

Structure Identification and Quality Assessment of Laccase (Lac InaCC) from *Neurospora crassa* by Using a Structure Prediction

Rini Kurniasih¹, Laksmi Ambarsari^{1*}, Setyanto Tri Wahyudi²

¹Department of Biochemistry, Bogor Agricultural University, Bogor, Indonesia

²Department of Physics, Bogor Agricultural University, Bogor, Indonesia

ARTICLE INFO

Article history:

Received October 19, 2019

Received in revised form November 15, 2020

Accepted November 20, 2020

KEYWORDS:

laccase,
Neurospora crassa,
sequence motifs,
structure prediction,
molecular dynamics simulation

ABSTRACT

Laccases are multi-copper oxidase enzyme, developed for being applied widely. The laccase gene in this study was isolated from local isolates of *Neurospora crassa* (LAC inaCC). The structure of this enzyme has not been known and there is no laccase structure of *Neurospora crassa* based on protein structure development in database. Here, we aimed to analyze the characteristics of the sequence and prediction structure, the structure quality after refinement through the molecular dynamics (MD) simulation method. LAC inaCC has been identified with typical sequence motifs (HWH, HSH, HXXH) which played role in copper-binding on 274(HWH)G-DG-T-CP on CBL-1, 314GT-WY(HSH)FS-QYG-G on CBL-2, and 607HPIHL on CBL-3. The four copper atoms have an important role in the catalytic activity. LAC inaCC is a multi-subunit enzyme consisted of three functional domains with structural motifs of Greek-key β barrel which is typical structure motif. Refinement in the prediction structure through the MD simulation showed that this method was proven to be able to improve the structure quality. The increase on the most favoured area on Ramachandran plot, clashcore percentile score, and molprobit score showed that the laccase structure headed to conformation change, to be more stable conformation with better resolution compared to earlier prediction structure.

1. Introduction

Laccases are oxidoreductase enzymes with EC 1.10.3.2, the enzyme that catalyzes substrate with polyphenol aromatic substance as the donor of electron and involves oxygen as the recipient of the electron to form water molecule. Laccases are enzymes found in bacteria, fungi, and plants. Laccases play different roles in each organism. In fungi, laccases act differently with ones in plants, where laccases play a role in lignin degradation contained in the cell wall of the host plants (Dwivedi *et al.* 2011). The differences in laccase roles in each organism caused laccases catalytic properties vary depending on the sources, such as various types of thermostability and substrate. Catalytic and functional properties of the enzyme are closely related to characteristics at the molecular level, i.e., the genetic sequence and the structure of the enzyme. Understanding the properties and functions of catalytic enzymes at the molecular level in the computational simulation is faster, more

efficient, and more cost-effective than the laboratory experiment.

Laccase analyzed in this study was from our previous research, where the laccase gene was isolated from *Neurospora crassa* code of inaCC F226 fungi. The structure of LAC inaCC has not been known and there is no laccase structure of *Neurospora crassa* fungi based on protein structure development in the laccase database of Biocatnet LccED V6.4 yet. The purposes of this research were to predict the structure of LAC inaCC, to identify the sequence characteristics, to identify the laccase prediction structure, and to ameliorate the prediction structure conformation obtained. The structure of LAC inaCC in this research was obtained from structure prediction by using the fold recognition method, I-TASSER.

Protein structure quality appraisal plays an important role in the structure prediction of protein, as the structure prediction of protein might contain a substantial error (Benkert *et al.* 2011). Refinement on such structures is necessary if the prediction has poor results (Benkert *et al.* 2011). The refinement of the structure prediction of LAC inaCC in this research used the molecular dynamics (MD) simulation method.

* Corresponding Author

E-mail Address: laksmi@apps.ipb.ac.id

The MD simulation method is known to be able to ameliorate the structure prediction of protein (Feig 2016; Feig and Mirjalili 2016).

2. Materials and Methods

2.1. Materials

The nucleotide sequence of laccase, the amino acid sequence of laccase from the LAC inaCC gene was translated by using the sequence translation sites, 17 amino acid sequences and PDB files consisting of laccase and the multicopper oxidase (MCO) enzymes from homologous sequences of LAC inaCC based on BLASTp homology analysis and the template of the modeling structure of LAC inaCC. The laccase's structure was predicted by I-TASSER.

2.2. Instruments

The high-performance computer (HPC) with Linux Ubuntu system 64 bit, identification of sequence and structure characteristics using the sequence translation site from the European Bioinformatics Institute (EBI), homology analysis BLASTp, the conserved domain architecture retrieval tool (CDART), and the multiple sequences alignment by Kalign. Assessment of protein structure quality before refinement was carried out using PDBsum and molprobity. Molecular dynamics simulation with Assisted model building with energy refinement (AMBER) system version 12 for assessment and refinement of the prediction structure of LAC inaCC. The Protein structure visualization was carried out using Pymol dan Visual Molecular Dynamics (VMD) version 1.9.2.

2.3. Homologous Analysis

The nucleotide sequence of LAC inaCC was translated into the amino acid sequence via the Emboss_Transseq sequence translation server from the European Bioinformatics Institute (EBI). Then, the homology of the amino acid sequence was analyzed using BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), to determine the similarity of the amino acid sequences to the amino acid sequences in the Protein Data Bank (PDB) database.

2.4. Amino Acids Sequences Characterization

The amino acid sequence was characterized to prove that it is a laccase gene. Characteristics analyzed were the sequence motif and the copper-binding site of laccase by using the multiple sequence alignment method, Kalign server. That sequence aligned with 17 amino acid sequence of laccase and other enzymes from the multicopper oxidase (MCO) superfamily. The determination of sequence motif and copper-binding site of LAC inaCC referred to Giardina *et al.* (2009).

Then, some sequences that play a role in the sequence motif and copper-binding site were visualized into sequence logos by using a web logo (<https://weblogo.berkeley.edu/logo.cgi>).

2.5. LAC inaCC Structure Prediction

The amino acid sequence resulted from the Emboss_Transseq EBI was used to predict the structure of LAC inaCC. The structure of LAC inaCC was predicted by using the fold recognition method, I-TASSER. The amino acid sequence was submitted to I-TASSER, <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>. The prediction is started by clicking the "RUN I-TASSER" icon. Then, the best prediction structure of LAC inaCC resulted in I-TASSER was chosen, model 1 was chosen as the best predictive structure. The quality of the prediction structure produced by I-TASSER was evaluated for the first time using the web server protein structure quality, including 1) PDBsum, <http://www.ebi.ac.uk/pdbsum/>, to see structural conformational stability through Ramachandran plots and percentage regions of most favored region and disallowed region, 2) proQ, <https://proq.bioinfo.se/ProQ/ProQ.html>, to predict protein structure quality by LGscore and MaxSub, 3) molprobity, <http://molprobity.biochem.duke.edu/>, to evaluate the model quality of protein at both the global and local level by molprobity score.

2.6. Structure Characterization

The structure prediction of LAC inaCC was conducted to analyze its characteristics such as domains, structural motifs, active sites, and copper-binding sites. The domain was analyzed using CDART, structural motifs were analyzed using Pymol and VMD, active site and copper-binding site were analyzed using the multiple sequence alignment method (Kalign) and PDBsum, then they were visualized by Pymol.

2.7. Refinement of The Prediction Structure of LAC inaCC by MD Simulation

The structure prediction of LAC inaCC was refined by molecular dynamics simulation method. This simulation was conducted using AMBER software, with the ff14SB force field that was recommended for protein. Lacase is a metalloprotein consisting of three copper atoms. Parameterization (preparation stage) of three copper atoms is carried out using the MCPB.py Version 2.0 method which has been released in AmberTools16. The explicit solvent model was chosen for laccase's structure simulation with water solvent (code: TIP3PBOX) without adding Cl⁻ and Na⁺. First, the structure was prepared in a box with a distance between protein and the box's wall was 20 Å. The structure was subsequently minimized using an all-atom force field

on Ca atoms with respect to the initial model. Applying constraints in minimization was necessary because the initial model may be energetically very unfavorable and could cause the conformation alteration due to high initial strain energies (Shen *et al.* 2015; Feig 2016). Then, the equilibration and heating step was conducted using NPT condition with temperature in an interval of 0 K to 310 K. On the final step, the structure simulation is conducted in NPT condition with a temperature of 310 K for 100 ns time simulation.

2.8. Assessment of The Prediction Structure of LAC inaCC Before and After Refinement by MD Simulation

Assessment of the structure quality of LAC inaCC was conducted using PDBsum to obtain the Ramachandran plot with the percentage of amino acid distribution regions, and molprobit to obtain claschore and molprobit score for describing the resolution of the structure prediction. Meanwhile, the structure prediction of laccase after refinement by MD simulation was analyzed by using PDBsum, molprobit, and some parameters of molecular dynamics such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), interaction energy, and the solvent-accessible surface area (SASA).

3. Results

The laccase (LAC inaCC) gene had the molecular weight of 1,900 bp with a similarity percentage of 95-97% towards laccase from a different strain of *Neurospora crassa* and *Neurospora tetrasperma*. The homology of the amino acid sequence from translation was not high. The highest similarity (Table 1) was ones to laccase of *Melanocarpus albomyces* and *Thielavia arenaria* with similarity percentage (identical) of

63%, this was because *Neurospora crassa* is a fungus from the same ordo with *Melanocarpus albomyces* and *Thielavia arenaria* (Phylum of Ascomycetes and ordo of Sordariales). Meanwhile, the similarity with laccase sequences from other organisms was only about 34-35%. Through multiple sequence alignment it is possible to identify enzyme groups by analyzing sequence motifs especially for sequences that have a low level of similarity such as LAC inaCC which only has the highest similarity of 63%.

This enzyme belongs to the multicopper oxidase (MCO) superfamily of enzyme and has three sequence motifs of H-X-H as shown in Figure 1. The sequence motifs of H-X-H consisting of HWH, HSH, and HXXH. The HWH motif is located on the tip of N area. Meanwhile, the HSH and HXXH are on the tip of C area. These three motifs H-X-H are part of amino acid residue interacting with copper atoms (CU). All of the sequence motifs of laccases were designed to be sequence logo which can be seen in Figure 2. The sequence motif is not only used to identify a group of the enzyme but also can be used to identify the functional site of the enzyme, such as catalytic site, binding site, structural motif, etc. The three H-X-H motifs are related to copper atom binding site involving the most conserved histidine site.

The determination of copper atom binding site was known by laccase amino acid sequence alignment. There were four loops of copper-binding sites involving the H-X-H motifs and a number of amino acids. There were three loops of copper-binding site found in LAC inaCC, named copper-binding loop (CBL), with a number of conserved amino acid residues in each loop, in Table 2 and Figure 3. The conserved amino acid residues in each copper-binding loop were ²⁷⁴(HWH) G---DG---T-CP on CBL-1, ³¹⁴GT-WY(HSH)FS-QYG-G--- on CBL-2, meanwhile the CBL-3 consisted of only five amino acids, they were ⁶⁰⁷HPIHL. These conserved

Table 1. Homology parameters of the amino acid sequence of LAC inaCC

Organism	Max score	Total score	Query cover %	E value	Ident %	Accession
Laccase <i>Melanocarpus albomyces</i> mutant L559A	599	599	72	0.0	63	3DKH_A
Laccase <i>Melanocarpus albomyces</i>	599	599	72	0.0	63	1GW0_A
Laccase <i>Melanocarpus albomyces</i> Near-atomic resolution structure	596	596	72	0.0	63	2Q90_A
Laccase <i>Thielavia arenaria</i>	585	585	73	0.0	63	3PPS_A
Laccase <i>Botrytis aclada</i> mutant L499M	274	274	81	7e-84	35	3V9E_A
Laccase <i>Botrytis aclada</i> at 1.67Å resolution	274	274	81	7e-84	35	3SQR_A
Multicopper oxidase <i>Aspergillus niger</i> mutant H253D	244	244	70	2e-72	34	5LWX_A
Laccase-like-Multicopper oxidase <i>Aspergillus niger</i> bound to zinc	243	243	70	2e-72	34	5LWW_A
Laccase-like-Multicopper oxidase <i>Aspergillus niger</i>	243	243	70	2e-72	34	5LM8_A
Laccase <i>Lentinus</i> sp. At 1.8Å resolution	179	179	62	3e-49	34	3X1B_A



Figure 1. Amino acid sequence alignment of LAC inaCC with laccases and other enzymes from the different organisms by Kalign. The other enzymes are shown with asterisks, they are the multicopper oxidase from *Aspergillus niger*. The LAC inaCC sequence is shown in bold. The black box is the laccase sequence motif (the conserved sequence)

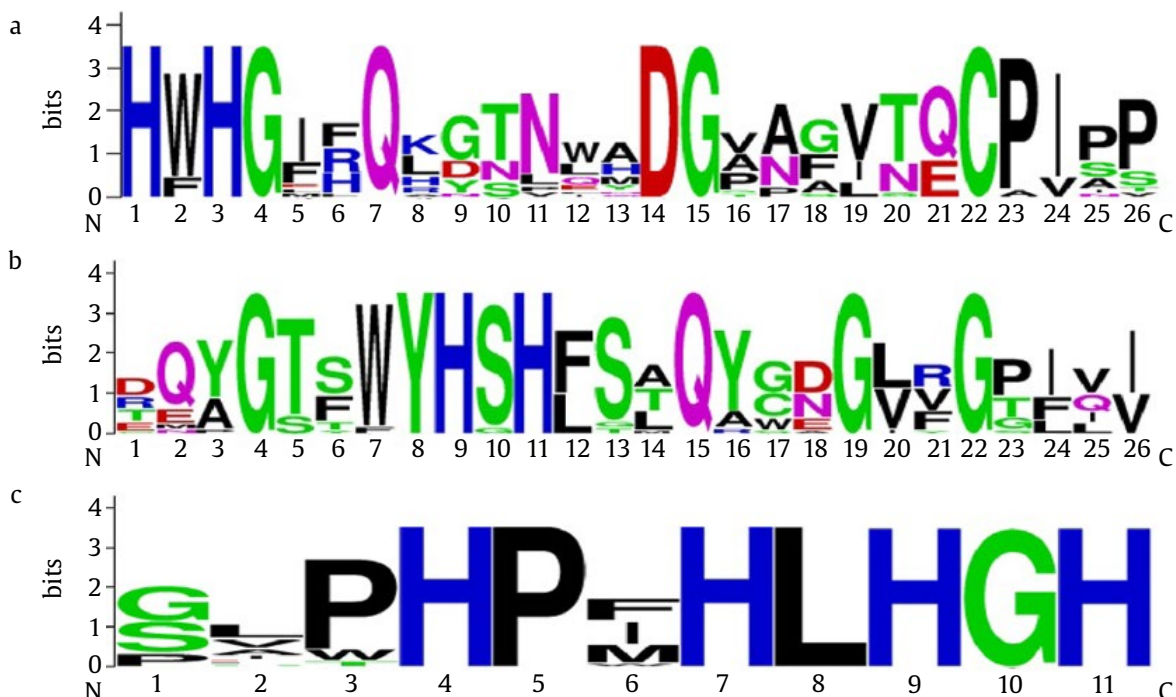


Figure 2. Logo sequences for the copper-binding residues of LAC inaCC. The first, second and third of copper-binding residues are shown in the a, b and c logo sequences. The sequence logos were created by using web logo

amino acids were expected to have an important role in the catalytic activity of laccase.

The structure prediction of LAC inaCC by using I-TASSER, showed that the enzyme is a multisubunit monomer enzyme consisting of three functional subunits known as cupredoxin-like domain. Three domains composing this enzyme are domain A (1-341), domain B (342-500), and domain C (501-611), in Figure 4. This enzyme contained four copper atoms, showed in Figure 4a. The four copper atoms showed that the enzyme belongs to a group of multicopper oxidase (MCO) enzyme.

In addition to domain analysis and number of copper atoms, several other characteristics were also analyzed to identify that the LAC inaCC structure predicted by iTASSER is a lacase structure such as the substrate pocket and the substrate-binding loop and the substrate pocket. The analysis of substrate-binding loop result showed that the LAC inaCC had four substrate-binding loops, they were loop I to loop IV and can be seen in Figure 5a. The conserved amino acid residues in the substrate-binding loops

of the protein were (X)AD(X) TS(X)NHF on loop I, VTFG(X)CG(X) on loop II, LD(X)FVW on loop III, and the residue of YWLIE(X)PHP(X)H(X) on loop IV.

The substrate pocket of LAC inaCC was composed of a number of amino acids located near to T1 copper (the site between domain A and B). The T1 copper and the site around it are the location of substrate oxidation of laccase (active site) in Figure 5. Most of the substrate pocket of this enzyme was composed of hydrophobic residues, they are Phe322, Phe351, Ala361, Val365, Leu366, Ala555, Ile556, Phe419, and Trp560, and the hydrophilic residues are Ser323, Asp362, Gln369, Asp545, and Asn557, showed at Figure 5b and c.

The quality of the structure prediction of LAC inaCC was analyzed through PDBsum and molprobit. In the Ramachandran plot, the red zone is the most favoured regions, the brown zone is the additional allowed regions, the yellow zone is the generously allowed regions, and the white zone is the disallowed regions. Blue dot is shown amino acid residue. Based on Ramachandran plot from PROCHEK on PDBsum, the structure prediction of

Table 2. The copper-binding loop of LAC inaCC

Organisme	Id	Copper-binding loop 1	Copper-binding loop 2	Copper-binding loop 3
<i>Neurospora crassa</i>	-	²⁷⁴ H ^a WHGMHQ ^R NSNIQDGVNGVTECP	³¹⁴ GTSWY HSH ^a FSAQYGN ^G IVGPI	⁶⁰⁷ HPI Ha L - - -
Zucchini	1AOZ	⁶⁰ HWHG ILQRGTP ^t iWADGTASISQCA	⁹⁹ GTTYFYHGH ^L GMQRSAGLYGSL	⁴⁴⁵ HPWHLHGH
<i>Botrytis aclada</i>	3SOR	¹²⁴ HWHG IRQLG SLEYDGVPGVTQCP	¹⁶³ GTTWYHSHFSLQYGDGLFGPL	⁴⁶³ HP I HLHGH
<i>Botrytis aclada</i>	3V9E	¹²⁴ HWHG IRQLG SLEYDGVPGVTQCP	¹⁶³ GTTWYHSHFSLQYGDGLFGPL	⁴⁶³ HP I HLHGH
<i>Corioloopsis galicca</i>	4A2H	⁶⁴ HWHGFFQHGTNWADGPAFVNQCP	¹⁰⁴ GTFWYHSHLSTQYCDGLRGPI	³⁹⁴ HPFHLHGH
<i>Lentinus.sp</i>	3X1B	⁸⁵ HWHGFFQHGTNWADGGAFVNQCP	¹²⁵ GTFWYHSHLSTQYCDGLRGPF	⁴¹⁷ HPFHLHGH
<i>M. albomyces</i>	3FU7	⁹³ HWHGI HQKDTNLHDGANGVTECP	¹³³ GTSWYHSHFSAQYGN ^G VVGTI	⁴³¹ HPMHLHGH
<i>M. albomyces</i>	1GW0	⁹³ HWHG IHQKDTNLHDGANGVTECP	¹³³ GTSWYHSHFSAQYGN ^G VVGTI	⁴³¹ HPMHLHGH
<i>M. albomyces</i>	2Q90	⁹³ HWHG IHQKDTNLHDGANGVTECP	¹³³ GTSWYHSHFSAQYGN ^G VVGTI	⁴³¹ HPMHLHGH
<i>M. albomyces</i>	3DKH	⁹³ HWHG IHQKDTNLHDGANGVTECP	¹³³ GTSWYHSHFSAQYGN ^G VVGTI	⁴³¹ HPMHLHGH
<i>Aspergillus niger</i>	5LWW	⁸¹ HFHG IRQLYNNQMDGVAALTQCP	¹²⁰ GSSWYHSHFSLQAWEGVFGGI	⁴²¹ HP I HLHGH
<i>Aspergillus niger</i>	5LM8	⁸⁰ HFHG IRQLYNNQMDGVAALTQCP	¹¹⁹ GSSWYHSHFSLQAWEGVFGGI	⁴²⁰ HP I HLHGH
<i>Aspergillus niger</i>	5LWX	¹⁰⁸ HFHG IRQLYNNQMDGVAALTQCP	¹¹⁷ GSSWYHSHFSLQAWEGVFGGI	⁴⁴⁸ HP I HLHGH
<i>S.ochraceum</i>	3T6V	⁶⁵ HWHGGEFQKGTNWADGPAFITQCP	¹⁰⁵ GTYWYHSHLTTQYCDGLRGPF	³⁹⁷ HPFHLHGH
<i>Trametes versicolor</i>	1GYC	⁶⁴ HWHGFFQAGTNWADGPAFVNQCP	¹⁰⁴ GTFWYHSHLSTQYCDGLRGPF	³⁹⁵ HPFHLHGH
<i>Trametes versicolor</i>	1KYA	⁶⁴ HWHGFFQKGTNWADGPAFINQCP	¹⁰⁴ GTFWYHSHLSTQYCDGLRGPF	³⁹⁵ HPFHLHGH
<i>Trametes hirsuta</i>	3FPX	⁶⁴ HWHGFFQHGTNWADGPAFINQCP	¹⁰⁴ GTFWYHSHLSTQYCDGLRGPF	³⁹⁵ HPFHLHGH
<i>Thielavia arenaria</i>	3PPS	¹³⁴ HWHGLRQLGNVFN ^D GANGVTECP	¹⁷⁴ GTSWYHSHFSAQYGN ^G VVGTI	⁴⁷² HPMHLHGH

^aThe copper-binding residue of LAC inaCC

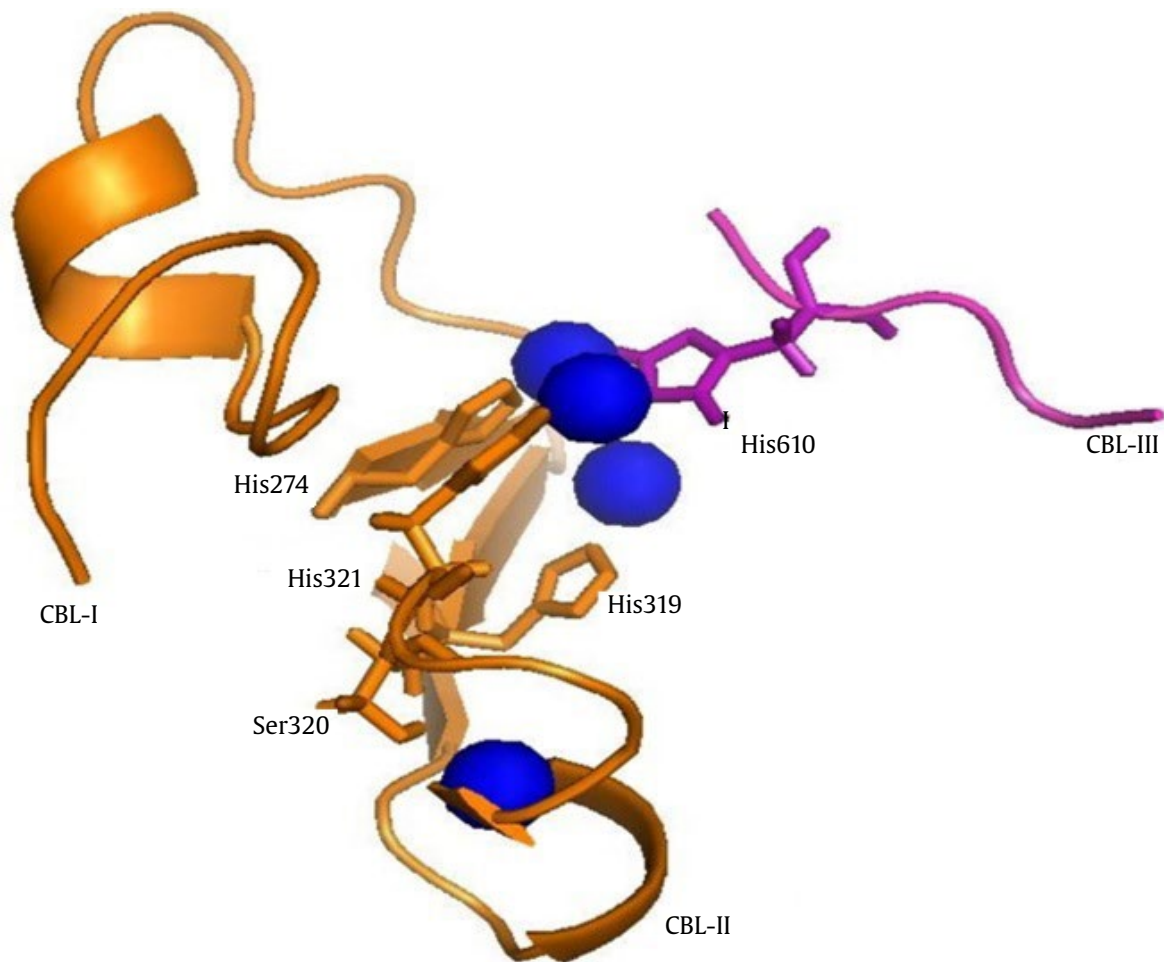


Figure 3. The interaction of the copper atoms with the conserved amino acid residues in LAC inaCC. The copper atoms are shown as a blue sphere, the dashed lines are hydrogen bond, and the colored sticks are the conserved amino acid residues

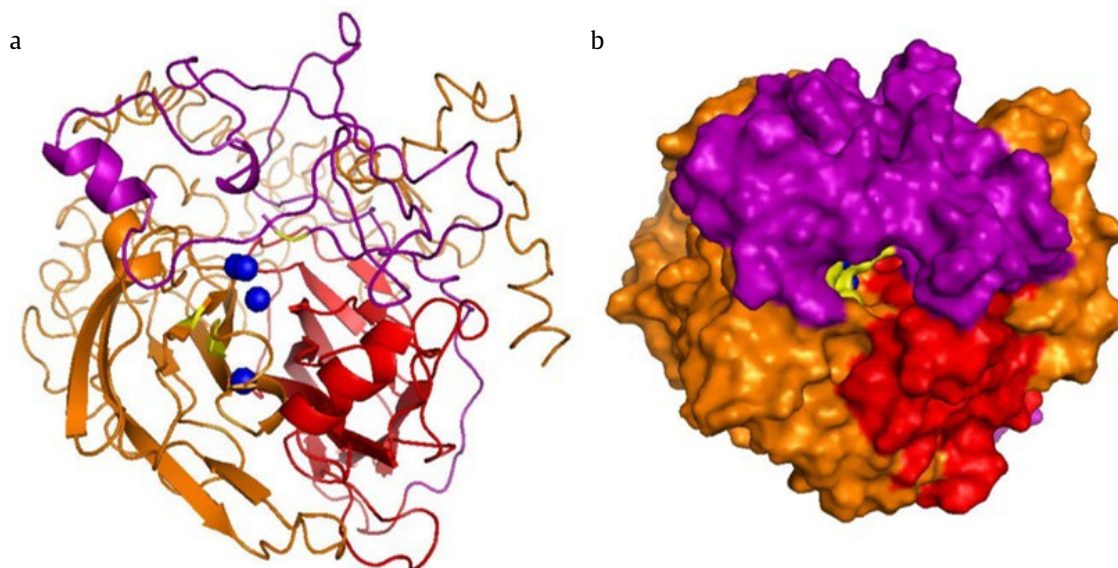


Figure 4. The structure prediction of LAC inaCC by using I-TASSER. (a) The structure prediction of the enzyme in new cartoon form, (b) the structure prediction of the enzyme in surface form. Domain A is colored by red, domain B is colored by orange, and domain C is colored by purple. The copper atoms are shown as a blue sphere

LAC inaCC had a low most favored region percentage of 55.6% with pretty-high disallowed regions percentage of 2.8% (in Table 3), described in Figure 6a. The percentage of the most favored region increased significantly into 78.7% after refinement through the MD simulation method, it was followed by an increase in the value of clashscore and molprobity (99th percentile and 89th percentile).

Other parameters that indicate improvement in the quality of the structure after refinement through the MD simulation method are root mean square deviation (RMSD), root mean square fluctuation (RMSF), interaction energy, and the solvent-accessible surface area (SASA). Based on the RMSD graphic value towards the simulation time in Figure 7a, there was significantly

increasing and fluctuation of RMSD value towards the simulation for 50 ns. The continuous increasing of RMSD value showed a continuous alteration of the structure. The RMSD value was inclined to be stable at the simulation of 80 ns until the end of the simulation (100 ns) with the RMSD value of 3.0 Å. It showed an alteration in structure conformation of LAC inaCC during the simulation towards the conformation stability. The fluctuation of amino acid positional alteration can be explained by using the SASA value. The SASA value showed the area of protein structure surface interaction towards solvents. The highest SASA value was obtained by the back-bone structure of protein as much as 4073 Å².

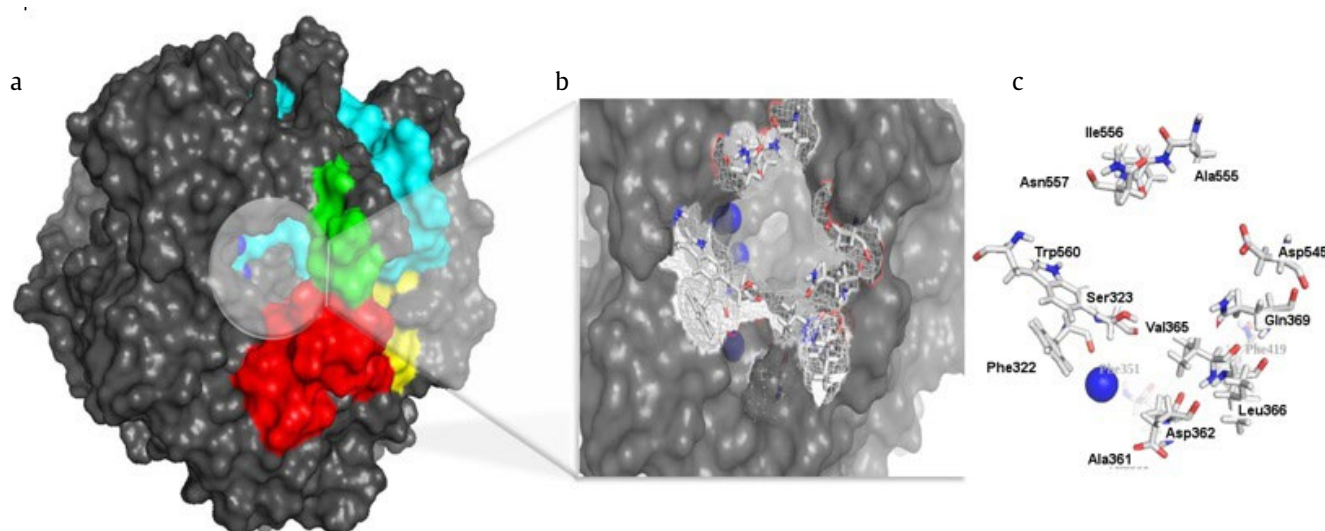


Figure 5. (a) The Substrate pocket and substrate binding loop of LAC inaCC. The Substrate-binding loop I is shown by red, loop II is shown by yellow, loop III is shown by green, and loop IV is shown by cyan, (b) the substrate pocket of the enzyme and their constituent amino acids, (c) the amino acid substrate pocket near the T1 copper site. The copper atoms are shown by a blue sphere, the amino acids are shown by stick form

Table 3. The structure quality parameter of the prediction structure of LAC inaCC before and after refinement by molecular dynamics simulation method (MD)

Structure quality parameter	Before refinement %	After refinement by MD %
Ramachandran Plot Statistics:		
Most favored regions	55.6	78.7
Additional allowed regions	32.7	18.1
Generously allowed regions	8.9	1.4
Disallowed regions	2.8	1.8
Molprobity Summary Statistics:		
Clashscore (all-atom contacts)	33.47 (12 th Percentile) ^a	0.75 (99 th Percentile) ^a
Molprobity score	3.77 (5 th Percentile) ^a	1.70 (89 th Percentile) ^a

^a100th percentile is the best structure of the resulting comparative structure resolution; 0th percentile is the structure with the lowest quality

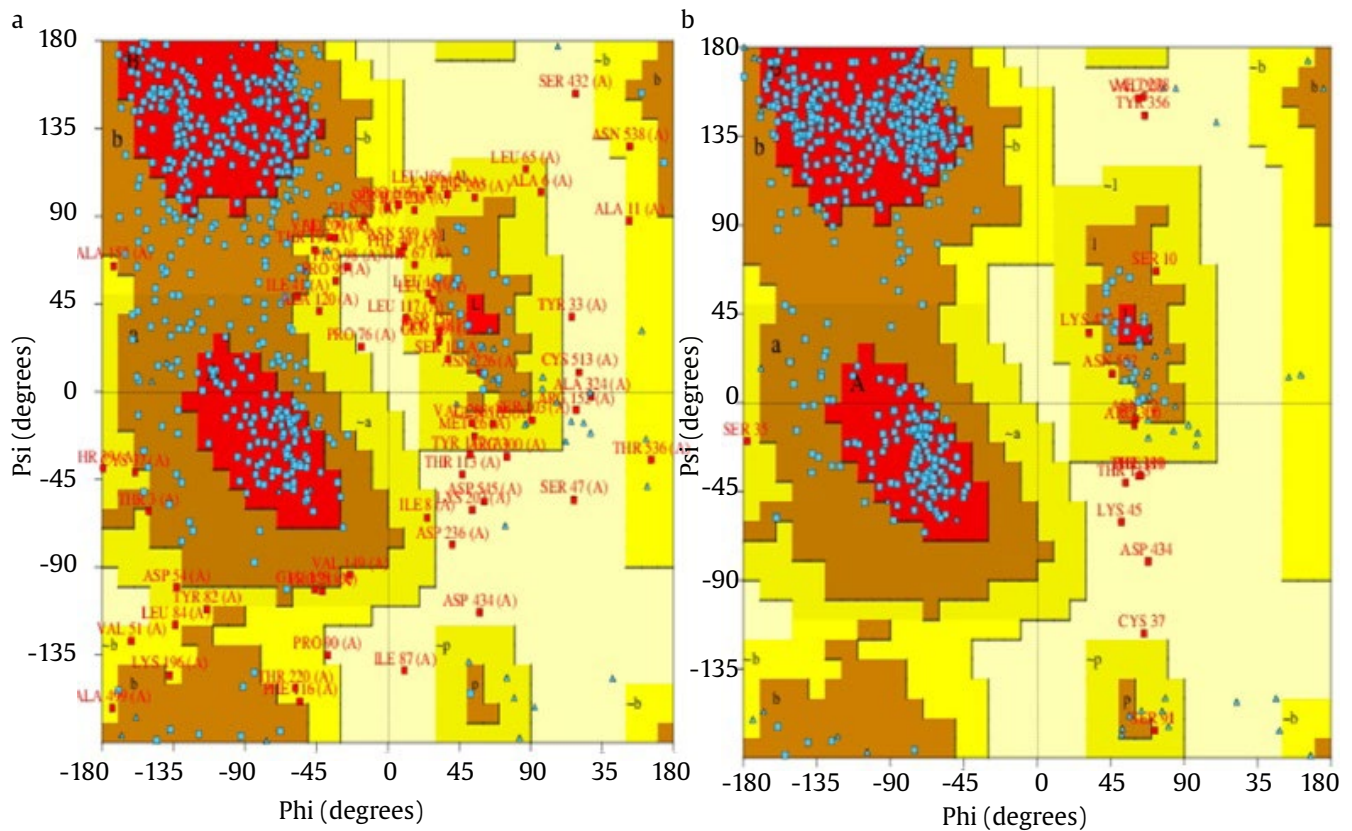


Figure 6. Distribution of torsion-angle of amino acids on the protein backbone of LAC inaCC by PROCHECK. a) The Ramachandran plot of the prediction structure of the enzyme before refinement by MD simulation method, b) the Ramachandran plot of the prediction structure of the enzyme after refinement by MD simulation method

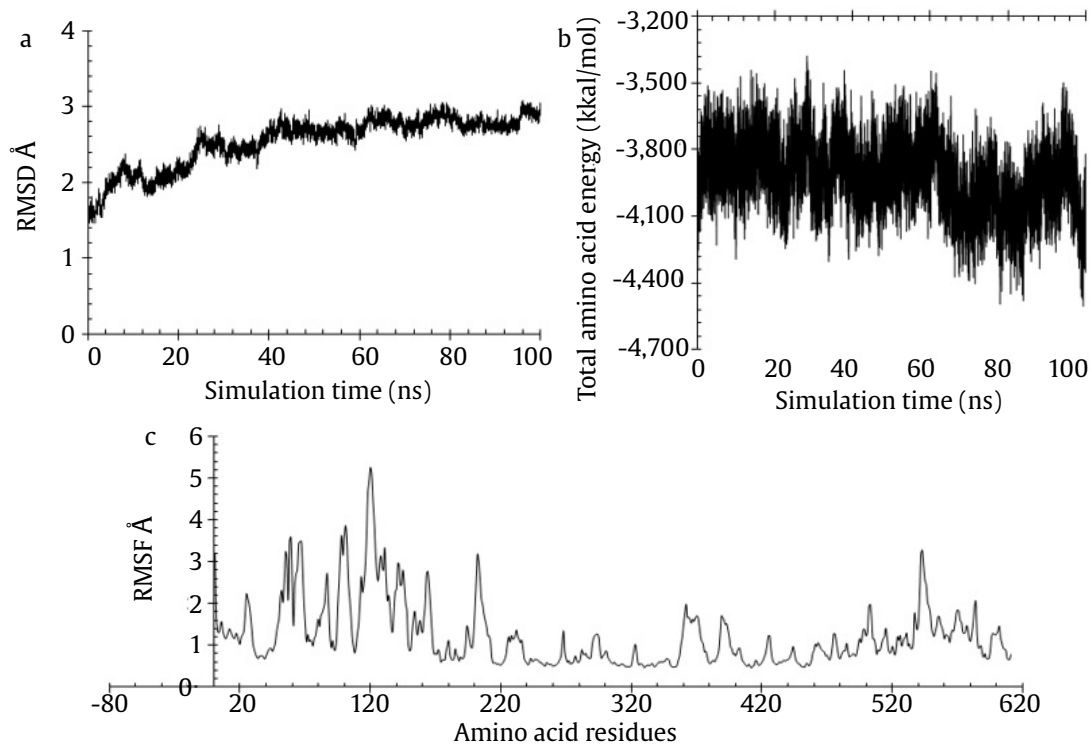


Figure 7. Molecular dynamic parameters of LAC inaCC for 100 ns time simulation. a) RMSD of C α atom, b) interaction energy of amino acids for 100 ns time simulation, c) RMSF of amino acids

4. Discussion

Based on the homology result, the homology of the amino acid sequences of LAC InaCC was not high, only 63% with *Melanocorpus albomyces* and *Thielavia arenaria*. Although the similarity percentage of its amino acid sequence belongs to the low homology category, the amino acid sequence of LAC inaCC can be concluded as homolog, based on sequence alignment in Figure 1. This was because the sequence alignment showed a number of sequences that were identical to the typical amino acid residue pattern called the sequence motif (as illustrated in Figure 1), in accordance with the provisions of Koonen and Galperin (2003). The sequence motif is an area consisting of a number of sequences that are conserved among proteins.

A conserved area means in each protein there is a certain amount of residue in the same typical pattern, which is a characteristic of a group of proteins (Ben-Hur and Brutlag 2006). Sequence motif analysis is necessary to ensure the class of an enzyme (Ben-Hur and Brutlag 2006), especially LAC inaCC that has a low level of protein homology.

The first and the third sequence motifs (HWH and HXXH) are known as the typical motif of laccases, commonly used to differentiate laccase from other enzymes in MCO superfamily (Kumar *et al.* 2003; Awatshi *et al.* 2014). Meanwhile, the second motif (HSH) is a typical motif of every enzyme in MCO superfamily. Based on the sequence alignment at Figure 1, tryptophan as a highly conserved residue was found between two histidines (HWH motif) in laccase sequence from Ascomycetes and Basidiomycetes fungi. Meanwhile, the phenylalanine residue was found in MCO enzyme from Ascomycetes fungi, *Aspergillus niger*. The tryptophan residue differentiates the laccase with other enzymes in the superfamily of multicopper oxidase. Those three sequence motifs found in LAC inaCC had the same sequence motifs with laccase from Ascomycetes fungi, *Xylaria polymorpha*, which plays a role as the binding site of copper atom (Liers *et al.* 2007). The existence of three sequence motifs with the typical pattern of H-X-H in the amino acid sequence of LAC inaCC proved that the isolated genes from *Neurospora crassa* inaCC F226 are genes encoding laccase based on typical characteristic of laccase sequences found (Kumar *et al.* 2003; Sirim *et al.* 2011; Garg *et al.* 2012; Awatshi *et al.* 2014).

On the sequence logo (Figure 2), the sequence motifs were seen. The conserved amino acid would have a high mark of bits in the sequence logo. Based on the results of Cooper-Binding Loop (CBL) analysis (Figure 3), proving that the sequence motifs has an important role in enzyme activity. All sequence motifs of LAC inaCC play a role in the binding of copper atoms (Awatshi *et al.* 2014). In the copper-binding loop, each copper atom was bond to histidine and serine which are sequence motifs of laccase through the hydrogen bond, forming the mononuclear site and the trinuclear site (Kumar *et al.* 2003; Giardina *et al.* 2009; Sirim *et al.* 2011; Awatshi *et al.* 2014), visualized in Figure 3.

On the other hand, LAC inaCC was proven to be an enzyme multisubunit consisting of three functional subunits known as cupredoxin-like domain (domain A, B, and C), in Figure 4. Each domain plays specific roles in the catalytic process of laccase that are interconnected between domains (Trubitsina *et al.* 2015), thus the diminishing of one of the domains would cause laccase to lose its catalytic function. The site between domain A and B are used as a location for substrate oxidation of laccase involving one copper atom (T1 copper), called as the mononuclear site (Giardina *et al.* 2009; Jones and Solomon 2015). Meanwhile, the site between domains A and C provides function as a location to reduce the oxygen involving three copper atoms (T2/T3 copper), called the trinuclear site (Giardina *et al.* 2009; Jones and Solomon 2015), visualized in Figure 4. The prediction structure of LAC inaCC showed typical structure motifs as laccase structure, it is the "Greek key β -barrel" structure motif at domains A and B (Hakulinen *et al.* 2002; Hakulinen *et al.* 2008; Giardina *et al.* 2009). Those characteristics proved that the prediction structure of LAC inaCC by using I-TASSER is the structure of laccase based on typical characteristics found.

The diversity of amino acid hydrophobicity in the substrate pocket (in Figure 5a and b) showed a wide substrate specificity (Ferraroni *et al.* 2012) in LAC inaCC. The wide substrate specificity caused the ability of laccase to oxidize various types of substrate, especially various polycyclic aromatic hydrocarbon (PAH) substances and also non-PAH ones, such as ABTS (Shraddha *et al.* 2011; Chrausia *et al.* 2013; Awatshi *et al.* 2014). The substrate-binding loop is closely related to the active site of the enzyme (Figure 5a). The substrate-binding loop plays a role in

substrate binding and becomes the in-and-out track of the substrate-product. The substrate-binding loop determination referred to the research conducted by Giardina *et al.* (2009), stating that there were four substrate-binding loops in laccase.

Several important roles of the substrate-binding loop are to form some wall on the active site of laccase, to be expected as location for substrate binding during the catalysis process of laccase (Giardina *et al.* 2009), to open a space as location for interaction between substrate and copper atom on the mononuclear site, trinuclear site, and amino acid residues on the active site, the hydrophobic residue composing the substrate-binding loop in LAC inaCC has important structural function to maintain the protein conformation and hydrophobic nuclear stabilization (Pravda *et al.* 2014), the aromatic amino acids (Phe/F, Trp/W, Tyr/Y and Pro/P) and the charged amino acids (Asp/D, Glu/E, Lys/K, Arg/R, His/H) in the substrate-binding loop play roles as "gatekeeper" organizing the in-and-out track of substrate-product between the active site and surface of the enzyme through conformation alteration (Pravda *et al.* 2014; Pavlova *et al.* 2009).

The amino acid residue located on the axial position from the T1 copper (Figure 5c) atom plays an important role to determine the value of reduction-oxidation (redox) potential of laccase (Mate *et al.* 2010; Cambria *et al.* 2012). The existence of Phenylalanine (Phe) as laccase axial-ligand T1 copper, which can be seen in Figure 5c, would increase the redox potential value of laccase (Mate *et al.* 2010; Cambria *et al.* 2012; Pardo and Camarero 2015; Wang *et al.* 2015). Therefore, the LAC inaCC was expected to have higher redox potential value like laccases from other types of fungi (Cambria *et al.* 2012) with redox potential value of 700-800 mV (Pardo and Camarero 2015). A high redox potential value would give an impact on the catalytic activity of LAC inaCC that can be developed in various fields of biotechnology (Fowler *et al.* 2017).

Based on Ramachandran plot result in the quality structure prediction assessment (Figure 6), the structure prediction of LAC inaCC had a low most favored region percentage of 55.6% with high disallowed regions percentage of 2.8%. This showed that the structure conformation stability was low, the forming of inappropriate negative torsion-angle of ϕ and ψ in $C\alpha$ atoms of some amino acids. The appropriate torsion-angles to form the

secondary structure are negative ϕ and ψ as much as -60° and -50° to form the helix, and torsion-angle ϕ and ψ as much as -140° and -130° to form the beta-sheet (Hovmoller *et al.* 2002). The inaccuracy of torsion-angle in some amino acids caused overlap between atoms. High clashcore value (Table 3) indicated high atomic-overlap between amino acids (unfavorable all-atom steric overlaps) which is unfavorable to the quality of the protein structure. High molprobity value (Table 3) showed the resolution of LAC inaCC structure was low. The structure refinement process is required to improve the quality of this enzyme structure.

The increasing of most favored region percentage and the decreasing of sites outside the most favored regions percentage (the additional allowed regions, generously allowed regions and disallowed regions) in laccase structure after refinement through the MD simulation method. The value of the most favored region percentage increased significantly into 78.7% and the disallowed regions percentage decreased to 1.8% (Table 3).

It showed that several amino acids were experiencing the conformation alteration, which can be seen in Figure 6. The clashcore and molprobity values also showed a significant increase, making each of them reached to 99th percentile and 89th percentile. It showed that the molecular dynamics simulation was able to improve the resolution and structure quality of LAC inaCC.

The increasing of RMSD value at the beginning of the simulation (Figure 7a) was supported by high fluctuation of positional alteration of some amino acid residues, in Figure 7b. It was caused by the conformation alteration from earlier structure which took place during the simulation to form several new secondary structures, such as forming four helical structures comprised of ¹⁰¹PSS, ¹²³VAGL, ¹⁵³QDSQ, and ³⁶⁴LVLTT helixes, and also five beta-sheets comprised of ²⁰APQ, ²⁷QQQ, ¹⁹⁹RV, ²⁰⁴SIL, and ²³²GVI beta-sheets. The conformation alteration during the simulation was followed by fluctuation of interaction energy of the amino acid residues obtained from the electrostatic energy calculation, the Van der Waals energy, and the non-binding energy in Figure 7c. The highest SASA value as much as 4073 Å², based on the back-bone structure of protein. It showed that there was a significant conformation alteration in $C\alpha$ atoms in each of amino acid residues during the simulation to form more stable conformation, causing the

increasing value of most favoured region percentage on Ramachandran plot. The SASA value for non-polar residue was 11839 Å², lower than the polar residues as much as 18073 Å². It showed a slightly strong interaction between non-polar amino acid residues, influencing the increasing of conformation stability of LAC inaCC, thus improving the quality of the structure.

5. Conclusion

The gene encoding LAC inaCC has been identified that the gene is the one coding the laccase enzyme based on the sequence motifs of H-X-H (HWH, HSH, HXXH) found. The sequence motifs play an important role in copper atom binding. The structure prediction of the enzyme has appropriate characteristics with laccase structure obtained through the experiment, although its level of similarity was low (65%). This enzyme is a monomer of multisubunit enzyme, containing four copper atoms, and possessing the structural motifs of "Greek-key beta-barrel".

LAC inaCC expected to have broad substrate specificity due to its hydrophobic diversity in substrate-pocket, and high redox potential value. The molecular dynamics simulation was proven to be able to improve the structure quality of LAC inaCC, resulted from the I-TASSER prediction.

Acknowledgements

The authors acknowledge the Directorate General of Higher Education, Ministry of Research Technology and Higher Education, Republic of Indonesia for funding in the research program of Penelitian Dasar Unggulan Perguruan Tinggi (PDUPT), No. 1760/IT3.11/PN/2018.

References

- Awatshi M *et al.* 2014. Molecular docking and dynamics simulation analyses unraveling the differential enzymatic catalysis by plant and fungal laccase with respect to lignin biosynthesis and degradation. *J Biomol Struct Dyn* 33:1835-1849.
- Ben-Hur A, Brutlag D 2006. Sequence motifs: highly predictive features of protein function. *StudFuzz* 207: 625-645.
- Benkert P *et al.* 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinform J* 27:343-350.
- Cambria MT *et al.* 2012. In silico study of structural determinants modulating the redox potential of *Rigidoporus lignosus* and other fungal laccases. *Biomol Struct Dynamics* 30:89-101.
- Chrausia PK *et al.* 2013. A review on mechanism of laccase action. *Res Rev Biosci* 7:66-71.
- Dwivedi UN *et al.* 2011. Structure function relationship among bacterial, fungal, and plant laccase. *J Mol Catal B:Enzym* 68:117-128.
- Feig M. 2016. Local protein structure refinement via molecular dynamics simulations with locPREFMD. *J Chem Inf Model* 56:1304-1312.
- Feig M, Mirjalili V. 2016. Protein structure refinement via molecular-dynamics simulations: what works and what does not?. *Prot* 84:282-292.
- Ferraroni M *et al.* 2012. Reaction intermediates and redox state changes in a blue laccase from *Steccherinum ochraceum* observed by crystallographic high/low X-ray dose experiments. *J Inorg Biochem* 111:203-209.
- Fowler NJ *et al.* 2017. Prediction of reduction potentials of copper proteins with continuum electrostatics and density functional theory. *Chem Eur J* 23:15436-15445.
- Garg N *et al.* 2012. Cloning, sequence analysis, expression of *Cyathus bulleri* laccase in *Pichia pastoris* and characterization of recombinant laccase. *BMC Biotechnol* 12:1-12.
- Giardina P *et al.* 2009. Laccase: a neverending story. *Cell Mol Life Sci* 67:369-385.
- Hakulinen N *et al.* 2002. Crystal structure of a laccase from *Melanocarpus albomyces* with an intact trinuclear copper site. *Nat Struct Biol* 9:601-605.
- Hakulinen N *et al.* 2008. A near atomic resolution structure of a *Melanocarpus albomyces* laccase. *J Struct Biol* 162:29-38.
- Hovmoller S *et al.* 2002. Conformations of amino acids in proteins. *Acta Cryst* 58:768-776.
- Jones SM, Solomon EI. 2015. Electron transfer and reaction mechanism of laccases. *Cell Mol Life Sci* 72:869-883.
- Koonin EV, Galperin MY. 2003. *Sequence-Evolution-Function: Computational Approaches in Comparative Genomics*. Boston: National Institutes of Health.
- Kumar SVS *et al.* 2003. Combined sequence and structure analysis of the fungal laccase family. *Biotechnol Bioeng* 83:386-384.
- Liers C *et al.* 2007. Production, purification and partial enzymatic and molecular characterization of a laccase from the wood-rotting ascomycete *Xylaria polymorpha*. *Enzyme Microb Technol* 41:785-793.
- Mate D *et al.* 2010. Laboratory evolution of high-redox potential laccases. *Chem Biol* 17:1030-1041.
- Pardo I, Camarero S. 2015. Laccase engineering by rational and evolutionary design. *Cell Mol Life Sci* 72:897-910.
- Pavlova M *et al.* 2009. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat Chem Biol* 5:727-733.
- Pravda L *et al.* 2014. Anatomy of enzyme channels. *BMC Bioinformatics* 15:1-8.
- Shen R *et al.* 2015. Structural refinement of proteins by restrained molecular dynamics simulations with non-interacting molecular fragments. *J PLOS Comput Biol* 27:1-19.
- Shraddha RS *et al.* 2011. Laccase: microbial sources, production, purification, and potential biotechnological applications. *Enzyme Res* 2011:1-11.
- Sirim D *et al.* 2011. The laccase engineering database: a classification and analysis system for laccases and related multicopper oxidases. *Database* 2011:1-7.
- Trubitsina LI *et al.* 2015. Structural and functional characterization of two-domain laccase from *Streptomyces viridochromogenes*. *Biochimie* 41:151-159.
- Wang J *et al.* 2015. Lignin engineering through laccase modification: a promising field for energy plant improvement. *Biotechnol Biofuels* 8:1-11.