCTGGAGTCTGTT
CAACCATAGCGAACTGTCGCTGCTGGAAGAGTCGCTTAAGGCCCTGCCGTGG
GACAGCTTGCTGGAAAATCCGCAGTTGGTGTTATTGCAGGCGTGGCTGATGC
AAAGCCAACATCGCTACGGCGAAGTTAACACCCTGCTAGCCCGTGCTGAA

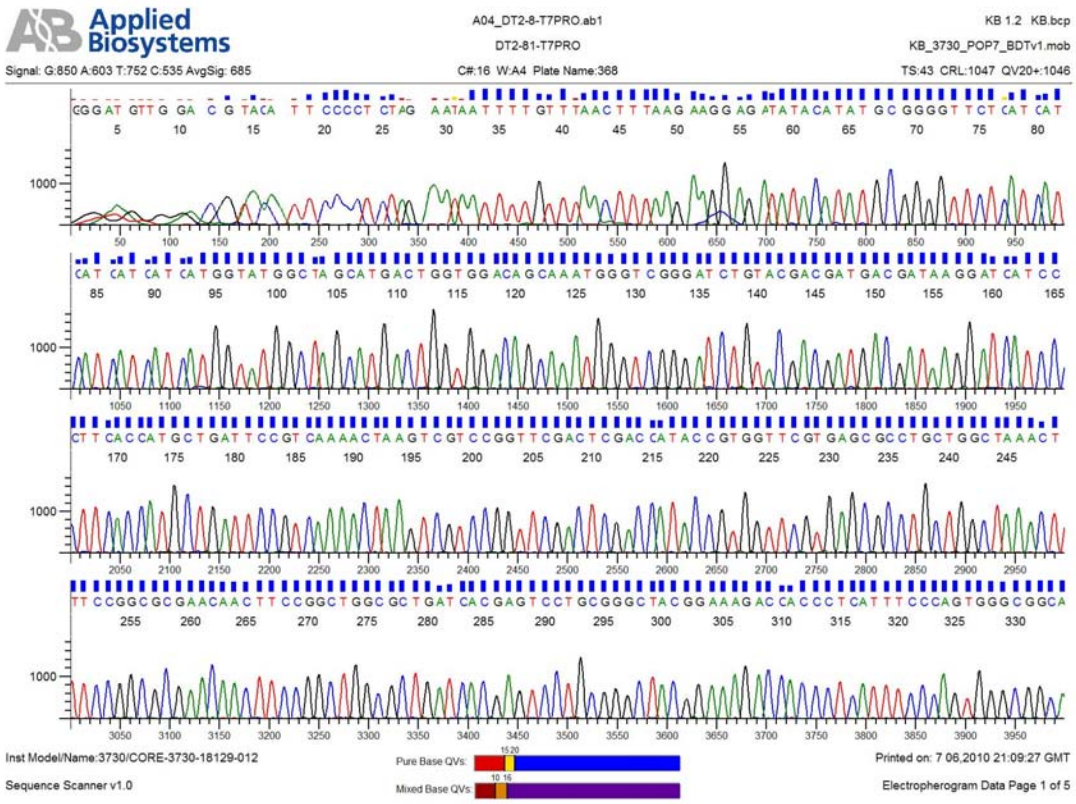
**6.9. Amino acid sequence of DT1-DT2 from *E. coli*.**

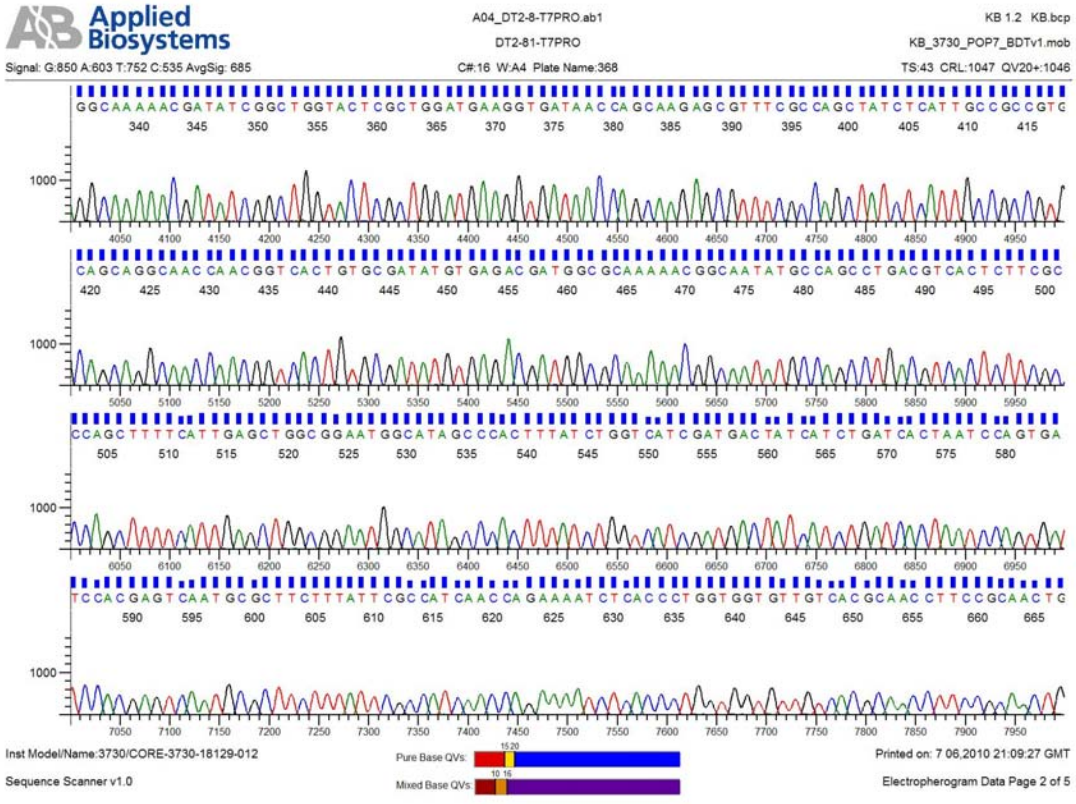
MLIPSKLSRPVRLDHTVVRERLLAKLSGANNFRLALITSPAGYGKTTLISQWAAG
KNDIGWYSLDEGDNQQERFASYLIAAVQQATNGHCAICETMAQKRQYASLTSLF
AQLFIELAEWHSPLYLVIDDYHLITNPVIHESMRFFIRHQPENLTLVVLSRNLPQLG
IANLRVRDQLLEIGSQQLAFTHQEAKQFFDCRLSSPIEAAESSRICDDVSGWATAL
QLIALSARQNTHSAHKSARRLAGINASHLSDYLVDEVLDNVDLATRHFLLKSAIL
RSMNDALITRVTGEENGQMRLEEIERQGLFLQRMDDTGEWFCYHPLFGNFLRQR
CQWELAAELPEIHRAAAESWMAQGFPSEAIHHALAAGDALMLRDILLNHAWSL
FNHSELSLLEESLKALPWDSLLENPQLVLLQAWLMQSQHRYGEVNTLLARAE

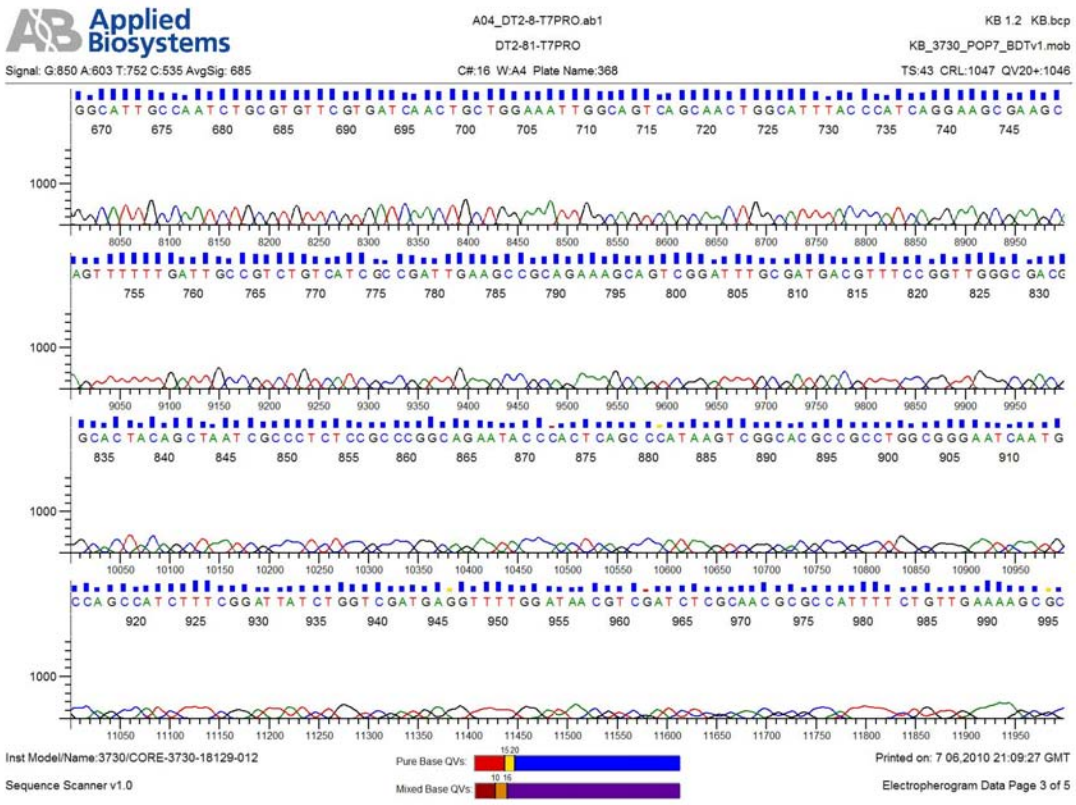
**6.10. The result of DNA sequencing of DT1-DT2.**

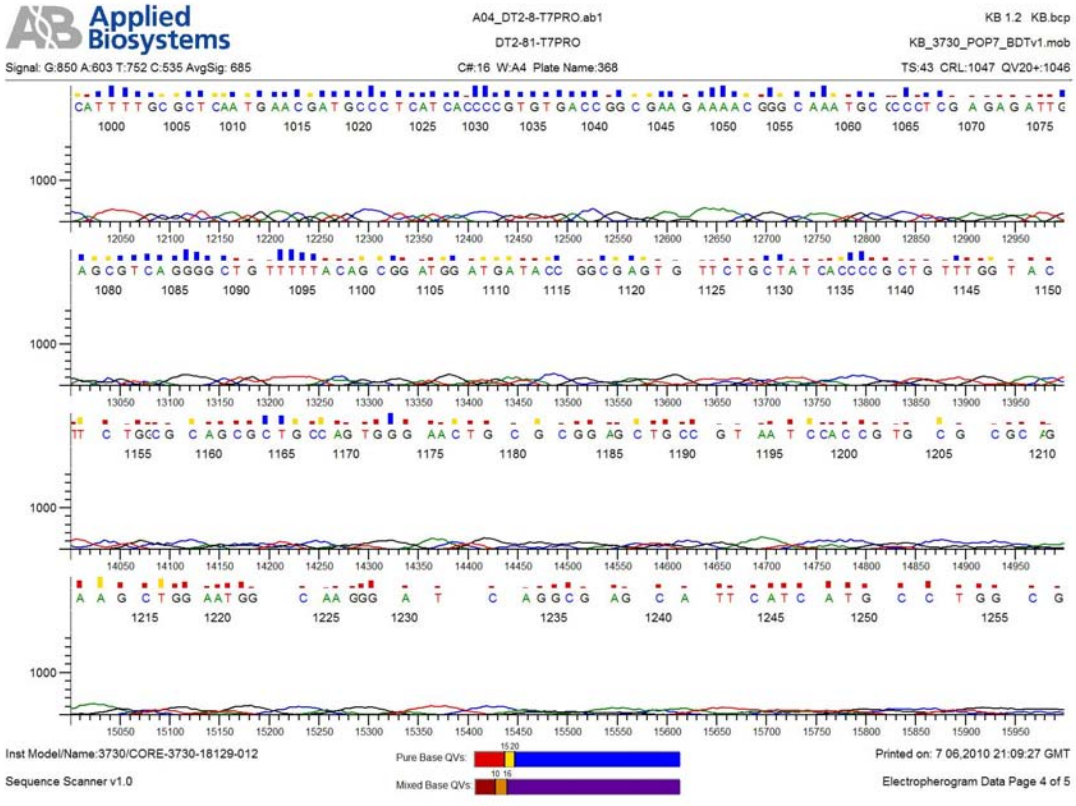
**T7 promoter**

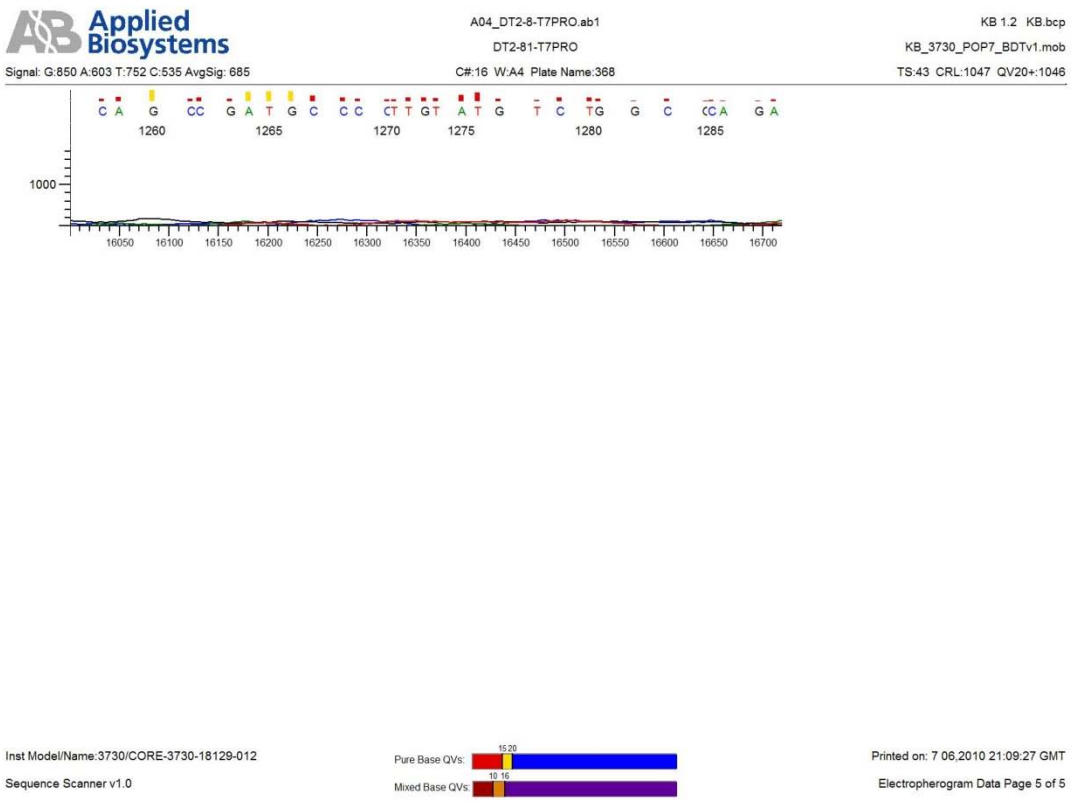
GGGATGTTGGACGTACATTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAG
GAGATATACATATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGC
ATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGG
ATCATCCCTTCACCATGCTGATTCCGTCAAAACTAAGTCGTCCGGTTCGACTC
GACCATACCGTGGTTCGTGAGCGCCTGCTGGCTAAACTTTCCGGCGCGAACA
ACTTCCGGCTGGCGCTGATCACGAGTCCTGCGGGCTACGGAAAGACCACCCT
CATTTCCCAGTGGGCGGCAGGCAAAAACGATATCGGCTGGTACTCGCTGGAT
GAAGGTGATAACCAGCAAGAGCGTTTCGCCAGCTATCTCATTGCCGCCGTGC
AGCAGGCAACCAACGGTCACTGTGCGATATGTGAGACGATGGCGCAAAAAC
GGCAATATGCCAGCCTGACGTCACTCTTCGCCCAGCTTTTCATTGAGCTGGCG
GAATGGCATAGCCCACTTTATCTGGTCATCGATGACTATCATCTGATCACTAA
TCCAGTGATCCACGAGTCAATGCGCTTCTTTATTCGCCATCAACCAGAAAATC
TCACCCTGGTGGTGTTGTCACGCAACCTTCCGCAACTGGGCATTGCCAATCTG
CGTGTTCGTGATCAACTGCTGGAAATTGGCAGTCAGCAACTGGCATTTACCCA
TCAGGAAGCGAAGCAGTTTTTTGATTGCCGTCTGTCATCGCCGATTGAAGCCG
CAGAAAGCAGTCGGATTTGCGATGACGTTTCCGGTTGGGCGACGGCACTACA
GCTAATCGCCCTCTCCGCCCGGCAGAATACCCACTCAGCCCATAAGTCGGCA
CGCCGCCTGGCGGGAATCAATGCCAGCCATCTTTCGGATTATCTGGTCGATGA
GGTTTTGGATAACGTCGATCTCGCAACGCGCCATTTTCTGTTGAAAAGCGCCA
TTTTGCGCTCAATGAACGATGCCCTCATCACCCCGTGTGACCGGCGAAGAAA
ACGGGCAAATGCGCCCTCGAGAGATTGAGCGTCAGGGGCTGTTTTTACAGCG
GATGGATGATACCGGCGAGTGTTCTGCTATCACCCCGCTGTTTGGTACTTCTG
GCGCAGCGCTGCCAGTGGGAACTGCGCGGAGCTGCCGTAATCCACCGTGCGC
GCAGAAGCTGGAATGGCAAGGGATCAGGCGAGCATTCATCATGCCTGGCGCA
GCCGATGCCCCTTGTATGTCTGGCGCAGA


Applied Biosystems
A04_DT2-8-T7PRO.ab1
KB 1.2 KB.bcp
DT2-81-T7PRO
KB_3730_POP7_BDTv1.mob
Signal: G:850 A:603 T:752 C:535 AvgSig: 685
C#:16 W:A4 Plate Name:368
TS:43 CRL:1047 QV20+:1046
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:09:27 GMT
Sequence Scanner v1.0
Electropherogram Data Page 1 of 5


Applied Biosystems
A04_DT2-8-T7PRO.ab1
KB 1.2 KB.bcp
DT2-81-T7PRO
KB_3730_POP7_BDTv1.mob
Signal: G:850 A:603 T:752 C:535 AvgSig: 685
C#:16 W:A4 Plate Name:368
TS:43 CRL:1047 QV20+:1046
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:09:27 GMT
Sequence Scanner v1.0
Electropherogram Data Page 2 of 5


Applied Biosystems
A04_DT2-8-T7PRO.ab1
KB 1.2 KB.bcp
DT2-81-T7PRO
KB_3730_POP7_BDTv1.mob
Signal: G:850 A:603 T:752 C:535 AvgSig: 685
C#:16 W:A4 Plate Name:368
TS:43 CRL:1047 QV20+:1046
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:09:27 GMT
Sequence Scanner v1.0
Electropherogram Data Page 3 of 5

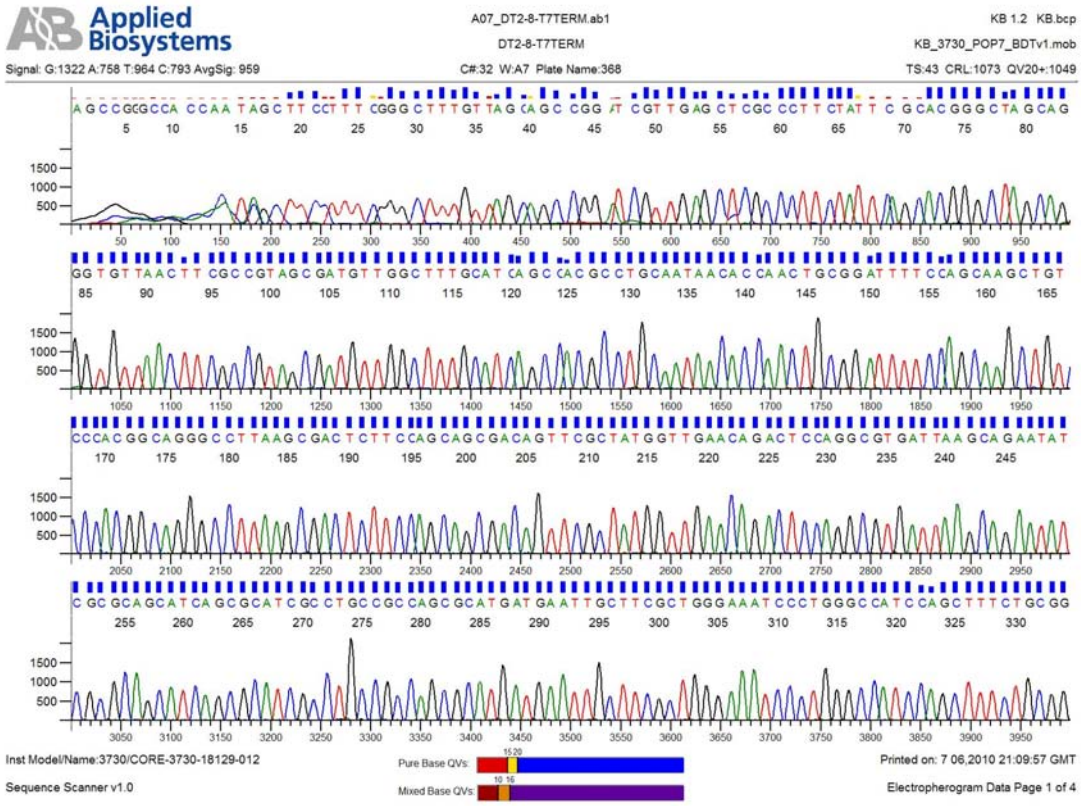

Applied Biosystems
A04_DT2-8-T7PRO.ab1
KB 1.2 KB.bcp
DT2-81-T7PRO
KB_3730_POP7_BDTv1.mob
Signal: G:850 A:603 T:752 C:535 AvgSig: 685
C#:16 W:A4 Plate Name:368
TS:43 CRL:1047 QV20+:1046
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:09:27 GMT
Sequence Scanner v1.0
Electropherogram Data Page 4 of 5

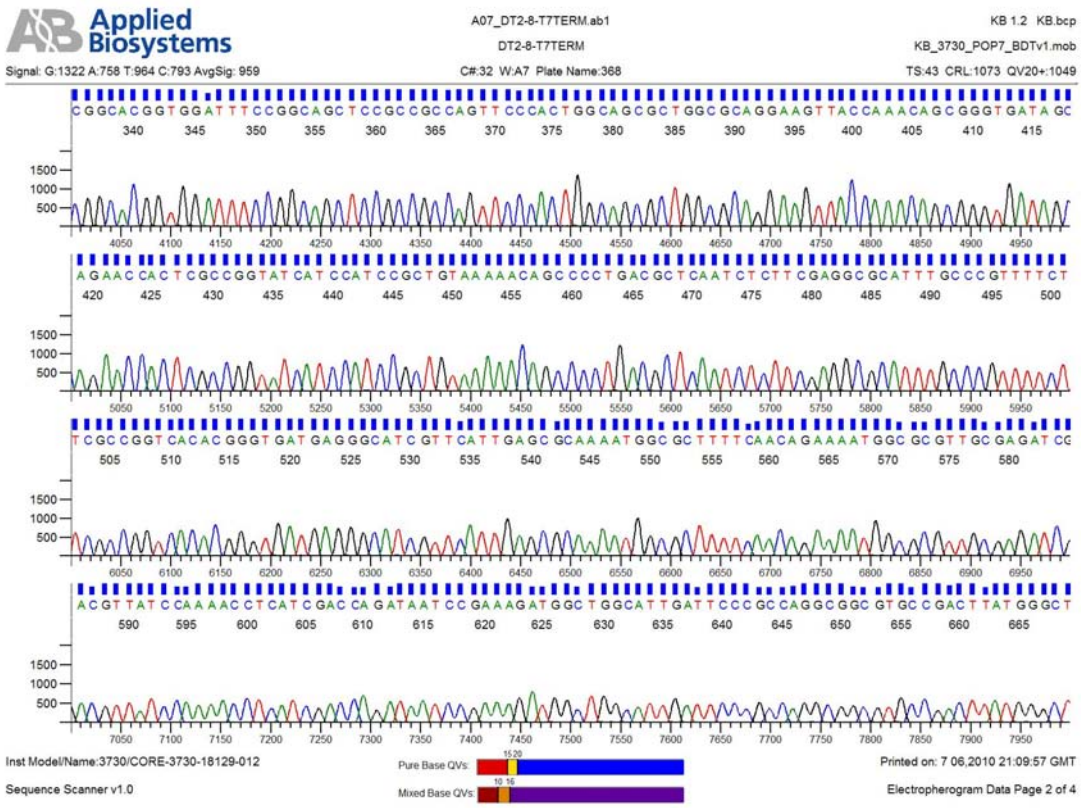


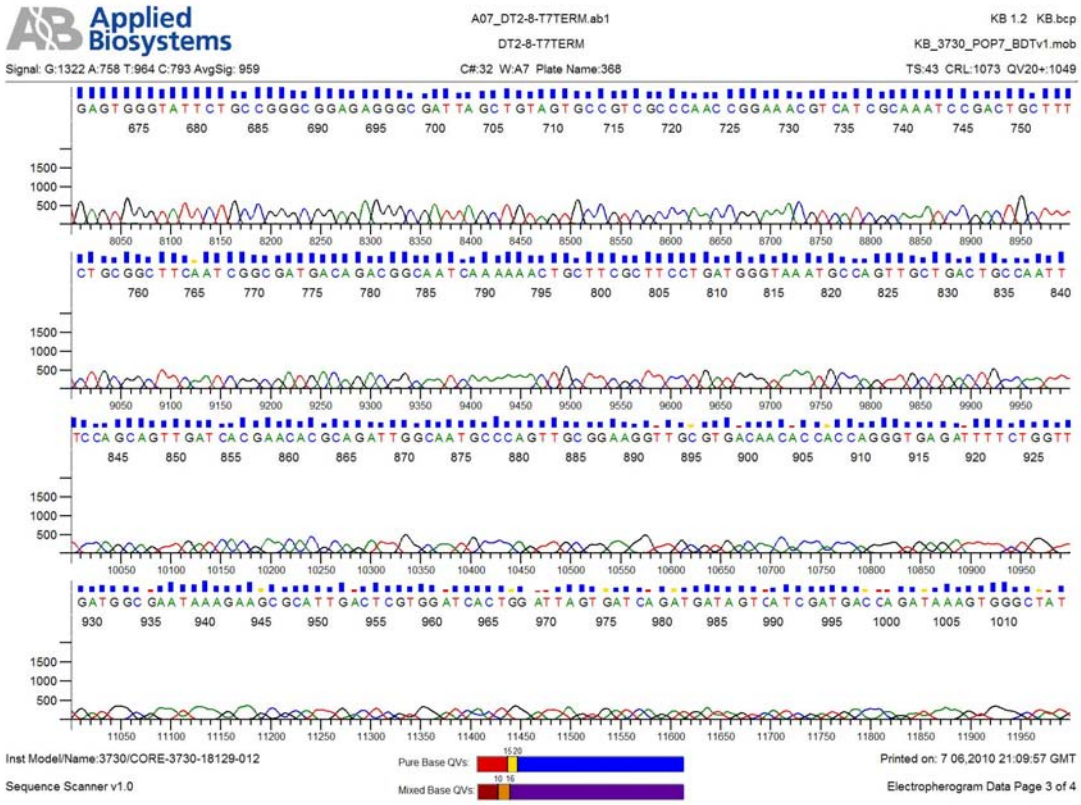

**T7 terminator**

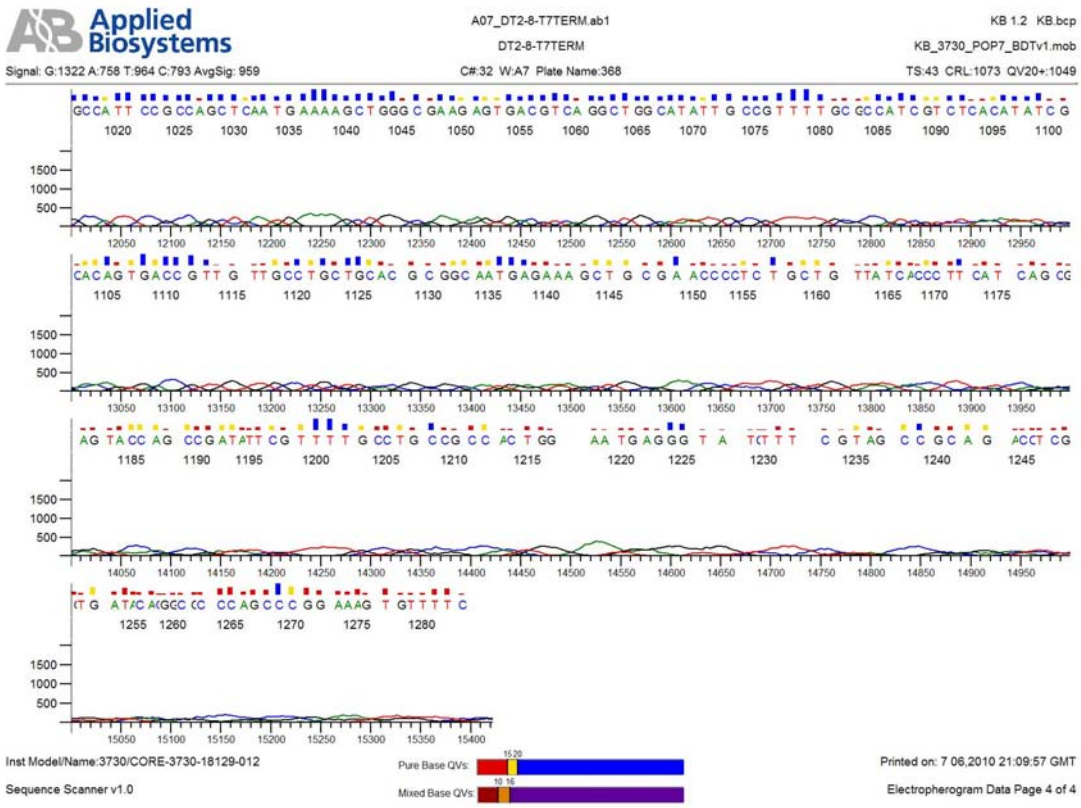
AGCCGGGCCACCAATAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCGTTGA
GCTCGCCCTTCTATTCGCACGGGCTAGCAGGGTGTTAACTTCGCCGTAGCGAT
GTTGGCTTTGCATCAGCCACGCCTGCAATAACACCAACTGCGGATTTTCCAGC
AAGCTGTCCCACGGCAGGGCCTTAAGCGACTCTTCCAGCAGCGACAGTTCGC
TATGGTTGAACAGACTCCAGGCGTGATTAAGCAGAATATCGCGCAGCATCAG
CGCATCGCCTGCCGCCAGCGCATGATGAATTGCTTCGCTGGGAAATCCCTGG
GCCATCCAGCTTTCTGCGGCGGCACGGTGGATTTCCGGCAGCTCCGCCGCCA
GTTCCCACTGGCAGCGCTGGCGCAGGAAGTTACCAAACAGCGGGTGATAGCA
GAACCACTCGCCGGTATCATCCATCCGCTGTAAAAACAGCCCCTGACGCTCA
ATCTCTTCGAGGCGCATTTGCCCGTTTTCTTCGCCGGTCACACGGGTGATGAG
GGCATCGTTCATTGAGCGCAAAATGGCGCTTTTCAACAGAAAATGGCGCGTT
GCGAGATCGACGTTATCCAAAACCTCATCGACCAGATAATCCGAAAGATGGC
TGGCATTGATTCCCGCCAGGCGGCGTGCCGACTTATGGGCTGAGTGGGTATTC
TGCCGGGCGGAGAGGGCGATTAGCTGTAGTGCCGTCGCCCAACCGGAAACGT
CATCGCAAATCCGACTGCTTTCTGCGGCTTCAATCGGCGATGACAGACGGCA
ATCAAAAAACTGCTTCGCTTCCTGATGGGTAAATGCCAGTTGCTGACTGCCAA
TTTCCAGCAGTTGATCACGAACACGCAGATTGGCAATGCCCAGTTGCGGAAG
GTTGCGTGACAACACCACCAGGGTGAGATTTTCTGGTTGATGGCGAATAAAG

AAGCGCATTGACTCGTGGATCACTGGATTAGTGATCAGATGATAGTCATCGA
TGACCAGATAAAGTGGGCTATGCCATTCCGCCAGCTCAATGAAAAGCTGGGC
GAAGAGTGACGTCAGGCTGGCATATTGCCGTTTTGCGCCATCGTCTCACATAT
CGCACAGTGACCGTTGTTGCCTGCTGCACGCGGCAATGAGAAAGCTGCGAAC
CCCTCTGCTGTTATCACCCTTCATCAGCGAGTACCAGCCGATATTCGTTTTGC
CTGCCGCCACTGGAATGAGGGTATCTTTCGTAGCCGCAGACCTCGGTGATAC
ACGGCGCCCAGCCCGGAAAGTGTTTTC




Applied Biosystems
A07_DT2-8-T7TERM.ab1
DT2-8-T7TERM
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:1322 A:758 T:964 C:793 AvgSig: 959
C#:32 W:A7 Plate Name:368
TS:43 CRL:1073 QV20+:1049
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: 7 06,2010 21:09:57 GMT
Electropherogram Data Page 2 of 4


Applied Biosystems
A07_DT2-8-T7TERM.ab1
DT2-8-T7TERM
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:1322 A:758 T:964 C:793 AvgSig: 959
C#:32 W:A7 Plate Name:368
TS:43 CRL:1073 QV20+:1049
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs
Mixed Base QVs
Printed on: 7 06,2010 21:09:57 GMT
Electropherogram Data Page 3 of 4




**6.11. DNA sequence of *malT* from *E. coli*.**

ATGCTGATTCCGTCAAAACTAAGTCGTCCGGTTCGACTCGACCATACCGTGGT
TCGTGAGCGCCTGCTGGCTAAACTTTCCGGCGCGAACAACTTCCGGCTGGCG
CTGATCACGAGTCCTGCGGGCTACGGAAAGACCACCCTCATTTCCCAGTGGG
CGGCAGGCAAAAACGATATCGGCTGGTACTCGCTGGATGAAGGTGATAACCA
GCAAGAGCGTTTCGCCAGCTATCTCATTGCCGCCGTGCAGCAGGCAACCAAC
GGTCACTGTGCGATATGTGAGACGATGGCGCAAAAACGGCAATATGCCAGCC
TGACGTCACTCTTCGCCCAGCTTTTCATTGAGCTGGCGGAATGGCATAGCCCA
CTTTATCTGGTCATCGATGACTATCATCTGATCACTAATCCAGTGATCCACGA
GTCAATGCGCTTCTTTATTCGCCATCAACCAGAAAATCTCACCCTGGTGGTGT
TGTCACGCAACCTTCCGCAACTGGGCATTGCCAATCTGCGTGTTCGTGATCAA
CTGCTGGAAATTGGCAGTCAGCAACTGGCATTTACCCATCAGGAAGCGAAGC
AGTTTTTTGATTGCCGTCTGTCATCGCCGATTGAAGCCGCAGAAAGCAGTCGG
ATTTGCGATGACGTTTCCGGTTGGGCGACGGCACTACAGCTAATCGCCCTCTC
CGCCCGGCAGAATACCCACTCAGCCCATAAGTCGGCACGCCGCCTGGCGGGA
ATCAATGCCAGCCATCTTTCGGATTATCTGGTCGATGAGGTTTTGGATAACGT
CGATCTCGCAACGCGCCATTTTCTGTTGAAAAGCGCCATTTTGCGCTCAATGA
ACGATGCCCTCATCACCCGTGTGACCGGCGAAGAAAACGGGCAAATGCGCCT

CGAAGAGATTGAGCGTCAGGGGCTGTTTTTACAGCGGATGGATGATACCGGC
GAGTGGTTCTGCTATCACCCGCTGTTTGGTAACTTCCTGCGCCAGCGCTGCCA
GTGGGAACTGGCGGCGGAGCTGCCGGAAATCCACCGTGCCGCCGCAGAAAG
CTGGATGGCCCAGGGATTTCCCAGCGAAGCAATTCATCATGCGCTGGCGGCA
GGCGATGCGCTGATGCTGCGCGATATTCTGCTTAATCACGCCTGGAGTCTGTT
CAACCATAGCGAACTGTCGCTGCTGGAAGAGTCGCTTAAGGCCCTGCCGTGG
GACAGCTTGCTGGAAAATCCGCAGTTGGTGTTATTGCAGGCGTGGCTGATGC
AAAGCCAACATCGCTACGGCGAAGTTAACACCCTGCTAGCCCGTGCTGAACA
TGAAATCAAGGACATCAGAGAAGACACCATGCACGCAGAATTTAACGCTCTG
CGCGCCCAGGTGGCGATTAACGATGGTAATCCGGATGAAGCGGAACGGCTGG
CAAAACTGGCACTGGAAGAGCTGCCGCCGGGCTGGTTCTATAGCCGCATTGT
GGCAACCTCGGTGCTGGGTGAAGTGCTGCACTGCAAAGGCGAATTGACCCGC
TCACTGGCGCTAATGCAGCAAACCGAACAGATGGCACGCCAGCACGATGTCT
GGCACTACGCTTTGTGGAGTTTAATCCAGCAAAGTGAAATTCTGTTTGCCCAA
GGGTTCCTGCAAACCGCGTGGGAAACGCAGGAAAAAGCATTCCAGCTGATCA
ACGAGCAGCATCTGGAACAGCTGCCAATGCATGAGTTTCTGGTGCGCATTCG
TGCGCAGCTGTTATGGGCCTGGGCGCGGCTGGATGAAGCCGAAGCGTCGGCG
CGTAGCGGGATTGAAGTCTTGTCGTCTTATCAGCCACAGCAACAGCTTCAGTG
CCTGGCAATGTTGATTCAATGCTCGCTGGCCCGTGGTGATTTAGATAACGCCC
GTAGCCAGCTGAACCGTCTGGAAAACCTGCTGGGGAATGGCAAATATCACAG
CGACTGGATCTCTAACGCCAACAAAGTCCGGGTGATTTACTGGCAAATGACC
GGCGATAAAGCCGCCGCTGCCAACTGGTTGCGTCATACGGCTAAACCAGAGT
TTGCGAACAACCACTTCCTGCAAGGTCAATGGCGCAACATTGCCCGTGCACA
AATCTTGCTGGGCGAGTTTGAACCGGCAGAAATTGTTCTCGAAGAACTCAAT
GAAAATGCCCGGAGTCTGCGGTTGATGAGCGATCTCAACCGTAACCTGTTGC
TGCTTAATCAACTGTACTGGCAGGCCGGACGTAAAAGTGACGCCCAGCGCGT
GTTGCTGGACGCATTAAAACTGGCGAATCGCACCGGATTTATCAGCCATTTTG
TCATCGAAGGCGAAGCGATGGCGCAACAACTGCGTCAGCTGATTCAGCTTAA
TACGCTGCCGGAACTGGAACAGCATCGCGCGCAGCGTATTCTGCGAGAAATC
AATCAACATCATCGGCATAAATTCGCCCATTTCGATGAGAATTTCGTTGAACG
TCTGCTAAATCATCCTGAAGTACCTGAACTGATCCGCACCAGCCCGCTGACG
CAACGTGAATGGCAGGTACTGGGGCTGATCTACTCTGGTTACAGCAATGAGC
AAATTGCCGGAGAACTGGAAGTCGCGGCAACCACCATCAAAACGCATATCCG
CAATCTGTATCAGAAACTCGGCGTGGCCCATCGCCAGGATGCGGTACAACAC
GCCCAGCAATTGCTGAAGATGATGGGGTACGGCGTGTAA

### 6.12. Amino acid sequence of MalT from *E. coli*.

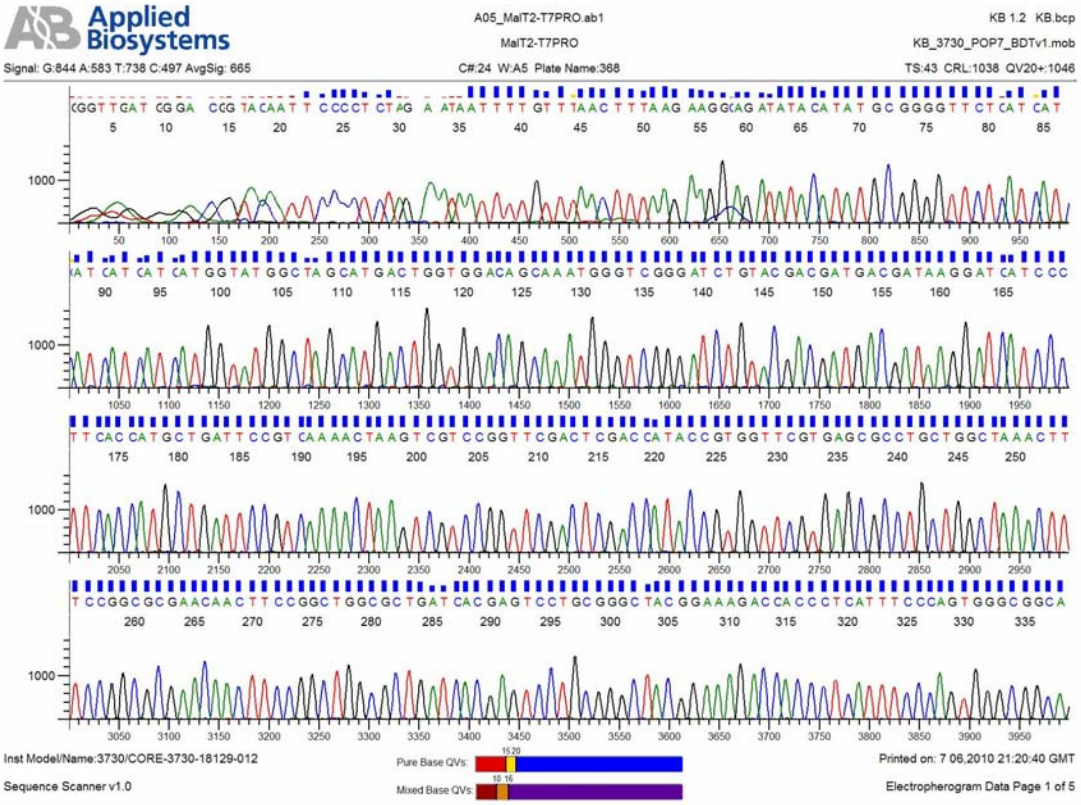
MLIPSKLSRPVRLDHTVVRERLLAKLSGANNFRLALITSPAGYGKTTLISQWAAG
KNDIGWYSLDEGDNQQERFASYLIAAVQQATNGHCAICETMAQKRQYASLTSLF
AQLFIELAEWHSPLYLVIDDYHLITNPVIHESMRFFIRHQPENLTLVVLSRNLPQLG

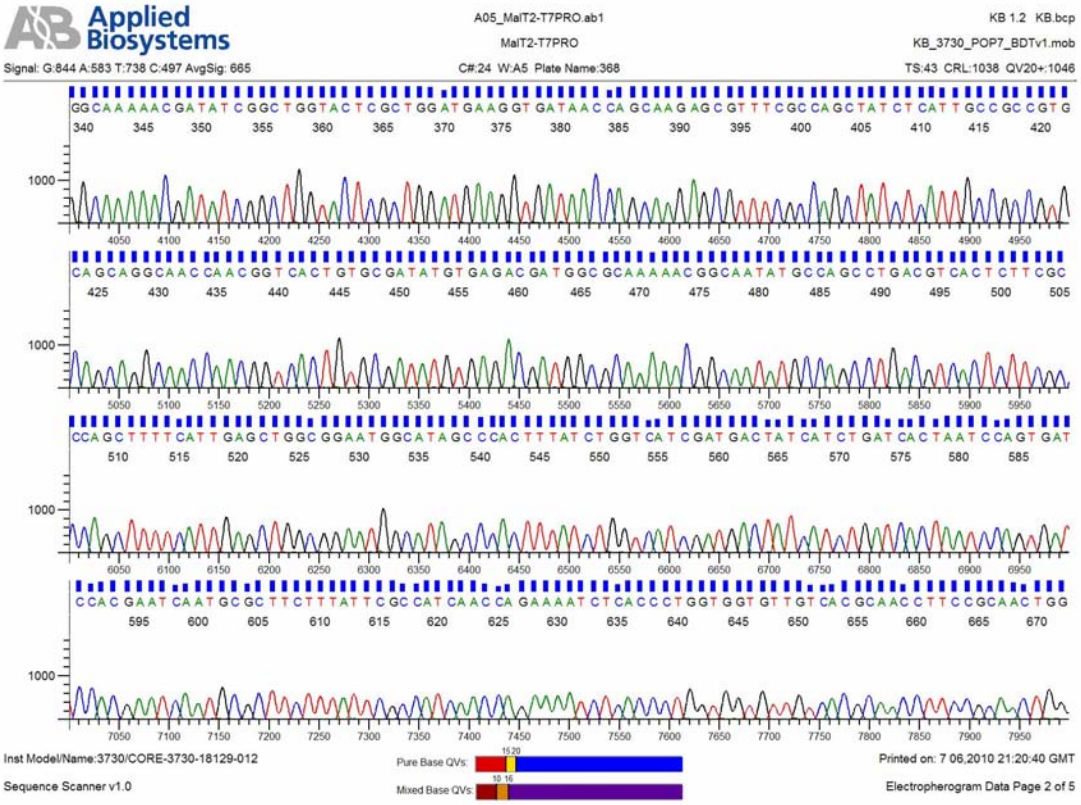
IANLRVRDQLLEIGSQQLAFTHQEAKQFFDCRLSSPIEAAESSRICDDVSGWATAL
QLIALSARQNTHSAHKSARRLAGINASHLSDYLVDEVLDNVDLATRHFLLKSAIL
RSMNDALITRVTGEENGQMRLEEIERQGLFLQRMDDTGEWFCYHPLFGNFLRQR
CQWELAAELPEIHRAAAESWMAQGFPSEAIHHALAAGDALMLRDILLNHAWSL
FNHSELSLLEESLKALPWDSLLENPQLVLLQAWLMQSQHRYGEVNTLLARAEHE
IKDIREDTMHAEFNALRAQVAINDGNPDEAERLAKLALEELPPGWFYSRIVATSV
LGEVLHCKGELTRSLALMQQTEQMARQHDVWHYALWSLIQQSEILFAQGFLQT
AWETQEKAFQLINEQHLEQLPMHEFLVRIRAQLLWAWARLDEAEASARSGIEVL
SSYQPQQQLQCLAMLIQCSLARGDLDNARSQLNRLENLLGNGKYHSDWISNAN
KVRVIYWQMTGDKAAAANWLRHTAKPEFANNHFLQGQWRNIARAQILLGEFEP
AEIVLEELNENARSLRLMSDLNRNLLLLNQLYWQAGRKSDAQRVLLDALKLAN
RTGFISHFVIEGEAMAQQLRQLIQLNTLPELEQHRAQRILREINQHHRHKFAHFDE
NFVERLLNHPEVPELIRTSPLTQREWQVLGLIYSGYSNEQIAGELEVAATTIKTHIR
NLYQKLGVAHRQDAVQHAQQLLKMMGYGV

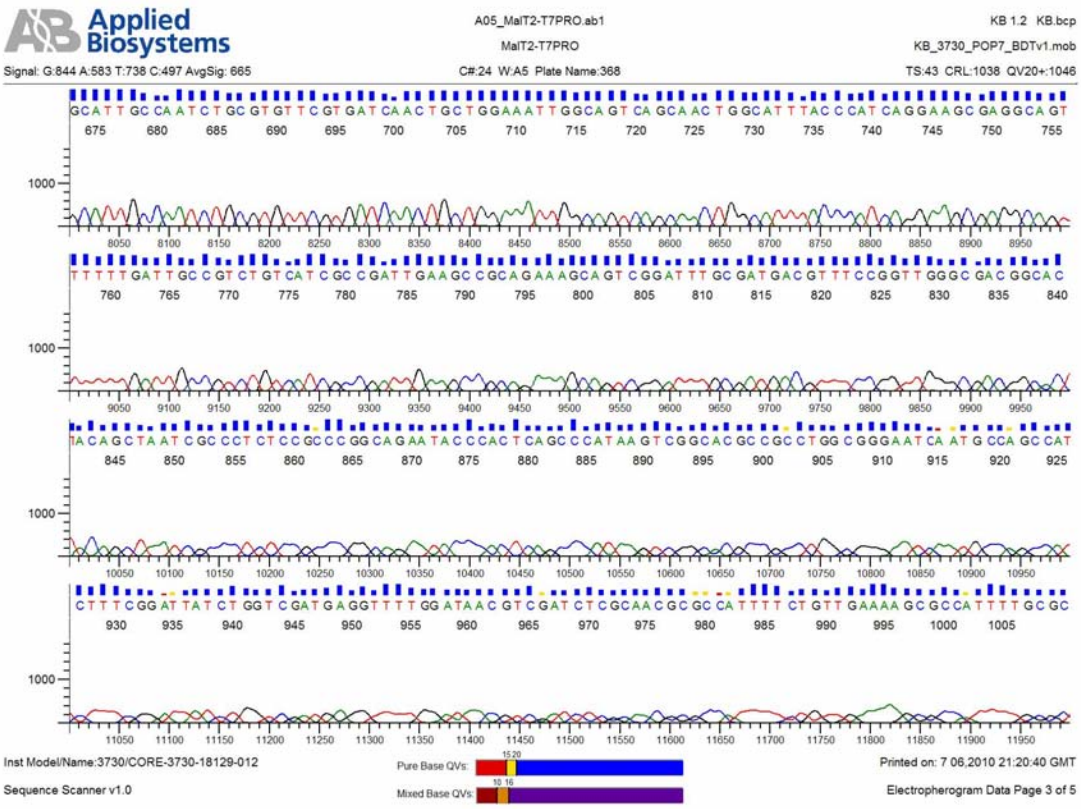
### 6.13. The result of DNA sequencing of *malT*.

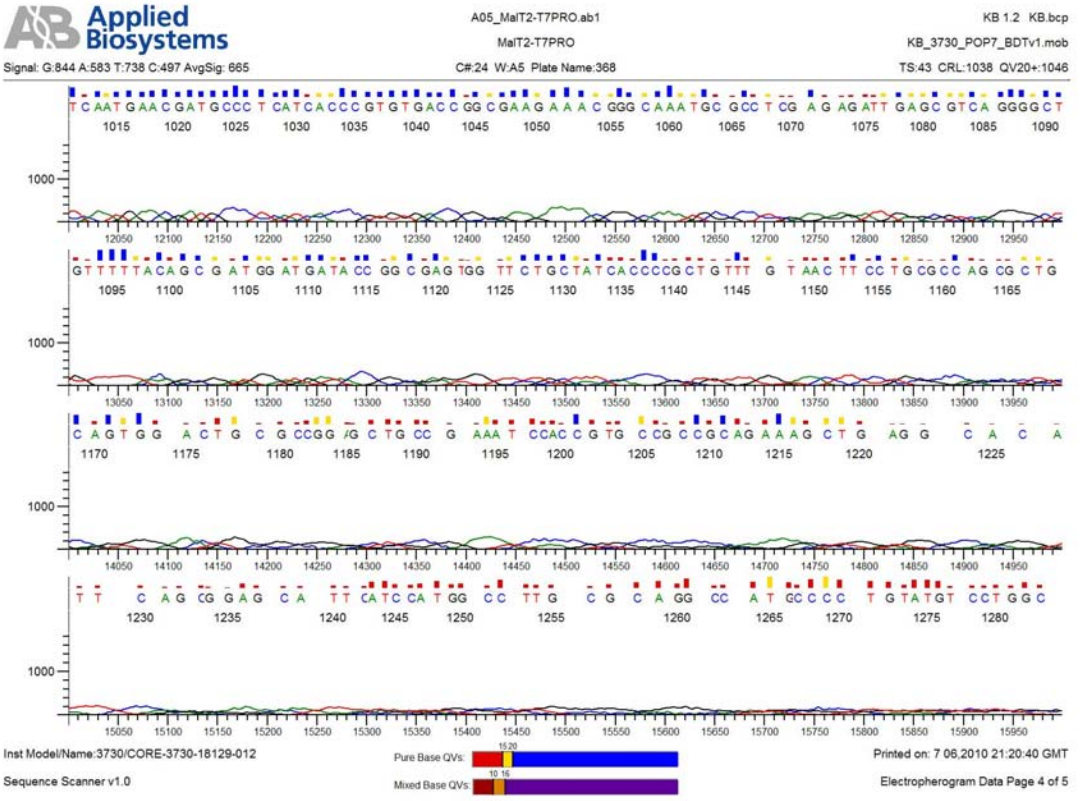
**T7 promoter**

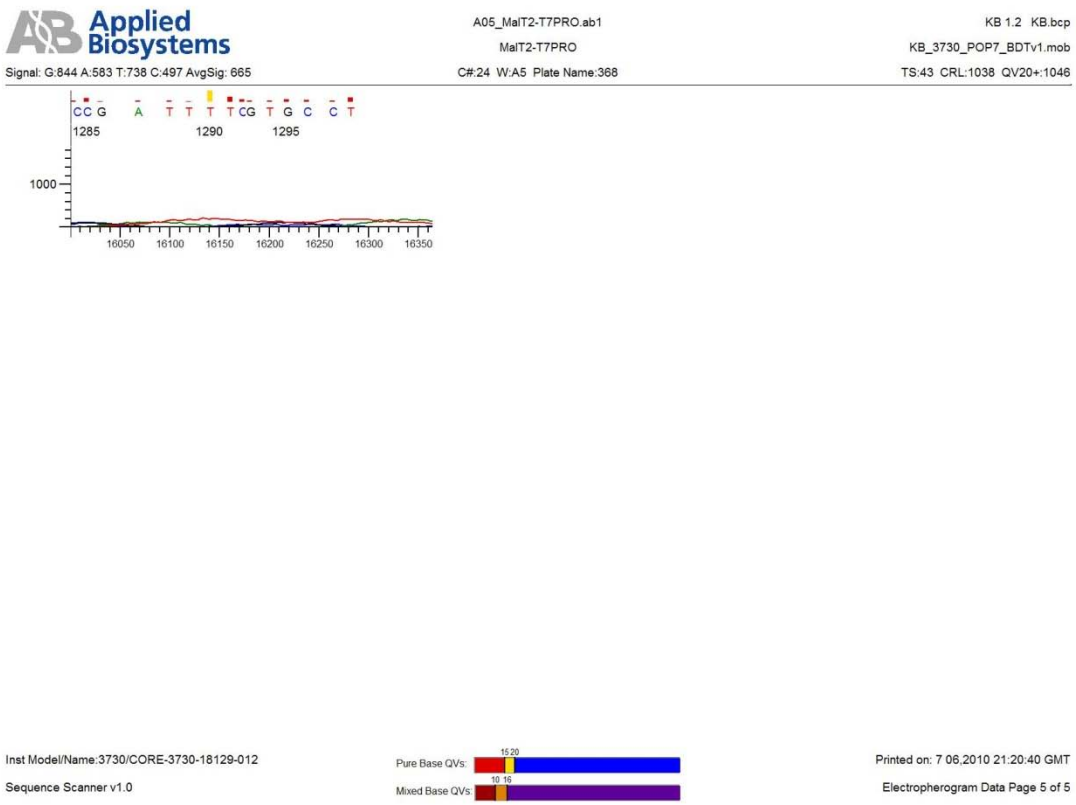
GGGTTGATGGGACGGTACAATTCCCCTCTAGAATAATTTTGTTTAACTTTAAG
AAGGCAGATATACATATGCGGGGTTCTCATCATCATCATCATCATGGTATGGC
TAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGACGATGACGAT
AAGGATCATCCCTTCACCATGCTGATTCCGTCAAAACTAAGTCGTCCGGTTCG
ACTCGACCATACCGTGGTTCGTGAGCGCCTGCTGGCTAAACTTTCCGGCGCG
AACAACTTCCGGCTGGCGCTGATCACGAGTCCTGCGGGCTACGGAAAGACCA
CCCTCATTTCCCAGTGGGCGGCAGGCAAAAACGATATCGGCTGGTACTCGCT
GGATGAAGGTGATAACCAGCAAGAGCGTTTCGCCAGCTATCTCATTGCCGCC
GTGCAGCAGGCAACCAACGGTCACTGTGCGATATGTGAGACGATGGCGCAAA
AACGGCAATATGCCAGCCTGACGTCACTCTTCGCCCAGCTTTTCATTGAGCTG
GCGGAATGGCATAGCCCACTTTATCTGGTCATCGATGACTATCATCTGATCAC
TAATCCAGTGATCCACGAATCAATGCGCTTCTTTATTCGCCATCAACCAGAAA
ATCTCACCCTGGTGGTGTTGTCACGCAACCTTCCGCAACTGGGCATTGCCAAT
CTGCGTGTTCGTGATCAACTGCTGGAAATTGGCAGTCAGCAACTGGCATTTAC
CCATCAGGAAGCGAGGCAGTTTTTTGATTGCCGTCTGTCATCGCCGATTGAAG
CCGCAGAAAGCAGTCGGATTTGCGATGACGTTTCCGGTTGGGCGACGGCACT
ACAGCTAATCGCCCTCTCCGCCCGGCAGAATACCCACTCAGCCCATAAGTCG
GCACGCCGCCTGGCGGGAATCAATGCCAGCCATCTTTCGGATTATCTGGTCG
ATGAGGTTTTGGATAACGTCGATCTCGCAACGCGCCATTTTCTGTTGAAAAGC
GCCATTTTGCGCTCAATGAACGATGCCCTCATCACCCGTGTGACCGGCGAAG
AAACGGGCAAATGCGCCTCGAGAGATTGAGCGTCAGGGGCTGTTTTTACAGC
GATGGATGATACCGGCGAGTGGTTCTGCTATCACCCCGCTGTTTGTAACTTCC
TGCGCCAGCGCTGCAGTGGACTGCGCCGGAGCTGCCGAAATCCACCGTGCCG
CCGCAGAAAGCTGAGGCACATTCAGCGGAGCATTCATCCATGGCCTTGCGCA
GGCCATGCCCCTGTATGTCCTGGCCCGATTTTCGTGCCT


Applied Biosystems
A05_MalT2-T7PRO.ab1
MalT2-T7PRO
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:844 A:583 T:738 C:497 AvgSig: 665
C#:24 W:A5 Plate Name:368
TS:43 CRL:1038 QV20+:1046
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:20:40 GMT
Sequence Scanner v1.0
Electropherogram Data Page 1 of 5


Applied Biosystems
A05_MalT2-T7PRO.ab1
MalT2-T7PRO
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:844 A:583 T:738 C:497 AvgSig: 665
C#:24 W:A5 Plate Name:368
TS:43 CRL:1038 QV20+:1046
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:20:40 GMT
Sequence Scanner v1.0
Electropherogram Data Page 2 of 5


Applied Biosystems
A05_MalT2-T7PRO.ab1
KB 1.2 KB.bcp
MalT2-T7PRO
KB_3730_POP7_BDTv1.mob
Signal: G:844 A:583 T:738 C:497 AvgSig: 665
C#:24 W:A5 Plate Name:368
TS:43 CRL:1038 QV20+:1046
GCATTGCCAATCTGCGTGTTCGTGATCAACTGCTGGAAATTGGCAGTCAGCAACTGGCATTTACCCATCAGGAAGCGAGGCAGT
TTTTTGATTGCCGTCTGTCATCGCCGATTGAAGCCGCAGAAAGCAGTCGGATTTGCGATGACGTTTCCGGTTGGGCGACGGCAC
TACAGCTAATCGCCCTCTCCGCCCGGCAGAATACCCACTCAGCCCATAAGTCGGCACGCCGCCTGGCGGGAATCAATGCCAGCCAT
CTTTCGGATTATCTGGTCGATGAGGTTTTGGATAACGTCGATCTCGCAACGCGCCATTTTCTGTTGAAAAGCGCCATTTTGCGC
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:20:40 GMT
Sequence Scanner v1.0
Electropherogram Data Page 3 of 5

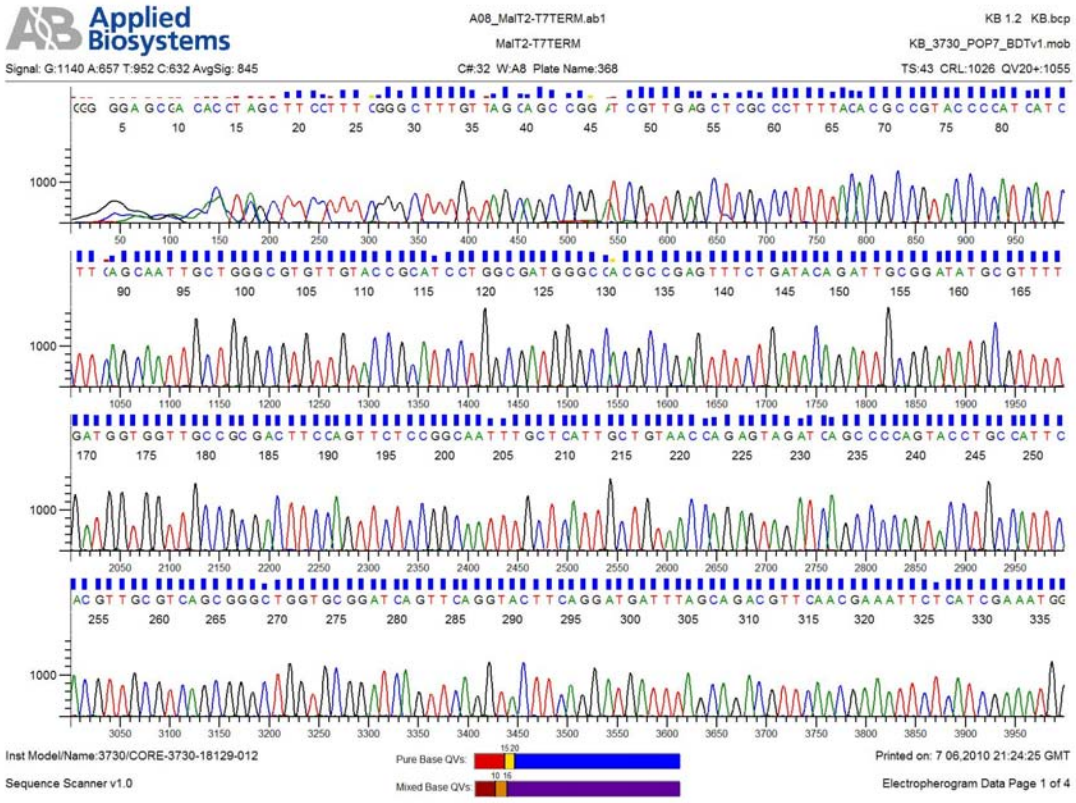

Applied Biosystems
A05_MalT2-T7PRO.ab1
KB 1.2 KB.bcp
MalT2-T7PRO
KB_3730_POP7_BDTv1.mob
Signal: G:844 A:583 T:738 C:497 AvgSig: 665
C#:24 W:A5 Plate Name:368
TS:43 CRL:1038 QV20+:1046
TCAATGAACGATGCCCTCATCACCCGTGTGACCGGCGAAGAAACGGGCAAATGCGCCTCGAGAGATTGAGCGTCAGGGGCT
GTTTTTACAGCGATGGATGATACCGGCGAGTGGTTCTGCTATCACCCCGCTGTTTGTAACTTCCTGCGCCAGCGCTG
CAGTGGACTGCGCCGGAGCTGCCGAAATCCACCGTGCCGCCGCAGAAAGCTGAGGCAA
TTCAGGAGCAATTCATCCATGGCCTTGCGCAGGCCATGCCCCTGTATGTCCTGGC
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:20:40 GMT
Sequence Scanner v1.0
Electropherogram Data Page 4 of 5

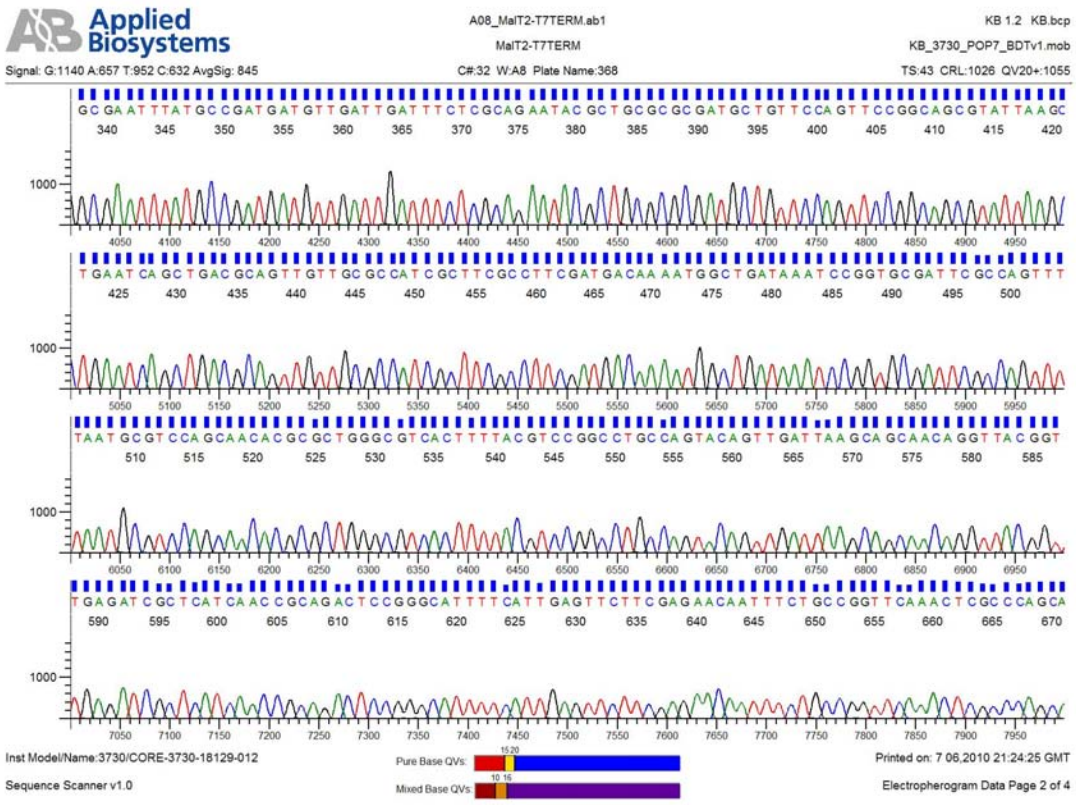


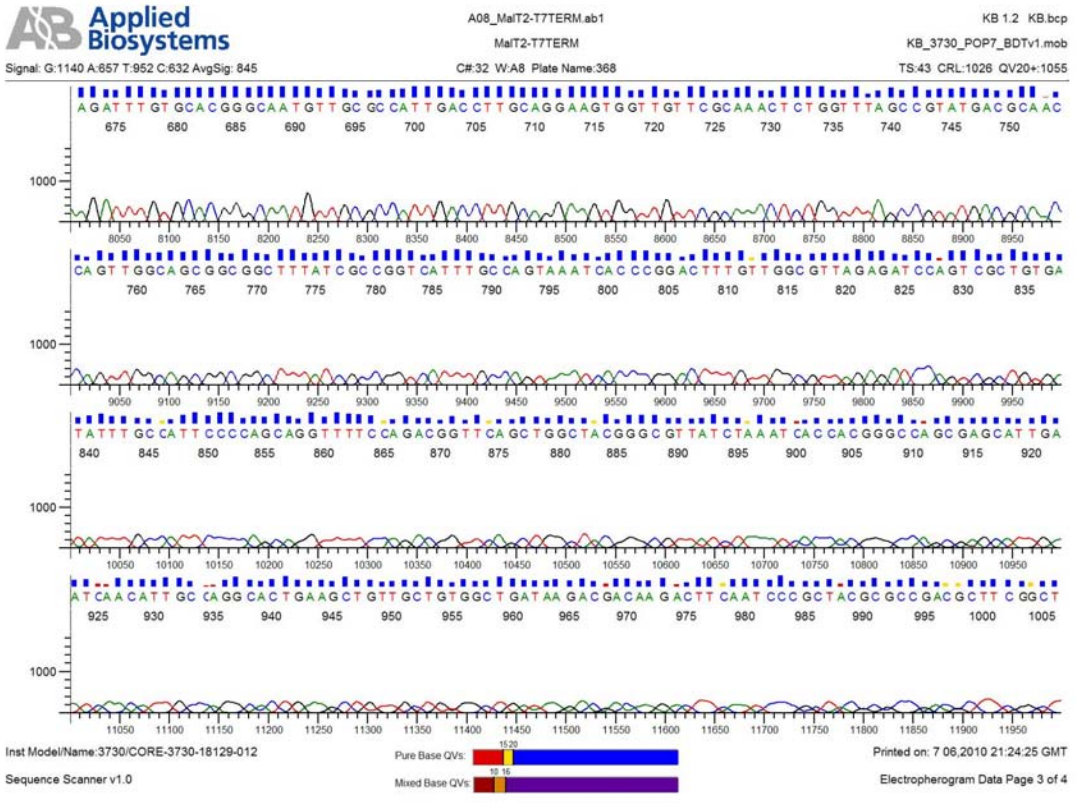

**T7 terminator**

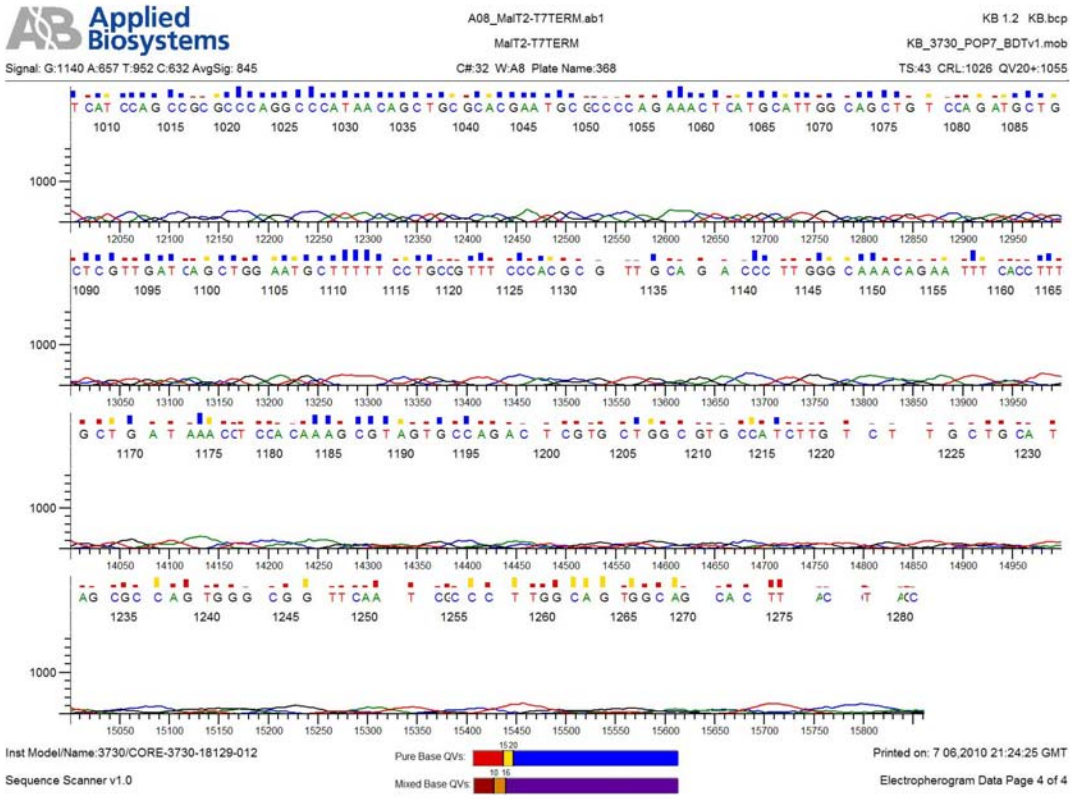
GGGGGAGCGACACCTAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCGTTGA
GCTCGCCCTTTTACACGCCGTACCCCATCATCTTCAGCAATTGCTGGGCGTGT
TGTACCGCATCCTGGCGATGGGCCACGCCGAGTTTCTGATACAGATTGCGGA
TATGCGTTTTGATGGTGGTTGCCGCGACTTCCAGTTCTCCGGCAATTTGCTCA
TTGCTGTAACCAGAGTAGATCAGCCCCAGTACCTGCCATTCACGTTGCGTCAG
CGGGCTGGTGCGGATCAGTTCAGGTACTTCAGGATGATTTAGCAGACGTTCA
ACGAAATTCTCATCGAAATGGGCGAATTTATGCCGATGATGTTGATTGATTTC
TCGCAGAATACGCTGCGCGCGATGCTGTTCCAGTTCCGGCAGCGTATTAAGCT
GAATCAGCTGACGCAGTTGTTGCGCCATCGCTTCGCCTTCGATGACAAAATG
GCTGATAAATCCGGTGCGATTCGCCAGTTTTAATGCGTCCAGCAACACGCGCT
GGGCGTCACTTTTACGTCCGGCCTGCCAGTACAGTTGATTAAGCAGCAACAG
GTTACGGTTGAGATCGCTCATCAACCGCAGACTCCGGGCATTTTCATTGAGTT
CTTCGAGAACAATTTCTGCCGGTTCAAACTCGCCCAGCAAGATTTGTGCACGG
GCAATGTTGCGCCATTGACCTTGCAGGAAGTGGTTGTTCGCAAACTCTGGTTT
AGCCGTATGACGCAACCAGTTGGCAGCGGCGGCTTTATCGCCGGTCATTTGC
CAGTAAATCACCCGGACTTTGTTGGCGTTAGAGATCCAGTCGCTGTGATATTT
GCCATTCCCCAGCAGGTTTTCCAGACGGTTCAGCTGGCTACGGGCGTTATCTA
AATCACCACGGGCCAGCGAGCATTGAATCAACATTGCCAGGCACTGAAGCTG
TTGCTGTGGCTGATAAGACGACAAGACTTCAATCCCGCTACGCGCCGACGCT

TCGGCTTCATCCAGCCGCGCCCAGGCCCATAACAGCTGCGCACGAATGCGCC
CCAGAAACTCATGCATTGGCAGCTGTCCAGATGCTGCTCGTTGATCAGCTGG
AATGCTTTTTCCTGCCGTTTCCCACGCGTTGCAGACCCTTGGGCAAACAGAAT
TTCACCTTTGCTGATAAACCTCCACAAAGCGTAGTGCCAGACTCGTGCTGGCG
TGCCATCTTGTCTTGCTGCATAGCGCCAGTGGGCGGTTCAATCGCCCTTGGCA
GTGGCAGCACTTACCTAGC




Applied Biosystems
A08_MalT2-T7TERM.ab1
KB 1.2 KB.bcp
MalT2-T7TERM
KB_3730_POP7_BDTv1.mob
Signal: G:1140 A:657 T:952 C:632 AvgSig: 845
C#:32 W:A8 Plate Name:368
TS:43 CRL:1026 QV20+:1055
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:24:25 GMT
Sequence Scanner v1.0
Electropherogram Data Page 2 of 4


Applied Biosystems
A08_MalT2-T7TERM.ab1
KB 1.2 KB.bcp
MalT2-T7TERM
KB_3730_POP7_BDTv1.mob
Signal: G:1140 A:657 T:952 C:632 AvgSig: 845
C#:32 W:A8 Plate Name:368
TS:43 CRL:1026 QV20+:1055
Inst Model/Name:3730/CORE-3730-18129-012
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:24:25 GMT
Sequence Scanner v1.0
Electropherogram Data Page 3 of 4


Applied Biosystems
A08_MalT2-T7TERM.ab1
MalT2-T7TERM
C#:32 W:A8 Plate Name:368
KB 1.2 KB.bcp
KB_3730_POP7_BDTv1.mob
Signal: G:1140 A:657 T:952 C:632 AvgSig: 845
TS:43 CRL:1026 QV20+:1055
Inst Model/Name:3730/CORE-3730-18129-012
Sequence Scanner v1.0
Pure Base QVs:
Mixed Base QVs:
Printed on: 7 06,2010 21:24:25 GMT
Electropherogram Data Page 4 of 4

VITA

Mamiko Nishida

Candidate for the Degree of

Doctor of Philosophy

Thesis: STRUCTURAL STUDIES OF ACETYL ESTERASE AND MALT FROM *ESCHERICHIA COLI*

Major Field: Chemistry

Biographical:

Personal Data: Born in Miyagi-ken, Japan, 1979.

Education: Received Bachelor of Science degree in Biology and Chemistry at Pittsburg State University, Pittsburg, Kansas in May 2005.

Completed the requirements for the Doctor of Philosophy degree in Chemistry at Oklahoma State University, Stillwater, Oklahoma in July 2010.

Experience: Teaching assistant in the Department of Chemistry, Oklahoma State University, Oklahoma from August 2005 to July 2010.

Professional Memberships: American Chemical Society and American Association of Pharmaceutical Scientists.

Name: Mamiko Nishida Date of Degree: July, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: STRUCTURAL STUDIES OF ACETYL ESTERASE AND MALT FROM *ESCHERICHIA COLI*

Pages in Study: 126 Candidate for the Degree of Doctor of Philosophy

Major Field: Chemistry

Scope and Method of Study: An acetyl esterase, also known as Aes from *Escherichia coli*, belongs to the hormone sensitive lipase family and down regulates MalT which is the transcriptional activator of the maltose regulon. Moreover, a recent study suggests an interaction between Aes and α-galactosidase which is also involved in carbohydrate metabolism. Since Aes interacts with several important proteins, it plays critical roles in carbohydrate metabolism in *E. coli*. Therefore, the purpose of this study was to determine crystal structures of Aes, DT1, DT1-DT2, and MalT in order to understand the remarkable maltose system in *E. coli*. To achieve this, cloning, over-expression, purification, and crystallization for each gene were carried out. Moreover, structural studies of Aes were performed.

Findings and Conclusions: The *E. coli aes*, DT1, DT1-DT2, and *malT* genes have been cloned with an N-terminal histidine tag and over-expressed. Aes, DT1, and MalT have been successfully purified and a crystal structure of Aes has been determined in this study. X-ray crystallography revealed that Aes contained an α/β hydrolase fold, the central β-strands being surrounded by α-helices. Moreover, the catalytic triad of Aes consisted of Ser-165, Asp-262, and His-292, which was stabilized by hydrogen bonds and was hidden in a shallow cleft. Since crystal screening showed some promising results for DT1 and MalT, further optimization in crystallization of these proteins will lead to crystal structure determination of them. Moreover, purification trials of DT1, DT1-DT2, and MalT should be carried out so as to find better conditions for crystal production.

ADVISER'S APPROVAL: Dr. Stacy D. Benson