JURNAL ILMIAH BERKALA SAINS DAN TERAPAN KIMIA

Jurnal Ilmiah Berkala: Sains dan Terapan Kimia
Vol. 16 No. 2, July 2022

Three Dimension Structure Modeling of The Superoxide Dismutase (SOD) of Rice (*Oryza sativa*) Using Fold Recognition Method Using Phyre^2 Web Server

Pemodelan Struktur Tiga Dimensi Enzim Superoxide Dismutase (SOD) Padi (*Oryza Sativa*) dengan Metode Fold Recognition Menggunakan Web Server Phyre^2

Noer Komari[1*], Khairil Anwar[1], Eko Suhartono[2]

[1]Chemistry Department, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarmasin, Indonesia
[2]Biochemistry Division, Faculty of Medicine, Lambung Mangkurat University, Banjarmasin, Indonesia

*Email: nkomari@ulm.ac.id

ABSTRACT

Determining the 3D structure of proteins using laboratory instrumentation is time-consuming and expensive. The in silico method can be used as an alternative to predict the 3D structure of proteins, such as the fold recognition method. This study aims to create a 3D structural model of rice's (Oryza sativa) protein superoxide dismutase (SOD). The 3D structure modeling of the protein was carried out with the Phyre2 web server. The protein sequence was obtained from the UniProt KB database with the code A0A6F8FUX1. The results showed that the suitable template used to build the model was the template with the code c1unfX. The c1unfX template has a coverage value of 80%, 100% confidence, and i.d. of 51%. Validation of the model with the PROCHECK program showed that the most favored area on the Ramachandran Plot was 87.8%, and the disallowed area was 1.1%. The disallowed area, which is still below 15%, indicates that the three-dimensional structure model of the SOD protein built from the c1unfX template has good a value.

Keywords: fold recognition, Phyre^2, 3D protein structure, superoxide dismutase

ABSTRAK

Penentuan struktur 3D protein dilakukan menggunakan instrumentasi laboratorium yang memakan waktu dan biaya yang mahal. Metode in silico dapat digunakan sebagai alternatif untuk memprediksi struktur 3D protein, misalnya metode fold recognition. Penelitian ini bertujuan membuat model struktur 3D protein superoxide dismutase (SOD) padi (Oryza sativa). Pemodelan struktur 3D protein dilakukan dengan web server Phyre^2. Sekuen protein didapat dari database UniProt KB dengan kode A0A6F8FUX1. Hasil penelitian menunjukkan bahwa template yang digunakan untuk membangun model adalah template dengan kode c1unfX. Template c1unfX memiliki nilai coverage 80%, confidence 100% dan i.d. 51%. Hasil validasi model menggunakan program PROCHECK menunjukkan daerah most favored pada Ramachandran Plot sebesar 87,8% dan daerah yang disallowed sebesar 1,1%. Daerah disallowed yang masih dibawah 15% menunjukkan model struktur tiga dimensi protein superoxide dismuatse yang dibangun dari template c1unfX bernilai baik.

Kata Kunci: fold recognition, Phyre^2, struktur 3D protein, superoksida dismutase

Received: December 9, 2021; **Accepted**: June 29, 2022 **Available online**: July 31, 2022

1. INTRODUCTION

The 3D structure of the protein was determined using X-ray diffraction instrumentation, Nuclear Magnetic Resonance (NMR), and electron microscopy (Westbrook et al., 2003). This method is time-consuming, expensive, and cannot be applied to some proteins that are difficult to crystallize (Xu et al., 2016). It causes the protein structures available from experiments to be incomparable or significantly smaller than the number of successfully sequenced protein sequences (Schwede, 2013). As of April 2021, the number of protein 3D structures that have been experimentally obtained and stored in the Protein Data Bank (RSCB PDB) is 176,247, while the number of protein sequences that have been mapped and stored in the UniProt KnowledgeBase (UniProt KB) is 546,227. It implies that there are still many other important and essential proteins whose 3D structures are not yet known (Saudale, 2020).

The 3D structure of proteins can be determined by the in-silico method based on the sequence of amino acids. There are three methods: homology/comparative, fold recognition, and ab initio. The homology method is easier and faster than the fold recognition and ab initio methods (Zaki & Brystoff, 2008; Komari et al., 2020). However, the homology method cannot produce an accurate model if the sequence identity between the target and template is low (sequence identity <30%) (Khor et al., 2015). Fold recognition is based on identifying the same fold fragment (fold recognition) from a protein database with evolutionarily related relationships (Saudale, 2020). Modeling with the fold recognition method can be done using the Phyre^2 web server. Phyre^2 is a web server for 3D protein structure prediction where targets are not only based on evolutionarily close but also remote homologous proteins (Kelley et al., 2015).

This study aimed to determine the 3D structure of rice's superoxide dismutase (SOD) protein. SOD is an antioxidant enzyme that plays a role in the metabolism of neutralizing free radicals in the form of Reactivity Oxygen Species (ROS) (Ashraf & Foolad, 2007). SOD is also often the target of research on rice plants (Widowati et al., 2005; Lee et al., 2013). Research to determine the 3D structure of the SOD enzyme has never been carried out, as evidenced by the unavailability of a 3D structure database of the SOD enzyme in the Protein Data Bank (RCSB PDB). Determination of the 3D structure of the SOD enzyme can be carried out in silico using the Phyre2 web server.

2. MATERIALS AND METHODS

2.1. Materials

The materials used were Rice SOD target protein sequences (*Oryza sativa*). The tools used were web server for protein which is UniProt KB database (https://www.uniprot.org/), the ProtParam web server (https://web.expasy.org/protparam/), and the Phyre^2 web server.

2.2. Target Protein Sequence

Rice SOD target protein sequences (*Oryza sativa*) were obtained from the UniProt KB database (https://www.uniprot.org/). The protein sequences of the search results were evaluated for their 3D structure in the Protein Data Bank (RSCB PDB). Selected protein sequences were saved in FASTA format.

2.3. Target Protein Analysis

Target protein analysis was performed on the ProtParam web server (https://web.expasy.org/protparam/). The analysis was carried out to determine the physical and chemical properties of the target protein.

2.4. Template Identification and Protein Modeling

Template identification and modeling were performed on the Phyre^2 web server (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index). The target protein sequences in FASTA format were submitted to the Phyre^2 web server for template identification and modeling. The resulting template will be selected based on the 100% confidence value and the i.d. value above 20%. The best template will be selected as the protein model.

2.5. Model Evaluation

The model evaluation was carried out on the Phyre^2 web server after the template

identification and protein modeling processes were completed (http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/ca601671cde736c0/summary.html) The model with the highest parameter value was selected and evaluated. The evaluation parameters of the model include the Quality model and the Function model. Quality model parameters were ProQ2 quality assessment, clashes, rotamers, Ramachandran analysis, alignment confidence, and disorder. Function model parameters, namely conservation, pocket detection, and mutational sensitivity.

2.6. Model Validation

Validation of the model used the PROCHECK program on the SAVESv6.0 web server (https://saves.mbi.ucla.edu/). The model file selected on the Phyre2 web server was downloaded in .pdb format and submitted to the SAVESv6.0 web server. Parameters of protein model validation included Ramachandran plot, chi1-chi2 plot, side-chain params, and residue properties.

3. RESULTS AND DISCUSSION

3.1. Selection of Target Protein Sequence

The protein sequence of rice SOD (*Oryza sativa*) on the UniProt KB web server (https://www.uniprot.org/) was searched by entering the keyword "SOD" and by selecting the organism in the other organism column "*Oryza sativa*." A total of eight different types of protein were obtained. Data accessed on April 2021 is depicted in Table 1.

The target protein sequence was selected from a protein that did not yet have a 3D structure, namely protein SOD (Fe-SOD) with UniProt code A0A6F8FUX1. The FASTA format of the SOD protein sequence is presented in Fig. 1.

Sequence data was used to create a 3D protein structure model with the Phyre^2 web server. The protein sequences were copied from the UniProt web server to the Phyre2 web server using the FASTA format, as shown in Fig. 2.

Table 1. SOD Oryza sativa protein on UniProt KB web server

Code	Protein name	Length of amino acid
A0A6F8FQT0	Superoxide dismutase [Cu-Zn]	211
A0A6F8FUX1	Superoxide dismutase	255
A0A6F8F8N3	Superoxide dismutase	231
Q43803	Superoxide dismutase	231
Q01JW6	OSIGBa0147H17.7 protein	316
Q7M238	Superoxide dismutase (Cu-Zn) IV, cy...	57
Q7M240	Superoxide dismutase (Cu-Zn) I, chl...	55
Q7M237	Superoxide dismutase (Cu-Zn) III, c...	48

```
>tr|A0A6F8FUX1|A0A6F8FUX1_ORYSA Superoxide dismutase OS=Oryza sativa OX=4530 GN=Fe-SOD PE=2 SV=1
MAAFASALRVLPSPPAAVPRRLRSREQRQGCRSRRYSKVVAYYALTTPPYKLDALEPYIS
KRTVELHWGKHQQDYVDSLNKQLATSMFYGYTLEELIKEAYNNGNPLPEYNNAAQVWNHH
FFWESMQPEGGGSPGRGVLQQIEKDFGSFTNFREEFIRSALSLLGSGWVWLVLKRKERKF
SVVHTQNAISPLALGDIPLINLDLWEHAYYLDYKDDRRMYVTNFIDHLVSWDTVTLRMMR
AEAFVNLGEPNIPVA
```

Figure 1. SOD Protein Sequences in FASTA format

3.2. Analysis of Protein Target

Protein analysis was carried out to determine the physical and chemical properties of the target protein using the ProtParam web server tool (https://web.expasy.org/protparam/). The physical and chemical properties of the target proteins are presented in Table 2.

Table 2 provides information on the number of amino acids in the target protein, namely 255, molecular weight of 29475.58, and chemical formula of C1338H2041N367O374S7. Its isoelectric point (pI) was 8.84. The pI is the pH value at which the total charge of the positively and negatively charged groups reaches stability (so that the total charge of the amino acid is zero). If the pH is above the isoelectric point, the protein is negatively charged, whereas if it is below the isoelectric point, the protein is positively charged (Thenawidjaja et al., 2017). The estimated half-life of the target protein in mammalian reticulocytes in vitro is 30 hours, yeast in vivo > 20 hours, and Escherichia coli in vivo > 10 hours. The half-life predicts the time required for half of the total protein in the cell to disappear after its synthesis in the cell (Walker, 2002).

The instability index (II) on the target protein was 52.32. The instability index is a protein measure that estimates a protein's stability. A protein whose instability index is less than 40 is predicted to be stable; if it is greater, it is likely to be unstable (Guruprasad et al., 1990). These results indicate that the target protein is unstable. It will affect the process of forming the target protein's 3D structure. The target protein aliphatic index was 82.24. The aliphatic index is the relative volume occupied by amino acids such as alanine, valine, isoleucine, and leucine which have aliphatic side chains in their structure and are considered factors that can increase protein thermostability (Walker, 2002).

GRAVY (grand average of hydropathicity) is a parameter that determines the hydrophobic nature of a protein. The more positive the hydropathic index value of the amino acid, the more hydrophobic the amino acid is. On the other hand, the more negative the hydropathic index value of the amino acid, the more hydrophilic the amino acid is (Nelson & Cox, 2007). The value of the grand average of hydropathicity (GRAVY) of the target protein was -0.413. This value indicates that the target protein is hydrophilic. It dramatically affects an amino acid's folding process in forming its tertiary structure. Generally, proteins rich in hydrophilic amino acid residues are difficult to fold into a good tertiary (3D) structure. The ten highest amino acid compositions of the target proteins are presented in Table 3. The highest amino acid composition is leucine (10.6%), alanine (7.8%), and arginine (7.5%).

Table 2. Physical and chemical properties of the target protein

Physical and chemical properties	Results
Chemical formula	$C_{1338}H_{2041}N_{367}O_{374}S_7$
Number of amino acids	255
Molecular weight	29475.58
Isoelectric point (pI)	8.84
Half-life	30 hours in mammalian reticulocytes (in vitro) >20 hours on yeast (in vivo) >10 hours on Escherichia coli (in vivo)
Instability index (II)	52.32
Aliphatic index	82.24
Grand average of hydropathicity (GRAVY)	-0.413

Table 3. The composition of the ten highest amino acids of rice SOD protein

Amino acid	One letter symbol	Three letter symbols	Amino acid composition	Percentage
Leucine	L	Leu	27	10.6%
Alanine	A	Ala	20	7.8%
Arginine	R	Arg	19	7.5%
Serine	S	Ser	17	6.7%
Valine	V	Val	17	6.7%
Asam glutamate	E	Glu	16	6.3%
Proline	P	Pro	15	5.9%
Glycine	G	Gly	14	5.5%
Tyrosine	Y	Tyr	14	5.5%
Aspargine	N	Asn	13	5.1%

3.3. Template Identification and Protein Modeling

There were 104 templates generated on the Phyre2 web server. The templates showed different confidence values, coverage values, and i.d. values. Templates 1-20 had a protein model, while the rest were not modeled successfully. Information on the 20 templates that have been successfully modeled is presented in Table 3.

The chosen template was the code c1unfX which had a coverage value of 80%, a confidence value of 100%, and an i.d. value of 51%. The model generated by the c1unfX template is visualized in Fig. 2.

The protein model of the SOD enzyme in rice with the c1unfX template comprised of 10 α-helix, 5 β-sheets, and 17 coils. The longest α-helix is shown in number (a1) with 24 amino acids, while the shortest α-helix is shown in (a8) and (a9) with three amino acids each. The longest β-sheet is shown in (b1) and (b2) with eight amino acids each, while the shortest β-sheet has 2 -sheets with one amino acid number each. The longest coil is shown in number (c2) with 23 amino acids, while the shortest has six coils with one amino acid number each.

The c1unfX template information can be found through RCSB PDB (https://www.rcsb.org/) by entering the keyword "1UNF" in the search field. The SOD enzyme in the c1unfX template is a FeSOD enzyme with Fe as a cofactor. This enzyme belongs to the type of oxidoreductase enzyme. The FeSOD enzyme is found in the organism *Vigna unguiculata*, also known as cowpea. The 3D structure of the FeSOD enzyme was discovered using an X-ray diffraction instrument with a resolution of 1.97. The SOD enzyme information on the c1unfX template has been published.

3.4. Model Evaluation

Model evaluation was carried out on the Phyre2 web server (http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/ca601671cde736c0/summary.html). The model with the selected parameters was the c1unfX template. The evaluation of the model was carried out with two parameters: quality and function. The quality model showed a value of 7, namely ProQ2 quality assessment, Clashes, Rotamers, Ramachandran analysis, Alignment confidence, and Disorder. The function model had three parameters: Conservation, Pocket detection, and Mutational sensitivity.

3.5. Quality Model

ProQ2 quality assessment showed a value of 0 to 1 (red to blue), where a value of 0 (red) indicates a good value amino acid. The ProQ2 score model shows that the value closest to 1 is 0.396 for amino acids 38 (K), and the value closest to 0 is -0.050 for amino acids 67 (H), 71 (H), 72 (Q), 122 (F), 167-171 (G, W, V, W, L), 199 (L), 202 (L), 204 (L), 205 (W), 230 (S), and 231 (W). Fig. 3 shows that the model is dominated by red, orange, and green colors in the -helix, -sheet, and coil regions, so the model has good value (Ray et al., 2012).

Table 3. Information on 20 successfully modeled templates

No.	Template code	Coverage value (%)	Confidence value (%)	i.d. value (%)
1.	c6bejA	77%	100%	40%
2.	c1xreB	78%	100%	40%
3.	c1gv3B	77%	100%	36%
4.	c1my6A	77%	100%	43%
5.	c3js4C	78%	100%	43%
6.	c6jovC	76%	100%	43%
7.	c4yioB	78%	100%	38%
8.	**c1unfX**	**80%**	**100%**	**51%**
9.	c1mngA	78%	100%	32%
10.	c1gn4B	79%	100%	32%
11.	c3ceiA	78%	100%	33%
12.	c5a9gB	80%	100%	32%
13.	c1y67D	78%	100%	41%
14.	c1ma1E	78%	100%	35%
15.	c1dt0A	79%	100%	43%
16.	c4f2nL	78%	100%	34%
17.	c2rcvA	77%	100%	39%
18.	c5n57B	78%	100%	36%
19.	c4h3eB	78%	100%	32%
20.	c4ffkA	80%	100%	33%

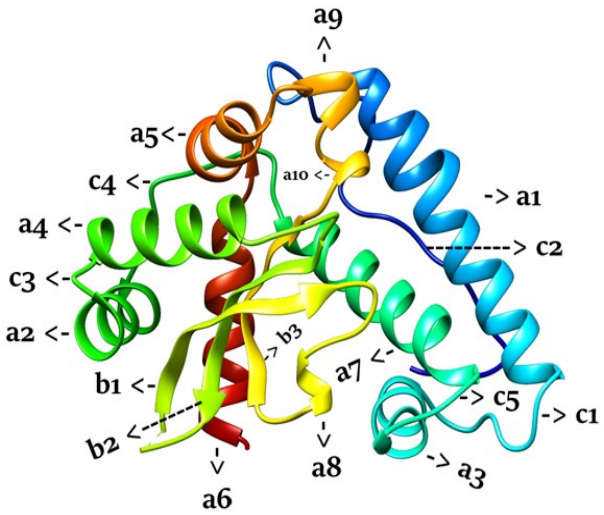

Figure 2. Model visualization with c1unfX template

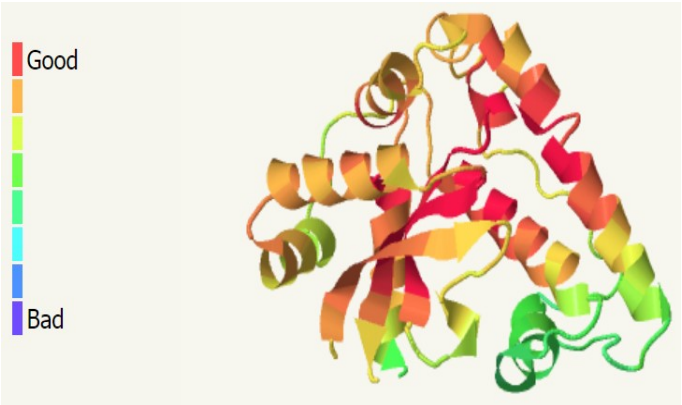

Figure 3. Model based on ProQ2 quality assessment

Clashes show residual clashes with low to high values (blue to red). The clashes model color indicator shows that there are seven amino acids whose color is close to red, namely amino acids 38 (K), 40 (V), 72 (Q), 83 (L), 160 (A), 169 (V) and 204 (L). The color towards purple indicates a residue that has good clashes value (Good). Fig. 4. shows a good value model because it is dominated by purple, light blue, and green in the α-helix, β-sheet, and coil regions.

Rotamers show the number of side chains that are not ideally modeled. From 38-244 residues, only one residue, 120 (H), is problematic. Fig. 5. shows the indicator in red where the residue is not ideally modeled (Bad). The blue indicator shows the ideally modeled residue area (Good). The model is of good value because it is dominated by blue in the α-helix, β-sheet, and coil regions.

The Ramachandran analysis in Fig. 6 shows the residues in the favored (favorable, blue), allowed (allowed, green), and not allowed (disallowed, red) areas. There are 39 allowed residues, namely (V), 49 (P), 70 (K), 85 (T), 103 (N), 136 (R), 187 (N), 194 (L), 195 (G) and 214 (K). In addition, there were 41 disallowed residues, namely (A), 88 (F), 105 (N), and 196 (D). The model is considered a good value because it is dominated by blue in the α-helix region, while green and red are only in some parts of the coil.

Alignment confidence shows the results of the alignment between the template and the target protein sequence, with a value of 0 to 1 (good to poor). Alignment is not good at residues 84 (A) and 85 (T), with a value of 0.54. The color indicator towards red indicates a good alignment confidence value (Good), and the color toward blue is not good (Bad). Fig. 7 shows that the model is dominated by red, yellow, and green colors on the residues in the α-helix and β-sheet regions. Therefore the model is stated to be of good value.

Prediction of disorder is vital to knowing residues that hinder the success of the crystallization of a protein. Values close to one indicate disorder, and values close to 0 indicate order. The disorder areas are found in residues 1-42, 128-135, and 249-255. The color indicator toward red is the residual disorder, while the indicator towards the blue is the order residue. Fig. 8 shows a good value model because the α-helix and β -sheet regions are dominated by blue, light blue, and green, while the areas on the coil are dominated by green and light blue.

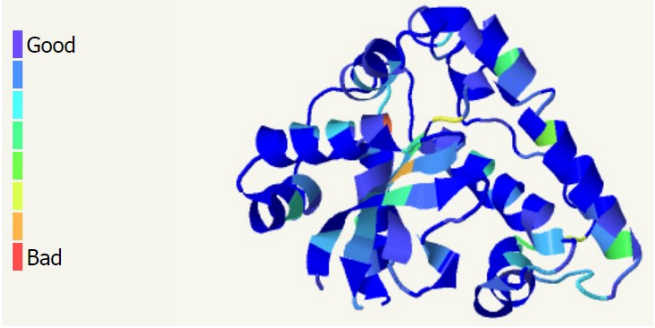

Figure 4. Model based on clashes

Figure 5. Model based on rotamers

Figure 6. Model based on Ramachandran analysis

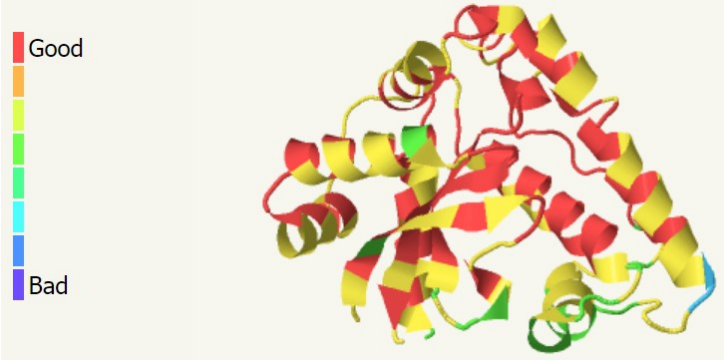

Figure 7. Model based on alignment confidence

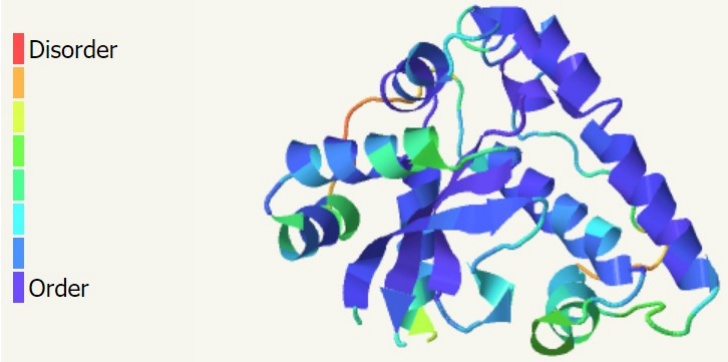

Figure 8. Model based on disorder

3.6. Model Function

Conservation provides information on the presence of functional residues. The color indicator towards red indicates a residue with higher conservation (High). The indicator color towards blue has a lower conservation value (Low). Fig. 9. shows a model dominated by purple, light blue, and green residues in the -helix, -sheet, and coil regions. The red color is found in the residue in one of the -sheets, namely residue 168 (W), and the orange color in one of the α-helix, β-sheet, and coil, namely residues 119 (H), 167 (G), 206 (E), 207 (H) and 209 (Y).

Pocket detection is used to predict amino acids at the active site. The largest pocket is an active site area. The largest pockets are marked in red, as shown in Fig. 10. The residues at the active site are 141 (Q), 144 (K), 145 (D), 173 (L), 175 (R), 176 (K), 178 (R) , 179 (K), 180 (F), 199 (L), 238 (M), 241 (A), and 242 (E).

Mutational sensitivity predicts the occurrence of missense mutations of specific amino acids in the target protein. The

mutation sensitivity score is from 0 to 9, where 0 indicates a low residual mutation value. The color indicator towards red indicates the mutation sensitivity value is getting higher (High). The color indicator towards blue has a lower mutation sensitivity value (Low). Fig. 11 shows the highest mutation value at residue 198 (P), which is indicated by the orange color on the β-sheet. The yellow color indicates the second-highest mutation value in several α-helix, β-sheet, and coils. The model is still dominated by purple, light blue, and green residues with low mutation values.

3.7. Model Validation

Model validation was carried out using the PROCHECK program on the SAVESv6.0 web server (https://saves.mbi.ucla.edu/). The validation results are presented in Table 4 and Fig. 12.

The quality of the protein structure depends on the number of non-glycine residues in the outlier/disallowed regions. The protein structure quality is of good value if the non-glycine residue in the outlier area is <15%, and the smaller it is, the better the protein structure quality (Lovell et al., 2003). Table 4 shows the number of amino acid residues in quadrant I most favored regions (87.8%), quadrant II additional allowed regions (8.9%), quadrant III generously allowed regions (2.2%), and quadrant IV disallowed regions (1.1%). It indicates that the structure of the model has good quality because it has non-glycine residues in the disallowed regions, which are below 15%, namely 1.1%. The quality and stability of the model are indicated by the presence of amino acid residues in the most favored regions, which are larger than the amino acid residues in the disallowed regions (Ho & Brasseur, 2005). Ramachandran plot is an indicator of the intrinsic quality of 3D structure (Petsko & Ringe, 2004).

Table 4 shows that in the All Ramachandran plot, there are ten labeled residues from 201 residues. Labeled is indicated by a red dot with a number above it. The red dot with the number indicates that the residue is in the unfavorable area of the graph. Parentheses containing a number next to the name of the amino acid indicate the number of amino acids present on the graph. The All Ramachandran plot in Fig. 12 shows which residues were labeled out of the ten labeled residues. The ten residues are alanine, aspartic acid, glutamic acid, leucine, lysine, phenylalanine, proline, threonine, and two amino acids valine.

Table 4 shows the results of chi1-chi2 plots that of the 137 residues, there are two labeled residues: histidine and tyrosine. In the side-chain parameters plot, there is one graph with an Inside description, four with BETTER description, and 0 indicating worse. Residual properties plots show how the geometric properties of the protein vary along the sequence of the target protein. This plot provides a visualization of which areas have poor and more normal geometry. Table 6 shows the results that there is a max. deviation of 10.6.

Table 4 shows the Overall G-factor value in the protein model, which is 0.01. The G-factor is a value that measures the stereochemistry of a protein model. A low G-factor value indicates that the protein model has a low conformational probability. The ideal G-factor value is above -0.5. A G-factor value below -1.0 is a non-ideal value (Ahmed et al., 2009). Based on Table 6, the overall G-factor value in the resulting protein model is ideal.

The RMS (Root Mean Square) distance is the distance from the planarity for the various planar groups in the structure. The dotted lines on the graph show different ideal values for aromatic rings (Phe, Tyr, Trp, His) and planar groups (Arg, Asn, Asp, Gln, Glu). The default values are 0.03Å and 0.02Å, respectively. Histogram bars crossing the dotted line are shown as highlighted. Table 4 shows that 0.0% highlighted indicates nothing is highlighted, and 100% within limits shows histogram bars do not cross the dotted line or are still within limits.

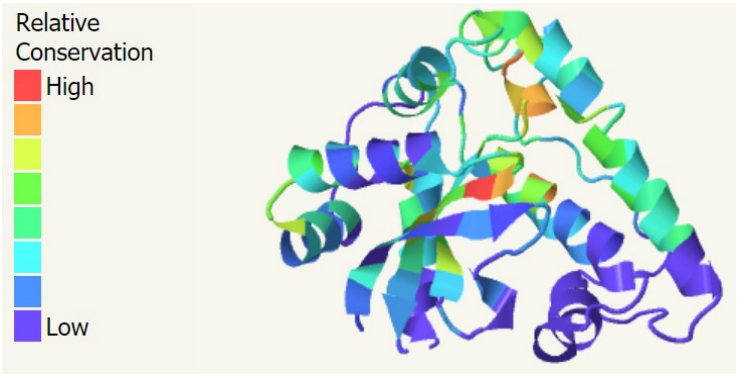

Figure 9. Model based on conservation

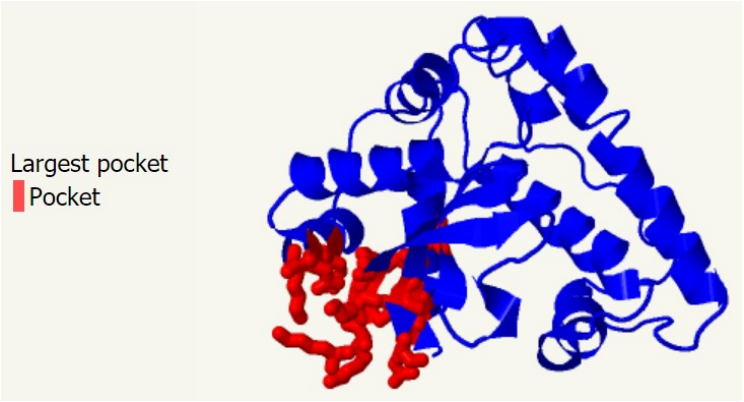

Figure 10. Model based on Pocket detection

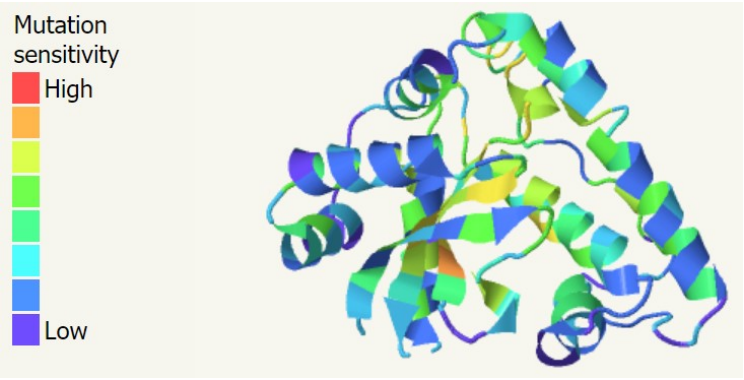

Figure 11. Model based on mutational sensitivity

Table 4. Model validation results using PROCHECK

Parameter	Results
Ramchandran plot	The residue is in the most favored area of 87.8% The residue is in an additional allowed area of 8.9% The residue is in the area that is generously allowed by 2.2% The residue is in the disallowed area of 1.1%
All Ramachandrans	10 labelled residues (out of 201 residues)
Chi1-chi2 plots	2 labelled residues (out of 137 residues)
Side-chain params	4 better, 1 inside, 0 worse
Residue properties	Max. deviation: 10.6
G-factor	Overall: 0.01
Planar groups	100% within limits, 0% highlighted

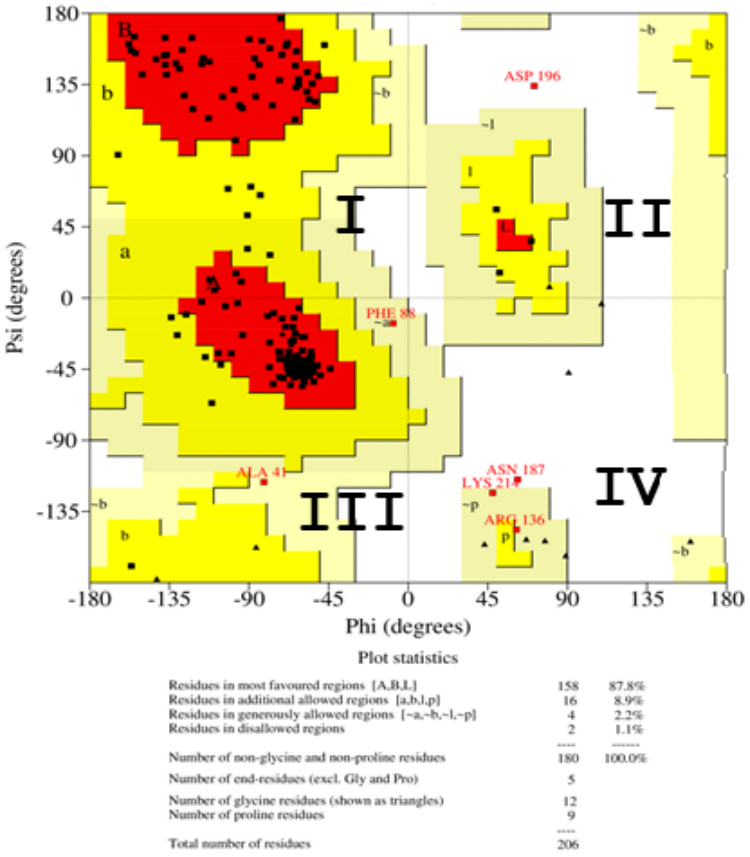

Figure 12. Ramachandran plot of PROCHECK model validation results

4. CONCLUSION

Modeling the 3D structure of the SOD enzyme using the Phyre2 web server produces a model based on the c1unfX template with a coverage value of 80%, a confidence value of 100%, and an i.d. value of 51%. The model's evaluation results include good values based on quality with ProQ2 quality assessment, clashes, rotamers, Ramachandran analysis, alignment confidence, and disorder dominated by good indicators (Good). Evaluation of the model based on function shows that each residue's conservation value and mutational sensitivity value are relatively low. The model validation results on the Ramachandran plot show that the residue in the most favored area was 87.8%, the additional allowed area was 8.9%, the generously allowed area was 2.2%, and the disallowed area was 1.1%. The resulting model is of good value because the residue in the disallowed area is still below 15% and smaller than the most favored area of 87.8%.

ACKNOWLEDGEMENT

Acknowledgments are conveyed to Mr. Sunardi, the head of the chemistry study program, FMIPA ULM, who has facilitated the use of the bioinformatics laboratory for this research.

LIST OF REFERENCES

Ahmed, A., Koller, A., & Gohlke, H. (2009). Multi-scale modelling of macromolecular conformational changes. In *Chemistry Central Journal,* 3(1). https://doi.org/10.1186/1752-153X-3-S1-P53

Ashraf, M., & Foolad, M. R. (2007). Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environmental and Experimental Botany*, *59*(2), 206–216. https://doi.org/10.1016/j.envexpbot.2005.12.006

Guruprasad, K., Reddy, B. V. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*. https://doi.org/10.1093/protein/4.2.155

Haryanto, T., & Budiman, B. (2015). Penggunaan Fitur Kimia Fisik dan Posisi Atom untuk Prediksi Struktur Sekunder Protein. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *1*(2). https://doi.org/10.26418/jp.v1i2.11919

Ho, B. K., & Brasseur, R. (2005). The Ramachandran plots of glycine and pre-proline. *BMC Structural Biology*, *5*, 1–11. https://doi.org/10.1186/1472-6807-5-14

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Khor, B. Y., Tye, G. J., Lim, T. S., & Choong, Y. S. (2015). General overview on structure prediction of twilight-zone proteins. *Theoretical Biology and Medical Modelling*, *12*(1). https://doi.org/10.1186/s12976-015-0014-1

Komari, N., Hadi, S., & Suhartono, E. (2020). *Pemodelan Protein dengan Homology Modeling menggunakan SWISS-MODEL Protein Modeling with Homology Modeling using SWISS-MODEL*, *2*(2), 65–70.

Lee, M. H., Cho, E. J., Wi, S. G., Bae, H., Kim, J. E., Cho, J. Y., Lee, S., Kim, J. H., & Chung, B. Y. (2013). Divergences in morphological changes and antioxidant responses in salt-tolerant and salt-sensitive rice seedlings after salt stress. *Plant Physiology and Biochemistry*, *70*. https://doi.org/10.1016/j.plaphy.2013.05.047

Lovell, S. C., Davis, I. W., Iii, W. B. A., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. *Proteins*, *50*(3).

Petsko, G. a, & Ringe, D. (2004). From Sequence to Function : Case Studies in Structural and Functional Genomics. In *Protein Structure and Function*.

Ray, A., Lindahl, E., & Wallner, B. (2012). Improved model quality assessment using ProQ2. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-13-224

Rosmawati, T. (2013). Lama perebusan terhadap kandungan protein pada kerang darah (Anadara granosa). *Jurnal Biology Science & Education*, *2*(2), 103–109.

Saudale, F. Z. (2020). Biokimia di Era Big Data Genomik: Tantangan, Aplikasi dan Peluang Inovasi. *Chem. Notes*, *1*(2), 21–43.

Schwede, T. (2013). Protein modeling: What happened to the "protein structure gap"?. In *Structure*. https://doi.org/10.1016/j.str.2013.08.007

Walker, J. M. (2002). Protein Protocols Handbook, The. In *Protein Protocols Handbook, The*. https://doi.org/10.1385/1592591698

Westbrook, J., Feng, Z., Chen, L., Yang, H., & Berman, H. M. (2003). The Protein Data Bank and structural genomics. *Nucleic Acids Research*, *31*(1), 489–491. https://doi.org/10.1093/nar/gkg068

Widowati, W., Safitri, R., Rumumpuk, R., & Siahaan, M. (2005). Penapisan Aktivitas Superoksida Dismutase pada Berbagai Tanaman. *JKM*, *5*(1).