


RESEARCH

Open Access



Production of codon-optimized Factor C fragment from *Tachypleus gigas* in the *Pichia pastoris* GS115 expression system for endotoxin detection

Zubaidi Bachtiar¹, Apon Zaenal Mustopa^{2*} , Rika Indri Astuti^{3,4}, Fauziyah Fauziyah^{5*}, Fatimah Fatimah², Rozirwan Rozirwan⁵, Tuah Nanda Merlia Wulandari⁶, Dina Permata Wijaya⁷, Fitri Agustriani⁵, Arwansyah Arwansyah⁸, Herman Irawan^{1,2} and Jendri Mamangkey^{2,9}

Abstract

Background Factor C (FC) is widely used as a standard material for endotoxin testing. It functions as a zymogenic serine protease and serve as a biosensor that detects lipopolysaccharides. Prior investigations involving molecular docking and molecular dynamics simulations of FC demonstrated an interaction between the C-type lectin domain (CLECT) and the ligand lipopolysaccharide (lipid A). In this study, our aim was to assess the stability of the interaction between fragment FC and the lipid A ligand using protein modeling approaches, molecular docking, molecular dynamics simulation, and gene construction into the pPIC9K expression vector.

Methods and results The FC structure was modelled by online tools. In this case, both molecular docking and MD simulations were applied to identify the interaction between protein and ligand (lipid A) including its complex stability. The FC structure model using three modeling websites has varied values, according to a Ramachandran plot study. When compared to other models, AlphaFold server modeling produced the best Ramachandran findings, with residues in the most advantageous area at 88.3%, followed by ERRAT values at 89.83% and 3D Verify at 71.93%. From the docking simulation of FC fragments with three ligands including diphosphoryl lipid A, FC-Core lipid A, and Kdo2 lipid A can be an activator of FC protein by binding to receptor regions to form ligand-receptor complexes. MD simulations were performed on all three complexes to assess their stability in water solvents showing that all complexes were stable during the simulation. The optimization of recombinant protein expression in *Pichia pastoris* was conducted by assessing the OD value and protease activity. Induction was carried out using 1% (v/v) methanol in BMMY media at 30°C for 72 h.

Conclusions Protein fragments of Factor C has been proven to detect endotoxins and serve as a potential biomarker. Molecular docking simulation and MD simulation were employed to study the complex formation of protein fragments FC with ligands. The expression of FC fragments was successfully achieved through heterologous expression. We propose optimizing the expression of FC fragments by inducing them with 1% methanol at 30°C and incubating

*Correspondence:

Apon Zaenal Mustopa
azmustopa@yahoo.com
Fauziyah Fauziyah
siti_fauziyah@yahoo.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

them for 72 h. These optimized conditions are well-suited for upscaling the production of recombinant FC fragments using a bioreactor.

Keywords *Tachypleus gigas*, Endotoxin, Factor C, pPIC9K

Background

Horseshoe crabs are exotic aquatic organisms belonging to the family Limulidae and are considered living fossil animals [1]. Currently, there are four known species of horseshoe crabs in the world, namely *Carcinoscorpius rotundicauda* [2], *Tachypleus gigas* [3], *Tachypleus tridentatus* [4], and *Limulus polyphemus* [5]. Among these, *Carcinoscorpius rotundicauda*, *Tachypleus gigas*, and *Tachypleus tridentatus* are commonly found in coastal areas of Indonesia, including Indonesia. They are collectively referred to as Asian horseshoe crabs. On the other hand, *Limulus polyphemus*, also known as the Atlantic horseshoe crab, is exclusively found along the Atlantic coast of North America [6].

Asian horseshoe crabs, specifically *C. rotundicauda* and *T. gigas* are protected animals in Indonesia, although their conservation status remains unclear and insufficient. These species have been categorized as Near Threatened (2010), Threatened (2014), and Data Deficient [7]. The “Data Deficient” category indicates the unavailability of adequate data to determine estimates of extinction risk, distribution, and population numbers. According to the Minister of Environment’s Regulation Number P.92/MENLHK/SETJEN/KUM.1/8/2018, all three types of horseshoe crabs are classified as protected animals [8]. Even though their conservation status seems contradictory, the Atlantic horseshoe crab or *L. polyphemus* produces *Limulus Amoebocyte Lysate* (LAL) in their bloodstream which is known as the current Factor C used as a standard material for endotoxin testing. Proenzyme serine proteases are the first proteins in the cascade reaction that lead to LAL coagulation upon contact with the endotoxin [9]. *Tachypleus* sp. produces *Tachypleusin amoebocyte lysate* (TAL) and *C. rotundicauda* produces *Carcinoscorpius amoebocyte lysate* (CAL) which can also be used to detect gram bacterial endotoxins negative, human blood endotoxins, and pathogens in medicinal products [10].

Gram-negative bacteria’s outer membrane is made of lipopolysaccharide (LPS), a complex macromolecule with three structural parts: core oligosaccharides, lipid A, and O-antigens [11]. LPS within the outer membrane has the ability to activate Factor C through autocatalysis, a protein responsible for endotoxins binding. Upon binding to endotoxins, Factor C interacts with serine protease to remain active [12, 13]. Once activated by LPS, Factor C activates coagulation Factor B, which

subsequently converts pro-clotting enzymes into active clotting enzymes. These enzymes then facilitate the conversion of proteolytic coagulants to coagulants, resulting in the spontaneous formation of insoluble polymers. In this manner, Factor C functions by binding to defense molecules from pathogens (PAMPs) such as LPS present on the cell walls of Gram-negative bacteria. This process ultimately forms a barrier at the site of microbial invasion [14].

Factor C is synthesized as a single polypeptide has a molecular weight of 123 kDa. It consists of two polypeptide chains connected by disulfide bonds: a heavy chain (80 kDa) and a light chain, (43 kDa) in *T. tridentatus*, while in *C. rotundicauda*, it consists of a heavy chain (80 kDa) and a light chain (52 kDa) [13, 15–17]. Factor C of *C. rotundicauda* has been cloned and expressed in *Escherichia coli* [18]. Muta et al. [12] cloned the Factor C-coding gene of the Japanese horseshoe crab (*T. tridentatus*) in two separates, overlapping partial fragments. Recombinant Factor C (rFC) has been expressed in different host cells, including, among others: *Pichia pastoris* [19, 20], *Saccharomyces cerevisiae* [21], mammalian cells [22, 23], and insect cells using the baculovirus system [24]. Factor C has also been cloned from the species *L. polyphemus* [25]. Our report on the analysis of docking simulations in silico and molecular dynamics of protein Factor C (Receptor) simulations with ligands of lipid A (lipopolysaccharides) shows that the active sites and binding pockets between proteins and lipopolysaccharide ligands of gram-negative bacteria, are present in Heavy Chain Fragment Factor C with Lectin Domain [26]. In this study, cloning and expression of Factor C gene fragments from *Tachypleus gigas* in the eukaryotic organism *Pichia pastoris* was carried out.

Methods

The Factor C gene from Horseshoe crab (*T. gigas*) has been analyzed for its active site and binding site on certain amino acid parts cut for recombinant protein studies by ligating Factor C gene fragments into the pPIC9K plasmid which is a plasmid derived from yeast. *E. coli* DH5 α bacteria are used as hosts for cell cloning. The cloning of recombinant plasmids is then cut on the side of the *SacI* restriction enzyme section to neutralize the recombinant plasmids so that they can enter and integrate into the expression host. Yeast with strain GS115 hosts expression of Factor C gene fragments.

Gene Factor C fragment expression was analyzed using SDS-PAGE. Bioinformatic studies were conducted by modeling protein from Factor C gene fragments using a web server, followed by protein model validation with Ramachandran, 3D verify, ERRAT, molecular docking to analyze the interaction between proteins and ligands, and MD simulation to determine the stability of molecular docking results with a simulation time of 100 ns.

Physicochemical analysis of factor C protein fragment

The gene sequence was translated using SnapGene to obtain amino acid data. The amino acid sequence was then analyzed for its physical and chemical parameters using the ExPASy—ProtParam tool, a web server. This program predicted the hydrophobicity of the amino acid sequence, isoelectric point, and molecular weight of the protein. Additionally, the protein solubility was analyzed using the Protein-sol sequence solubility web server (Protein-sol sequence solubility (manchester.ac.uk) [27]).

Protein modeling analysis and molecular docking

Protein modeling was initiated by entering the files (fasta) into several web-based servers including AlphaFold (AlphaFold2.ipynb-Colaboratory (google.com), Robetta (<https://robetta.bakerlab.org>), and I-Tasser (<https://zhanggroup.org/I-TASSER/>). A web-based protein refine model server or Galaxy refine (GalaxyWEB (seoklab.org) was used to correct some missing residues. The results of the refined protein model were validated using several validity parameters including Ramachandran Plot, 3D Verify Structure, and ERRAT (www.saves.mbi.ucla.edu). Investigation on the protein fragment Factor C as a receptor to obtain ligand lipid A binding pockets, we simulated molecular docking using a web-based hdock server. Ligands were retrieved from the Pubchem database. The three ligand component structures were then downloaded in SDF form. The SDF file is then converted into a PDB file using OpenBabel 2.4.1 Program Packages. Geometry optimization on ligands was carried out using Avogadro software. Docking simulations were performed with Factor C fragment proteins as receptors to obtain binding pockets and lipid A as ligands.

The website, <http://hdock.phys.hust.edu.cn/> was used to execute HDock-based molecular docking [28]. The docking parameters were calculated using the docking online tool's defaults. Then, Pymol software and PLIP online tools were used to visualize the interaction between the protein and its ligands. The bond stability between receptor proteins and lipid A ligands resulting from molecular docking, molecular dynamics simulations were carried out for 100 ns using AMBER 20 program, the best modeling outcomes will be confirmed. The methodologies for molecular docking and MD modeling

are identical to those used in our earlier work, which was described in the previous study [29].

Molecular dynamics simulation

To assess the stability of the protein-ligand complex within a TIP3P water solvent, an all-atom molecular dynamics (MD) simulation was conducted using the Amber20 software package [30]. The protein's force field parameters were determined using the AMBER14 force field, while the ligand's parameters were computed using the general amber force field (GAFF) [31]. Electrostatic interactions and hydrogen bonds were managed using the Particle Mesh Ewald (PME) algorithm [32] and the SHAKE algorithm [33]. The cut-off distance and time step were set to 10 Å and 2 fs, respectively, for the entire system. The MD simulation involved multiple steps: energy minimization, followed by heating the system from 0 to 300 K over 2 ns using the NVT ensemble. To maintain system temperature and pressure, the Langevin thermostat [34] and the isotropic position scaling algorithm were applied during the simulation. Subsequently, the NPT ensemble was employed to equilibrate the system for 100 ns. Trajectories were collected at intervals of 5000 steps for each MD simulation. The CPPTRAJ software package was used to analyze the MD trajectories [35]. The MD simulation protocols closely followed our previous studies [36, 37].

Cloning preparation of gene fragment Factor C

Factor C gene sequences have been carried out active site topology analysis and pocket binding in previous studies, obtained sequences of Factor C (Lectin-serine) gene fragments that play an important role in silico Factor C studies for protein expression analysis. The sequence of these fragments was assembled in silico and we make the gene into a synthetic gene that has been inserted into the plasmid. pPIC9K was selected to be the expression vector in the host *Pichia pastoris* cell strain GS115. The gene cloning stage is introduced into *E. coli* DH5α (competent cells). *E. coli* transformants were selected on LB agar media supplemented with 20 μL of kanamycin per plate.

Linearization of recombinant plasmids

Positive clones (plasmids carrying genes) were grown on 5 mL of liquid LB media for isolation using plasmid Midi Kit (Qiagen) protocol. The results of plasmid isolation were linearized using the restriction enzyme SacI. The linearization results were verified using 1% agarose gel (marker 1 kb). Through the use of QIAquick PCR Purification Kit (Qiagen) procedure, the generated DNA bands were purified. Plasmids that had been purified were kept at −20°C for further study.

Transformation of recombinant plasmids with *Pichia pastoris* GS115

The linearized recombinant plasmids were then introduced into *P. pastoris* GS115 cells (competent cells) by electroporation method. A total of 80 μ l of competent *P. pastoris* GS115 cells were mixed with 5 μ g of linear plasmids in a microtube. Competent cells and plasmids are inserted into electroporated cuvettes (2 mm gap) that have been cooled at -4°C . The sample was incubated in ice for 5 min. Electroporation was carried out under conditions of 2 kV/400 Ω on the electroporator machine. After electroporation, 1 ml of cold sorbitol solution (1 M) was added to the cuvette, the contents of the cuvette were transferred into a 1.5-mL sterile microtube and shaker incubator at 30°C for 2 h. A total of 100, 150, and 200 μ L of transformed *P. pastoris* GS115 cells were spread on YPDS agar media containing 0.25 mg/ml of geneticin on the plate incubated for 2–3 days at 30°C until colonies formed. The colonies formed were purified by scratching method on fresh YPD or YPDS agar media containing geneticin.

Expression of Factor C gene fragment in *Pichia pastoris*

Single colonies of recombinant *P. pastoris* GS115 were cultured overnight in BMGY medium at 30°C under agitation of $250\times g$ until reaching the logarithmic phase ($\text{OD}_{600} = 2-6$). The culture was then centrifuged for 5 min at $1500\times g$ and 4°C . The obtained pellet was resuspended in 25 mL of BMMY medium supplemented with 0.5% methanol as an inducer to achieve an OD_{600} of 1.0. Induction was repeated during a 24-h cultivation period. Harvesting was performed by centrifugation for 5 min at room temperature and $1500\times g$ to obtain the supernatant containing Factor C protein fragment. Subsequently, SDS-PAGE was carried out to analyze the successful expression of Factor C in *P. pastoris* GS115.

Optimization of culture conditions

Selected colonies from the prior results were then expressed in a shake-flask fermentation with aeration (1:10). The recombinant yeasts were grown in buffered glycerol complex medium (BMGY) to achieve high biomass production. The culture was then transferred to buffered methanol complex medium (BMMY) to determine the optimal methanol concentration. Methanol was added to the expression medium at final concentrations of 0.5%, 1%, and 2% (v/v) and incubated at various temperatures (25°C , 30°C , 37°C) with a shaker speed of $250\times g$ for 24 h. The kinetics of growth were analyzed by measuring the OD_{600} value every 24 h until 96 h.

SDS-PAGE analysis

Acrylamide gel electrophoresis was performed using a separating gel (12%) composed of 3.2 mL of ddH₂O, 4 mL of 30% acrylamide, 2.6 mL of 1.5 M Tris-HCl pH 8.8, 100 μ L of 10% sodium dodecyl sulfate (SDS), 100 μ L of 10% ammonium persulfate (APS), and 10 μ L of TEMED. All reagents were thoroughly mixed, and 5 mL of the mixture was poured into the gel mold. Then, 1 mL of 70% ethanol was added. The separating gel was allowed to polymerize for 20–30 min until solidified. After solidification, the ethanol layer on top of the separating gel was discarded. For the stacking gel (4%), 2.93 mL of ddH₂O, 0.67 mL of 30% acrylamide, 1.3 mL of 0.5 M Tris-HCl pH 6.8, 50 μ L of 10% SDS, 500 μ L of 10% APS, and 5 μ L of TEMED were mixed to ensure homogeneous mixing. A total of 30 μ L of the sample was mixed with 15 μ L of loading buffer in a 1.5-mL tube. The sample was then denatured at 95°C for 10 min. Subsequently, 30 μ L of the sample and 5 μ L of the marker were loaded into the gel wells. Electrophoresis was run at voltages of 80 V and 110 V for 130 min. The SDS-PAGE gel was stained with Coomassie brilliant blue (CBB) and incubated overnight. The gel was then rinsed (destained) until clear protein bands appeared.

Zymogram

According to Bencsik et al. [38], the zymogram test contained 8% (v/v) separating gel and 4.5% (v/v) stacking gel with a 1.5% casein addition. The gel is incubated in 2.5% Triton X-100 at room temperature for 1 h following the separation procedure. The Gel was then treated with a 10 mM pH 8 Tris HCl buffer for an overnight period. For 2 to 3 h, the gel was dyed with Sigma's 0.05% Coomassie brilliant blue G-250. The next step was rinse (destained) the sample until clear protein bands were visible.

Protease activity test

Protease activity was measured using a modified version of Cupp-Enyard's [39] technique. In a 96-well microplate, the protease test was performed at 540 nm.

Results

Solubility of Factor C protein

The solubility of the recombinant protein was analyzed using the ProtParam ExPASy webserver, which yielded an isoelectric point (pI) of 6.57. Proteins with isoelectric point (pI) values falling between 5 and 7 are classified as exhibiting pronounced solubility. Moreover, a greater abundance of negatively charged amino acid residues corresponds to an elevated degree of protein solubility [37]. The recombinant protein's isoelectric point falls

within the 5–7 range, confirming its elevated solubility characteristics.

Factor C gene fragment modeling

The results of the three-dimensional protein structure modeling display the protein structure of the Factor C fragment using three different web servers: Alphafold, Robetta, and I-Tasser. The models presented by these three protein modeling servers consist of α -helices, β -sheets, and loop/coil structures. The three-dimensional protein model structures can be observed in Fig. 1.

One of the most fundamental ideas in structural biology is the Ramachandran plot, which is frequently applied to validate protein structure models. On Fig. 2, a Ramachandran plot analysis of the *T. gigas* Factor C protein structure is presented in this work

AlphaFold modeling exhibits the best Ramachandran results when compared to other models because it has residues in the most advantageous region of 95.8% (quadrant 1), the allowable additional region of

3.5% (quadrant 2), the generously permitted region of 0.0% (quadrant 3), and the area that is not permitted of 0.7% (quadrant 4). Ramachandran plot analysis demonstrates that Factor C structure model with three modeling websites (Fig. 2) has different values. Better quality and stability of protein structure are correlated with higher percentages of amino acid residues in the most favored area and lower percentages of residue in the not authorized region [37]. The AlphaFold modeling of the *T. gigas* Factor C structure reveals that it has a medium quality and a tendency to be stable. In this case, when a protein structure has a preferred region score of >90% and a forbidden area score of 2%, it is said to be of excellent quality [40].

Validation analysis of the 3D model structure was also performed using the ERRAT and Verify 3D scores for the Factor C protein fragment. ERRAT serves as an overall quality factor for non-bonded atom interactions, indicating that higher scores correspond to better protein quality. High-resolution structures typically have

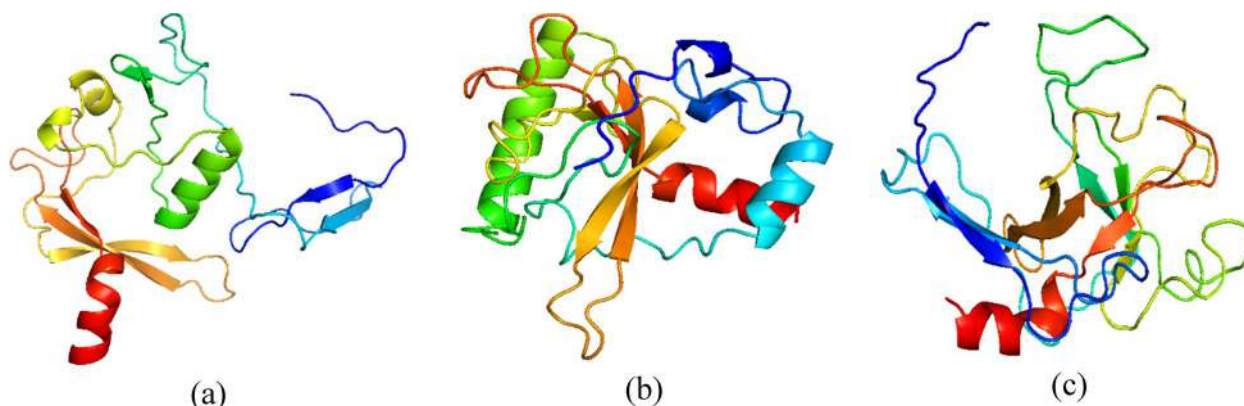


Fig. 1 3D Structure of lectin-serine protease Factor C modeled using **a** alphafold, **b** robetta, and **c** i-tasser

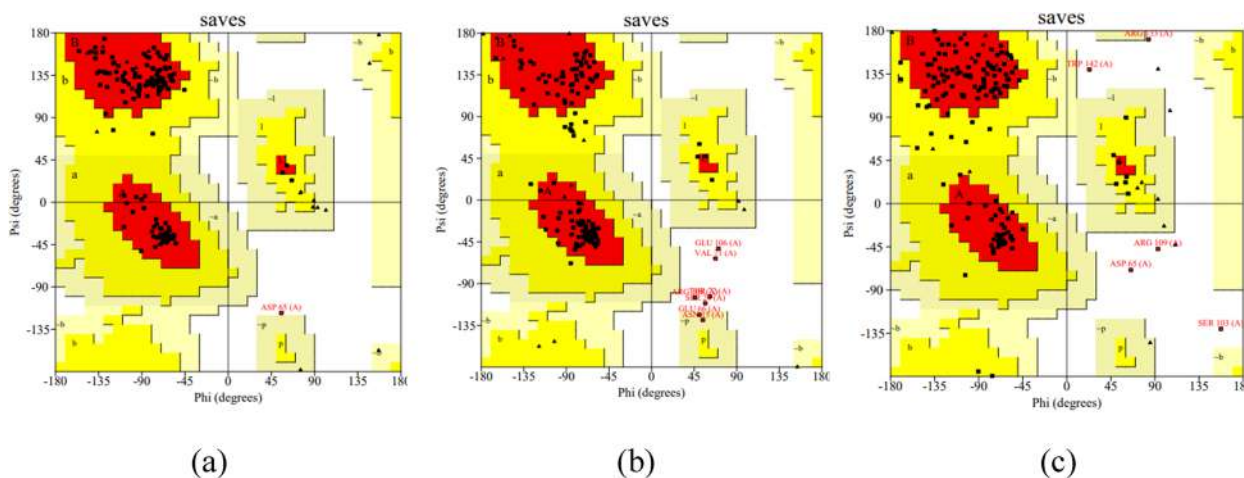


Fig. 2 Ramachandran plot analysis of *T. gigas* Factor C **a** model by alphafold, **b** model by robetta, and **c** model by i-tasser

ERRAT scores above 95% and Verify 3D scores exceeding 80% (Table 2).

Molecular docking

Molecular docking simulation is a method used to identify the binding site of active compounds to the protein’s catalytic site [41]. This method is advantageous

and widely applied in the design of new drugs as it can elucidate how specific small molecules bind to proteins using computational algorithms and scoring functions [42]. The docking results explain how the ligand can bind to the receptor protein with several observed interactions of amino acid residues, as shown in Fig. 3.

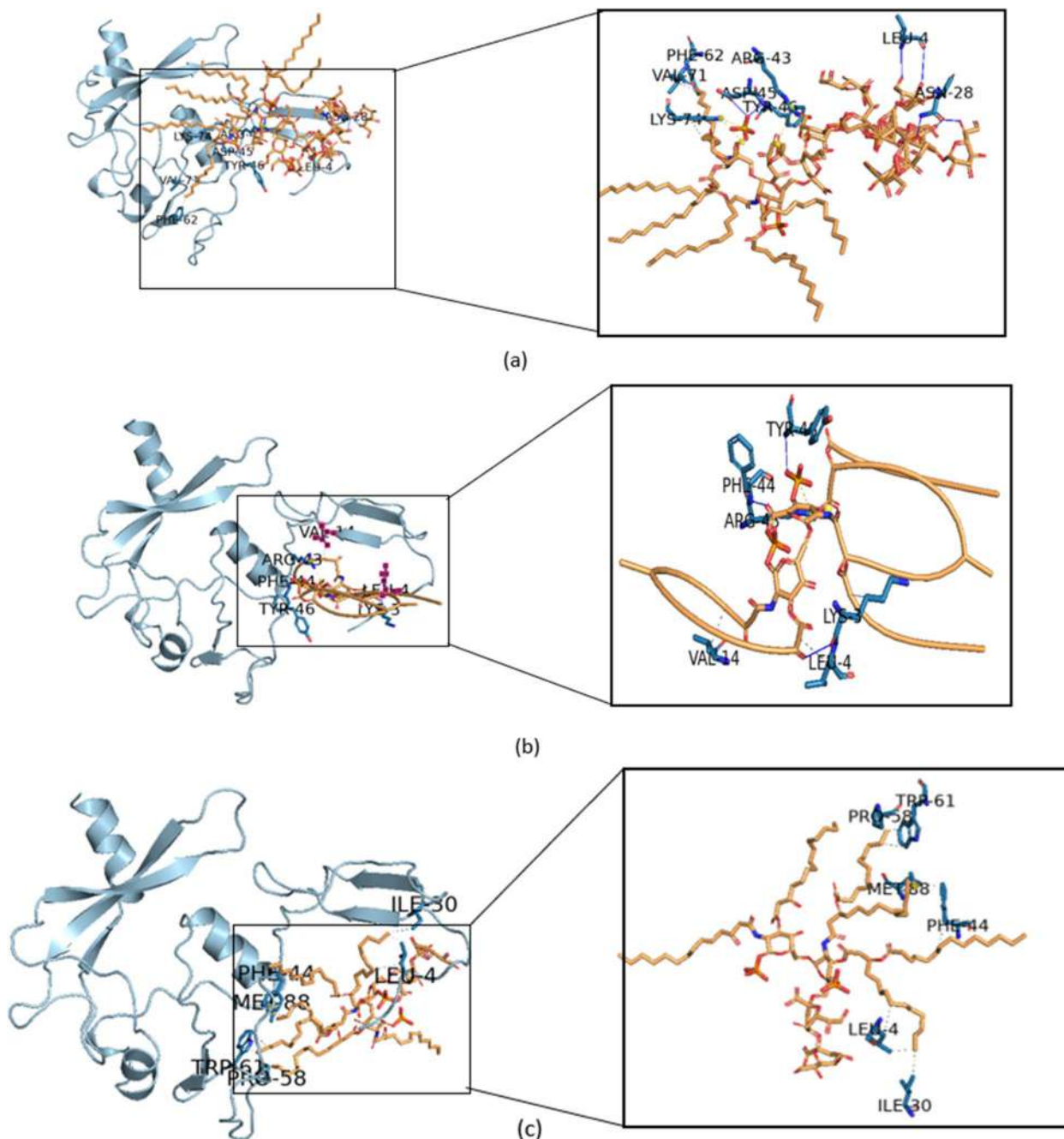


Fig. 3 Binding of the ligand to protein Factor C receptor site. **a** Complex 1 (protein-core lipid A ligand), **b** Complex 2 (protein-diphosphoryl lipid A ligand), and **c** complex 3 (protein-Kdo2 lipid A ligand). The conformational poses of each complex were visualized with the pymol program

Some amino acid residues that bind to the receptor protein with the ligand lipid A in complex 1 (core lipid A) include Phe22, Ile30, Ser25, Tyr46, Ser48, Tyr70, Ser73, Lys74, Asp127, Arg129, and Arg132. In complex 2 (diphosphoryl lipid A), the amino acid residues Met1 and Tyr46 are involved in binding, and in complex 3 (Kdo2 lipid A), the binding involves the amino acid residues Gly2, Phe44, Tyr46, Ser49, and Asp59. All three complexes exhibit the same binding pattern with the amino acid residue Tyr46. The interaction results of the amino acid residues from the three complexes indicate that the receptor protein with the lipid A ligand forms hydrogen interactions Tables 1 and 2.

The protein and ligand obtained from docking using the HDOCK web server suggest that the ligand can bind to the receptor site when the complex exhibits negative binding energy. The binding energies obtained from the simulations of the three complexes are listed in Table 3. All ligands are capable of binding to the receptor since all binding energy scores are negative. These results indicate that all ligands can form complexes with the receptor Fig. 4.

The docking scores for each complex have negative values with varying binding energy values. In complex 1, the docking score is the highest with a binding energy

Table 1 Solubility of Factor C protein fragment with isoelectric point

Amino acid	Composition	Percentage
Ala (A)	11	6.40%
Arg (R)	10	5.80%
Asn (N)	8	4.70%
Asp (D)	11	6.40%
Cys (C)	5	2.90%
Gln (Q)	4	2.30%
Glu (E)	8	4.70%
Gly (G)	19	11.10%
His (H)	1	0.60%
Ile (I)	6	3.50%
Leu (L)	13	7.60%
Lys (K)	9	5.30%
Met (M)	2	1.20%
Phe (F)	8	4.70%
Pro (P)	6	3.50%
Ser (S)	21	12.30%
Thr (T)	6	3.50%
Trp (W)	5	2.90%
Tyr (Y)	4	2.30%
Val (V)	14	8.20%
Pyl (O)	0	0.00%
Sec (U)	0	0.00%

Table 2 Quality evaluation of Factor C protein structure with ERRAT and verify 3D webserver

Modelling structure of Factor C	ERRAT (%)	Verify 3D (%)
AlphaFold	89.83	71.93
Robetta	89.79	73.10
I-Tasser	78.76	60.23

of -467.93 and a high confidence score of 0.9983. Complex 2 has a docking score of -302.86 and a confidence score of 0.9551, while complex 3 has a docking score of -322.98 and a confidence score of 0.9695. The more negative the binding energy value, the more stable of the ligand-receptor complex formation. This finding suggested that the most stable complex could be found in Complex 1 then followed by Complex 3 and Complex 2, respectively.

Molecular dynamics simulations of Factor C protein fragment

The parameters of each MD (RMSD, RMSE, SASA, and hydrogen bond) were analyzed to determine the mean value as shown in Table 4.

Construction and cloning of the Factor C gene fragment

The construction of the Factor C fragment gene and the pPIC9K vector was designed using the SnapGene application, involving the integration of the gene into the pPIC9K vector at the multiple cloning site (MCS). The MCS site was cleaved using AvrII and NotI restriction enzymes, followed by ligating the gene into the vector. The Factor C fragment gene carried by the pPIC9K vector was then sequence-aligned for synthesis. The synthetic gene was amplified through transformation into *E. coli* DH5 α . Successfully cloned synthetic genes were confirmed through plasmid isolation. To verify the presence of the inserted gene in the recombinant plasmid, the isolated plasmid underwent single digestion with restriction enzymes [43]. The pPIC9K vector has a sequence length of 9276 bp, while the Factor C fragment gene comprises a sequence of 516 bp. Gel electrophoresis results showed a thick band with a size of 9792 bp. (Fig. 5B).

The results of the single digest of the recombinant pPIC9K vector were transformed on the *P. P. pastoris* (GS115) expression host cells. Furthermore, the success of the transformants was confirmed by employing colony PCR and conducting sequencing analysis with designated primers (Table 5) Fig. 6.

The pPIC9K vector is commonly used as a general expression vector for expression in *P. pastoris*. This expression vector contains the AOX1 nucleotide sequence at the 3' end, the AOX1 promoter sequence

Table 3 Docking outcomes between the protein fragment Factor C and lipid A utilizing the HDock website

Complex	Docking score	Confidence score
Complex 1 (fragment FC-Core lipid A)	-467.93	0.9983
Complex 2 (fragment FC-diphosphorylLipid A)	-302.86	0.9551
Complex 3 (fragment FC-Kdo2 lipid A)	-322.98	0.9695

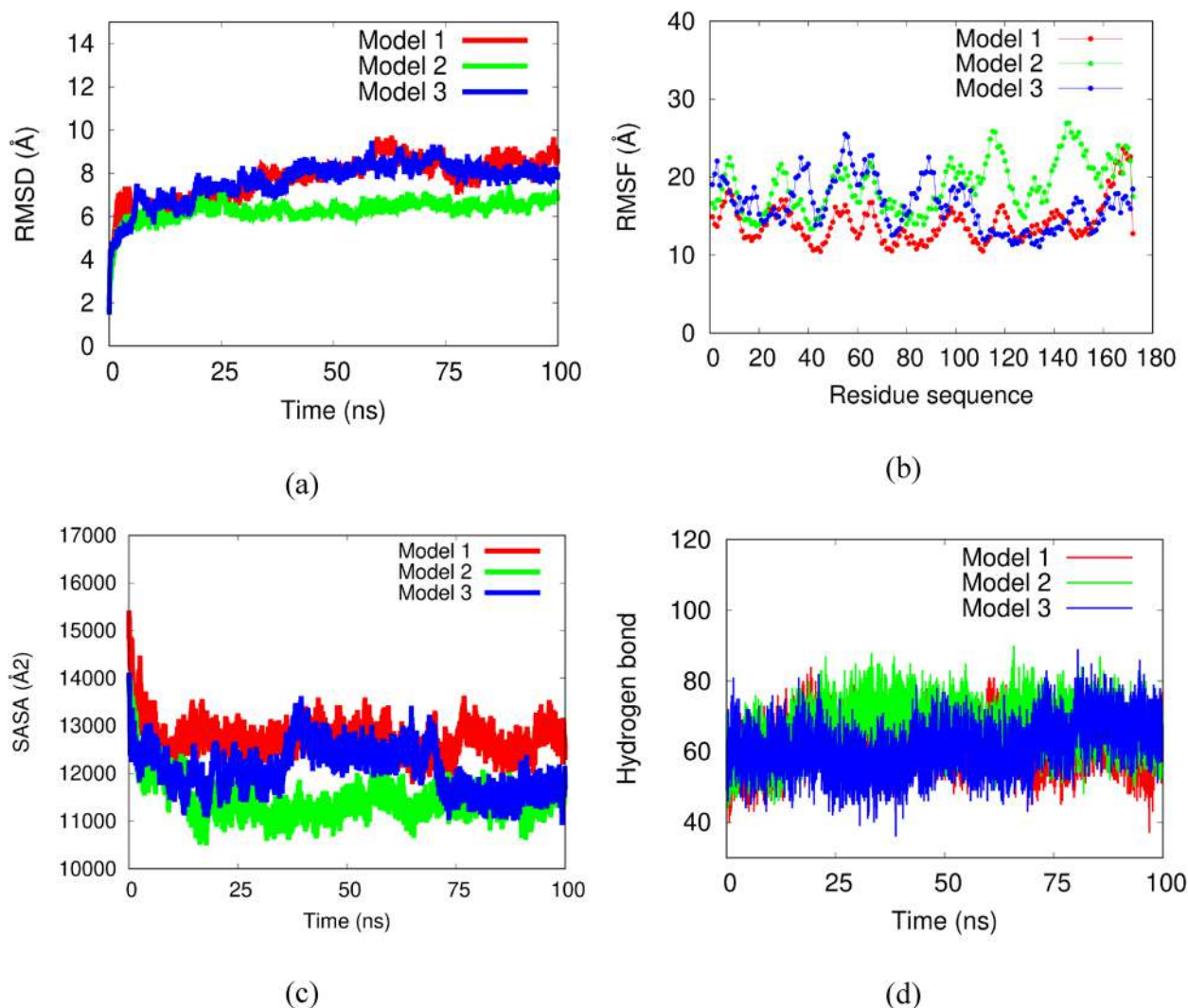


Fig. 4 The MD validation metrics consisted of **a** RMSD value of the complexes, **b** RMSF descriptor, **c** protein surface area measured from SASA, and **d** hydrogen bonds

Table 4 Mean values of the parameters of each MD analysis

Complex	RMSD	RMSF	SASA	Hydrogen bond
Complex 1	7.7	14.1	12734.5	61.7
Complex 2	6.3	19	11453	66.4
Complex 3	7.7	16.4	12083.2	61.5

at the 5' end, and the HIS4 marker gene [44]. The insert DNA is typically integrated into the *P. pastoris* chromosome to obtain a stable expression system. Integration of the insert DNA into the *P. pastoris* chromosome can be achieved through two strategies: single crossover homologous recombination and double crossover

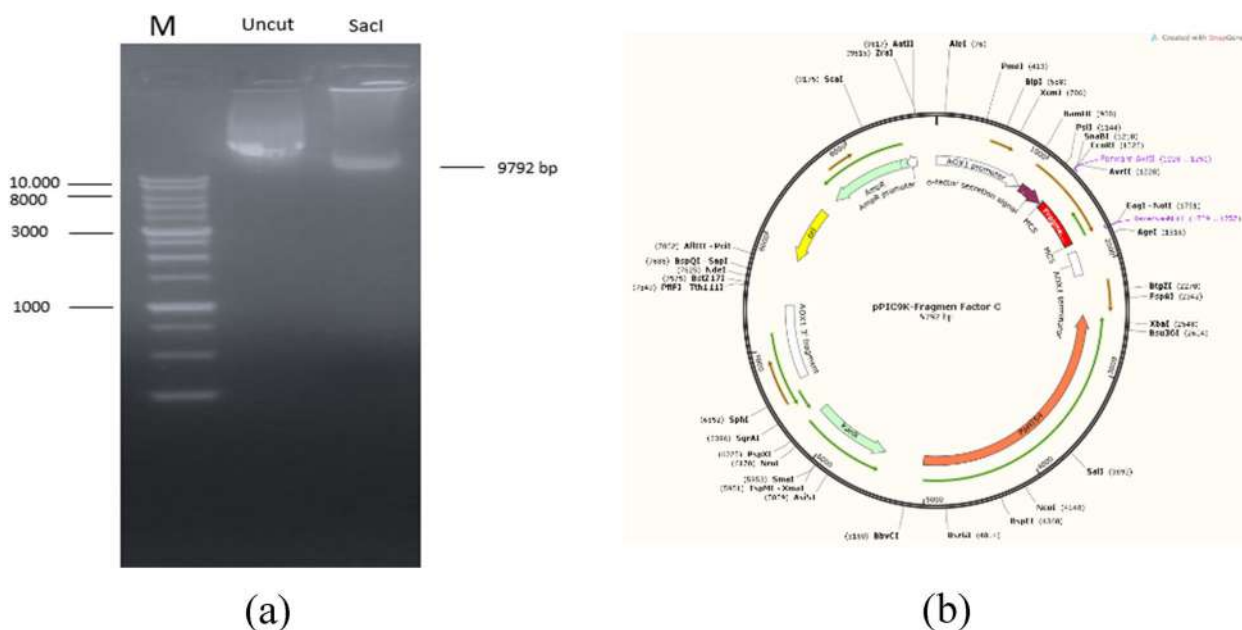


Fig. 5 **a** Digestion results of the recombinant plasmid for fragment factor C; lane M = marker, uncut = pPIC9K plasmid - fragment FC, SacI = digestion of the plasmid by the restriction enzyme. **b** Construction of FC gene fragment in pPIC9K

Table 5 Information on primer employed

Primer	Utility	DNA sequence	Annealing (Tm)	Amplicon size (bp)
AOX-Forward	Gene amplification	GACTGGTCCAATTGACAAGC	60	895
FC-Reverse		TCTGCTGCCACTTTAACAC		
FC-Forward	Gene amplification	ATGGGCAAGCTTCCAACTCT	63	516
FC-Reverse		TCTGCTGCCACTTTAACAC		

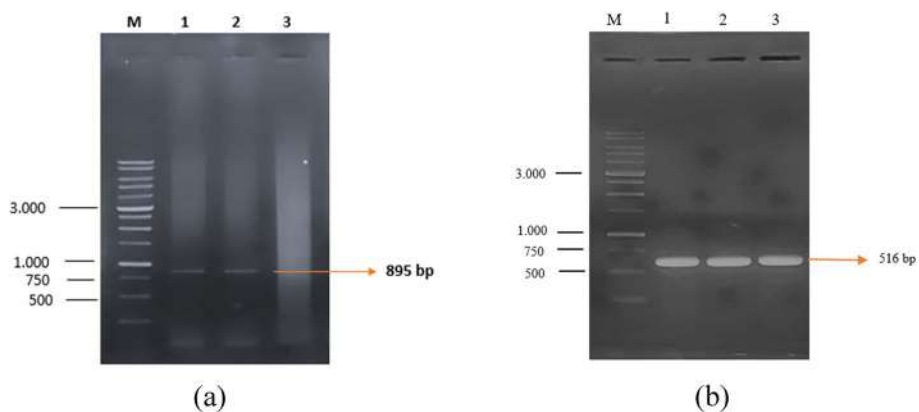


Fig. 6 Transformed colony PCR. **a** Primer AOX-forward (pPIC9K) and FC-reverse from Factor C gene fragment. **b** Transformed colony PCR using a pair of FC-forward and FC-reverse from Factor C gene fragment

homologous recombination, also known as gene transplacement [45].

Optimization of Factor C protein fragment expression

For the optimization of protein expression of the Factor C fragment, the clone was induced with methanol at concentrations of 0.5%, 1%, and 2%. The induction was performed in a medium with a pH of 6 at various temperatures (25°C, 30°C, and 37°C). The optimization results showed that the highest OD value of 16.83 was obtained when the induction was carried out with 1% methanol at a temperature of 30°C for 72 h of incubation. This OD value was higher compared to the other treatment conditions (Fig. 7).

As illustrated in Fig. 7, the growth pattern of *P. pastoris* cultured in BMMY media was compared at three different temperature levels. A lower cell density was observed at the end of fermentation when grown at 37°C compared to 25°C and 30°C. This observation supports the findings reported by Li et al. [45], who also observed a decrease in viable cell densities after three days of fermentation at 30°C, compared to 23°C.

SDS-PAGE and zymogram analyses was employed to measure the weight of the molecular (MW) of the recombinant Factor C protein fragment expressed in *P. pastoris* GS115. The results showed a distinct protein band indicating the MW of the Factor C protein fragment (Fig. 8). Visualization using SDS-PAGE revealed the MW of the Factor C protein fragment to be approximately 18.6 kDa (Fig. 8A). This was further confirmed by zymogram analysis using a casein substrate, which showed that the Factor C protein fragment, belonging to the serine protease group, successfully hydrolyzed the casein substrate when

induced with 1% methanol at 30°C for 72 h of fermentation (Fig. 8B).

The *P. pastoris* GS115 expression system, which offers several benefits, was used to express the recombinant protein. This system allows for easy manipulation at the molecular level, post-translational modifications, protein folding, and extracellular protein secretion [46]. The ease of production, relatively inexpensive production equipment, and scalability make the *P. pastoris* expression system highly desirable for producing recombinant proteins [47]. The scalability of the *P. pastoris* expression system allows for larger production compared to the *E. coli* expression system [48]. In some cases, the expression system can achieve higher levels of recombinant protein expression compared to conventional yeast expression systems using *S. cerevisiae* [49].

Methanol was added to the expression medium at final concentrations of 0.5%, 1%, and 2% for the best methanol concentration. The results demonstrated that 1% methanol concentration had the maximum protease activity. At greater methanol concentrations, the protease activity, however, sharply decreased [Fig. 8]. These results are consistent with the cell density measurements taken by monitoring OD₆₀₀ nm every 24 h, which indicated that cell growth reached the logarithmic phase after 72 h of fermentation [Fig. 7]. Subsequently, the expression of the Factor C fragment was tested at 25°C, 30°C, and 37°C to determine the optimal temperature for enzyme production. In our study, we successfully achieved high levels of Factor C fragment protein expression in the methylotrophic yeast *P. pastoris* after 72 h of induction under optimal conditions (30°C and 1% methanol concentration). The Factor C was secreted into the medium at a high level and protease activity was 3254 U/ml [Fig. 8].

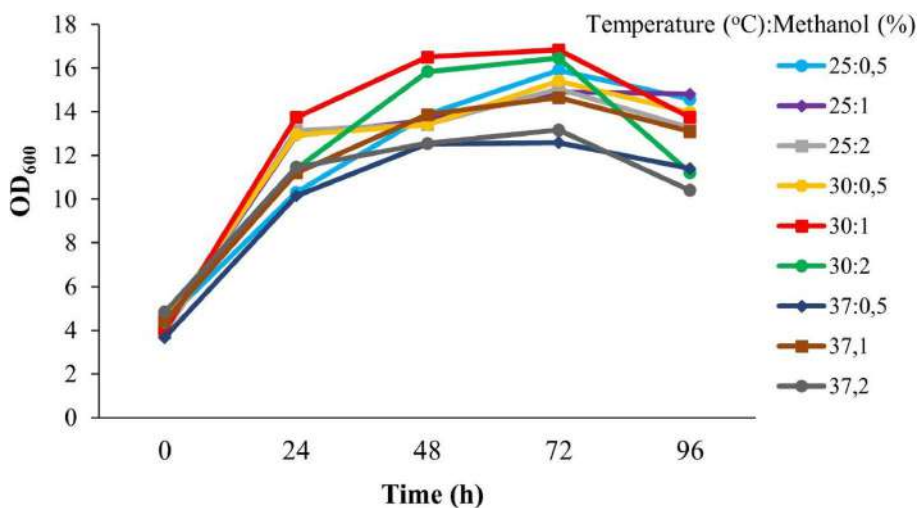


Fig. 7 Growth curve of recombinant *P. pastoris* GS115 under optimized MeOH induction and temperature conditions

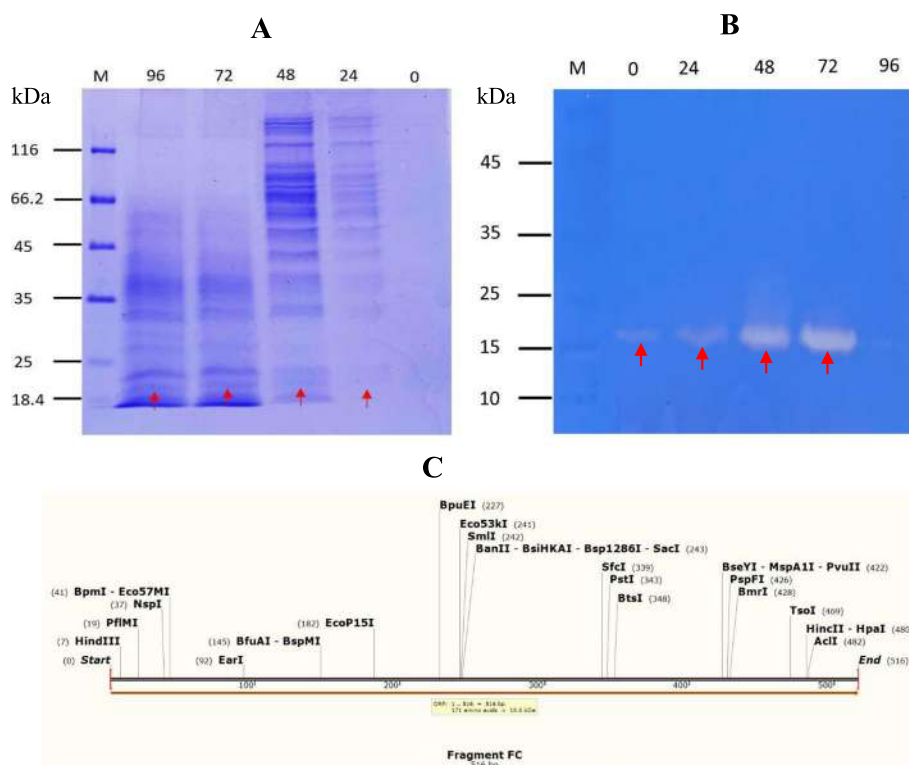


Fig. 8 The profile of the Factor C protein fragment from *P. pastoris* GS115. **A** SDS-PAGE analysis of serine protease induction with 1% MeOH at 30°C. M, marker; Lane 0, 0 h (no induction); lane 24, 24 h induction with 1% MeOH; lane 48, 48 h induction with 1% MeOH; lane 72, 72 h induction with 1% MeOH; lane 96, 96 h induction with 1% MeOH. **B** Zymogram analysis of serine protease induction with 1% MeOH at 30°C. M, marker; lane 0, 0 h (no induction); lane 24, 24 h induction with 1% MeOH; lane 48, 48 h induction with 1% MeOH, lane 72, 72 h induction with 1% MeOH, lane 96, 96 h induction with 1% MeOH. **C** Mapping of gene sequence and molecular weight of the Factor C protein (18.6 kDa)

Discussion

Due to its obvious benefits, including as inducible expression utilizing adjustable promoters and the ease of medium conditions, *P. pastoris* has become more popular as a host for heterologous protein production [50]. By employing codon optimization techniques, we achieved successful expression of the Factor C protein fragment at higher levels in this unicellular eukaryotic system. Subsequently, we conducted observations on the solubility of the Factor C protein fragment. Protein solubility is a crucial property in recombinant protein production and biotechnological development. Solubility prediction based on protein sequence provides a straightforward approach. The results, presented in Table 3, reveal that the solubility of the Factor C fragment protein is influenced by negatively charged amino acids Asp and Glu, with 6.4% and 4.7% composition, respectively. On the other hand, positively charged amino acids Arg and Lys contribute to the solubility with 5.8% and 5.3% composition, respectively. Additionally, hydrophobic amino acids Val, Leu, and Ile account for 8.2%, 7.6%, and 3.5%, respectively, while aromatic amino acids Phe, Tyr, and Trp contribute 4.7%, 2.3%, and 2.9%, respectively. Aromatic residues and

hydrophobic interactions have a minor impact on protein aggregation formation. Further confirmation through the Protein-Sol webserver yielded a solubility value of 0.454. The Protein-Sol analysis relies on the Population average Solubility (PopAvrSol) value, where a value greater than 0.45 indicates higher solubility, while values below 0.45 suggest lower solubility [51].

Molecular docking is a method used to examine how tiny molecules behave inside a protein's binding pocket and to forecast the potential affinities during binding [52]. Lipopolysaccharide (LPS) consists of three main components: O antigen, lipid A, and core oligosaccharide. Factor C recognizes LPS based on endotoxin activity. Among the three components of LPS, lipid A is considered the main constituent and plays a crucial role. Therefore, docking studies between Factor C and lipid A were performed to explore the binding site activity. Lipid A is of particular importance as it exhibits strong endotoxin detection capabilities. By binding to the receptor's area to create a ligand-receptor complex, three ligands diphosphoryl lipid A, core lipid A, and Kdo2 lipid A could behave as activators of the Factor C protein, according to our docking simulation. As a result, three

structures (Complexes 2, 3, and 4) were chosen for additional examination [25].

In the previous results, 3 models (core lipid A, diphosphoryl lipid A, and Kdo2 lipid A) were identified by using molecular docking. We then performed MD simulation on these three complexes to assess their stability in water solvent. Four validation metrics including root mean square deviation (RMSD), root mean square deviation (RMSF), solvent accessible surface area (SASA), and hydrogen bonds were calculated for all models. Figure 4a showed the RMSD as a function of time for all models. In this figure, all models were involved large fluctuation at 0–5 ns, but after this time, the RMSD value reached the stable state along the simulation. The small fluctuation during simulation is usually found at the protein-ligand complex due to some interactions in the system such as hydrogen bonds, hydrophobic interaction, electrostatic interaction, and water molecules. These interactions contribute to the structural rearrangement protein/ligand complex to achieve the equilibrium state. From the RMSD graph trends, all complexes were stable during the simulation. To know the flexibility of protein-ligand complex, RMSF profile was estimated. Models 2 and 3 seem to be more flexible as provided in Fig. 4b, suggesting the ligands could be easily bound to the site of proteins. In Fig. 4c, SASA value was calculated for all models. SASA profile is used to know the volume of the protein-ligand complex along the simulation. All complexes are not significant difference of SASA, indicating all protein-ligand complex formation might be stable during the simulations. Furthermore, hydrogen bonds were analyzed for all models. This interaction was also involved to maintain the structural interface of the protein-ligand. Figure 4d provided the hydrogen bond ranging from 40–80 for all complexes and the graphic trends tends to similar form. This finding suggested that hydrogen bond contributed in holding a stable structure of protein-ligand complex during the MD simulation.

The expression of Factor C from *T. gigas* using *P. pastoris* (GS115) as the host cell is reported here for the first time. In the case of producing different recombinant proteins, the *Pichia* expression system has several advantages [53]. Since *P. pastoris* produces less endogenous secretory proteins in this expression system, it is simpler to purify recombinant proteins [54]. To maximize the expression of the Factor C fragment, the time course of methanol-induced peptide synthesis was adjusted in this investigation. We employed AOX1, a potent methanol-induced promoter, and examined the effects of various methanol concentrations on the expression of heterologous proteins. The primary carbon source and gene expression inducer in the majority of *P. pastoris* fermentation techniques is methanol. Our findings demonstrated that

increased cell growth of *P. pastoris* was seen when treated with 1% methanol in BMMY medium and incubated at 30°C for 72 h; higher cell growth of *P. pastoris* was observed ($OD_{600}=16.83$). The production of recombinant proteins in *P. pastoris* is known to be influenced by culture and induction conditions [55–57]. The concentration of methanol, as an inducer, directly affects protein production due to the use of an inducible promoter [58]. While low levels of methanol may not effectively induce the AOX1 promoter, excessive amounts of methanol can be toxic to the cells. Protease activity was found to significantly decrease at higher methanol concentrations as the inducer.

Following the optimization of the Factor C fragment gene and cultivation of recombinant *P. pastoris* GS115, we successfully detected the 18.6 kDa Factor C fragment protein using SDS-PAGE and zymogram techniques. We report that the Factor C fragment consists of 516 bp, which translates to 171 amino acids. Based on the translation, the calculated molecular weight of the Factor C protein is 18.6 kDa [Fig. 9C]. The appropriate methanol concentration varied between 0.1 and 3% (v/v) [59] based on different foreign proteins. The alcohol oxidase (AOX) enzyme is encoded by two genes in the *P. pastoris* GS115 strain (AOX1 and AOX2). Methanol induces the transcription of these genes, which leads to the production of a significant quantity of AOX enzyme [60]. Methanol cannot efficiently stimulate the AOX1 promoter at low concentrations, but excessive levels of methanol can be harmful to cells [61]. The growth of *P. pastoris* GS115 cells exhibited a significant decline at higher concentrations of methanol used as an inducer (Fig. 7). Higher methanol concentrations appear to promote misfolded proteins while also accelerating cell growth and protein expression [62]. Protein concentration and activity rose at the highest concentration of methanol (2%), probably as a result of cell death and the subsequent release of intracellular proteins in medium culture.

According to various protocols, the optimal temperature for *P. pastoris* cell growth is 30°C. In this study, maximum cell growth of *P. pastoris* GS115 was also observed at 30°C (Fig. 7). An increase in cell numbers was also observed at 25°C after 72 h of induction with 0.5% methanol. Similarly, Naseem et al. [63] reported that maximum expression of rMgTx was achieved when biomass was induced with 0.5% methanol. Due to the sensitivity of *P. pastoris* cells to temperature, high temperatures can negatively impact cell growth and protein expression. Since the recombinant Factor C fragment from *P. pastoris* is intended for industrial-scale applications, a relatively fast protein expression time is desirable. Therefore, the recommended expression time is 72 h of induction with 1% methanol at 30°C.

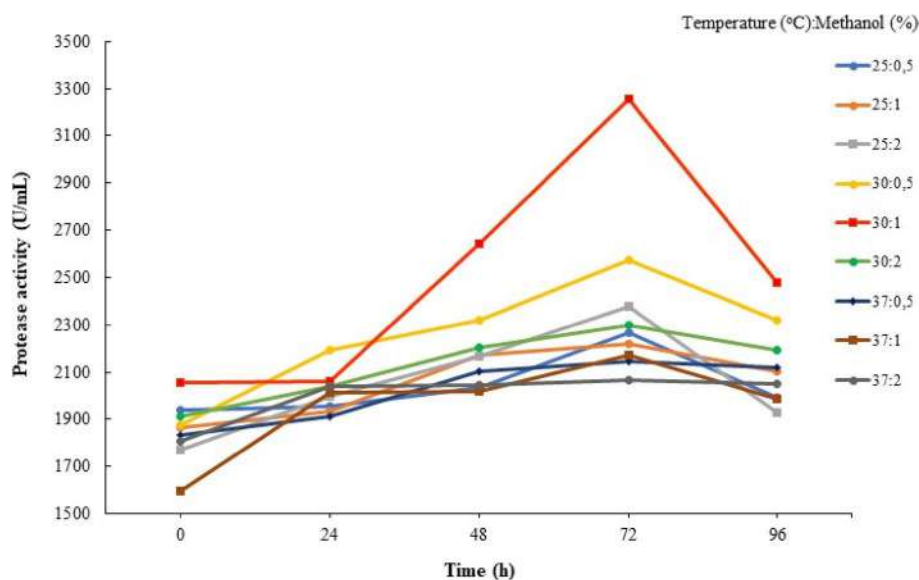


Fig. 9 The Profile of protease activity of recombinant *P. pastoris* GS115 expressing the Factor C protein after induction with varying concentrations of MeOH at different temperatures

Conclusions

Protein binding in several complexes there are amino acid residues that bind in the same region to the amino acid Tyr46 so that it can be estimated that recombinant protein fragment Factor C has activity. Molecular dynamics (MD) simulations were conducted on all three complexes to evaluate their stability in water solvent. The results showed that all complexes are stable along the simulation. We have also successfully demonstrated the expression of the Factor C gene fragment of *T. gigas* in the eukaryotic organism *P. pastoris*, as a vector system. Based on our findings, we recommend optimizing the expression of Factor C gene fragments by inducing them with 1% methanol at a temperature of 30°C and incubating them for 72 h. This optimized condition can be used for the production of recombinant Factor C gene fragments on a larger scale in a biofermentor.

Acknowledgements

This work was supported by Research Center for Genetic Engineering, National Research and Innovation Agency (BRIN), Bogor, 16911, we are grateful to have anonymous reviewers for final version of the manuscript.

Authors' contributions

ZB data analysis, interpretation, and wrote the manuscript. AZM planned, designed the research, wrote the manuscript, and performed experiments. RIA data analysis and interpretation. FF conceptualization and review. FF conceptualization and review. RR sampling. TNMW data analysis and interpretation. DPW data analysis and interpretation. FA sampling. AA molecular docking and molecular dynamic factor C protein. HI conducted laboratory work. JM conducted laboratory work, writing, reviewing, and editing.

Funding

This study is fully funded by National Research and Innovation Agency/*Badan Riset dan Inovasi Nasional (BRIN)* under scheme of Program Riset dan Inovasi untuk Indonesia Maju (RIIM) with the contract number: No. 12/II.7/HK/2023 and

Universitas Sriwijaya as the local host for the project with the contract number: 0256/UN9/3.1/PT/2023.

Availability of data and materials

Not applicable

Declarations

Ethics approval and consent to participate

Animal used in this study complied guidelines and approved by Ethical Committee "Komite etik Penelitian Kedokteran dan Kesehatan (KEPKK)" Faculty of medicine University of Sriwijaya with Protocol No 126-2021.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Post Graduate Program of Biotechnology, Bogor Agricultural University, Bogor, Indonesia. ²Research Center for Genetic Engineering, Research Organization for Life Sciences and Environment, National Research and Innovation Agency (BRIN), KST Soekarno, Cibinong, Bogor 16911, Indonesia. ³Department of Biology, IPB University, Bogor, West Java 16680, Indonesia. ⁴Biotechnology Research Center, IPB University, Bogor, West Java, Indonesia. ⁵Marine Science Study Program, Faculty of Mathematics and Natural Science, University of Sriwijaya, Palembang, Indonesia. ⁶Research Center for Conservation of Marine and Inland Water Resources, National Research and Innovation Agency, Cibinong 16911, Indonesia. ⁷Department of Pharmacy, Faculty of Science, Sriwijaya University, Indralaya, South Sumatera, Indonesia. ⁸Department of Chemistry Education, Faculty of Teacher Training and Education, Tadulako University, Palu, Indonesia. ⁹Department of Biology Education, Faculty of Education and Teacher Training, Universitas Kristen Indonesia, Jl. Mayjen Sutoyo No. 2, Cawang, Jakarta Timur 13630 Jakarta, Indonesia.

Received: 26 June 2023 Accepted: 8 October 2023

Published online: 17 October 2023

References

- Eldredge N, Stanley SM. 1984. Living fossils: introduction to the casebook. In: Eldredge, N., Stanley, S.M. (eds) Living fossils. Earth Sci. Springer NY. https://doi.org/10.1007/978-1-4613-8271-3_1
- Latreille PA. 1802. Histoire naturelle, générale et particulière des crustacés et des insectes, tome troisième 1-467
- Müller OF. 1785. Entomostraca seu insecta testacea, quae in aquis daniae et norvegiae reperit, descriptis et iconibus illustravit otho fridericus müller. F.W. Thiele, Lipsiae & Havniae. 1-134
- Leach WE. 1819. [Fam. IX. Formicidae. Leach.]. Pp. 272-273 in: Samouelle G. 1819. The entomologist's useful compendium. London: Thomas Boys, 496 pp
- Linnaeus C. 1758. Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata. Holmiae [= Stockholm]: L. Salvii, 824 pp
- Walls E, Berkson J, Smith S (2002) The horseshoe crab, *Limulus polyphemus*: 200 million years of existence, 100 years of study. Rev Fish Sci 10:39–73. <https://doi.org/10.1080/20026491051677>
- IUCN (International union for conservation of nature). 2015. Red list of threatened species. Retrieved on June 2023. <http://www.iucnredlist.org>
- KSDAE. 2018. Statistik direktorat jenderal konservasi sumber daya alam dan ekosistem tahun 2017. Jakarta: kementerian lingkungan hidup dan kehutanan
- Ding JL, Ho B (2010) Endotoxin detection—from limulus amebocyte lysate to recombinant factor C. Subcell Biochem 53:187–208. https://doi.org/10.1007/978-90-481-9078-2_9
- Novitsky TJ (1994) Limulus amebocyte lysate (LAL) detection of endotoxin in human blood. J Endotoxin Res 1:253–263
- Clifton LA, Ciesielski F, Skoda MWA, Paracini N, Holt SA, Lakey JH (2016) The effect of lipopolysaccharide core oligosaccharide size on the electrostatic binding of antimicrobial proteins to models of the gram negative bacterial outer membrane. Langmuir 32:3485–3494. <https://doi.org/10.1021/acs.langmuir.6b00240>
- Muta T, Miyata T, Misumi Y, Tokunaga F, Nakamura T, Toh Y, Ikehara Y, Iwanaga S (1991) Limulus factor C. An endotoxin-sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like, and lectin-like domains. J Biol Chem 10:6554–6561
- Abate W, Sattar AA, Liu J, Conway ME, Jackson SK (2017) Evaluation of recombinant factor C assay for the detection of divergent lipopolysaccharide structural species and comparison with limulus amebocyte lysate-based assays and a human monocyte activity assay. J Med Microbiol 66:888–897. <https://doi.org/10.1099/jmm.0.000510>
- Iwanaga S (2002) The molecular basis of innate immunity in horseshoe crab. Curr Opin Immunol 14:87–95
- Ding JL, Navas MA 3rd, Ho B (1993) Two forms of factor C from the amoebocytes of *Carcinoscorpius rotundicauda*: purification and characterisation. Biochim Biophys Acta 1202:149–156. [https://doi.org/10.1016/0167-4838\(93\)90076-4](https://doi.org/10.1016/0167-4838(93)90076-4)
- Ariki S, Koori K, Osaki T, Motoyama K, Inamori K, Kawabata S (2004) A serine protease zymogen functions as a pattern-recognition receptor for lipopolysaccharides. Proc Natl Acad Sci 101:953–958. <https://doi.org/10.1073/pnas.0306904101>
- Ding JL, Navas MA 3rd, Ho B (1995) Molecular cloning and sequence analysis of Factor C cDNA from the Singapore horseshoe crab *Carcinoscorpius rotundicauda*. Mol Mar Biol Biotechnol 4:90–103
- Roopashree SD, Ho B, Ding JL (1996) Expression of *Carcinoscorpius rotundicauda* Factor C in *Pichia pastoris*. Mol Mar Biol Biotech 5:334–343
- Pui AWM, Ho B, Ding JL (1997) Yeast recombinant Factor C from horseshoe crab binds endotoxin and causes bacteriostasis. J Endotoxin Res 4:391–400
- Ding R, West RR, Morphew DM, Oakley BR, McIntosh JR (1997) The spindle pole body of *Schizosaccharomyces pombe* enters and leaves the nuclear envelope as the cell cycle proceeds. Mol Biol Cell 8:1461–1479
- Roopashree SD, Ho B, Ding JL (1997) Recombinant COS-1 cells express *Carcinoscorpius rotundicauda* Factor C. Biotech Lett 19:357–361
- Mizumura H, Ogura N, Aketagawa J, Aizawa M, Kobayashi Y, Kawabata S, Oda T (2017) Genetic engineering approach to develop next-generation reagents for endotoxin quantification. Innate Immun 23:136–146. <https://doi.org/10.1177/1753425916681074>
- Ding JL, Ho B (2001) A new era in pyrogen testing. Trends Biotechnol 19:277–281
- McClymont K (2017) Structural studies of the endotoxin sensing protein Factor C. [Doctoral dissertation, University of Glasgow]. University of Glasgow Research Repository. <https://theses.gla.ac.uk/8058/>
- Mustopa AZ, Izaki AF, Suharsono S, Fatimah F, Fauziyah F, Damarani R, Arwansyah A, Wahyudi ST, Sari SS, Rozirwan R, Bachtiar Z (2023) Characterization, protein modeling, and molecular docking of Factor C from Indonesian horseshoe crab (*Tachypleus gigas*). J Genet Eng Biotechnol 21:44. <https://doi.org/10.1186/s43141-023-00496-8>
- Higgins DR, Cregg JM (1998) Introduction to *Pichia pastoris*. Methods Mol Biol. 103:1–15. <https://doi.org/10.1385/0-89603-421-6:1>
- Yan Y, Tao H, He J, Huang S (2020) The HDock server for integrated protein–protein docking. Nat Protoc 15:1829–1852. <https://doi.org/10.1038/s41596-020-0312-x>
- Arwansyah A, Arif AR, Kade A, Taiyeb M, Ramli I, Santoso T, Ningsih P, Natir H, Tahril T, Uday Kumar K (2022) Molecular modelling on multiepitope based vaccine against SARS-CoV-2 using immunoinformatics, molecular docking and molecular dynamics simulation. SAR QSAR Environ Res. <https://doi.org/10.1080/1062936X.2022.2117846>
- Salomon-Ferrer R, Case DA, Walker RC. 2013. An overview of the amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci. 1-13. <https://doi.org/10.1002/wcms.1121>
- Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 11:3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh ewald method. J Chem Phys 103:8577–8593
- Ryckaert JP, Ciccotti G, Berendsen HJC (1997) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
- Loncharich RJ, Brooks BR, Pastor RW (1992) Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanine-N-methylamide. Biopolymers 32:523–535
- Roe DR, Cheatham TE (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 9:3084–3095
- Fitrasari D, Arwansyah MS, Kawaguchi K, Purqon A, Nagao H (2022) Theoretical study of dissociation process of plastocyanins by pacs-md simulation. J Phys Conf Ser 2207:1–5
- Arwansyah A, Arif AR, Ramli I, Hasrianti H, Kurniawan I, Ambarsari L, Sumaryada TI, Taiyeb M (2022) Investigation of active compounds of brucea javanica in treating hypertension using a network pharmacology-based analysis combined with homology modeling, molecular docking and molecular dynamics simulation. Chem Select 7:e202102801
- Arwansyah A, Arif AR, Kade A, Taiyeb M, Ramli I, Santoso T, Ningsih P, Natir H, Tahril T, Uday Kumar K (2022) Molecular modelling on multiepitope-based vaccine against SARS-CoV-2 using immunoinformatics, molecular docking, and molecular dynamics simulation. SAR QSAR Environ Res 33:649–675. <https://doi.org/10.1080/1062936X.2022.2117846>
- Bencsik P, Bartekova M, Görbe A, Kiss K, Pálóczi J, Radosinska J, Szűcs G, Ferdinandy P (2017) MMP activity detection in zymograms. Methods Mol Biol 1626:53–70. https://doi.org/10.1007/978-1-4939-7111-4_6
- Cupp-Enyard C (2008) Sigma's non-specific protease activity assay casein as a substrate. J Visual Exp 19:1–2
- Sobolev O, Afonine PV, Moriarty NW, Hekkelman ML, Joosten RP, Perrakis A, Adams PD (2020) A global Ramachandran score identifies protein structures with unlikely stereochemistry. Structure 28:1249–1258. <https://doi.org/10.1101/2020.03.26.010587>
- Meng XY, Zhang HX, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7:146–57. <https://doi.org/10.2174/157340911795677602>
- Arwansyah A, Aris RA, Ramli I, Kurniawan I, Sukarti S, Nur Alam M, Illing I, Farid Lewa A, Manguntungi B (2021) Molecular modelling on SARS-CoV-2 papain-like protease: an integrated study with homology modelling, molecular docking, and molecular dynamics simulations. SAR QSAR Environ Res 32:699–718. <https://doi.org/10.1080/1062936X.2021.196060>
- Mustaffa ABS, Wagiran A (2013) Molecular cloning and bioinformatic analysis of endosperm specific. J Teknol 2:1–5

44. Burdychova R, Ruzicka V, Bartos M (2002) PCR-based method for identification of integration events in the *Pichia pastoris* genom. *Biotechniques* 33:1214–1218
45. Li Z, Xiong F, Lin Q, d'Anjou M, Daugulis AJ, Yang DSC, Hew CL (2001) Low-temperature increases the yield of biologically active herring anti-freeze protein in *Pichia pastoris*. *Protein Exp Purif* 21:438–445
46. Balamurugan V, Reddy G, Suryanarayana V (2007) *Pichia pastoris*: a notable heterologous expression system for the production of foreign proteins-vaccines. *Indian J Biotechnol* 6:175–186
47. Voza LA, Wittwer L, Higgins DR, Purcell TJ, Bergseid M, Collins-Racie LA, LaVallie ER, Hoeffler JP (1996) Production of a recombinant bovine enterokinase catalytic subunit in the methylotrophic yeast *Pichia pastoris*. *Biotechnol* 14:77–81. <https://doi.org/10.1038/nbt0196-77>
48. Fickers P (2014) *Pichia pastoris*: a workhorse for recombinant protein production. *Curr Res Microbiol Biotechnol* 2:354–363
49. Patrick SM, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast J* 22:249–270
50. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J (2017) Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinform* 33:3098–3100
51. Roy K, Kar S, Das RN (2015) Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Academic Press, Cambridge. <https://doi.org/10.1016/B978-0-12-801505-6.00010-7>
52. Pansar T, Poso A (2018) Binding affinity via docking: fact and fiction. *Molecules*. <https://doi.org/10.3390/molecules23081899>
53. Karbalaeei M, Rezaee SA, Farsiani H (2020) *Pichia pastoris*: a highly successful expression system for optimal synthesis of heterologous proteins. *J Cell Physiol* 235:5867–5881. <https://doi.org/10.1002/jcp.29583>
54. Alias NI, Mahadi NM, Murad AMA, Bakar FDA, Rabu A, Illias RM (2011) Expression optimization of recombinant α -L arabinofuranosidase from *Aspergillus niger* ATCC 120120 in *Pichia pastoris* and its biochemical characterization. *J Biotechnol* 10:6700–6710
55. Damaso MCT, Almeida MS, Kurtenbach E, Martins OB, Pereira J, Andrade CM, Albano RM (2003) Optimized expression of a thermostable xylanase from *Thermomyces lanuginosus* in *Pichia pastoris*. *Appl Environ Microbiol* 69:6064–6072
56. Thiruvengadam G, Init I, Fong MY, Lau YL (2011) Optimization of the expression of surface antigen SAG1/2 of *Toxoplasma gondii* in the yeast *Pichia pastoris*. *Trop Biomed* 28:506–513
57. Minning S, Serrano A, Ferrer P, Sola C, Schmid RD, Valero F (2001) Optimization of the high-level production of *Rhizopus oryzae* lipase in *Pichia pastoris*. *J Biotechnol* 86:59–70
58. Jafari R, Sundström BE, Holm P (2011) Optimization of production of the anti-keratin single-chain Fv Ts1-218 in *Pichia pastoris* using design of experiments. *Microb Cell Fact* 10:1–8
59. Minjie G, Zhongping S (2013) Process control and optimization for heterologous protein production by methylotrophic *Pichia pastoris*. *Chin J Chem Eng* 21:216–226
60. Vanz A, Lünsdorf H, Adnan A, Nimtz M, Gurrankonda C, Khanna N, Rinas U (2012) Physiological response of *Pichia pastoris* GS115 to methanol-induced high-level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. *Microb Cell Factories* 11:103
61. Dehnavi E, Siadat SOR, Roudsari MF, Khajeh K (2016) Cloning and high-level expression of β -xylosidase from *Selenomonas ruminantium* in *Pichia pastoris* by optimizing of pH, methanol concentration and temperature conditions. *Protein Expr Purif* 124:55–61
62. Hong F, Meinander NQ, Jonsson LJ (2002) Fermentation strategies for improved heterologous expression of laccase in *Pichia pastoris*. *Biotechnol Bioeng* 79:438–449
63. Naseem MU, Tajti G, Gaspar A, Szanto TG, Borrego J, Panyi G (2021) Optimization of *Pichia pastoris* expression system for high-level production of margatoxin. *Front Pharmacol* 12:733610. <https://doi.org/10.3389/fphar.2021.733610>
64. Nakamura H, Morita T, Sato C (1986) Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp Cell Res* 165:291–297

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)