Primer Design for Amplification of Lignin Peroxidase H8 (lipH8) from Phanerochaete chrysosporium using Ugene Software as an Effort to Accelerate Lignocellulose-based Bioenergy Production Technology

Noor I A[1], *Purwani* N N[2], *Wahab* R A[3], *and Susanti* E[4,4]

[4]Biotechnology Program, Department of Applied Sciences, Faculty of Mathematics and Science, State University of Malang, Malang, Indonesia
[2] Faculty of Vocational , Universitas Airlangga, Surabaya, Indonesia
[3]Departement of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, Malaysia
[1]Chemistry Program, Faculty of Mathematic and Science, State University of Malang, Malang, Indonesia

Abstract. Ligninase enzymes play a crucial role in the delignification stage of lignocellulose-based bioenergy production, especially in bioethanol production, providing an environmentally friendly and cost-efficient process with minimal byproducts. However, the practical application of this strategy is limited by the low expression level of wild-type ligninase from *Phanerochaete chrysosporium*. Heterologous protein expression provides a promising solution to overcome this limitation. Primer design is a key step in recombinant DNA assembly for successful heterologous expression. Using bioinformatic tool, in silico primer design enables pre-laboratory optimization to enhance experimental success rate. This study aimed to design primers for the *lipH8* gene from *Phanerochaete chrysosporium* through sequence homology analysis using the Ugene software. Lignin peroxidase isozyme H8 (lipH8) showed the highest catalytic activity in lignin degradation among the ligninase family. Primer selection was based on amplification size, self-dimer formation, and open reading frame consideration. The optimal primer pair identified was a 20-base pair forward primer (‘5-GCATGGTGGGGTGAAATACG-‘3) and reverse primer (‘5-TGTGAGACGAGTCGGTGATG-‘3), predicted to amplify a 1630 bp fragment of *lipH8*. This primer pair showed no self-dimer formation and exhibited superior open reading frame coverage compared to other candidates, making it suitable for further experimental validation in heterologous expression system.

[4] Corresponding author: evi.susanti.fmipa@um.ac.id

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

1 Introduction

Lignocellulosic biomass, particularly derived from agricultural and forestry waste, represents an abundant and renewable source for bioethanol production [1]. However, its utilization as a bioenergy feedstock remains limited by the presence of lignin, a complex and recalcitrant polymer that inhibits enzymatic access to fermentable sugars. This biomass process includes several stages such as pretreatment, saccharification, and fermentation. Pretreatment is a critical prerequisite to modify the lignocellulosic structure by removing or redistributing lignin, thereby exposing the cellulose and hemicellulose fractions. Following that, saccharification is carried out to transform cellulose and hemicellulose into monosaccharides such as glucose and xylose, finally, fermentation transforms these monosaccharides into ethanol [2]. Researchers have traditionally employed non-enzymatic pretreatment methods such as the use of high pressure and temperature which requires considerable energy, while the use of chemicals such as NaOH and H_2SO_4 causes side reactions and waste that need to be considered, consequently, enzymatic biomass pretreatment is a solution to reduce lignin content in the conversion of lignin biomass into bioethanol [1].

Phanerochaete chrysosporium is classified as white-rot fungus known for its ability to degrade wood indicating presence of certain enzymes that play a role in the process of decomposition of lignin tissue [3]. One of the main enzymes is lignin peroxidase (EC 1.11.1.14) which utilizes oxidation catalytic mechanism to depolymerize lignin structure tissue. Specifically, the isozyme *lipH8* has demonstrated stability under high temperatures and acidic condition[9], make it a highly promising biocatalyst for industrial biomass processing [4]. Despite its potential, the application of native *lipH8* lignin peroxidase enzyme from *Phanerochaete chrysosporium* is hindered by the metabolic complexity of the expression of wild-type crude enzyme. The presence of 17 isozymes biosynthesized by *Phanerochaete chrysosporium* lead to the purification process that is technically challenging and economically unfeasible for implementation at an industrial scale [5]. To overcome these limitations, heterologous protein expression via recombinant DNA technology offers a strategic pathway to produce high-purity, selective isozymes with enhanced catalytic activity compared to wild-type enzyme expression with minimal purification process [6]. A critical initial step in this recombinant workflow is the amplification of the target gene, which necessitates the precise design of primers specific to the lignin peroxidase encoding DNA sequence.

The primer design stage is crucial in biotechnology research for the production of heterologous enzymes. This stage replicates the target DNA template that encodes a protein to express enzymes in another microorganism. Primer design software, as one of bioinformatic tools, has been used over the past decade for primer design. Bioinformatics software enhances research efficiency by providing rapid and convenient data analysis, in addition, the algorithms of bioinformatic software can calculate important parameters in the success of a primer such as, primer length, melting temperature (Tm), Primer %GC and thermodynamics, as a result, failure in laboratory wet practices can be minimized [7]. To achieve amplification of the molecularly crowded *lipH8* gene amidst its homologous isozymes, we implemented a rigorous multi-tier in silico screening pipeline. Moving beyond the limitations of standard standalone or web-based tools, this framework integrates genome-wide specificity mapping with PCR product evaluation to effectively eliminate cross-reactivity within the *Phanerochaete chrysosporium* RP-78 genome and its primer products. By leveraging the whole-genome sequence as a primary reference, this study aims an optimized primer strategy validated through integrated bioinformatics software.

2 Research Method

The research was conducted using computer equipped with an AMD Ryzen™ 5 7530U (6 Cores 12 Threads) with 16GB RAM, no dedicated GPU card was used. The initial stage of the study involved retrieving the whole genome of *Phanerochaete chrysosporium*-RP78, this data was sourced from the NCBI Nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide). The bioinformatic software used was Ugene, equipped a with a local BLAST plug-in to identify the *lip*H8 isozyme coding region, followed by primer design using Primer3 program built-in Ugene software.

3 Result and Discussion

This research utilized whole genome database from *Phanerochaete chrysosporium*-RP78 rather than a single sequence of DNA lignin peroxidase coding protein. The whole genome of *Phanerochaete chrysosporium*-RP78 was retrieved from National Center for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov/nucleotide. The results of whole genome analysis are presented in Table 1.

Table 1 Genome Information *Phanerochaete chrysosporium*-RP78

Genome	*Phanerochaete chrysosporium*-RP78
Genom Size (bp)	29,855,776
Taxonomy Id	273507
Amount of open reading frame	197,346
Amount of *lipH8* coding DNA	389

The National Center for Biotechnology Information (NCBI) database provides a whole genome sequence data from *Phanerochaete chrysosporium*-RP78 spanning 29 million DNA base pairs. As a eukaryotic wood-decay fungus, its genome is characterized by a complex architecture of exons and introns, containing a vast array of protein-coding sequences beyond those dedicated to lignin peroxidase. To identify potential coding regions, an Open Reading Frame (ORF) analysis was performed, screening for sequences initiated by start codons (ATG, TTG, or CTG) and terminated by stop codons (TAA, TAG, or TGA). This analysis identified 197,436 potential ORFs. To specifically isolate the DNA sequences encoding lignin peroxidase within this expansive genomic data, a local BLAST (Basic Local Alignment Search Tool) search was conducted. The processed genome data *Phanero chrysosporium* RP-78 (Table 1) was queried against a specialized database comprising 48 known *lip* isozyme sequences to ensure precise identification of the target gene.

Analysis of the *Phanerochaete chrysosporium* RP-78 whole-genome sequence identified 389 putative *lip* isozyme-encoding DNA regions, these data then futher processed using Microsoft Excel™ to categorize and identify the most prominent *lipH8*. The results confirmed the presence of *lipH8* coding regions within the *Phanerochaete chrysosporium* RP-78 genome, exhibiting 100% sequence identity with the reference database of 48 *lip* isozyme sequences. Subsequently, the specific *lipH8* coding sequence was isolated, and a secondary ORF analysis was performed to precisely delineate the exon-intron region. These sequence data were then utilized as a template in the Primer3 program to design and optimize primers for target gene amplification (Table 2).

Table 2 Analysis result BLAST search against *lipH8 Phanerochaete chrysosporium – RP78*

Isozyme Type	No	Genomic coordinates		Base pair length	%Match	Area open reading frame
		Start	End			
Lignin Peroxidase H8	1	8963484	8963563	80	80/80 (100%)	1
	2	9052961	9053872	912	716/941 (76.09%)	5
	3	9052961	9053872	912	715/940 (76.06%)	5
	4	8995120	8995894	775	663/791 (83.82%)	6
	5	8995120	8995894	775	661/790 (83.67%)	6
	6	9083151	9083960	810	651/822 (79.2%)	6
	7	8962892	8962957	66	64/66 (96.97%)	1
	8	9083214	9083961	748	609/759 (80.24%)	5
	9	8963981	8964037	57	56/57 (98.25%)	1
	10	8963617	8963659	43	43/43 (100%)	0
	11	8963006	8963434	429	417/430 (96.98%)	2
	12	8963887	8963927	41	39/41 (95.12%)	1
	13	8995345	8995773	429	383/430 (89.07%)	4
	14	9070217	9070641	425	377/429 (87.88%)	2
	15	8997587	8998014	428	372/431 (86.31%)	1
	16	8978742	8979166	425	363/427 (85.01%)	2
	17	9005104	9005505	402	349/404 (86.39%)	1
	18	9053936	9054362	427	334/431 (77.49%)	3
	19	21229932	21230239	308	253/314 (80.57%)	3
	20	21229932	21230239	308	249/313 (79.55%)	3
	21	8962526	8964625	2100	2100/2100 (100%)	14
	22	8962632	8964484	1853	1811/1875 (96.59%)	11
	23	21230053	21230239	187	156/188 (82.98%)	2
	24	9006233	9006403	171	140/180 (77.78%)	2
	25	8963712	8963851	140	140/140 (100%)	2
	26	9006233	9006403	171	139/180 (77.22%)	2
	27	8964091	8964221	131	131/131 (100%)	0
	28	8997468	8999025	1558	1250/1601 (78.08%)	9
	29	8977786	8979288	1503	1211/1546 (78.33%)	7
	30	9053337	9053471	135	121/135 (89.63%)	1
	31	9069285	9070764	1480	1200/1522 (78.84%)	9
	32	8964363	8964482	120	120/120 (100%)	2
	33	9069285	9070764	1480	1197/1525 (78.49%)	9
	34	8977786	8979170	1385	1117/1416 (78.88%)	7
	35	8997468	8998807	1340	1096/1370 (80%)	7
	36	9004309	9005653	1345	1067/1378 (77.43%)	9
	37	9004309	9005653	1345	1060/1376 (77.03%)	9

Following appropriate target of *lipH8* coding results, the *lipH8* coding sequence was localized within the genomic coordinate 8962526 – 8964625 (Entry 21, Table 2), encompassing 14 potential ORF regions for DNA amplification. To ensure optimal primer design, parameters

within the Primer3 program were calibrated according to the specific experimental requirements. The optimized parameter values, which were used to generate three primer candidates (Table 4) and their corresponding specifications (Table 5), are summarized in Table 3. Key parameters influencing the fidelity and efficiency of target DNA amplification include primer length, melting temperature (Tm), GC content (%GC), and annealing temperature (Ta). The optimal primer size was restricted to 18–25 base pairs (bp) to ensure high specificity for the target sequence. While shorter sequences within this range are often preferred for their binding kinetics, the selected lengths were balanced to maintain stable annealing at the required temperatures.

Table 3 Primer design Parameter setup

Parameter(s)	Value(s)
Product size range	200 – 2000
Primer size	18 – 25
Primer Tm (°C)	57 – 63
Primer GC%	40 – 60
Max Tm Difference (°C)	2,00
Annealing Temperature (°C)	60
Table of thermodynamic parameter	Santalucia 1998

Table 4 Candidate primer sequence

No	Primer Name	Forward/Reverse	Sequence
1	PCP-78-1	*Forward* ('5 → '3)	'5-CGAGCTTGGACTGGTTGTTG-'3
		Reverse ('5 ← '3)	'5-TTTTGGCAGCGTCCTCATTG-'3
2	PCP-78-2	*Forward* ('5 → '3)	'5-TGAGGGCGAGGAAGATGAAC-'3
		Reverse ('5 ← '3)	'5-TGTGAGACGAGTCGGTGATG-'3
3	PCP-78-3	*Forward* ('5 → '3)	'5-GCATGGTGGGGTGAAATACG-'3
		Reverse ('5 ← '3)	'5-TGTGAGACGAGTCGGTGATG-'3

Table 5 Candidate primer specification

No	Primer Name	Tm (°C)	%GC	*Amplification Size*(bp)	*Self Dimer* (bp)	*Hetero Dimer* (bp)	Ta (°C)	*open reading frame*
1	PCP-78-1	59.41 59.40	55 50	1387	4	0	67.51	9
2	PCP-78-2	59.46 59.48	55 55	1476	0	0	67.55	9
3	PCP-78-3	59.26 59.48	55 55	1630	0	0	67.62	10

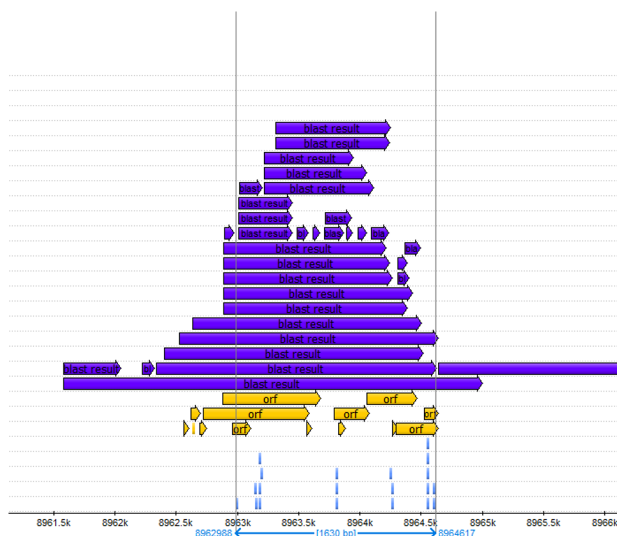

Fig. 1. Amplification sequence target area DNA *lip H8* with primer PCP-78-3

The %GC content in the primer maintained between 40% and 60% to ensure stronger hydrogen bonds, and determines the overall stability of the primer, excessive of %GC content must be avoided, as it can promote primer dimer formation, which in turn inhibits the amplification process [7]. Furthermore, GC clamp has also met the criteria, namely the presence of guanine (G) or Cytosine (C) bases at the end of 5 'or 3' resulting match product amplification sequence (Figure 1). Besides that, the values of ΔGº, ΔHº, ΔSº must be considered because PCR is closely related to the thermodynamics and indicates the quality of the amplification product. In this case, PCR "system" is defined as the PCR apparatus used, such as a PCR test tube sample consisting of nucleic acid strands, solvents, buffers, buffer salts, and also other chemical compounds, whereas the "environment" refers to the thermal conditions generated by the PCR thermocycler. Consequently, the simulation of the reaction process that occurs in PCR can be predicted generally. The ΔGº value can be determined by the following formula:

$$\Delta G_T^o = \frac{\Delta H^o \times 1000 - T \times \Delta S^o}{1000} \quad (1)$$

Primer Tm (melting temperature) is defined as the temperature at which half of the DNA strand is in a double-helical state and the other half is in a random-coil state or also known as spatial denaturation. The determination of the Tm value is determined by the "Wallace-rule" provisions with the following equation:

$$T_m\ (^oC) = 2(A+T) + 4(G+C) \quad (2)$$

While the "Wallace Rule" provides a basic estimation of Tm based on the total count of (A + T) and (G + C) nucleotide residues, it is subject to significant limitations. Specifically, this mathematical model fails to account for critical experimental variables such as DNA template concentration and salt molarity. To achieve greater predictive accuracy, the melting temperature was then determined using Nearest-Neighbor (NN) thermodynamic theory. This approach utilizes

an interaction table (Table 6) to derive specific enthalpy (ΔH°) and entropy (ΔS°) values based on adjacent base-pair sequences [8]. By incorporating these parameters into the following equation, the Tm can be calculated with adjustments for the specific chemical environment of the PCR mixture:

Table 6 ΔH° and ΔS° value from *Neighbor Nearest Theory*

Sequence	ΔH° kcal/mol	ΔS° cal/kmol
AA/TT	-7.9	-22.2
AT/TA	-7.2	-20.4
TA/AT	-7.2	-21.3
CA/GT	-8.5	-22.7
GT/CA	-8.4	-22.4
CT/GA	-7.8	-21.0
GA/CT	-8.2	-22.2
CG/GC	-10.6	-27.2
GC/CG	-9.8	-24.4
GG/CC	-8.0	-19.9
Init. w/term G•C	0.1	-2.8
Init. w/tem A•T	2.3	4.1
Symmetry correction	0	-1.4

$$Tm = \Delta H^o / (\Delta S^o + R \ln Ct) \quad (3)$$

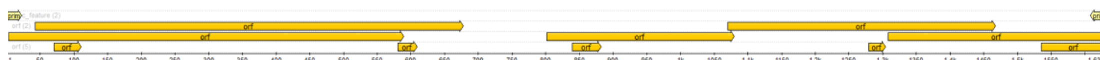

Fig 2. Amplification result of primer PCP-78-3 In silico PCR .

Evaluation of three primers candidates revealed that each met basic criteria of Tm, %GC content, and Ta. The 3rd primer (PCP-RP-78-3) has the highest open reading frame among the other candidates, and successfully simulate in in silico PCR (Figure 2), in contrast, the first primary candidate (PCP-78-1) is not recommended to proceed to the wet lab stage due to formation of self-dimer involving 4 base pairs (Figure 3), this formation can significally diminish amplification results and interfere accurate identification gene amplification product, Consequently, the third primer candidate was selected as the optimal choice, offering the highest probability of successful *lipH8* expression while minimizing potential experimental errors during laboratory trial.

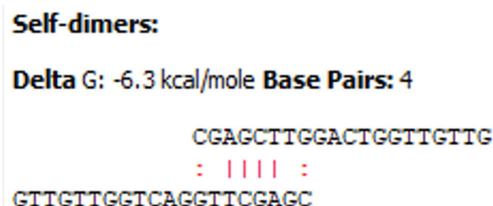

Fig 3. Self-Dimer on primer PCP-78-1

4 Conclusion

In silico analysis obtained the optimum primer for amplification *lip H8* was choosen was PCP-78-3 (Table 5) forward-reverse primer with a size of 20 base pairs, with the nucleotide sequence forward 5'-GCATGGTGGGGTGAAATACG-3' and reverse 5'TGTGAGACGAGTC GGTGATG-3'. This primer have melting point (Tm) was 59,26 and 59,48°C; %GC was 55; predicted to amplify a 1630 bp fragment of *lipH8*; no self-dimer and hetero dimer, Temperature annealing (Ta) was 67,62 °C, with 10 open reading frames for more successful probability in wet laboratory trial.

References

[1] Placido J & Capareda S. 2015. Ligninolytic enzymes: a biotechnological alternative for bioethanol production. Bioresources and Bioprocessing 2:23. doi: 10.1186/s40643-015-0049-5.
[2] Nigam P, Gupta N, Anthwal A. 2009. Pre-treatment of agro-industrial residues. In: Nigam P and Pandey A. Biotechnology for Agro-Industrial Residues Utilisation, Springer Netherlands, p. 13–33.
[3] Dashtban M, Schraft H and Qin W. 2010. Fungal bioconversion of lignocellulosic residues; opportunities & perspective. Int J Biol Sci; 5:578-595. doi: 10.7150/ijbs.5.578.
[4] Falade A.O. Mabinya L.V, Okoh A.I. Nwodo U.U. 2019, Biochemical and molecular characterization of a novel dye-decolourizing peroxidase from *Raoultella ornithinolytica* OKOH-1, Int. J. Biol. Macromol. 121 454–462. doi: 10.1016/j.ijbiomac.2018.10.045.
[5] Passardi F, Theiler G, Zamocky M, Cosio C, Rouhier N, Teixera F, Margis Pinheiro M, Ioannidis V, Penel C, Falquet L. Dunand C. 2007. PeroxiBase: the peroxidase database, Phytochemistry 68 (12) 1605–1611. doi: 10.1016/j.phytochem.2007.04.005.
[6] Sung H J, Khan F M, Kim H Y. 2019. Recombinant lignin peroxidase - catalyzed decolorization of melanin using in-situ generated H_2O_2 for application in whitening cosmetics. International Journal of Biological Macromolecules 136 20-26. doi: https://doi.org/10.1016/j.ijbiomac.2019.06.026.
[7] Basu C. 2022. Methods in Molecular Biology: PCR Primer Design. California State University, Humana Press.
[8] Santalucia J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest neighbor thermodynamics. Proc. Natl. Acad. Sci. Vol. 95, pp. 1460–1465. doi: 10.1073/pnas.95.4.1460.
[9] Pham, L. T. M., Seo, H., Kim, K. J., & Kim, Y. H. 2018. In silico-designed lignin peroxidase from *Phanerochaete chrysosporium* shows enhanced acid stability for depolymerization of lignin. Biotechnology for Biofuels, *11*(1). https://doi.org/10.1186/s13068-018-1324-4.