# Synthetic Gene-Based Heterologous Expression, Proteolytic, and Structural Characterization of Caseinolytic Protease of *Lactobacillus plantarum* IIA-1A5

**M. Yusuf\*, C. Budiman, I. I. Arief, & C. Sumantri**

Department of Animal Production and Technology, Faculty of Animal Science, IPB University
Jalan Agatis, Kampus IPB Darmaga Bogor 16680, Indonesia
*Corresponding author: m.yusuf08@yahoo.com
*(Received 05-01-2021; Revised 07-06-2021; Accepted 08-06-2021)*

## ABSTRACT

Genome sequence of Indonesian probiotic of *Lactobacillus plantarum* II1A5 contains a gene encoding a proteolytic subunit of caseinolytic protease, designated as ClpP_LP. This study aims to express the Clp gene heterological and apply its proteolytic activity to some livestock products. To address this, the gene encoding ClpP_LP was optimized *in silico* by improving its Codon Adaptation Index and GC content to 0.94 and 53.62%, respectively. The optimized gene was then inserted into pET28a, transformed into *Escherichia coli* BL21(DE3), and over-expressed by induction of 1 mM Isopropyl β-D-1-thiogalactopyranoside at 37°C. The result showed that ClpP_LP was successfully over-expressed in a fully soluble form with the specific activity towards milk casein was 7739.89 AU mg$^{-1}$. This activity was significantly greater than that of chymotrypsin. Further, the three-dimensional model of ClpP_LP was built using SWISS MODEL, which showed that this protein formed a homo-tetradecameric (14-mer) structure with each monomer consisting of 7 $\alpha$-helix and 10 $\beta$-sheets. The identification of the active side showed that the active side of ClpP_LP is Ser-97, His-122, Asp-171, and forms a substrate-binding cavity with a size of about 29.5 Å. Overall, our approach can serve as an appropriate platform for the production of ClpP_LP in a large-scale production for various applications in dairy products and derivatives.

*Keywords: Clp-Protease; heterologous expression; Lactobacillus plantarum IIA-1A5; structural homology modelling; synthetic gene*

## INTRODUCTION

Protease is an enzyme that has the ability to hydrolyze peptide bonds in proteins (Sulthoniyah *et al.,* 2015). This enzyme is widely used by various bio-industries such as detergent making, leather tanning, food processing (Aruna *et al.,* 2014; Miguel *et al.,* 2013), drug manufacture, and biocatalysts in organic synthesis (Nigam, 2013). Protease sales occupy 60% of total commercial enzyme sales globally and are derived from microorganisms (Singh *et al.,* 2016a). The use of microorganisms as an enzyme-producing source is more beneficial than plants and animals since microorganisms are bred more easily at relatively more affordable costs (Raveendran *et al.,* 2018; Singh *et al.,* 2016b). In addition, the production of enzymes from microorganisms is easier to optimize through regulating conditions of growth and genetic engineering (Gurung *et al.,* 2013).

In general, the production of protease enzymes from bacteria is dominated by non-local bacteria, whereas Indonesia is rich in biodiversity that can act as a source of enzyme-producing microorganisms. Lactic acid bacteria (BAL), including *L. plantarum*, is one type of bacteria that has the potential as a source of protease. Bacteria *L. plantarum* IIA-1A5 is a probiotic isolated from fresh beef derived from a traditional market around Bogor, West Java, Indonesia (Arief *et al.,* 2015). Besides being a probiotic, these bacteria can also produce lactic acid and are able to produce bacteriocin plantarisin, which can inhibit the growth of pathogenic bacteria such as *E. coli*, *Salmonella typhimurium*, *Bacillus cereus*, and *S. aureus* (Arief *et al.,* 2013).

Afiyah *et al.* (2015) reported that the *L. plantarum* IIA-1A5 could hydrolyze meat protein from sarcoplasmic and myofibrillar fractions. This capacity is presumably because this strain secreted some proteases which can hydrolyze proteins in the muscles. Budiman *et al.* (2021) indicated that the genome sequence of this strain harbors some functional genes for foldases and proteases. One protease gene encodes a caseinolytic protease (Clp) which is known to be an essential protease involved in cell division, stress responses, and pathogenicity of bacteria (Frees *et al.,* 2014). The ability of Clp as a proteolytic system is due to the presence of proteolytic subunit (clpP), which belongs to a member of the serine protease family, with the canonical catalytic triad of Ser-

His-Asp (Florentin *et al.*, 2020). Apart from its cellular function for *L. plantarum* IIA-1A5, ClpP is promising to be harnessed as a bacterial protease for some applications. Nevertheless, the production of ClpP as a bacterial protease through direct extraction and purification is a considerably lengthy and tedious process. An alternative approach is to utilize recombinant technology, in which the gene is inserted in an appropriate expression vector, transformed into technically and economically affordable host cells (preferably *Escherichia coli*), over-expressed, and purified. Cardoso *et al.* (2020) confirmed that recombinant protein production is generally a simple and cost-effective production strategy.

Nevertheless, there is no study for the production of ClpP from *L. plantarum* group so far. Indeed, no protease from Indonesian LAB was produced through recombinant technology. The production of ClpP of *L. plantarum* IIA-1A5 is therefore considered a novel recombinant ClpP produced from LAB. In addition, protein production under recombinant technology is often halted by expression and solubility issues due to the incompatibility of the gene interest, the existence of rare codon, un-ideal GC ratio, or other related issues (Rosano & Ceccarelli, 2014; Razali *et al.*, 2021). Accordingly, the codon optimization strategy was introduced a few decades ago in which the sequence of gene interest is changed to mimic the codon preferences of the host (Rosano & Ceccarelli, 2014). This approach was reported to increase the expression of recombinant proteins by more than 1,000-fold (Mauro, 2018).

This study aims to construct the expression system and determine the expressibility of ClpP from *L. plantarum* IIA-1A5 (ClpP_LP) in *E. coli* BL21 (DE3) and its ability to degrade protein from livestock products, and predictions of its three-dimensional structure will also be performed. For the construction of the expression system, the gene encoding ClpP_LP was optimized in its codon to meet the ideal condition of *E. coli* expression. This is, to date, the first ClpP produced from the optimized codon approach and considered a novel approach in the field.

## MATERIALS AND METHODS

### Construction and Codon Optimization of ClpP_LP

The ClpP_LP synthetic gene was designed to be cloned into pET28a (+) expression vector, using two restriction enzymes. This process was carried out by GenScript (New Jersey, USA). The sequence of ClpP_LP gene from the RAST server was then used for the optimization process. The optimization process was carried out using the Optimum™ system from GenScript (New Jersey, USA). The optimization process was directed to three parameters: codon adaptation index (CAI), GC content, and frequency of codon usage (FOC). The optimized genes were then synthesized chemically and inserted into the pET28a (+) plasmid expression vector (Novagen) using *Nde*1 and *Bam*H1 restriction enzymes. *E. coli* host cells BL21 (DE3) were used for recombinant protein production. Recombinant plasmids were cut by two restriction enzymes *Mlu*1 (A|CGCGT) and *Bam*H1

(G|GATC C), to evaluate the success of target DNA insertion in the pET28a (+) plasmid. Restriction fragments were then observed using an agarose gel.

### Transformation of Recombinant Plasmids and Screening of Positive Transformants

The transformation of recombinant plasmids was performed according to the methods previously described by Murwantoko *et al*. (2016) and Rahimzadeh *et al.* (2016). A total of 5 μL recombinant plasmids DNA was dissolved in 100 μL of a competent cell suspension then homogenized. The mixture was incubated in the ice for 30 min, then heat-shocked at 42°C for 90 seconds using heat block, then placed back in ice for 30 min and incubated at 37°C for 1 hour by adding 1 mL of LB medium (without antibiotics). After 1 hour, 0.3 mL of the transformed cells were cultured in LB medium containing kanamycin antibiotics (35 μg/mL) to select positive transformants. The positive transformants were used for further expression experiments.

### Heterologous Expression of ClpP_LP in *E. coli* BL21 (DE3)

The recombinant plasmids (pET28a_ClpP_LP) were expressed in *E. coli* BL21 (DE3) according to the method described by Kang *et al.* (2015). Recombinant *E. coli* were cultured in 5 mL of LB (Luria Bertani) broth medium at 37°C containing kanamycin antibiotics at the final concentration of 35μg/mL overnight. The culture was transferred in a larger new medium (250 mL) containing kanamycin antibiotics at the final concentration of 35 μg/mL and incubated with a water-bath shaker at 37°C for 1 hour. The protein expression was induced by isopropyl-D-1-thiogalactopyranoside (IPTG) at the final concentration of 1 mM. The culture was further incubated at 37°C for 5 hours, and every 1 hour, as much as 50 mL of the culture was harvested. The cells were harvested by centrifugation at 13,000 rpm for 5 minutes at 4°C. The cell pellets were dissolved into a lysis buffer (10 mM Tris-HCl, pH 7.5) and were sonicated at 30 kHz amplitude for 30 min. Centrifugation at 13,000 rpm was used to separate between cell-free supernatant and insoluble protein aggregates. The resulting supernatant (soluble protein) and precipitate (insoluble/aggregated proteins) were analyzed by 15% SDS-PAGE.

### Analysis of Expression of Heterologous Proteins by SDS-PAGE

SDS-PAGE also analyzed the expression of the recombinant plasmids to validate the molecular weight of the protein produced following the procedure described by Fatmarani *et al*. (2018). The concentration of polyacrylamide used was 15% for the resolving gel and 5% for the stacking gel. Each sample (100 μL) was mixed with 200 μL of sample buffer and heated for 4 min at 95°C. Electrophoresis was performed with Mini Protean (Bio-Rad) at 120 V until the bromophenol blue marker reached the bottom of the gel (about 45 min). The gels were then stained for 40 minutes in 0.1% (w/v)

Coomassie Brilliant Blue R-250 (Sigma, St. Louis, MO, USA) in methanol: acetic acid: water (40: 10: 50) and then de-stained in methanol: acetic acid: distilled water (10: 10: 80) for 24 hours.

## Measurement of Protein Concentration

Proteins were analyzed based on the method described by Fatmarani *et al.* (2018). Each sample of 100 µL was homogenized with 3.9 mL of distilled water, and then 5.5 mL of reagent A containing 0.1 N NaOH; 2% $Na_2CO_3$; $CUSO_4$ 0.5%; and Tartat 1% was added. The mixture was then homogenized and allowed to stand for 15 minutes at room temperature (RT). The mixture of reagent B was added (folin-ciocalteu, 1:10) as much as 0.5 mL, then stirred immediately and allowed to stand for 30 minutes until the sample was blue. The concentration of protein was measured by spectrophotometer at a wavelength $\lambda$= 650 nm. Bovine Serum Albumin (BSA) was diluted sequentially with a concentration of 0.25 mg/mL, and it was used as the standard.

## Analysis of Proteolytic Activity

The analysis of proteolytic activity was carried out by the good diffusion method according to Rohin *et al.* (2012) with modifications. The extract of crude protease enzyme was put in sterile petri dishes containing LB medium that was enriched with 5% of specific substrate (milk and meat protein). A total of 60 µL of crude enzyme extract was dispensed in wells bored in the agar plates, then incubated at 37°C. Proteolytic activity is indicated by a clear visible zone around the wellbore. Furthermore, proteolytic activity was calculated by measuring the diameter of the clear area formed by using calipers and expressed as a protein-specific activity (AU $mg^{-1}$).

## Homology Modelling and Model Evaluation

The three-dimensional (3D) structure of ClpP_LP was constructed using the homology modeling method according to the method described by Waterhouse *et al.* (2018). SWISS-MODEL server (http://swissmodel.expasy.org/) was used for the ClpP_LP structure prediction. SWISS-MODEL is automatically equipped with a template selection feature by searching in parallel the best templates with BLAST (Camacho *et al.*, 2009) and Hhblits (Remmert *et al.*, 2011; Bienert *et al.*, 2017).

The protein model obtained was then evaluated by GMQE (Biasini *et al.*, 2014), QMEAN, and PROCHECK (Haddad *et al.*, 2020). GMQE is a quality estimation that combines properties from the target–template alignment and the template search method. QMEAN was used to evaluate the model based on different geometrical properties and provides both local (per residue) and global (for the entire structure) absolute quality estimates of different regions within the model (Haddad *et al.*, 2020). PROCHECK (Ramachandran plot) was used to analyze the stereochemical and overall quality of the structure (Shin *et al.*, 2017).

## RESULTS

### Codon Optimization and Cloning of ClpP_LP

The Codon sequence of ClpP_LP was analyzed using the Optimum™ system from GenScript (New Jersey, USA) (Figure 1). The result showed that its GC content was 48.08 %, with CAI of 0.51. Furthermore, the codons were optimized to match the *E. coli* codon preference. The result of optimization showed an increase of GC content to 53.62 % with CAI of 0.93.

The validation of recombinant plasmid was performed with 1% agarose gel electrophoresis. Double digest with restriction enzymes *Mlu*1 and *Bam*H1 on the recombinant plasmid yielded two DNA fragments. Figure 1 showed the presence of the inserted gene of ClpP in the expression system. The target gene was 591 bp, in its size. However, as the inserted check used *MIu*I and *Bam*HI, the digestion product had an accumulative size of about 1,500 bp. Figure 2(b), Lane 2, with the size of about 4500 bp showed two obvious bands with the apparent sizes of about 4500 and 1500 bp, which corresponded to the digested pET28a and ClpP_LP gene (with the addition of *Mlu*I and *Bam*HI fragments with the size of about 1000 bp). This confirmed that the recombinant plasmid indeed contained the inserted ClpP_LP gene. Meanwhile, the undigested recombinant plasmids showed multiples bands with apparent sizes in the range of 4000-6000 bp (Figure 2(b), Lane 1). This was due to the possible conformation of the plasmid in either supercoiled or open-circular. Under this conformation, the apparent size of the recombinant plasmid did not reflect its true size.

### Screening of Positive Transformants

The recombinant plasmids were expressed in the host cell of *E. coli* BL21 (DE3). Screening of positive transformants was done to ensure that the cell successfully captured the recombinant plasmids. The screening was done by growing *E. coli* in an LB medium containing kanamycin antibiotics of 35 µg/mL because plasmid pET28a (+) was equipped with a marker selection of genes that were resistant to kanamycin antibiotics. The positive transformant will grow to form colonies, as shown in Figure 3. Colonies that grew in media that have been added to kanamycin antibiotics were then collected and stored in the form of glycerol stock for further expression analysis.

### Analysis of Expression of Heterologous Proteins by SDS-PAGE

The results showed that adding isopropyl-D-1-thiogalactopyranoside (IPTG) as an inducer in a final concentration of 1 mM was successful in expressing the heterologous proteins in *E. coli* Bl21 cell hosts. In this study, *E. coli* Bl21 (DE3) was able to express the ClpP_LP at 5-hour induction as indicated by the presence of a new band that appeared only after IPTG induction (Figure 4A). In addition, the band only appeared in the soluble fraction, which indicated that ClpP_LP was
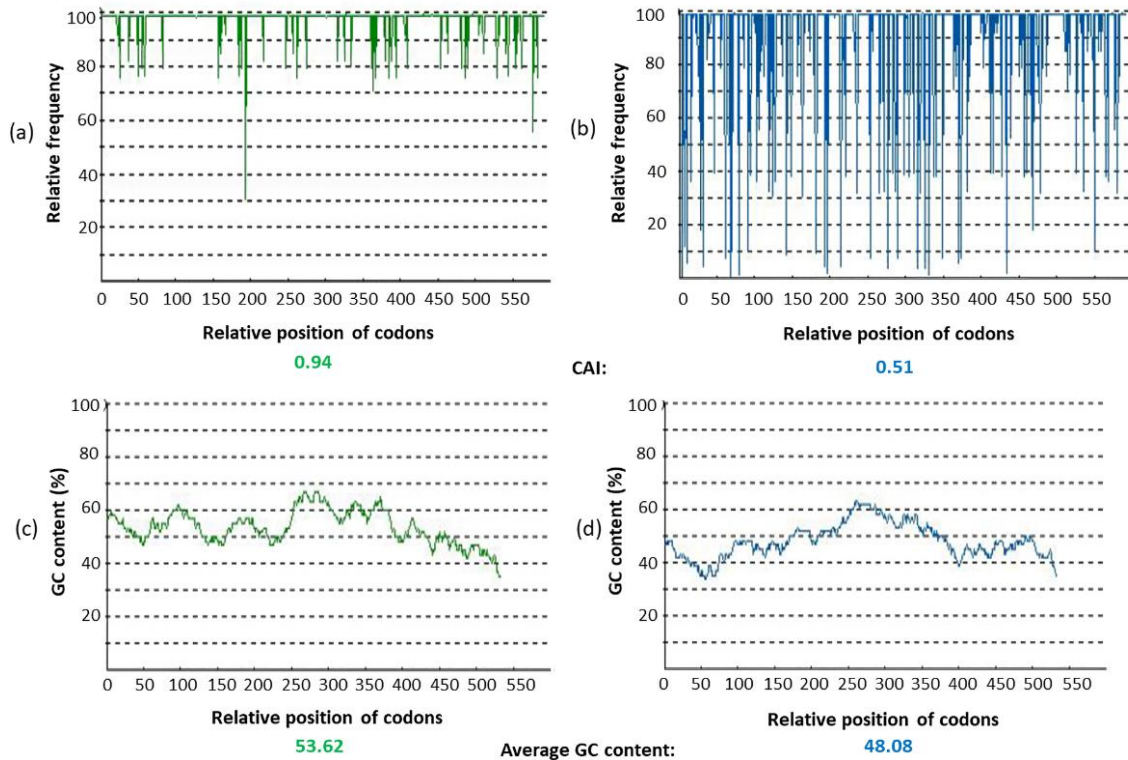
Figure 1. The value of Codon Adaptation Index (CAI) and GC content of ClpP_LP. (a) CAI after optimization, (b) CAI before optimization, (c) GC content after optimization, and (d) GC content before optimization.
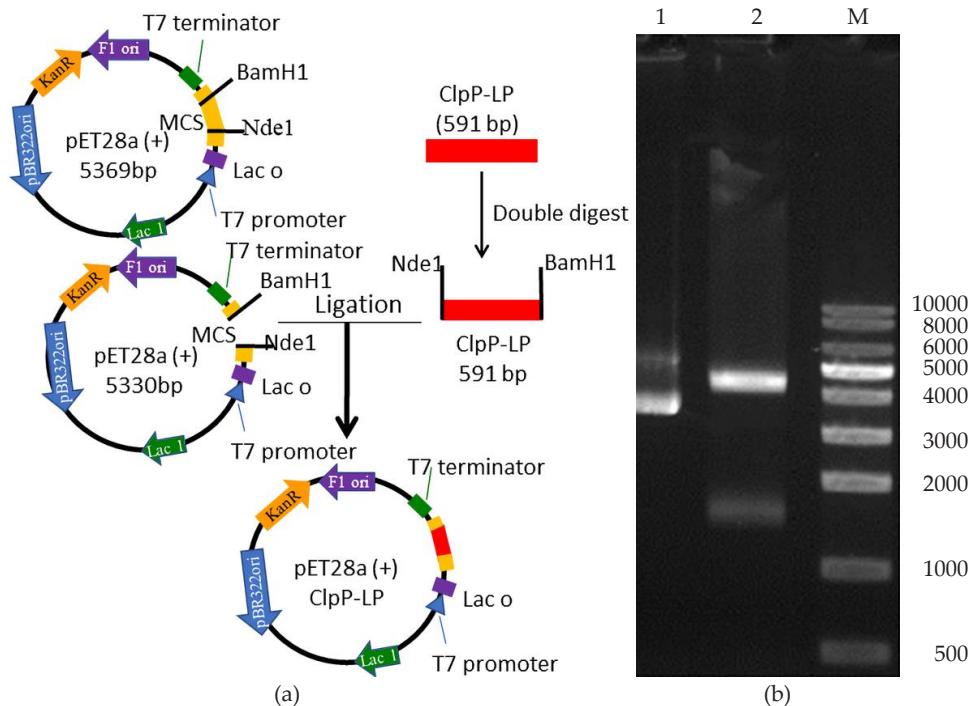


Figure 2. Construction of pET28a_ClpP_LP (a), Restriction analysis of recombinant plasmids by using 1% agarose gel electrophoresis (b). Line 1 is a non-digesting recombinant plasmid, line 2 is a recombinant plasmid digest with *Mul*1 and *Bam*H1, line M is a KB Ladder marker.

expressed in a fully soluble form. To note, Figure 4A indicated that ClpP_LP appeared as a double band on the SDS-PAGE, in which the upper and lower bands have relative mobility (Rf) in the gel of 0.55 and 0.56, respectively. Calculation of the apparent molecular sizes of the band of ClpP_LP based on the Rf values revealed that both bands had similar sizes 25 kDa (upper band) and 24.1 kDa (lower band). These sizes were slightly larger than the theoretical size of ClpP_LP, which was calculated based on its amino acid sequence (23.2 kDa).

However, this small difference is acceptable and often happens due to the unique behavior of proteins during migration in the gel.

### Protein Specific Activity

The analysis showed that the specific activity of ClpP_LP protein was higher than chymotrypsin and *E. Coli* (Figure 5). On the other hand, ClpP_LP and chymotrypsin also showed a trend of increasing protein-specific activity along with incubation time. In this study, the control treatment with *E. coli* extracts without the ClpP_LP gene showed no specific protein activity. This meant that the specific protein activity of ClpP_LP purely came from the ClpP_LP encoding gene.

### Homology Modelling and Model Evaluation

The target sequence of ClpP_LP with 196 amino acids was retrieved from the Rapid Annotation using Subsystem Technology (RAST) Annotation Server. The structure was modeled using SWISS-MODEL on the web at http://swissmodel.expasy.org. The struc-

tural evaluation and stereo-chemical analysis of the selected model were performed using QMEAN and PROCHECK, respectively. The identification of the template using BLAST and HHblits has been performed against the SWISS-MODEL template library, which produced as many as 638 templates. The results of template identification produced the best template in the form of 3TT7 protein (Figure 6). The selection of the best template structure was based not only on the similarity of the sequence but also on other factors such as experimental quality, bound substrate molecules, or different template conformations (Waterhouse *et al.*, 2018).

Evaluation of the protein model showed that the average QMEAN (Qualitative Model Energy Analysis) model was 0.92, which was generally greater than the minimum QMEAN value for a structure model. This indicated that the model was of good structure and acceptable. The Ramachandran plot generated by PROCHECK showed 94.2% amino acids were in the favored region, and 0.3% were in the outlier region based on Phi and Psi angles (Figure 7). Altogether, nearly 100% of the residues were in the favored and allowed regions indicating that the protein model was good.

### The ClpP_LP Active Side Analysis and Substrate Cutting

The analysis of the active side of ClpP_LP was carried out by aligning the amino acid constituents of ClpP_LP with the template protein (3TT7) and Clp from several different bacteria whose active side was previously known. The results of ClpP_LP alignment by Clustal Omega showed that the active side of ClpP_LP consisted of three amino acids (triad catalytic). The triad catalytic of ClpP_LP was located in the Ser97, His122, and Asp171 (Figure 10). The homo tetradecameric structure of ClpP_LP formed a barrel-like structure that had two tunnels opening (top and bottom) for entry of protein and exit of the products (Figure 11). The measurement of the diameter of the tunnel in the ClpP_LP homo-heptameric structure was 29.5 Å. This size was considered for several types of proteins substrate to be degraded by these proteases in the cells.
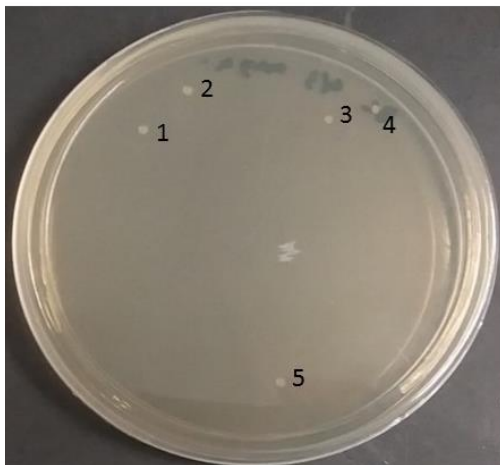


Figure 3. Screening of *E. coli* Bl21 (DE3) harboring recombinant protease gene onto LB agar plates containing kanamycin antibiotics (35 µg/mL).
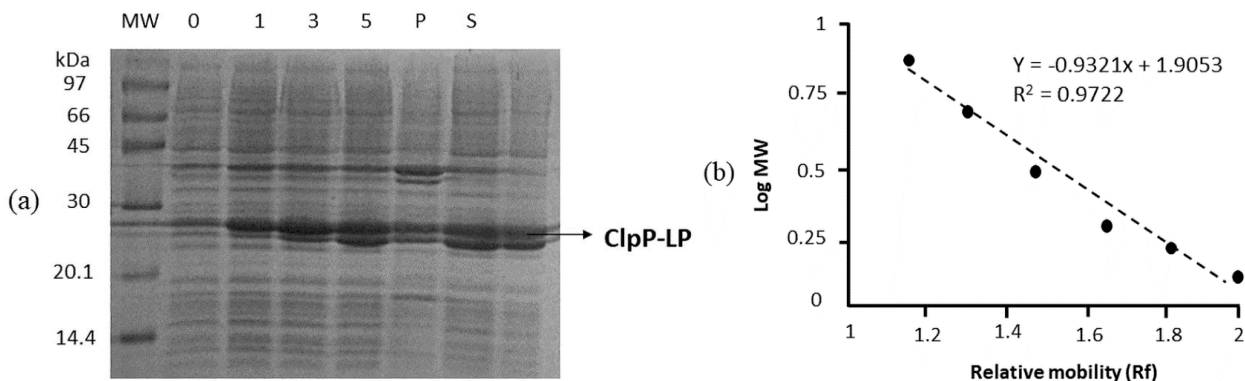


Figure 4. (a). Expression profile of ClpP_LP under SDS-PAGE. MW is a marker. Lane 0 is *E. coli* BL21(DE3) without IPTG induction. Lanes 1, 3, and 5 correspond to *E. coli* BL21(DE3) after being induced by IPTG for 1, 3, and 5 hours. Lanes P and S correspond to pellet and soluble fractions of the 5-hour induced *E. coli* BL21(DE3) after being lysed. (b). Relationship of the relative mobility (Rf) and molecular weight of protein markers derived from SDS-PAGE. The linear equation was used for the calculation of apparent size of ClpP_LP from SDS-PAGE.
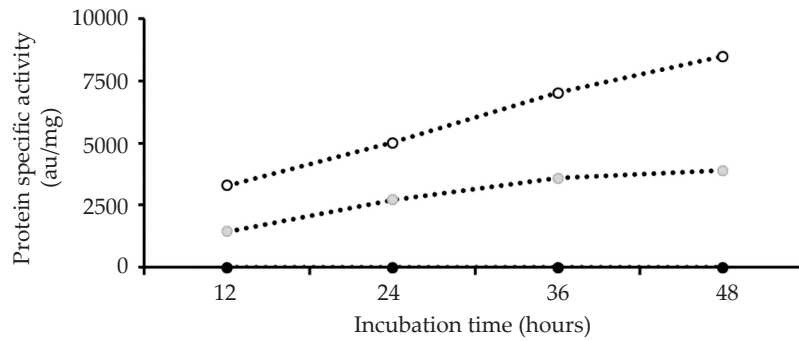
Figure 5. Specific activity of ClpP_LP protein in LB medium containing 5% skim milk;
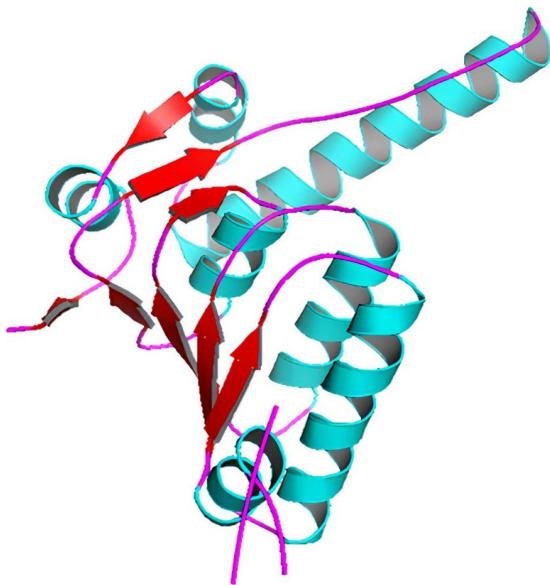ClpP_LP (…o…), Chymotrypsin (…●…), Protein extract of *E. coli* (…●…).



Figure 6. Three-dimensional structure of the 3TT7 template protein
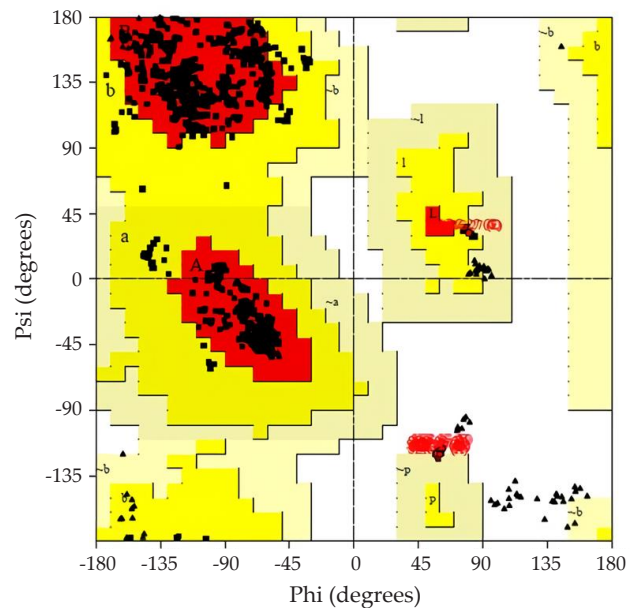


Figure 7. Ramachandran plot of ClpP_LP obtained through the modelling

## DISCUSSION

In this study, the expression level of the optimization of genes synthesis is quite good because it has a CAI value greater than 0.8. The GC content of the ClpP_LP recombinant genes is also quite good, where the percentage of GC ideally ranges from 30%-70%. The Codon Adaptation Index (CAI) is the primary index used to predict gene expression levels of heterologous proteins (Fu *et al.*, 2020), with the highest values being 1.0 and CAI values above 0.8 being good. Heterologous proteins will not be expressed as well if the target gene is found in many rare codons used in the host cells. It also induces a strong host cell metabolic burden (Rahmen *et al.*, 2015).

Codon optimization through the in-silico method is a common strategy to increase the expression of heterologous proteins without changing the amino acid sequence of the encoded protein (Gaspar *et al.*, 2012). The in silico (bioinformatics) method is a computational-based analysis method combined with mathematical,

chemical, and biological algorithms to predict and analyze a protein's structure and activity (Kuhlman & Bradely, 2019). The application of this technology has been utilized to generate recombinant proteins for use in pharmaceuticals (Liu *et al.*, 2018). It is more effective and efficient than direct gene isolation and can prevent disease transmission (Hughes *et al.*, 2011). In addition, the in-silico approach can save time and costs needed in a study (Yaraguppi *et al.*, 2012).

The cloning strategy was depicted in Figure 2 (a). In this study, restriction enzymes used were *Nde*1 and *Bam*H1, and both enzymes can be recognized restriction sites in the Multiple Cloning Site (MCS) of plasmid pET28a (+). The *Nde*1 restriction enzyme cuts the DNA fragment in the 5'-CA↓TATG-3' sequence and its complement 3'-GTAT↓AC-5' (Watson *et al.*, 1982), meanwhile the restriction enzyme *Bam*H1 intersects the DNA fragment in the 5'-G↓GGATCC-3' sequence and its complement of 3'-CCTAG↓G-5' (Newman *et al.*, 1994). Plasmid and synthesis DNA cutting with both enzymes will produce a fragment with complementary

Figure 8. The predicted model of ClpP_LP in the forms of barrel-shaped homo-heptameric. The red structure indicates protein monomer.
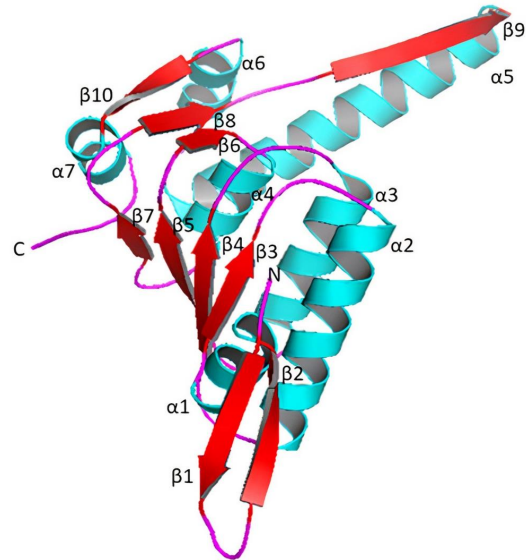


Figure 9. ClpP_LP monomer is dominated by β-sheet structures. Each of the structure identified and labeled as α-helix and β-sheet. Both of N-terminal and C-terminal labeled as N and C, respectively.
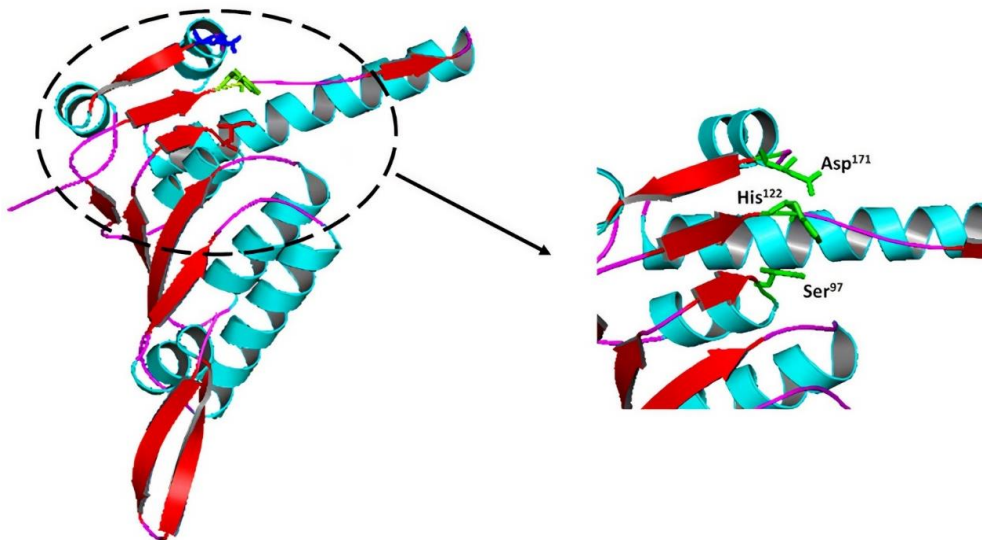


Figure 10. The residues catalytic triad in ClpP_LP. The black circle indicates the position substrate binding to the catalytic triads.

sticky tips so that DNA can be inserted into the plasmid to form a relatively stable hydrogen bond. Sticky ends are unpaired nucleotides at the ends of DNA molecules that can associate to link DNA segments with a plasmid (Ban & Picu, 2013). It will also facilitate the performance of ligase enzymes in connecting DNA with plasmids through phosphodiester bonds.

The success of heterologous proteins expression was influenced by many factors such as the good expression vector choice, the suitable cell hosts, using induction, time of induction, and codon optimization. In this study, we used the pET28a (+) (Novagen) as the expression vector. The pET expression system is based on the T7 promoter of T7 bacteriophage that is common for recombinant proteins expression (Fiege & Dinkel, 2020).

In this system, the gene of interest is cloned behind a promoter recognized by the phage T7 RNA polymerase (T7 RNAP), so that is why we use *E. coli* BL21 (DE3) as a cell host. The concentration of inducer used plays an important role in the expression process where the use of inducers in low concentrations causes inefficient induction, whereas if using high concentrations can also cause toxic effects that inhibit the growth of the host cells (Ramirez *et al.*, 1994; Browning *et al.*, 2019). However, low concentrations are more beneficial because they will reduce the cost of producing these proteins. In this study, a concentration of 1 mM IPTG is enough to induce the expression of these genes. This concentration is considered normal because many recombinant proteins are expressed using IPTG at this concentration.
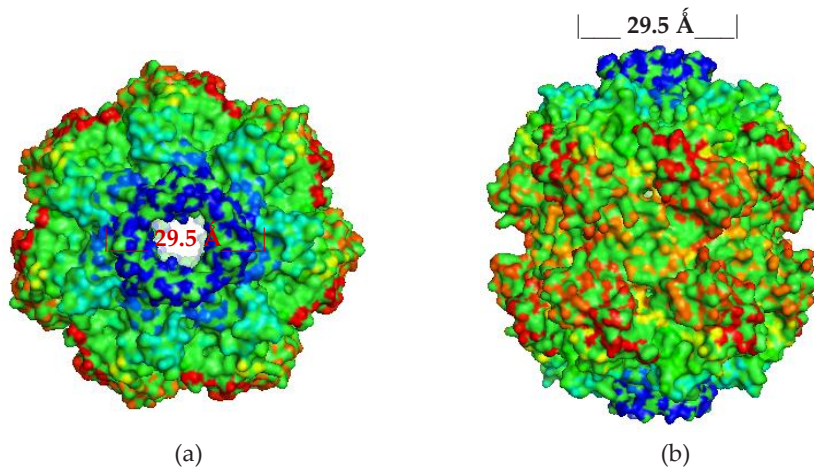
29.5 Å

29.5

(a)                                        (b)

Figure 11. Three-dimensional structure of 14-mer ClpP_LP. It shows the upper view (a) and side view (b) of the homo-heptameric ClpP_LP.

| Name of enzyme | No. of cleavages | Positions of cleavage sites |
|---|---|---|
| ClpP-LP | 13 | 3 29 34 37 42 45 55 89 107 118 132 133 137 |

```
                                              ClpP-LP
                                          ClpP-LP  |
                                   ClpP-LP   |    |
                              ClpP-LP    |   |    |
 ClpP-LP              ClpP-LP     |     |   |    |      ClpP-LP
    |                   |         |     |   |    |         |
    MMKSFFLVVTILALTLPFLGAQEQNQEQRICCEKDERFFDDKIAKYIPIQYVLSRYPSYG
  1 ---------+---------+---------+---------+---------+---------+   60


                        ClpP-LP           ClpP-LP        ClpP-LP
                           |                 |              |
    LNYYQQRPVALINNQFLPYPYYAKPVAVRSPAQTLQWQVLPNAVPAKSCQDQPTAMARHP
 61 ---------+---------+---------+---------+---------+---------+   120

                   ClpP-LP
             ClpP-LP   |
             ClpP-LP   |

               ||   |
    HPHLSFMAIPPKKDQDKTEIPAINTIASAEPTVHSTPTTEAVVNAVDNPEASSESIASAP
121 ---------+---------+---------+---------+---------+---------+   180



    ETNTAQVTSTEV
181 ---------+--   192
```
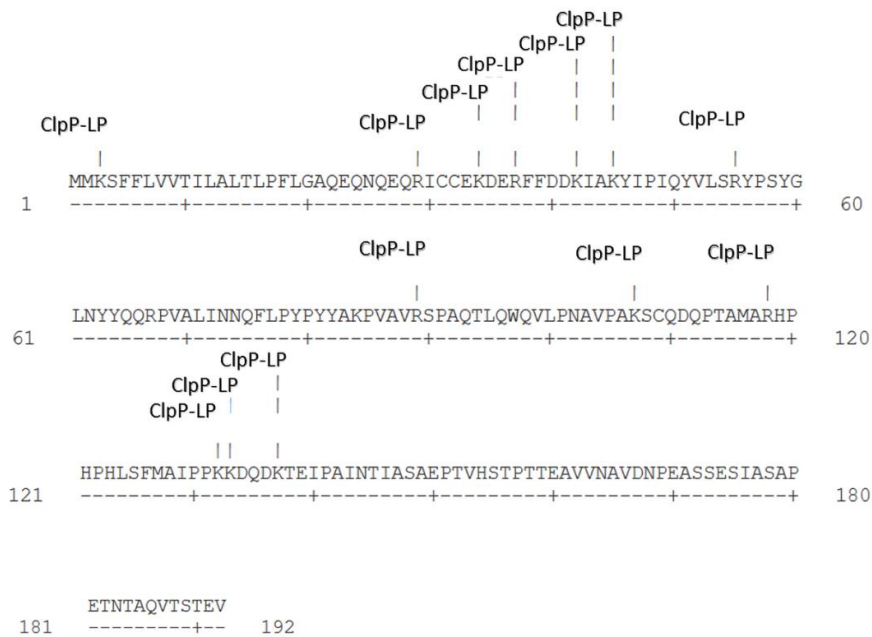
Figure 12. Map of ClpP_LP cutting in sequences of casein

In this study, the specific activity of proteins from ClpP_LP was better than control because skim milk contained a lot of casein protein, a specific substrate of ClpP_LP. Casein is the major protein in milk, divided into α-, β- and κ-casein (Rehan *et al.*, 2019). To note, casein is known to be the natural substrate for ClpP, which is therefore named as 'caseinolytic' protease. Nevertheless, the structural homology modeling indicates that the tunnel diameter of ClpP_LP is smaller than the micelle structure of casein (50-500 nm). Accordingly, it is not possible for ClpP_LP to degrade casein in its micelle-like structures due to size constrain. It is, therefore, proposed that ClpP_LP accommodates casein in its sub-micelle structure. Previously, crude protein extract of *L. plantarum* IIA-1A5 showed the ability to degrade myofibrillar proteins and sarcoplasmic meat (Afiyah *et al.*, 2015). This nevertheless remains to be confirmed whether the ability is modulated by ClpP_LP or other proteases of this strain.

The model generated by SWISS-MODEL server used ClpP of *Bacillus subtilis* (PDB ID: 3TT7), which was obtained from X-ray diffraction at 2.56 Å (Lee *et al.*, 2011). The template showed a sequence identity of 65.13% with ClpP_LP, with coverage of 99%. The value of sequence identity between the target protein and the template is the first assessment that must be done on

the quality of the model (Bordoli *et al.*, 2009). Overall, ClpP of *Bacillus subtilis* is considered as an acceptable template for structural modeling of ClpP_LP.

The prediction model of the target protein (ClpP_LP) using the SWISS-MODEL server produced a three-dimensional (3D) structure of the target protein, as was described in Figure 8. This protein model also showed the possibility of this protein forming an oligomeric structure composed of 14 identical protein monomers (homo-tetradecameric). The model showed that ClpP_LP was in the form of a barrel, which was the same as the 3TT7 template protein (Lee *et al.*, 2011). The structure of ClpP resembles a barrel, where the protein that will be degraded first is translocated into a protease and will be released in the form of chopped amino acids.

The monomer of ClpP_LP is dominated by the β-sheet structures (Figure 9). Seven α-helixes were found in this protease. The first α-helix was extended from amino acid of $I_{19}$ to $D_{26}$, the second α-helix was extended from amino acid of $D_{37}$ to $Q_{53}$, the third α-helix was extended from amino acid of $I_{70}$ to $F_{82}$, the fourth α-helix was extended from the amino acid of $M_{98}$ to $G_{105}$, the fifth α-helix was extended from amino acid of $Q_{132}$ to $T_{157}$, the sixth α-helix was extended from amino acid of $I_{161}$ to $T_{168}$, and the seventh α-helix was extended from amino acid of $A_{176}$ to $Y_{182}$. Ten β-sheet were found in this protease. The first β-sheet was extended from amino acid of $T_5$ to $T_{10}$, the second β-sheet was extended from amino acid of $G_{13}$ to $D_{18}$, the third β-sheet was extended from amino acid of $I_{28}$ to $L_{31}$, the fourth β-sheet was extended from amino acid of $I_{59}$ to $S_{65}$, the fifth β-sheet was extended from amino acid of $V_{87}$ to $A_{91}$, the sixth β-sheet was extended from amino acid of $M_{94}$ to $A_{96}$, the seventh β-sheet was extended from amino acid of $R_{111}$ to $A_{113}$, the eighth β-sheet was extended from amino acid of $T_{118}$ to $I_{121}$, the ninth β-sheet was extended from amino acid of $G_{127}$ to $Q_{129}$, and the tenth β-sheet was extended from amino acid of $H_{172}$ to $S_{175}$. The β-sheets of β-1 and β-2 were found to be in the antiparallel position, and this condition was also found in β-8 and β-10. Meanwhile, β-3, β-4, β-5, and β-7 were found to be in parallel position. The antiparallel β-sheets are more stable than parallel β-sheets as the hydrogen-bonding pattern in the parallel sheet is more optimal (Roeters *et al.*, 2017). The existence of antiparallel position of β-sheets is therefore required to maintain the structural stability of the barrel form of ClpP_LP. The barrel form of ClpP_LP is functionally important to serve as a degradation containment area. The hydrophobic environment inside the barrel is also functionally required for facilitating the binding to the substrate through hydrophobic interactions.

Figure 10 indicates that the catalytic site of ClpP_LP is highly conserved to the other serine proteases. This indicates that the ClpP_LP utilizes similar catalytic mechanisms to the other serine proteases, where Serine 97 plays a role as a nucleophile. Meanwhile, His122 plays an important role in the deprotonation of Ser97 for nucleophilic formation. The role of His122 is supported by the presence of Asp171, which forms a hydrogen bond with the imidazole ring of His122 so that it is in the proper orientation for deprotonation of Ser97. The

presence of Serine amino acids in the catalytic site of this protein makes ClpP_LP classified as a serine-protease group (Li *et al.*, 2013).

In addition, the results of the ClpP_LP intersection of the casein protein substrate using the Peptide Cutter server showed that ClpP_LP only cuts peptide bonds in the amino acid Lysin (K) or Arginine (R), which both amino acids are many found in the casein protein. The process of cutting peptide bonds by ClpP_LP can be seen simply in Fig. 12. The number of amino acids Lysine and Arginine in casein protein is 13 residues, so ClpP_LP will cut the casein substrate at 13 points. These cuttings consist of $K_3$, $R_{29}$, $K_{34}$, $R_{37}$, $K_{42}$, $R_{55}$, $R_{89}$, $K_{107}$, $R_{118}$, $K_{132}$, $K_{133}$, and $K_{137}$.

## CONCLUSION

This study showed that the optimization codon approach and synthesis genes could be used in the production of ClpP_LP proteins using inexpensive host cells, *E. coli* BL21 (DE3). The ClpP_LP produced in this study has a molecular weight of about 25 kDa and showed a real activity against milk casein. Three-dimensional structure modeling using the SWISS_MODEL server showed that this protein is composed of a combination of 7 α-helix and 10 β-sheets and formed a 14-mer oligomer in the form of a barrel. The active side analysis using Clustal Omega showed that ClpP_LP belongs to the serine protease, with the active side of this protein located at Serine-97, Histidine-122, and Aspartic Acid-171 and only recognizing specific cutting sites on amino acids Lysine (K) or Arginine (R).

## CONFLICT OF INTEREST

The authors (C. Budiman, I. I. Arief, & C. Sumantri) serve as editorial boards of the Tropical Animal Science Journal, but they have no role in the decision to publish this article. We also certify that no conflict of interest with any financial, personal, or other relationships with other people or organizations related to the material discussed in the manuscript.

## ACKNOWLEDGEMENT

## REFERENCES

**Afiyah, D. N., I. I. Arief, & C. Budiman**. 2015. Proteolytic characterization of trimmed beef fermented sausages inoculated by Indonesian probiotics: *Lactobacillus plantarum* IIA-2C12 and *Lactobacillus acidophilus* IIA-2B4. Adv. J. Food. Sci. Technol. 8:27-35. https://doi.org/10.19026/ajfst.8.1459

**Arief, I. I., B. S. L. Jenie, M. Astawan, K. Fujiyama, & A. B. Witarto.** 2015. Identification and probiotic characteristics of lactic acid bacteria isolate from Indonesian local beef. Asian J. Anim. Sci. 9:25-36. https://doi.org/10.3923/ajas.2015.25.36

**Arief, I. I., Jakaria, T. Suryati, Z. Wulandari, & E. Andreas.**

2013. Isolation and characterization of plantaricin produced by *Lactobacillus plantarum* Strain (IIA-1A5, IIA-1B1, IIA-2B2). Med. Pet. 36:91-100. https://doi.org/10.5398/medpet.2013.36.2.91

**Aruna, K., J. Shah, & R. Birmole.** 2014. Production and partial char-acterization of alkaline protease from Bacillus tequilensis strains CSGAB 0139 isolated from spoilt cottage cheese. Int. J. Appl. Biol. Pharm. 5:201-21.

**Ban, E. & C. R. Picu.** 2013. Strength of DNA sticky end links. Biomacromolecules 15:143–149. https://doi.org/10.1021/bm401425k

**Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, & L. Bordoli.** 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42: W252-W258. https://doi.org/10.1093/nar/gku340

**Bienert, S., A. Waterhouse, T. A. P. de Beer, G. Tauriello, G. Studer, L. Bordoli, & T. Schwede.** 2017. The SWISS-MODEL repository--new features and functionality. Nucleic Acids Res. 45:D313-D319. https://doi.org/10.1093/nar/gkw1132

**Bordoli, L., F. Kiefer, K. Arnold, P. Benkert, J. Battey, & T. Schwede.** 2009. Protein structure homology modeling using SWISS-MODEL workspace. Nat. Protoc. 4:1-13. https://doi.org/10.1038/nprot.2008.197

**Browning, D. F., R. E. Godfrey, K. L. Richards, C. Robinson, & S. J. W. Busby.** 2019. Exploitation of the *Escherichia coli lac* operon promoter for controlled recombinant protein production. Biochem. Soc. Trans. 47:755–763. https://doi.org/10.1042/BST20190059

**Budiman, C., I. I. Arief, F. Opook, & M. Yusuf.** 2021. A meat-derived lactic acid bacteria, *Lactobacillus plantarum* IIA, expresses a functional parvulin-like protein with unique structural property. Online J. Biol. Sci. 21:120-135. https://doi.org/10.3844/ojbsci.2021.120.135

**Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, & T. L. Madden.** 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10, 421. https://doi.org/10.1186/1471-2105-10-421

**Cardoso, V. M., G. Campani, M. P. Santos, G. G. Silva, M. C. Pires, V. M. Gonçalves, R. de C. Giordano, C. R. Sargo, A. C. L. Horta, & T. C. Zangirolami.** 2020. Cost analysis based on bioreactor cultivation conditions: Production of a soluble recombinant protein using *Escherichia coli* BL21(DE3). Biotechnol. Rep. 26:1-13. https://doi.org/10.1016/j.btre.2020.e00441

**Fatmarani, R., I. I. Arief, & C. Budiman.** 2018. Purification of bacteriocin from *Lactobacillus plantarum* IIA-1A5 grown in various whey cheese media under freeze dried condition. Trop. Anim. Sci. J. 41:191-199. https://doi.org/10.5398/tasj.2018.41.1.53

**Fiege, K., & N. F. Dinkel.** 2020. Construction of a new T7 promoter compatible *Escherichia coli* Nissle 1917 strain for recombinant production of heme dependent proteins. Microb. Cell Fact. 19:190. https://doi.org/10.1186/s12934-020-01447-5

**Florentin, A., D. R. Stephens, C. F. Brooks, R. P. Baptista, & V. Muralidharan.** 2020. Plastid biogenesis in malaria parasites requires the interactions and catalytic activity of the Clp proteolytic system. PNAS. https://doi.org/10.1073/pnas.1919501117

**Frees, D., U. Gerth, & H. Ingmer.** 2014. Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of *Staphylococcus aureus*. Int. J. Med. Microbiol. 304:142-149. https://doi.org/10.1016/j.ijmm.2013.11.009

**Fu, H., Y. Liang, X. Zhong, Z. Pan, L. Huang, H. Zhang, Y. Xu, W. Zhou, & Z. Liu.** 2020. Codon optimization with deep learning to enhance protein. Sci. Rep. 10: Article number 17617. https://doi.org/10.1038/s41598-020-74091-z

**Gaspar, P., J. L. Oliveira, J. Frommlet, M. A. Santos, & G. Moura.** 2012. EuGene: maximizing synthetic gene design for heterologous expression. Bioinformatics 28:683-2684. https://doi.org/10.1093/bioinformatics/bts465

**Gurung, N., S. Ray, S. Bose, & V. Rai.** 2013. A broader view: Microbial enzymes and their relevance in industries, medicine, and beyond. Biomed Res. Int. 2013:1-18. https://doi.org/10.1155/2013/329121

**Haddad, Y., V. Adam, & Z. Heger.** 2020. Ten quick tips for homology modeling of highresolution protein 3D structures. Plos Comput. Biol. 16:e1007449. https://doi.org/10.1371/journal.pcbi.1007449

**Hughes, R. A., A. E. Miklos, & A. D. Ellington.** 2011. Gene synthesis: methods and applications. Methods Enzymol. 498:277-309. https://doi.org/10.1016/B978-0-12-385120-8.00012-7

**Kang, Y. S., J. A. Song, K. Y. Han, & J. Lee.** 2015. *Escherichia coli* EDA is a novel fusion expression partner to improve solubility of aggregation-prone heterologous proteins. J. Biotechnol. 194:39-47. https://doi.org/10.1016/j.jbiotec.2014.11.025

**Kuhlman, B. & P. Bradley.** 2019. Advances in protein structure prediction and design. Nat. Rev. Mol. Cell Biol. 20:681-697. https://doi.org/10.1038/s41580-019-0163-x

**Lee, B. G., M. K. Kim, & H. K. Song.** 2011. Structural insights into the conformational diversity of ClpP from *Bacillus subtilis*. Mol. Cells. 32:589-595. https://doi.org/10.1007/s10059-011-0197-1

**Li, Q., L, Yi, P. Marek, & B. L. Iverson.** 2013. Commercial proteases: Present and future. FEBS Lett. 587:1155-1163. https://doi.org/10.1016/j.febslet.2012.12.019

**Liu, B., Q. Kong, D. Zhang, & L. Yan.** 2018. Codon optimization significantly enhanced the expression of human 37-kDa iLRP in *Escherichia coli*. 3 Biotech 8:Article number 210. https://doi.org/10.1007/s13205-018-1234-y

**Mauro, V. P.** 2018. Codon optimization in the production of recombinant biotherapeutics: Potential risks and considerations. BioDrugs 32:69-81. https://doi.org/10.1007/s40259-018-0261-x

**Miguel, Â. S. M., T. S. Martins-Meyer, E. Veríssimo da Costa Figue-iredo, B. W. P. Lobo, & G. M. Dellamora-Ortiz.** 2013. Enzymes in bakery: Current and future trends. In: Muzzalupo I (Eds). Food Industry. InTech, Rijeka, Croatia.

**Murwantoko., C. K. Fusianto, & Triyanto.** 2016. Gene cloning and protein expression of koi herpesvirus ORF25. Hayati 23:143-149. https://doi.org/10.1016/j.hjb.2016.10.001

**Newman, M., T. Strzelecka, L. F. Dorner, Schildkraut, & A. K. Aggarwal.** 1994. Structure of endonuclease BamH1 and its relationship to EcoR1. Nature 368:660-664. https://doi.org/10.1038/368660a0

**Nigam, P. S.** 2013. Microbial enzymes with special characteristics for biotechnological applications. Biomolecules 3:597-611. https://doi.org/10.3390/biom3030597

**Rahimzadeh, M., M. Sadeghizadeh, F. Najafi, S. S. Arab, & H. Mobasheri.** 2016. Impact of heat shock step on bacterial transformation efficiency. Mol. Biol. Res. Commun. 5:257-261.

**Rahmen, N., C. D. Schlupp, H. Mitsunaga, A. Fulton, T. Aryani, L. Esch, U. Schaffrath, E. Fukuzaki, K. E. Jaeger, & J. Büchs.** 2015. A particular silent codon exchange in a recombinant gene greatly influences host cell metabolic activity. Microb. Cell. Fact. 14:156. https://doi.org/10.1186/s12934-015-0348-8

**Ramirez, O., R. Zamora, G. Espinosa, E. Merino, F. Bolivar, & R. Quintero.** 1994. Kinetic study of penicillin acylase production by recombinant *E. coli* in batch cultures. Process Biochem. 29:197-206. https://doi.org/10.1016/0032-9592(94)85004-6

**Raveendran, S., B. Parameswaran, S. B. Ummalyma, A.**

Abraham, A. K. Mathew, A. Madhavan, S. Rebello, & A. Pandey. 2018. Applications of microbial enzymesin food industry. Food Technol. Biotechnol. 56:16-30. https://doi.org/10.17113/ftb.56.01.18.5491

Razali, R., C. Budiman, K. A. Kamaruzaman, & V. K. Subbiah. 2021. Soluble expression and catalytic properties of codon-optimized recombinant bromelain from MD2 pineapple in *Escherichia coli*. Protein J. 40:406-418. https://doi.org/10.1007/s10930-021-09974-9

Rehan, F., N. Ahemada, & M. Gupta. 2019. Casein nanomicelle as an emerging biomaterial-A comprehensive review. Colloids Surf. B: Biointerfaces. 179:280-292. https://doi.org/10.1016/j.colsurfb.2019.03.051

Remmert, M., A. Biegert, A. Hauser, & J. Soding. 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods. 9:173-175. https://doi.org/10.1038/nmeth.1818

Roeters, S. J., A. Iyer, G. Pletikapić, V. Kogan, V. Subramaniam, & S. Woutersen. 2017. Evidence for intramolecular anti-parallel beta-sheet structure in alpha-synuclein fibrils from a combination of two-dimensional infrared spectroscopy and atomic force microscopy. Sci. Rep. 7:4105. https://doi.org/10.1038/srep41051

Rohin, M. A. K., A. Bakar, C. Abdullah, & A. M. Ali. 2012. Antibacterial activity of flesh and peel methanol fractions of red pitaya, white pitaya and papaya on selected food microorganisms. Int. J. Pharm. Pharm. Sci. 4:185-190.

Rosano, G. L., & E. A. Ceccarelli. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front.

Microbiol. 5:172. https://doi.org/10.3389/fmicb.2014.00172

Shin, W. H., X. Kang, J. Zhang, & D. Kihara. 2017. Prediction of local quality of protein structure models considering spatial neighbors in graphical models. Sci. Rep. 7:40629. https://doi.org/10.1038/srep40629

Singh, R., A. Mittal, M. Kumar, & P. K. Mehta. 2016. Microbial proteases in commercial applications. J. Pharm. Chem. Biol. Sci. 4:365-374.

Singh, R., M. Kumar, A. Mittal, & P. K. Mehta. 2016. Microbial enzymes: Industrial progress in 21st century. 3 Biotech 6:174. https://doi.org/10.1007/s13205-016-0485-8

Sulthoniyah, S. T. M., Hardoko, & H. Nursyam. 2015. Characterization of extracellular protease lactic acid bacteria from shrimp paste. J. Life Sci. Biomed. 5:01-05.

Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, & T. Schwede. 2018. SWISS- MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46:W296-W303. https://doi.org/10.1093/nar/gky427

Watson, R. J., I. Schildkraut, B. Q. Qiang, S. M. Martin, & L. P. Visentin. 1982. NdeI: A restriction endonuclease from Neisseria denitrificans which cleaves DNA at 5′-CATATG-3′ sequences. Febs letters. 150:114-116. https://doi.org/10.1016/0014-5793(82)81315-X

Yaraguppi, D. A., B. B. Udapudi, L. R. Patil, V. S. Hombalimath, & A. R. Shet. 2012. In-silico analysis for predicting protein ligand interaction for snake venom protein. J. Adv. Bioinforma. Appl. Res. 3:345-356.