

Codon Optimization and Chaperone Assisted Solubilization of Recombinant Human Prethrombin-2 Expressed in *Escherichia coli*

SARONOM SILABAN^{1,2}, IMAN PERMANA MAKSUM¹,
SHABARNI GHAFAR¹, KHOMAINI HASAN³, SUTARYA ENUS⁴, TOTO SUBROTO¹,
AND SOETIJOSO SOEMITRO^{1*}

¹Biochemistry Laboratory, Department of Chemistry, Universitas Padjadjaran, Bandung, Indonesia;

²Department of Chemistry, Universitas Negeri Medan, Medan, Indonesia;

³Bioscience and Biotechnology Research Center, Institut Teknologi Bandung, Indonesia;

⁴National Eye Hospital, Rumah Sakit Mata Cicendo, Bandung, Indonesia;

Prethrombin-2 (PT2) is a thrombin precursor, which plays a role in the conversion of fibrinogen into fibrin during blood clotting process. Previous study reported that the expression of human prethrombin-2 (rhPT2) in *Escherichia coli* formed inclusion bodies. The aim of this study was to establish a strategy to express a soluble rhPT2 in *E. coli*. This study was aimed to design and optimize the codon of human prethrombin-2 gene as well as to optimize the expression analyses condition using four strains of *E. coli*. The codon adaptation index (CAI) of the unoptimized *hpt2* gene was 0.336, with 56.8% GC content. After optimization, the CAI of optimized *hpt2* became 1.000 with 53.1% GC content. The optimized gene was successfully cloned into pTWIN1 expression vector. Expression analysis indicated that only *E. coli* ArcticExpress strain could successfully express a soluble rhPT2 protein, with only small part of rhPT2 being expressed in insoluble form. However, the rest of the *E. coli* strains used in the experiments failed to express the rhPT2 in soluble form. We are deducing that the success in achieving soluble expression was not only due to the availability of chaperones Cpn60/Cpn10 in psychrophilic environment, which played a crucial role in the protein folding in *E. coli* ArcticExpress strain, but also due to the codon optimization of *hpt2* gene.

Key words: chaperons, codon optimization, *E. coli*, inclusion bodies, prethrombin-2

Prethrombin-2 (PT2) adalah prekursor trombin yang berperan dalam mengubah fibrinogen menjadi fibrin selama proses pembekuan darah. Penelitian sebelumnya melaporkan bahwa ekspresi prethrombin-2 manusia rekombinan (rhPT2) dalam inang *Escherichia coli* selalu menghasilkan badan inklusi. Tujuan penelitian ini adalah mencari strategi dalam mengekspresikan rhPT2 dalam bentuk terlarut dalam inang *E. coli*. Desain eksperimen meliputi perancangan dan optimasi kodon prethrombin-2 manusia (*hpt2*) serta mengoptimasikan ekspresi rhPT2 terlarut dengan menggunakan empat galur *E. coli*. Gen *hpt2* menunjukkan *codon adaptation index* (CAI) sebesar 0,336 dengan presentase GC 56,8%, sementara hasil optimasi menunjukkan berturut-turut sebesar 1,000 dan 53,1%. Selanjutnya, gen *hpt2* berhasil diklon ke dalam vektor ekspresi pTWIN1. Uji ekspresi protein rhPT2 terhadap empat galur inang *E. coli* menunjukkan bahwa galur ArcticExpress mampu mengekspresikan protein rekombinan dalam bentuk terlarut walaupun sejumlah protein juga terekspresi dalam bentuk badan inklusi. Sementara galur lain menunjukkan hasil ekspresi protein rekombinan selalu dalam bentuk badan inklusi. Selain preferensi kodon *hpt2* yang sesuai dengan *E. coli*, peran chaperon Cpn60/Cpn10 dalam lingkungan psikrofilik yang terdapat pada *E. coli* ArcticExpress memiliki peran yang sangat krusial dalam membantu pelipatan rhPT2, sehingga dapat diekspresikan dalam bentuk terlarut.

Kata kunci: badan inklusi, *chaperon*, *E. coli*, optimasi kodon, prethrombin-2

Prethrombin-2 (PT2) is a thrombin precursor that structurally has one glycosylation site and four disulphide bonds. As one of the fibrin glue components, thrombin is able to stick and to cover the wound. So, Thrombin can be potentially applied in replacing suture technic (Spotnitz and Prabhu 2005). Eventhough the fibrin glue has widespread

applications, commercial fibrin glue has several disadvantages, *i.e.* allergic effect; pathogenic contamination, since the use of any plasma-derived product in the surgical setting carries a potential risk of viral transmission; and expensive protein therapeutics since must be imported (Spotnitz 2001; Enus *et al.* 2011).

Generally, glycosylated protein therapeutics are produced from mammalian cells. Chinese Hamster Ovary (CHO) is the most commonly used mammalian

*Corresponding author; Phone: +62-818423339, Fax: +62-22-4218292; Email: s_soemitro@unpad.ac.id

cells due to safety consideration (Septisetyani 2014). The first registered recombinant human thrombin in the United States was produced from its precursor, prethrombin-1, by using CHO as a host (Bishop *et al.* 2006). Clinical study revealed that this product had homeostatic effect and was highly tolerance toward proteolytic digestion (Burnouf 2011). However, this expression host is very expensive and has low economic value.

Escherichia coli remains the system of first-choice for expressing proteins, as it is cheap, easy to be handled, and has short-life cycle (Cabrita *et al.* 2006). Additionally, the genome of *E. coli* well studied, so it can be easily to be manipulated. However, inability of *E. coli* to express proteins with high molecular weight, disulphide bond rich, or posttranslational modification become a bottleneck. Application of *E. coli* as a host on pathogenic free-thrombin production from prethrombin-2 had been reported. Prethrombin-2 had previously been expressed in *E. coli*, however only small amount of thrombin was active due to the formation of inclusion bodies (Soejima. *et al* 2001; Freydel *et al.* 2007).

In this study, several strategies were implemented to prevent the expression of rhPT2 as inclusion bodies in *E. coli*. We used codon optimization, host optimization and chaperone co-expression to express rhPT2 in soluble form. So, we established the expression system of soluble rhPT2 in *E. coli*.

MATERIALS AND METHODS

Strains, Chemicals, Vector, and Medium.

Transformation and cloning was performed using *E. coli* TOP10F' (Invitrogen, USA). *E. coli* ER2566 expression host was purchased from New England (New England, Biolabs, USA). *E. coli* BL21(DE3), BL21(DE3)_Rosetta and BL21(DE3)_ArcticExpress were kind gifts from Dr. Jiri Damborsky (Masaryk University, Brno, Czech Republic). The hosts were cultivated in Luria Bertani medium, (1% tryptone, 0.5% yeast extract, and 1% sodium chloride) supplemented by appropriate antibiotics (tetracycline 100 $\mu\text{g mL}^{-1}$ or ampicillin 100 $\mu\text{g mL}^{-1}$). Luria Bertani medium with addition of 2% agar was used as solid medium. All restriction enzymes, T4 DNA ligase, and pTWIN1 expression vector were purchased from New England (New England Biolabs, USA). Prethrombin-2 gene was commercially synthesized by GeneArt (Life Technologies, Jerman). Isolation kit (Roche Applied Science, USA) and gene extraction kit

(Geneaid, Taiwan) were purchased commercially. Isopropyl- β -D-thiogalactoside (IPTG) dan β -mercaptoethanol (β ME) were from Sigma Aldrich (Sigma Aldrich, USA). Polyacrilamide and Commasie Brilliant Blue were from Biorad (BioRad, Richmond, USA).

Construction and codon optimization of *hpt2*.

The *hpt2* synthetic gene was designed to be cloned into pTWIN1 expression vector, using two restriction sites at 5' and 3' ends, *Xho*I and *Bam*HI, respectively (Fig 1). We extracted the sequence database at *GenBank* with Accession number: NM_000506.3:

TATSEYQTFNPRTFGSGEADCGLRPLFEKKSLE
DKTERELLESYIDGRIVEGSDAEIGMSPWQVML
FRKSPQELLCGASLISDRWVLTAAHCLLYPPWD
KNFTENDLLVRIGKHSRTRYERNIEKISMLEKIY
IHPRYNWRENLDRIALMKLKKPVAFSDYIHPV
CLPDRETAASLLQAGYKGRVTGWGNLKETWT
ANVGKGQPSVLQVVNLPIVERPVCKDSTRIRIT
DNMFCAGYKPDEGKRGDACEGDSGGPFVMKS
PFNNRWYQMGIVSWGEGCDRDGKYGFYTHVF
RLKKWIQKVIDQFGE (308 amino acids).

Afterwards, the sequence was optimized by OPTIMIZER free software (<http://gnomes.urv.es/OPTIMIZER>; Puigbo *et al.* 2007) based on *Codon Usage Database* (<http://www.kazusa.or.jp/codon/>) and *E. coli* codon was used as codon preference. Subsequently, optimized codon was analyzed by *Graphical Codon Usage Analyzer* (GCUA) (<http://gcu.schoedl.de/>; McInerney 1998).

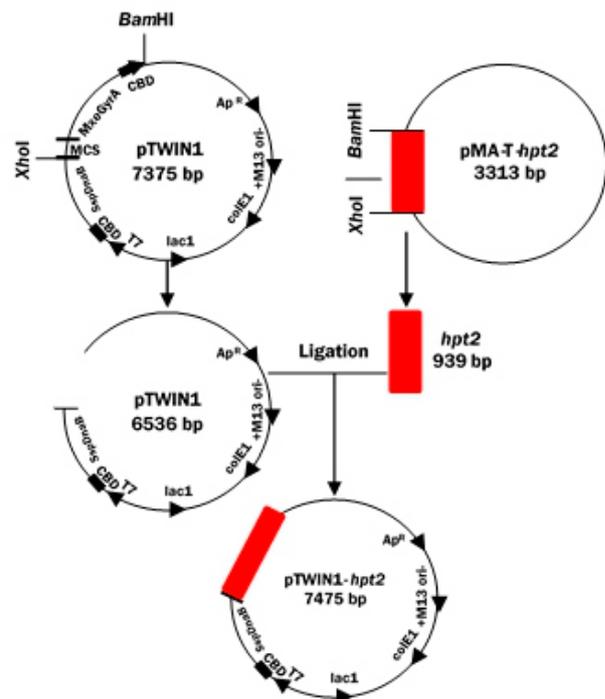


Fig 1 Construction of pTWIN1-*hpt2*. *hpt* and pTWIN1 were cut using restriction enzymes *Xho*I and *Bam*HI to obtain *hpt2* and pTWIN1 fragments. *hpt2* fragment was ligated to pTWIN1 using T4 DNA ligase.

Cloning. General molecular biology experiment was performed according to Sambrook and Russel (2001). pMA-T-*hpt2* was cut by *Xho*I and *Bam*HI. In parallel, pTWIN1 was also cut by same restriction enzymes. The restriction fragments were characterized using 1% agarose gel electrophoresis. Then, *hpt2* and pTWIN1 fragments were extracted from the agarose. The extracted *hpt2* was ligated into pTWIN1 by T4 DNA ligase resulting pTWIN1-*hpt2*. The cloned gene was verified by DNA sequencing (MacroGene, Korea).

Expression Analyses and Characterization. hPT2 expression was tested in four different *E. coli* hosts, *i.e.* BL21(DE3) ArcticExpress, ER2566, BL21(DE3), and BL21(DE3) Rosetta. The protein expression was induced by IPTG. Single transformant colony was grown overnight at 37 °C in 5 mL LB medium supplemented with the appropriate antibiotics. One mL of overnight culture was then transferred into 100 mL of LB medium supplemented with appropriate antibiotic. Subsequently, the culture was incubated for 3-4 h with shaking at 150 rpm until its optical density at 600 nm (OD₆₀₀) reached 0.5. Before induction, one mL culture was taken and labelled as 'before induction' sample. The culture was then induced with 0.1 mM of IPTG. After induction, ER2566, BL21(DE3) and BL21(DE3) Rosetta strains continued to be incubated at 30 °C for 4 h, while the BL21(DE3) ArcticExpress were incubated at 12 °C overnight. One mL was taken and labelled as 'after induction' sample. The cell was harvested by centrifugation at 4 °C, 7500 *g*. Then, the cell was resuspended with 2 mL of 50 mM glycine buffer pH 7.5 and sonicated for 15 min. Forty mL of the clear supernatant was kept and labelled as 'crude extract'. 50 µL of 8 M urea was added into pellet and then the sample was boiled for 15 min at 100 °C. The sample was labelled as 'insoluble fraction'. The labelled samples were analysed by 12% SDS PAGE according to Laemmli (1970).

RESULTS

Codon optimization. Codon sequence of *hpt2* was analysed using OPTIMIZER. The result showed that GC percentage was 56.8% with CAI 0.336. Analysis of the codons by *Graphical Codon Usage Analyzer* (GCUA) using *E. coli* codon preference indicated that 10 amino acids were encoded by non-preference codon. After codon optimization, GC percentage of *hpt2* was 53.1% with CAI of 1.000.

Cloning. The cloning strategy was depicted in Fig 1. The *hpt2* gene was cut from plasmid pMA-T-*hpt2* using *Xho*I and *Bam*HI. In parallel, plasmid pTWIN1 was cut with the same restriction enzymes. The agarose gel electrophoresis showed that *hpt2* fragment was successfully cut from pMA-T. Two bands with molecular weight approximately 939 bp and 2.374 bp were indicated as *hpt2* and pMA-T, respectively (Fig 2). pTWIN1 digest also produced two bands with molecular weight 6.536 bp and 839 bp, corresponding to pTWIN1 dan MCS (Multi Cloning Site), respectively (Fig 2).

Furthermore, the *hpt2* and pTWIN1 bands were extracted and purified from the gel. Agarose gel electrophoresis proved that *hpt2* and pTWIN1 were successfully extracted and purified (Fig 3). Two bands appeared with molecular weight 939 bp and 6.536 bp indicating *hpt2* and pTWIN1.

Table 1 Pattern of rhPT2 expression in *E. coli*

<i>E. coli</i> strain	CBD-Intein SspDnaB-rhPT2	
	Soluble	Insoluble
BL21(DE3) ArcticExpress	+++++	++
Er2566	+	+++
BL21(DE3)	-	+++
BL21(DE3) Rosetta	-	+

Then, the *hpt2* fragment was ligated into pTWIN1 vector (Figure 1) using T4 DNA ligase. To confirm the ligation process, the sequence pTWIN1-*hpt2* was verified by DNA sequencing. The nucleotide alignment revealed that *hpt2* was successfully inserted into pTWIN1.

Protein Expression Analyses. rhPT2 expression was tested in four different *E. coli* strains (Table 1). The rhPT2 was designed as fusion protein. The total molecular weight of fusion protein was 63.36 kDa. The protein expression was analyzed by 12.5% SDS-PAGE. The results showed that *E. coli* BL21(DE3) ArcticExpress was able to express soluble CBD-Intein-rhPT2 fusion protein (Fig 4A). In contrast, ER2566 (Fig 4B), BL21(DE3) (Fig 4C) and BL21(DE3) Rosetta (Fig 4D) failed to express rhPT2-fusion protein in soluble form, instead, these strains expressed the protein target inclusion bodies.

DISCUSSION

Several parameters may affect the protein expression and solubilisation efficiency, for instance

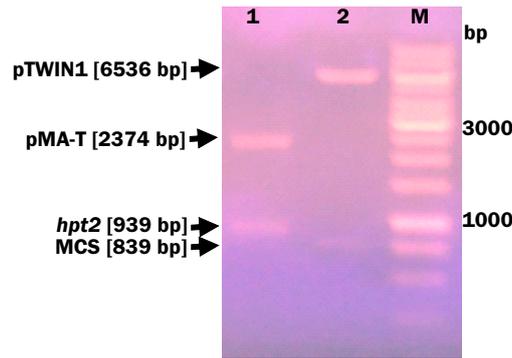


Fig 2 Restriction analysis of pMA-T-*hpt2* and pTWIN1 vector by using 1% agarose gel electrophoresis. Line 1: pMA-T-*hpt2* was cut with *Xho*I and *Bam*HI; line 2: pTWIN1 vector was cut with the same restriction enzymes; M: 1 kb DNA marker.

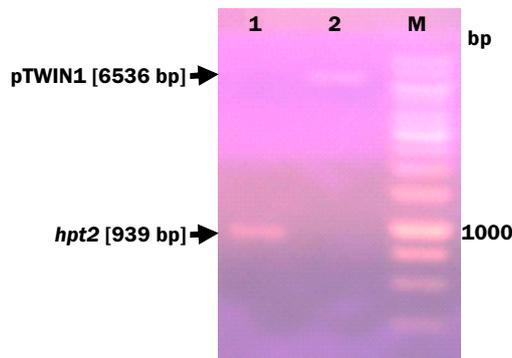


Fig 3 Extraction analysis of *hpt2* and pTWIN1 by using 1% agarose gel electrophoresis. Line 1: *hpt2* extracted from agarose gel; line 2: pTWIN1 extracted from agarose gel; M: 1 kb DNA marker.

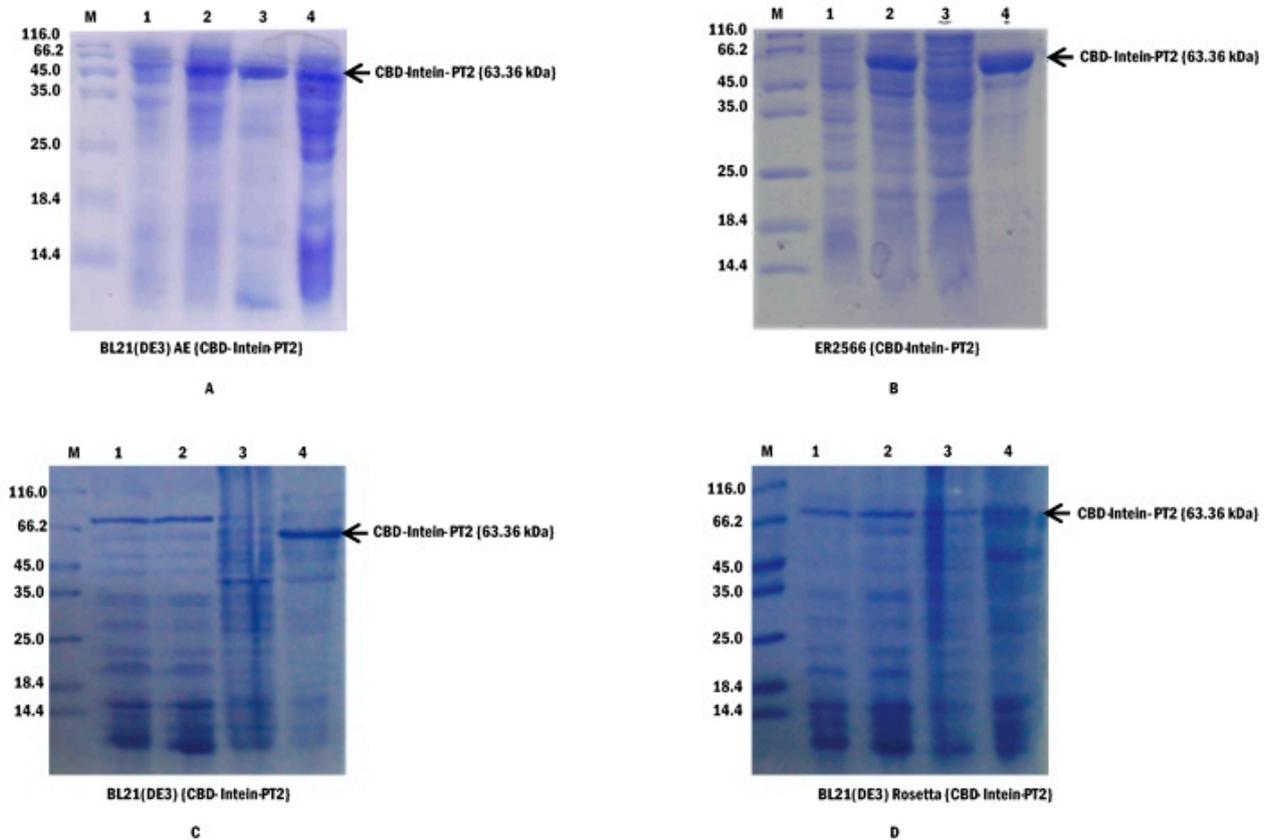


Fig 4 SDS-PAGE analysis of protein expression of CBD-Intein-rhPT2 in 4 strains of *E. coli*. (A) BL21 (DE3) ArcticExpress, (B) ER2566, (C)BL21(DE3) and (D) BL21(DE3) Rosetta. Line 1: cells before induction; line 2: cells after induction; line 3: soluble protein; line 4: insoluble protein; M: molecular weight marker.

codon sequence, choice of host, chaperonins, or environment. The compatibility between expression system and host is important to express a soluble protein recombinant. In order to achieve that goal, we used a commercial pTWIN1 vector. This vector has two advantages: first is as expression vector and second, the vector carried a highly specific purification tag, chitin-binding domain with self cleavable-intein system. These advantages will be used for mass production and purification of rhPT2. The pTWIN1 vector which carries T7 promoter, which has been used to express large quantity of soluble recombinant protein (Studier *et al.* 1991; Smith *et al.* 2012).

Differences between the host codon preference and the original *hpt2* codon created a problem in the expression of heterologous protein. The *hpt2* contains rare codons for *E. coli*, thus inefficient and/or miss-translation of the gene might occur (Merkl 2003; Gustafsson *et al.* 2004). Rare codons for *E. coli*, such as AGG or AGA encoding arginine, is common in human gene (Rinehart 2005).

Formation of inclusion bodies in *E. coli* is a major problem in the expression of recombinant protein (Freydell *et al.* 2007). Although it may also offer an advantage, that it is easily purified, as long as the protein can latter be refolded (Singh and Panda 2005). However, refolding involves multistep processes and the recovery of correctly folded and active enzyme are still uncertain (Soejima *et al.* 2001).

The first step in this study was codon optimization of *hpt2* gene according to the codon preference of *E. coli* (Gustafsson *et al.* 2004). Rarely used codons were replaced with high frequency codons (Xiong *et al.* 2008). Every gene in the genome has numeric number called codon adaptation index (CAI). The CAI can be used predict the expression of heterologous protein. (Carbone *et al.* 2003). The optimized codon was then chemically synthesized. This technology is so much more effective and efficient than gene isolation and, also, can prevent from disease transmission (Hughes *et al.* 2011).

It is well-known that *E. coli* is the most widely used and suitable host of heterologous proteins at 21-49 °C, with an optimum at about 37 °C (Ferrer *et al.* 2003), but when high expression of protein is needed, the *E. coli*'s capability to accurately express recombinant proteins decreased, and inclusion bodies can be formed. Low temperature has been proposed to improve protein solubility, but slower growth and low synthesis rates may downgrade protein yields (Sorensen and Mortensen 2005).

To eliminate this obstacle, *E. coli* BL21(DE3)_ArcticExpress, an engineered *E. coli* strain capable of growing at low temperature and expressing the psychrophilic bacterium *Oleispira antarctica* chaperones Cpn60/10, was employed. These chaperones display high refolding activities and govern growth of *E. coli* at low temperatures (Ferrer *et al.* 2004). As expected, expression analyses demonstrated that *E. coli* BL21(DE3)_ArcticExpress could produce hPT2 in soluble form.

It can be concluded that the synergy among optimized-codon usage, *E. coli* BL21(DE3)_ArcticExpress as a host, and psychrophilic environment, could be used to successfully produce a soluble form of recombinant hPT2.

ACKNOWLEDGMENT

This work is supported by grant "Penelitian Unggulan Strategis Nasional (PUSNAS)" from Directorate of Higher Education, Ministry of National Education, Indonesia.

REFERENCES

- Bishop PD, Lewis KB, Schultz J, Walker KM. 2006. Comparison of recombinant human thrombin and plasma-derived human α -thrombin. *Semin Thromb Hemost.* 32:086-097. doi: 10.1055/s-2006-939558.
- Burnouf T. 2011. Recombinant plasma proteins. *Vox Sang.* 100:60-83. doi:10.1111/j.1423-0410.2010.01384.x.
- Cabrita LD, Dai W, Bottomley SP. 2006. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* 6:1-8. doi:10.1186/1472-6750-6-12.
- Carbone A, Zinovyev A, Képès F. 2003. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics.* 19:2005-2015. doi: 10.1093/bioinformatics/btg272.
- Enus S, Natadisastra G, Shahib MN, Sulaeman R. 2011. Peran lem fibrin otologus pada penempelan tandur konjungtiva bulbi mata kelinci terhadap ekspresi gen fibronektin dan integrin. *MKB.* 43:183-188.
- Ferrer M, Chernikova TN, Yamikov MM, Golyshin PN, Timmis KN. 2003. Chaperonin govern growth of *E. coli* at low temp. *Nat Biotechnol.* 21:1266-1267. doi:10.1038/nbt1103-1266.
- Ferrer M, Chernikova TN, Timmis KN, Golyshin, PN. 2004. Expression of temperature-sensitive esterase in a novel chaperone-based *E. coli* strain. *Appl Environ Microbiol.* 70: 4499-4504. doi: 10.1128/AEM.70.8.4499-4504.2004.

- Freydell EJ, Ottens M, Eppink M, van Dedam G, van der Wielen L. 2007. Efficient solubilization of inclusion bodies. *J Biotechnol.* 2:678-684. Doi: 10.1002/biot.200700046.
- Gustafsson C, Govindrajana S, Minshull J. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol.* 22:346-353. doi:10.1016/j.tibtech.2004.04.006.
- Hughes RA, Miklos AE, Ellington AD. 2011. Gene synthesis: methods and applications. *Methods Enzymol.* 498: 277-309. doi: 10.1016/B978-0-12-385120-8.00012-7.
- Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.
- McInerney JO. 1998. GCUA: General Codon Usage Analysis. *Bioinformatics.* 14:372-373.
- Merkel R. 2003. A survey of codon and amino acid frequency bias in microbial genomes focusing on translational efficiency. *J Mol Evol.* 57:453-466. doi:10.1007/s00239-003-2499-1.
- Puigbo P, Guzman E, Romeu A, Vallve SG. 2007. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucl Acids Res.* 35:W126-W131. doi:10.1093/nar/gkm219.
- Rinehart CA. 2005. Introduction to recombinant genetics-biology 350. <http://bioweb.wku.edu/courses/biol350/syllabus.html>.
- Sambrook JEF, Russell DW. 2001. *Molecular cloning: A laboratory manual*, 3rd Ed. Cold Spring Harbor Laboratory Press. New York.
- Septisetyani EP. 2014. Perkembangan sistem ekspresi protein rekombinan yang bertarget pada modifikasi pola glikosilasi. <http://u.lipi.go.id/1403389335>.
- Singh SM, Panda AK. 2005. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng.* 99: 303-310. doi:10.1263/jbb.99.303.
- Smith LE, Yang J, Goodman L, Huang X, Huang R, Dressman J, Morris J, Silva RAGD, Davidson WS, Cavigiolio G. 2012. High yield expression and purification of recombinant human apolipoprotein A-II in *Escherichia coli*. *J Lipid Res.* 53:1708-1715. doi: 10.1194/jlr.D028043.
- Soejima K, Mimura N, Yonemura H, Nakatake H, Imamura T, Nozaki C. 2001. An efficient refolding method for the preparation of recombinant human prethrombin-2 and characterization of the recombinant-derived α -thrombin. *J Biochem.* 130:269-277. doi: 10.1093/oxfordjournals.jbchem.a002982.
- Sorensen HP, Mortensen KK. 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb Cell Factories.* 4:1. doi: 10.1186/1475-2859-41.
- Spotnitz WD. 2001. Commercial fibrin sealants in surgical care. *T Amer Jo Surg.* 182 (Supplement 1): S8-S14. doi: 10.1016/S0002-9610(01)00771-1.
- Spotnitz WD, Prabhu R. 2005. Fibrin sealant tissue adhesive-review and update. *J Long Term Eff Med Implants.* 15:245-270. doi: 10.1615/JLongTermEffMedImplants.
- Studier FW. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Bio.* 219:37-44. doi:10.1016/0022-2836(91)90855-Z.
- Xiong AS, Peng RH, Zhuang J, Gao F, Li Y, Cheng ZM, Yao QH. 2008. Chemical gene synthesis: strategies, softwares, error corrections, and applications. *FEMS Microbiol.* 32:522-540. doi:10.1111/j.1574-6976.2008.00109.x.