

### 3D STRUCTURE AND FUNCTION ANALYSIS OF RECOMBINANT ALDII PROTEIN FROM *Uncultured Acidilobus sp.* USING I-TASSER

Nishia Waya Meray<sup>1\*</sup>, Suharti<sup>2</sup>, Akhmaloka<sup>3</sup>

<sup>1</sup>Department of Pharmacy, Faculty of Humaniora and Health, Universitas Mulia, Jl. Letjend ZA Maulani No.9, Damai Bahagia, South Balikpapan, East Kalimantan, Indonesia

<sup>2</sup>Department of Chemistry, Faculty of Science and Computer Science, Pertamina University, Jakarta, Indonesia

<sup>3</sup> Biochemistry Research Team, Faculty of Mathematics and Natural Sciences, Bandung Institute of Technology, Bandung, Indonesia

\*Email: nishia@universitasmulia.ac.id

Received 20 March 2024

Accepted 17 May 2024

#### Abstract

The AldII protein is a new recombinant protein produced from a novel gene obtained via a metagenome approach. Previous studies showed that this protein is strong and has the same metal binding aspect as the class II Aldolase enzyme. Aldolase is a valuable enzyme used in pharmaceuticals, food processing, and biochemistry. Further investigation is required to comprehend the structure and function of the AldII protein due to its potential. Researchers will conduct sophisticated bioinformatic analysis on the 3D shape and function prediction of AldII using the I-TASSER webserver from Zhanlab. The I-TASSER server is an online tool for the automated prediction of protein structure and annotation of functions based on structure. Analysis of the AldII protein using the I-TASSER webserver shows that this protein has a stable structure with the closest structural homology to deoxyribose-phosphate aldolase from *Bacillus thuringiensis* with PDB code 6btdA. Additionally, the biological structure analysis shows that this protein shares the biological function of the enzyme L-fuculose-1-phosphate aldolase, which is part of the class II Aldolase enzyme that plays a role in the catabolism of arabinose, L-Fuculose, and Rhamnose. The results align with prior research that states the AldII protein is a stable protein with a catalytic side that is homologous to the class II Aldolase enzyme.

**Keywords:** *Acidilobus sp.*, AldII; aldolase class II, I-TASSER

#### Introduction

The AldII protein is a recombinant thermostable class II Aldolase protein derived from *Archaea* through the production of a novel gene using a metagenomic method. The previous studies of the gene which encodes the AldII protein show that the phylogenetic tree analysis indicates a close evolutionary relationship between the AldII sequence and the ribulose-5-phosphate-4-epimerase enzyme found in uncultured *Acidilobus sp.* The phylogenetic tree analysis was determined by the Jones-Taylor-Thornton (JTT) model with Mega516. Homology

examination of the AldII amino acid sequence performed by the ClustalX program reveals similarities with class II Aldolase in the conserved region, active site, polypeptide bond, and Zn<sup>2+</sup> binding site (Suharti *et al.*, 2015). AldII protein has a molecular weight of around 21.2 kilodaltons and a molecular formula C<sub>940</sub>H<sub>1539</sub>N<sub>261</sub>O<sub>281</sub>S<sub>8</sub>. The ProtParam webserver on the ExPASy site indicates 22 negatively charged residues (Asp + Glu) and 18 positively charged residues (Arg + Lys) based on the biochemical characteristics analysis. AldII has a theoretical isoelectric point (pI) of 5.86 and a stability index of 36.61,



categorizing it as a stable protein. The stability index is determined by the length of the amino acid sequence and dipeptide composition by the algorithm available on this web server. A protein is considered stable if its Stability Index value is less than 40 (Meray *et al.*, 2021).

Aldolase is a vital enzyme that plays a crucial role in various industries and applications. It has garnered significant interest due to its versatility across multiple sectors. In the pharmaceutical and biotechnology fields, Aldolase serves as a biocatalyst, enhancing the efficient synthesis of complex compounds. Through directed evolution techniques, scientists have successfully engineered Aldolases, improving their usefulness in the development of medicinal drugs and renewable fuels. Extensive research has also been conducted on the catalytic mechanisms of Aldolase, contributing to advancements in medical treatments (Rocha *et al.*, 2022; Windle *et al.*, 2014). In conclusion, Aldolase is a significant enzyme that has broad uses in pharmaceuticals, food production, and biochemistry, contributing to developments and breakthroughs in numerous sectors.

Further investigation of the structure and function of the recombinant AldII protein from uncultured *Acidilobus sp.* is essential due to its potential. AldII protein is a protein produced from a novel gene, so info on structure and function is done through in silico analysis using the I-TASSER webserver. Researchers will conduct further bioinformatics analysis on the 3D structure and function prediction of AldII using the I-TASSER webserver from Zhanglab. The I-TASSER server is an online application that automates the prediction of protein structure and function annotation using structural information. The objective is to provide precise protein structure and function predictions utilizing sophisticated algorithms that are continuously enhancing. This program uses a

hierarchical methodology that starts with the amino acid sequence to produce structure predictions and functional annotations. I-TASSER has recently integrated advanced methods for refining atomic-level structures, estimating neighbouring structures accurately, and predicting biological functions. The I-TASSER server is an essential tool for the clinical community, utilizing advanced algorithms and ongoing enhancements to forecast protein systems and their characteristics (Roy *et al.*, 2010; Yang *et al.*, 2014; Yang and Zhang, 2015; Zhang, 2008).

## Research Methods

### Materials

The research conducted in this study utilized the gene sequence of uncultured *Acidilobus sp.* which had been successfully sequenced in previous research. The aldII gene sequence was already available in the NCBI database under the identity number KP893071 (Suharti *et al.*, 2015).

### Translation of the aldII gene sequence

The *aldII* gene was translated in silico using the BLASTp program provided by NCBI on the website <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. The resulting AldII amino acid sequence was then stored in FASTA format for later analysis.

### Prediction of 3D structure and function of AldII protein

The prediction of the 3D structure and function of AldII protein was carried out using the I-TASSER webserver provided on the website <https://seq2fun.dcmf.med.umich.edu/I-TASSER/>. The parameters analyzed on this web server were 3D structure analysis, homology analysis of protein type proximity based on 3D structure, biological function analysis, and homology analysis of protein type proximity based on biological function.

Visualization of the predicted structure was done using the Pymol program. Biological function analysis was also performed using Expasy ENZYME and Gene Ontolog provided by I-TASSER (Zheng *et al.*, 2021).

## Results and Discussion

### 3D structure analysis of AldII

The AldII 3D structure prediction was conducted in a prior study utilizing the *ab initio* method in the QUARK program, and it is available at the website <http://zhanglab.ccmb.med.umich.edu/QUARK>. This tool can forecast the 3D configuration of proteins independently of pre-existing protein structures. The AldII protein deduction predicts a 3D structure with 6  $\beta$ -sheet structures and 6  $\alpha$ -helix structures (Meray *et al.*, 2021). The I-TASSER modeling process starts by choosing the structural data available in

the PDB library using the LOMETS mechanism, which incorporates many threading applications. I-TASSER selects the most important templates from different template alignments produced by each threading software using their Z-score, which measures the discrepancy between raw and average scores in standard deviation units. The top 10 templates with the highest Z-score from the LOMETS threading programs are chosen for this round. The template with the greatest Z-score is often selected, and the programs are then graded based on their performance in comprehensive benchmark tests. A Normalized Z-score greater than 1 indicates a good alignment, whereas a score less than 1 indicates a poor alignment. The greater the elevation, the more desirable. This can be utilized to assess if the protein is a simple or a challenging target.

**Table 1.** Z-score from each threading program

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score
1	6btdA	0.28	0.28	0.97	3.80
2	1dzuA	0.27	0.30	0.97	3.64
3	7x78A	0.28	0.29	0.95	4.53
4	6btd	0.28	0.28	0.97	2.88
5	4c24	0.27	0.28	0.98	2.14
6	6btdA	0.28	0.28	0.97	3.90
7	4c24	0.27	0.28	0.96	3.22
8	6btdA	0.29	0.28	0.97	3.80
9	6btdA	0.28	0.28	0.97	3.96
10	7x78A	0.28	0.29	0.95	2.98

Note:

The ranking of templates shows the ten most commonly used threading templates in I-TASSER. Iden1 represents the percentage of sequence identity between the templates and the query sequence. Iden2 represents the percentage of sequence identity between the template chains and the query sequence.

Cov represents the coverage of the threading alignment and is determined by dividing the number of aligned residues by the length of the query protein.

Norm. The Z-score represents the standardized value of the threading alignments.

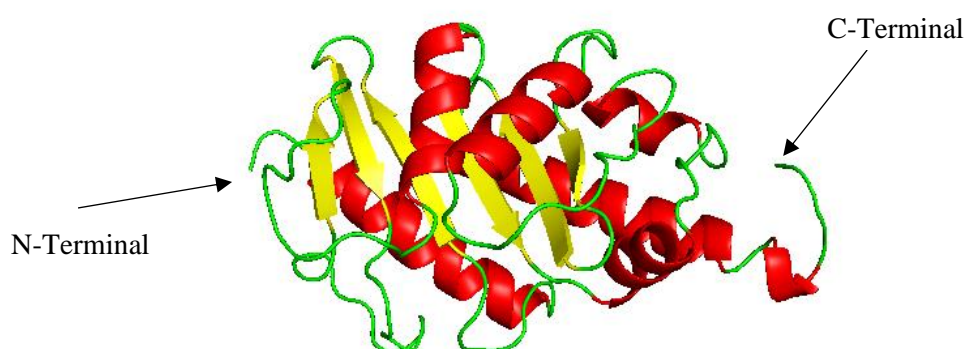
(Based on I-TASSER data result)

The optimal template for predicting the 3D structure of AldII is the deoxyribose-phosphate aldolase protein structure from *Bacillus thuringiensis*, identified by the PDB code 6btdA as shown in Table 1. The 6btdA structure exhibits superior Iden1,

Iden2, Cov, and Z-score values compared to comparable protein structures. The I-TASSER software predicts that the 3D structure of the AldII protein has a C-score of 1.39, TM-score of  $0.91 \pm 0.06$ , and RMSD of  $2.6 \pm 1.9 \text{ \AA}$ , as shown in Figure 1.

The C-score is a confidence metric utilized to evaluate the satisfaction of predicted models produced by I-TASSER. The calculation relies on the importance of aligning threading templates alignments and the convergence parameters of the structure assembly simulations. The C-score usually ranges from -5 to 2, with a higher number suggesting greater confidence in the model and a lower value indicating lower confidence. The TM-score scale measures the similarity in structure between different structures. The TM-score considers the significance of tiny distances over big ones, unlike RMSD, which is influenced by local mistakes.

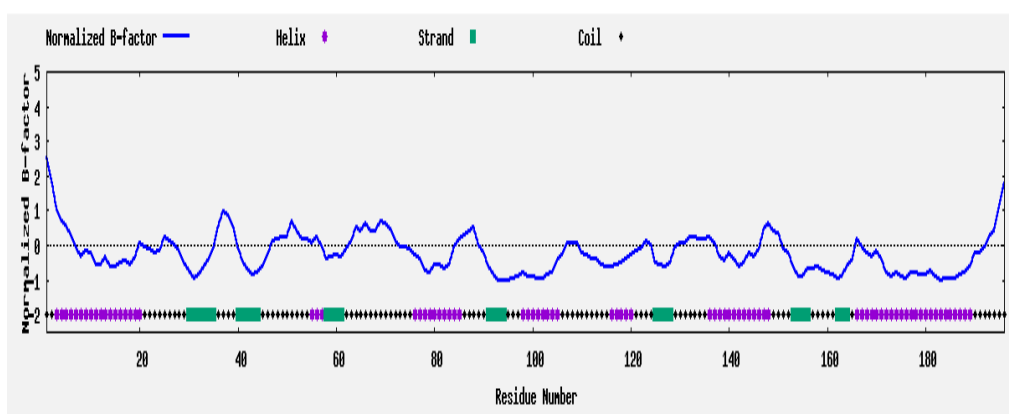
This enhances precision and reduces susceptibility to adjacent modeling inaccuracies. A TM-score above 0.5 signifies accurate topology, whilst a score below 0.17 shows insignificant resemblance. The cut-off values are independent of the protein's length under analysis (Roy *et al.*, 2010; Yang *et al.*, 2014; Y. Zhang, 2008). Z-score analysis data shows that the appropriate template to use for AldII protein structure prediction is the structure of deoxyribose-phosphate aldolase protein structure from *Bacillus thuringiensis*, with high confidence in the model proved by the C-score and the TM-score.



**Figure 1.** Predicted model by I-TASSER

The I-TASSER website was used to evaluate the stability of the predicted 3D configuration of the AldII protein. The B-factor notation was used to assess the thermal mobility of amino acids or atoms in proteins. The value is calculated by aligning the structure proteins template

from the PDB with sequence profiles generated from sequence databases in I-TASSER (Yang *et al.*, 2016). The B-factor profile shown in Figure 2 represents the standardized B-factor of the protein of interest.

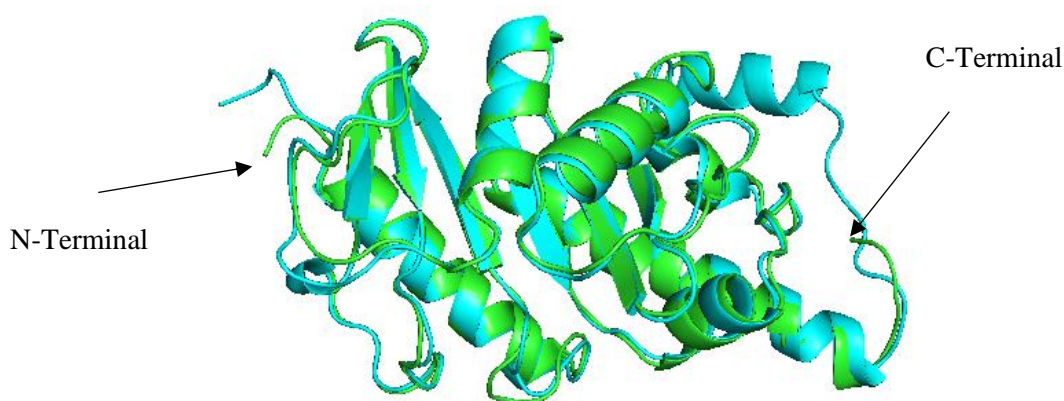


**Figure 2.** Predicted normalized B-factor

Negative values of B-factor indicate that the residue is more stable within the structure. The B-factor normalization graph in Figure 2 indicates that the 3D structure of AldII is stable. It is a sturdy construction. This confirms prior research findings that indicate the AldII protein is stable based on the Stability Index on the Expsay web server (Meray *et al.*, 2021).

I-TASSER aligns the initial model using the TM-align application after simulating the assembly structure. Due to their structural resemblance to the target, these proteins typically have comparable

traits. Users should utilize the data in the next section, 'anticipated feature using teach', to ascertain the role of the target protein. The educate function is taught to infer biological functions from several sequence and structural features, which generally yield more accurate results compared to function annotations from global structural analysis. TM-align analysis revealed that the structure of AldII closely resembles the structure of L-fucose 1-phosphate aldolase from *Streptococcus pneumoniae* with PDB code 4C24, as depicted in Figure 3.



**Figure 3.** Overlay of the AldII protein structure (green) with L-fucose 1-phosphate aldolase (cyan).

L-fucose 1-phosphate aldolase is an enzyme that facilitates the reversible breakdown of L-fucose 1-phosphate into dihydroxyacetone phosphate (DHAP) and L-lactaldehyde. It is classified as an aldehyde-lyase, a type of lyase enzyme that breaks carbon-carbon bonds (Higgins *et al.*, 2014).

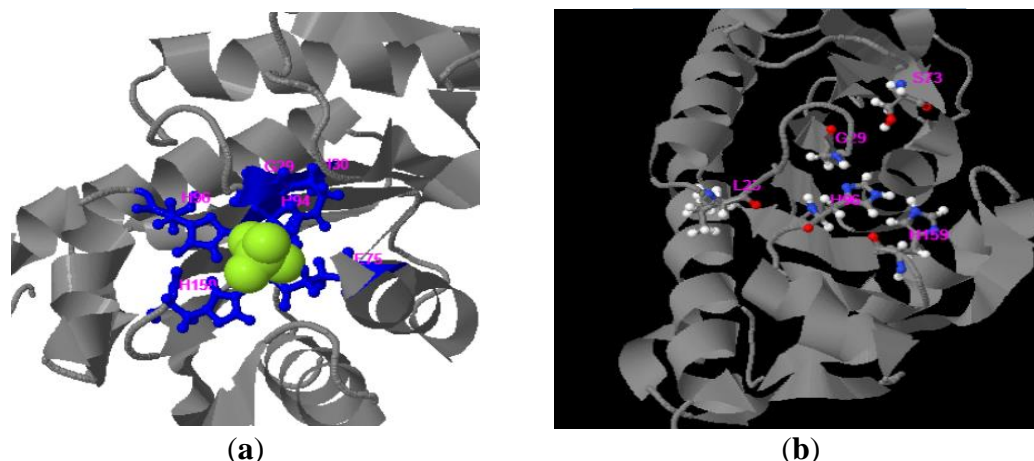
#### *Biological structure analysis of AldII*

This section contains details regarding the biological characteristics of the target protein. The annotations were generated by COFACTOR and COACH based on the I-TASSER structural prediction COFACTOR determines protein

functions, such as ligand-binding sites, enzyme classification, and gene ontology, by examining the protein's structure and studying protein-protein networks. COACH is a meta-server method that integrates function annotation outcomes from COFACTOR, TM-SITE, and S-SITE programs, focusing on ligand-binding sites.

COFACTOR, TM-SITE, and S-SITE are programs integrated with the I-TASSER webserver that can show the binding site of the ligand and substrate in the predicted structure. The structure with binding ligand and substrate is shown in Figure 4.





**Figure 4.** This is a figure of the Ligand Binding Site (a) and Active Site (b) of AldII. (a) Predicted binding ligand is represented by a green-yellow sphere, while the binding residues are depicted as blue ball-and-stick models; (b) Predicted active site residues are shown in colored ball and stick.

The residues Gly29, Asn30, Glu75, His94, His96, and His159 are identified by I-TASSER as being involved in ligand interaction with the ligand Phosphate (Figure 4a). The catalytic active residues are Leu23, Gly29, Ser73, His96, and His159 (Figure 4b), with a projected EC number of 5.1.3.4 (Yang and Zhang, 2015). Exspasy states that EC 5.1.3.4 refers to a group of L-ribulose-5-phosphate 4-epimerase enzymes that facilitate the conversion of L-ribulose 5-phosphate to D-xylulose 5-phosphate through an isomerase reaction. When examining the protein structure used to predict EC numbers, specifically the L-ribulose-5-phosphate 4-epimerase enzyme with PDB code 1k0wA, it is evident that this structure closely resembles that of the L-fuculose-1-phosphate aldolase enzyme with EC Numbers 4.1.2.17, belonging to class II Aldolase protein (Luo *et al.*, 2001).

A prediction of Gene Ontology (GO) for the AldII protein was performed to identify three GO categories: molecular functions, biological processes, and cell components. This protein is predicted to have a molecular function of zinc ion binding and phosphoprotein binding and exhibit L-fuculose-1-phosphate aldolase and rhamnulose-1-phosphate aldolase enzyme activities according to the

Consensus prediction of GO keywords. These enzymes are involved in the catabolic processes of D-arabinose, L-fucose, and rhamnose in biological systems. The AldII protein functions in the cytoplasmic membrane under the cellular component category.

### Conclusions

The AldII protein was analyzed using the I-TASSER website, revealing a stable structure closely resembling deoxyribose-phosphate aldolase from *Bacillus thuringiensis* with PDB code 6btdA. The protein's biological structure analysis reveals its function as the enzyme L-fuculose-1-phosphate aldolase, classified under Class II Aldolase. This enzyme is involved in the catabolism of arabinose, L-fuculose, and Rhamnose. The study's findings align with and corroborate prior investigations, indicating that the AldII protein is a stable protein with a catalytic site that is similar to the Class II Aldolase enzyme.

### Acknowledgment

We thank and gratefully acknowledge the funding provided by the LPPM of Mulia University.

## References

- Higgins, M. A., Suits, M. D., Marsters, C., and Boraston, A. B., 2014. Structural and Functional Analysis of Fucose-Processing Enzymes from *Streptococcus pneumoniae*. *Journal of Molecular Biology*, 426(7), 1469–1482.
- Luo, Y., Samuel, J., Mosimann, S. C., Lee, J. E., Tanner, M. E., and Strynadka, N. C. J., 2001. The Structure L-Ribulose-5-Phosphate 4-Epimerase: An Aldolase-like Platform for Epimerization. *Biochemistry*, 40(49), 14763–14771.
- Meray, N. W., Suharti, S., and Akhmaloka, A., 2021. Studi Bioinformasi Urutan Asam Amino dan Struktur 3D Protein Aldolase Kelas II (AldII) dari Uncultured *Acidilobus sp.* *Jurnal Kimia Riset*, 6(2), 117.
- Rocha, J. F., Sousa, S. F., and Cerqueira, N. M. F. S. A., 2022. Computational Studies Devoted to the Catalytic Mechanism of Threonine Aldolase, a Critical Enzyme in the Pharmaceutical Industry to Synthesize  $\beta$ -Hydroxy- $\alpha$ -amino Acids. *ACS Catalysis*, 12(9), 4990–4999.
- Roy, A., Kucukural, A., and Zhang, Y., 2010. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738.
- Suharti, Meray, N. W., Nurbaiti, S., and Akhmaloka., 2015. Cloning and expression of a novel gene encoded thermostable *archaeal* aldolase class II from natural Sample. *Biosciences Biotechnology Research Asia*, 12(2), 1023–1029.
- Windle, C. L., Müller, M., Nelson, A., and Berry, A., 2014. Engineering aldolases as biocatalysts. *Current Opinion in Chemical Biology*, 19, 25–33.
- Yang, J., Wang, Y., and Zhang, Y., 2016. ResQ: An Approach to Unified Estimation of B-Factor and Residue-Specific Error in Protein Structure Prediction. *Journal of Molecular Biology*, 428(4), 693–701.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y., 2014. The I-TASSER suite: Protein structure and function prediction. In *Nature Methods* (Vol. 12, Issue 1, pp. 7–8). Nature Publishing Group.
- Yang, J., and Zhang, Y., 2015. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, 43(W1), W174–W181.
- Zhang, C., Freddolino, P. L., and Zhang, Y., 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Research*, 45(W1), W291–W299.
- Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9.
- Zheng, W., Zhang, C., Li, Y., Pearce, R., Bell, E. W., and Zhang, Y., 2021. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Reports Methods*, 1(3), 100014.