Research Article

Exploratory *in silico* Assessment of Selected Carbohydrate-active Protein Structures from *Cellvibrio japonicus*

Ahmad Mulyadi Sirodjul Munir[1*], Grin Tommy Panggabean[1], Yuli Arif Tribudi[2], Ahmad Fahrudin Husen[3], Silvi Zakiyatul Ilmiyah[4], Ja'far Umar[5]

[1]Aquatic Resource Management Study Program, Faculty of Agriculture, Universitas Tanjungpura, Pontianak 78114, West Kalimantan, Indonesia; [2]Department of Animal Science, Faculty of Agriculture, Universitas Tanjungpura, Pontianak 78114, West Kalimantan, Indonesia; [3]Faculty of Animal Science and Technology, Universitas Brawijaya, Malang 65145, East Java, Indonesia; [4]Doctoral Program in Medical Sciences, Faculty of Medicine, Universitas Brawijaya, Malang 65145, East Java, Indonesia; [5]Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang 65145, East Java, Indonesia.

Abstract | *Cellvibrio japonicus* produces carbohydrate-active enzymes involved in plant-polysaccharide utilization, but predictions of protein–ligand interactions require careful consideration of enzyme identity and ligand chemistry. This exploratory *in silico* study aimed to assess three crystallographic protein templates and two PubChem compounds using molecular docking, two-dimensional contact analysis, and coarse-grained protein-flexibility profiles. The Protein Data Bank (PDB) entries 5OYD and 6HA9 represent different structural states of the GH5 endo-xyloglucanase Cel5D (CJA_3010), whereas entry 5I24 represents the GH31 α-transglucosylase Agd31B (CJA_3248). The analyzed compounds were the XXLG xyloglucan oligosaccharide (PubChem CID 52940123) and CID 16211032, whose PubChem label is "DEAE-cellulose" despite a displayed nitrogen-free molecular formula. Structures were prepared using BIOVIA Discovery Studio Visualizer and Open Babel in PyRx; docking was performed with AutoDock Vina. Retained poses were inspected in PyMOL and Discovery Studio, and protein root mean square fluctuation (RMSF) profiles were assessed with CABS-flex 2.0. The reported docking scores were −11.5 kcal/mol for 5OYD–XXLG, −8.5 kcal/mol for 6HA9–CID 16211032, and −6.3 kcal/mol for 5I24–CID 16211032. The shortest tabulated favorable protein contacts were ASN216 (2.20548 Å), SER510 (1.95772 Å), and SER198 (1.96980 Å), respectively; the 5OYD output also included a contact with crystallographic sulfate. The available RMSF plot showed values of approximately 0.1–3.0 Å over indices 1–201, without a full-length or aligned comparison. The 5OYD–XXLG result supports a testable structural hypothesis, but the data do not demonstrate substrate hydrolysis, comparative catalytic activity, or peatland function. Chemical verification of CID 16211032, chain-specific residue mapping, pose validation, and biochemical assays are needed to test these predictions.

Received | September 25, 2026; **Revised** | October 16, 2026; **Accepted** | October 24, 2026; **Published** | November 14, 2026
***Correspondence** | Ahmad Mulyadi Sirodjul Munir, Aquatic Resource Management Study Program, Faculty of Agriculture, Universitas Tanjungpura, Pontianak 78114, West Kalimantan, Indonesia; **Email:** ahmadmulyadi@faperta.untan.ac.id
DOI | https://dx.doi.org/10.17582/journal.nrmj/2026/20260925000046
Keywords: *Cellvibrio japonicus*; Carbohydrate-active enzymes; Molecular docking; Protein flexibility; Xyloglucan; Peatland

Copyright: 2026 by the authors. Licensee ResearchersLinks Ltd, England, UK.
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Introduction

Tropical peatlands contain large reservoirs of partially decomposed organic matter and make substantial contributions to regional and global carbon cycling. Extensive peat deposits occur in Indonesia, including Kalimantan (Miettinen and Liew, 2010; Page *et al.*, 2011). Decomposition of plant-derived cellulose and hemicellulose depends partly on microbial carbohydrate-active enzymes (CAZymes), which differ in substrate recognition, reaction mechanism, and catalytic specificity (DeBoy *et al.*, 2008; Drula *et al.*, 2022). However, substrate degradation under peatland conditions is also influenced by acidity, oxygen availability, water saturation, nutrient supply, and accessibility of complex organic matter (Dehaen *et al.*, 2025). Structural predictions therefore provide only one component of the evidence needed to understand the environmental polysaccharide transformation.

Members of the genus *Cellvibrio* are relevant models for examining plant-polysaccharide utilization. A culture-based investigation of peat soil from Parit Banjar, West Kalimantan, reported *Cellvibrio* among the bacterial genera associated with cellulose degradation (Khairiah *et al.*, 2013). The reference species *C. japonicus* Ueda107 possesses a diverse repertoire of glycoside hydrolases and carbohydrate-binding modules (DeBoy *et al.*, 2008). For example, Cel5D (CJA_3010) is a GH5_4 endo-xyloglucanase implicated in xyloglucan backbone cleavage (Attia *et al.*, 2018), while Agd31B (CJA_3248) is a GH31 enzyme biochemically characterized as an α-transglucosylase (Larsbrink *et al.*, 2012). These distinct functions require different interpretations of docking against carbohydrate-derived ligands.

Molecular docking estimates the scoring-function preference and geometrical compatibility of the candidate binding poses, rather than a catalytic turnover or an experimentally measured affinity (Trott and Olson, 2010; Eberhardt *et al.*, 2021). Blind-docking benchmarks additionally indicate that numerical rankings can be misleading when pose accuracy is not independently validated (Agboola *et al.*, 2026). Coarse-grained tools such as CABS-flex 2.0 can describe predicted residue mobility, but their root mean square fluctuation (RMSF) output is not equivalent to ligand-bound, all-atom molecular dynamics (Kuriata *et al.*, 2018). In this context, the objective of this study was to characterize the docking scores, residue-level contact patterns, and available RMSF profiles of three PDB-derived *C. japonicus* structure–ligand combinations: 5OYD–XXLG, 6HA9–CID 16211032, and 5I24–CID 16211032. This database-based analysis did not include the isolation of peatland microorganisms or direct measurements of enzyme activity.

Materials and Methods

Study design and scope

This exploratory, database-based *in silico* study examined three protein–ligand systems through molecular docking, interaction-map interpretation, and coarse-grained protein-flexibility analysis. Protein templates were obtained from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), and compound records were retrieved from PubChem. The analysis comprised the reported AutoDock Vina scores, pose-based contacts, and CABS-flex 2.0 RMSF profiles. Results were evaluated descriptively, no wet-laboratory assays, independent binding-affinity measurements, or inferential comparisons of enzyme activity were performed. Figure 1 summarizes the current computational study design.

Protein structures and ligand records

Three crystallographic protein records were identified in the RCSB Protein Data Bank (https://www.rcsb.org/; accessed August 24, 2026). Entries 5OYD and 6HA9 corresponded to Cel5D/CJA_3010 of *C. japonicus* Ueda107, each with two deposited chains of 396 amino acids. Entry 5OYD is an unmutated endo-xyloglucanase structure (2.10 Å resolution) associated with crystallographic xyloglucan-derived oligosaccharides, while 6HA9 is a mutated Cel5D structure (2.00 Å resolution) deposited with a mechanism-based xyloglucan-derived inhibitor (Attia *et al.*, 2018; Jain *et al.*, 2018). Entry 5I24 (1.85 Å resolution) corresponds to the unmutated GH31 enzyme Agd31B/CJA_3248, with a deposited sequence length of 836 amino acids and a crystal structure obtained in the presence of a cyclophellitol-aziridine probe (Jiang *et al.*, 2016). The experimentally determined structures are templates for computation, not evidence of enzyme occurrence in the Indonesian peat samples.

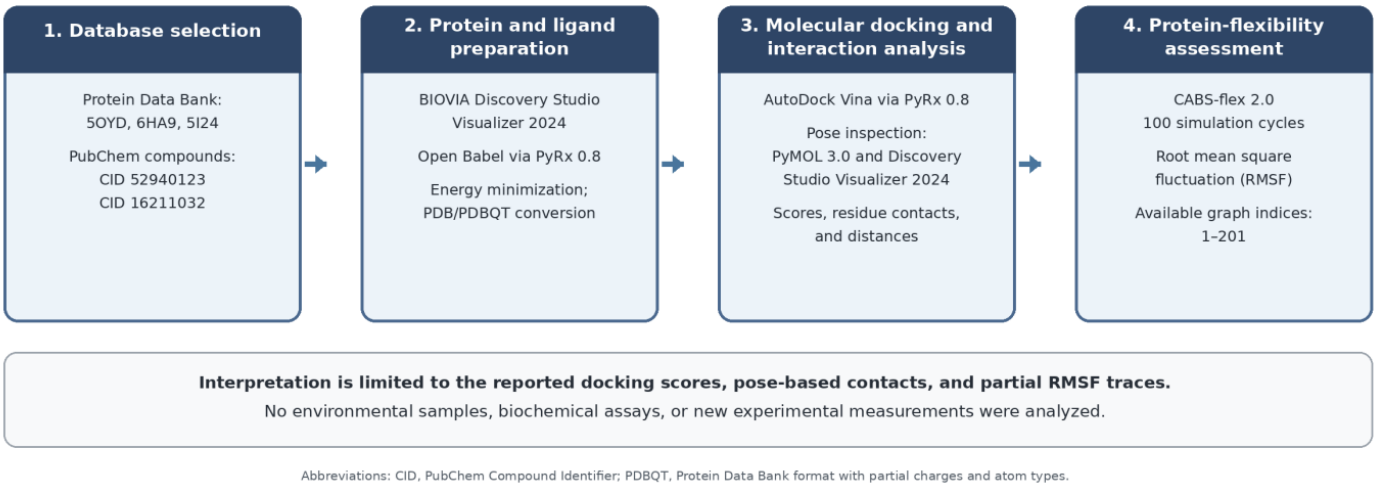


Figure 1: *High-resolution computational workflow for database-guided molecular docking, pose-based interaction analysis, and coarse-grained protein-flexibility assessment. Where; PDB: Protein Data Bank, CID: PubChem Compound Identifier, PDBQT: Protein Data Bank format with partial charges and atom types, RMSF: root mean square fluctuation. No environmental samples or wet-laboratory measurements were analyzed.*

Compound identities were examined using PubChem (https://pubchem.ncbi.nlm.nih.gov/; accessed August 24, 2026) (Kim *et al.*, 2021). CID 52940123 is the XXLG xyloglucan oligosaccharide (GlyTouCan G18132XG; C45H76O38). CID 16211032 carries the database label "DEAE-cellulose," but its displayed formula, C12H22O11, contains no nitrogen, although a diethylaminoethyl substituent necessarily contains nitrogen. The PubChem structure therefore cannot be assumed to represent authentic DEAE-substituted cellulose. Accordingly, using this compound, the two systems are identified by CID 16211032, and conclusions about actual DEAE-cellulose are withheld pending verification of the docked coordinates and atom connectivity.

Protein and ligand preparation

The protein models were inspected in BIOVIA Discovery Studio Visualizer 2024. The docking-preparation procedure included removal of solvent and designated nonessential crystallographic ligands before export of receptor structures in PDB format. Ligand coordinates were processed and energy-minimized using Open Babel through PyRx version 0.8, followed by conversion to PDBQT format (O'Boyle *et al.*, 2011; Dallakyan and Olson, 2015). Because the 5OYD interaction output included a contact with crystallographic sulfate (SO4 B:511), the contact map alone cannot confirm that all nonprotein components were excluded from the interaction environment. The exact provenance of the input coordinates, including the atom connectivity of CID 16211032, could not be confirmed from the available records.

Molecular docking and pose visualization

Blind docking was performed with the PyRx version 0.8 Vina Wizard using the AutoDock Vina scoring function (Trott and Olson, 2010). The three receptor–compound combinations were 5OYD–XXLG, 6HA9–CID 16211032, and 5I24–CID 16211032. For each combination, the reported lowest-scoring pose was used for visualization and contact analysis and its Vina score was expressed in kcal/mol. Docking scores were interpreted as scoring-function outputs for the individual modeled systems, not as experimental free energies or directly comparable catalytic measurements. Detailed grid settings, docking random seeds, and pose-validation statistics were not available in the study records and consequently were not inferred.

The docking poses were visualized with PyMOL version 3.0, and the two-dimensional interaction diagrams and residue-contact lists were examined using BIOVIA Discovery Studio Visualizer 2024. Conventional hydrogen bonds and carbon–hydrogen contacts were classified according to the exported interaction annotations. The distances in Table 1 reproduce the numerical

precision of the software output and are not independent experimental measurements. Contacts with nonprotein crystallographic components were distinguished from contacts with amino-acid residues. Residue labels are reported as exported; labels exceeding the length of a single Cel5D chain require verification against chain identifiers and deposited residue-numbering schemes before mechanistic interpretation (Burley *et al.*, 2021).

Table 1. *Reported pose-based contacts and distances for the three analyzed systems*

System	Residue / component-ligand	Category	Interaction type	Distance (Angstrom)
5OYD-XXLG	XXLG-PHE440	Hydrogen bond	Conventional hydrogen bond	2.31218
5OYD-XXLG	XXLG-HIS208	Hydrogen bond	Conventional hydrogen bond	2.21381
5OYD-XXLG	XXLG-GLU255	Hydrogen bond	Conventional hydrogen bond	2.29298
5OYD-XXLG	XXLG-ASN216	Hydrogen bond	Conventional hydrogen bond	2.20548
5OYD-XXLG	XXLG-ASN216	Hydrogen bond	Conventional hydrogen bond	2.85612
5OYD-XXLG	ASN132-XXLG	Hydrogen bond	Conventional hydrogen bond	3.24998
5OYD-XXLG	ASN254-XXLG	Hydrogen bond	Conventional hydrogen bond	3.25992
5OYD-XXLG	XXLG-SO4 B:511	Nonprotein sulfate	Reported carbon–hydrogen contact	3.51492
6HA9–CID 16211032	CID 16211032–THR509	Hydrogen bond	Conventional hydrogen bond	2.59537
6HA9–CID 16211032	CID 16211032–SER510	Hydrogen bond	Conventional hydrogen bond	1.95772
6HA9–CID 16211032	CID 16211032–ILE740	Hydrogen bond	Conventional hydrogen bond	2.40658
6HA9–CID 16211032	LYS643–CID 16211032	Hydrogen bond	Conventional hydrogen bond	3.15934
6HA9–CID 16211032	CID 16211032–GLU600	Hydrogen bond	Carbon-hydrogen bond	3.09088
6HA9–CID 16211032	CID 16211032–ASP467	Hydrogen bond	Carbon-hydrogen bond	3.36978
6HA9–CID 16211032	CID 16211032–ASP467	Hydrogen bond	Carbon-hydrogen bond	3.10457
6HA9–CID 16211032	ALA739–CID 16211032	Hydrogen bond	Carbon-hydrogen bond	3.71331
5I24–CID 16211032	CID 16211032–ARG123	Hydrogen bond	Conventional hydrogen bond	2.05645
5I24–CID 16211032	CID 16211032–SER198	Hydrogen bond	Conventional hydrogen bond	1.96980
5I24–CID 16211032	CID 16211032–GLU597	Hydrogen bond	Conventional hydrogen bond	2.27480
5I24–CID 16211032	CID 16211032–GLN200	Hydrogen bond	Carbon-hydrogen bond	3.40001
5I24–CID 16211032	CID 16211032–GLN200	Hydrogen bond	Carbon-hydrogen bond	3.73255

Where; SO_4 B:511 is a crystallographic sulfate component, not an amino-acid residue, and is excluded from the protein-contact count for 5OYD–XXLG. "DEAE-cellulose" is the source label for PubChem CID 16211032; its displayed formula (C12H22O11) lacks nitrogen and does not establish the identity of a DEAE-substituted cellulose polymer. Residue identifiers exported for the Cel5D structures require chain-specific verification.

Coarse-grained protein-flexibility assessment

The protein coordinate files associated with the three modeled systems were assessed using CABS-flex 2.0, with 100 simulation cycles and the relevant protein chain selected; other parameters were left at server defaults (Kuriata *et al.*, 2018). Residue-level RMSF was interpreted as a coarse-grained indicator of predicted local mobility. Only a combined graph displaying indices 1–201 is available. These indices have not been shown to represent homologous positions across Cel5D and Agd31B and do not cover the complete 396- and 836-residue deposited sequences. Because chain identity and ligand

retention during the calculations were not independently documented, the profiles were not treated as evidence of complex stability or as a substitute for all-atom molecular dynamics.

Descriptive interpretation

The three docking systems were summarized descriptively using the Vina scores, reported favorable-contact counts, individual contact distances, and plotted RMSF trends. For 5OYD–XXLG, the sulfate-associated annotation was recorded individually from the protein–ligand contacts. No significance testing or comparison of enzymatic rate constants was performed. Differences in protein template, ligand composition, and structural state precluded a controlled comparison of catalytic efficiency among the three combinations.

Results

Verified identities of the analyzed systems

The three crystallographic records represent two enzyme identities from the *C. japonicus* Ueda107 reference strain, not proteins obtained from a local peat isolate (Table 2). Entries 5OYD and 6HA9 are structural templates of Cel5D/CJA_3010; 6HA9 is a mutated structure deposited with a mechanism-based inhibitor. Entry 5I24 represents the GH31 α-transglucosylase Agd31B/CJA_3248, which is functionally distinct from a classical β-1,4-cellulose hydrolase. PubChem identifies CID 52940123 as XXLG. Although CID 16211032 is labelled "DEAE-cellulose," its displayed nitrogen-free formula does not support identifying the docked structure as authentic DEAE-substituted cellulose.

Table 2 . *Database annotations and compound identities of the three modeled systems*

ID	PDB	Database annotation	Gene / family	Structure	Analyzed compound
A	5OYD	GH5 endo-xyloglucanase Cel5D	*cel5D / CJA_3010; GH5_4*	Chains A and B; 396 aa per chain; 2.10 Å; unmutated	CID 52940123; XXLG xyloglucan oligosaccharide
B	6HA9	Endo-xyloglucanase Cel5D; inhibitor-complex template	*cel5D / CJA_3010; GH5_4*	Chains A and B; 396 aa per chain; 2.00 Å; mutated	CID 16211032; PubChem name "DEAE-cellulose"; displayed formula lacks N
C	5I24	Oligosaccharide 4-alpha-D-glucosyltransferase Agd31B	*agd31B / CJA_3248; GH31; EC 2.4.1.161*	Chain A; 836 aa; 1.85 Angstrom; no mutation	CID 16211032; PubChem name "DEAE-cellulose"; displayed formula lacks N

Where; PDB, Protein Data Bank; GH, glycoside hydrolase; EC, Enzyme Commission; CID, PubChem Compound Identifier; aa, amino acids; XXLG, xyloglucan oligosaccharide; DEAE, diethylaminoethyl. The PubChem label "DEAE-cellulose" for CID 16211032 does not establish that the displayed structure represents a nitrogen-containing DEAE-substituted cellulose.

Docking scores and interaction summaries

The reported AutoDock Vina scores are −11.5 kcal/mol for 5OYD–XXLG, −8.5 kcal/mol for 6HA9–CID 16211032, and −6.3 kcal/mol for 5I24–CID 16211032 (Table 3). The most negative score is observed for the Cel5D–XXLG combination. Because the systems differed in ligand identity and structural context, this ordering does not establish a superior substrate affinity, substrate hydrolysis, or relative catalytic efficiency. Of the eight favorable annotations in the 5OYD–XXLG output, seven are assigned to protein residues and one to crystallographic sulfate.

Where; CID, PubChem Compound Identifier; Vina, AutoDock Vina; Å, ångström. Contact counts are based on exported pose annotations and are not experimentally validated binding measurements.

In the 5OYD–XXLG output, seven favorable annotations are assigned to protein residues: PHE440, HIS208, GLU255, ASN216 (two contacts), ASN132, and ASN254 (Table 1). The shortest listed protein contact is ASN216–XXLG at 2.20548 Å. An additional carbon–hydrogen contact at 3.51492 Å has involved a crystallographic sulfate

Table 3 . *AutoDock Vina scores and reported favorable interaction annotations*

Protein-ligand system	Docking score (kcal/mol)	Favorable contact annotations	Nearest tabulated contact	Distance range (Angstrom)
5OYD-XXLG	-11.5	7 protein + 1 sulfate	ASN216–XXLG; 2.20548 Å	2.20548-3.51492
6HA9–CID 16211032	-8.5	8	SER510–CID 16211032; 1.95772 Å	1.95772-3.71331
5I24–CID 16211032	-6.3	5	SER198–CID 16211032; 1.96980 Å	1.96980-3.73255

SO4 B:511, not a protein residue, and was therefore excluded from the protein-contact count. The interaction diagram has also displayed unfavorable contacts involving ASN148 and TRP149, although distances are not available. Because some exported labels have exceeded the length of a single 396-residue Cel5D chain, residue identifiers must be reconciled with the deposited chain and author numbering schemes before mechanistic interpretation (Burley *et al.*, 2021).

For 6HA9–CID 16211032, four conventional hydrogen-bond annotations have involved THR509, SER510, ILE740, and LYS643, while four carbon–hydrogen annotations have involved GLU600, ASP467 (twice), and ALA739 (Table 1). SER510 has the shortest tabulated contact (1.95772 Å), whereas ALA739 has the longest (3.71331 Å). Unfavorable contacts involving GLU738 and ASN686 are also observed. Several residue labels exceed the length of a single Cel5D chain, where their chain assignment and numbering therefore require verification before they can be linked to known catalytic residues.

For 5I24–CID 16211032, conventional hydrogen-bond annotations have involved ARG123, SER198, and GLU597, while two carbon–hydrogen contacts have involved GLN200 at 3.40001 and 3.73255 Å (Table 1). SER198 has the shortest tabulated contact (1.96980 Å). No additional unfavorable contact is evident in the available diagram. Fig. 2 presents the modeled poses and interaction maps. These distances indicate geometric proximity within the modeled poses, not direct measurements of binding affinity or catalytic mechanism.

Protein-flexibility profiles

The CABS-flex 2.0 graph has displayed RMSF values for indices 1–201 across the three coordinate-based calculations (Fig. 3). The values are approximately 0.1–3.0 Å, with local minima around indices 20–31, 66–72, 104–109, and 180–188, and more variable regions around 55–61, 79–91, 113–121, 145–160, and 169–177. These index positions describe the graph only, but they have not been established as homologous or functionally equivalent sites across the proteins.

The displayed interval does not span the complete 396-residue Cel5D or 836-residue Agd31B sequences, and correspondence between graph indices, input-chain identity, and deposited residue numbering has not been demonstrated. Therefore, Fig. 3 supports only a descriptive view of the displayed fluctuations, not a full-protein mobility ranking or a site-specific structure–function comparison. The available traces do not provide ligand RMSD, contact occupancy, binding free energy, or experimental stability measurements.

Discussion

This study examined selected several carbohydrate-derived compounds docked to two functionally distinct enzymes from *C. japonicus* Ueda107. The most negative AutoDock Vina score has been observed for 5OYD–XXLG (-11.5 kcal/mol), followed by 6HA9–CID 16211032 (-8.5 kcal/mol), and 5I24–CID 16211032 (-6.3 kcal/mol). These values describe scoring-function outputs for specific modeled configurations, not experimentally measured binding free energies, substrate turnover, or environmental performance. Docking scores can help prioritize poses, but a favorable score alone does not confirm pose accuracy or catalytic activity

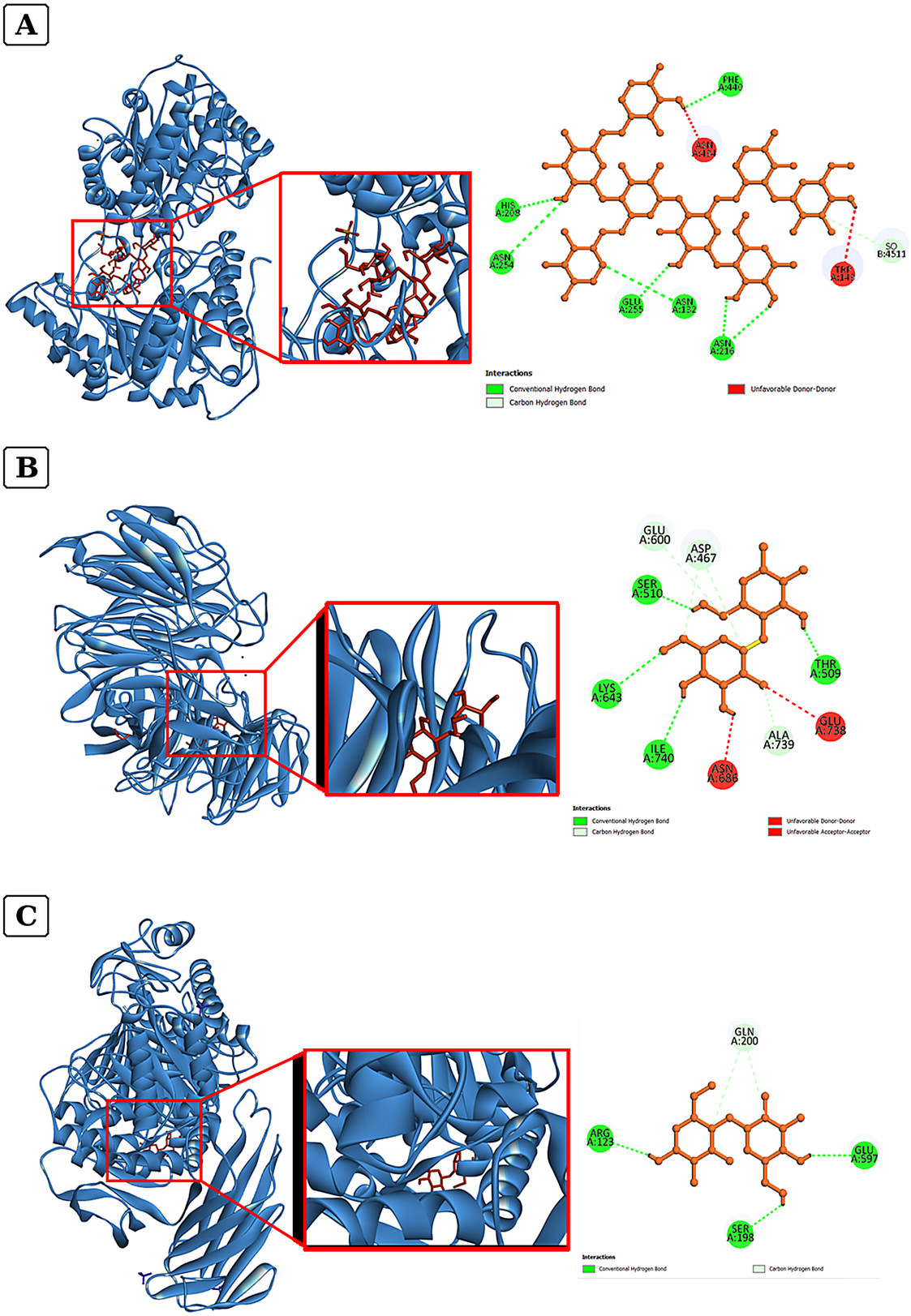


Figure 2: *Modeled docking poses and two-dimensional contact diagrams. (A) Cel5D structure 5OYD with the XXLG xyloglucan oligosaccharide (CID 52940123); (B) Cel5D structure 6HA9 with CID 16211032; and (C) Agd31B structure 5I24 with CID 16211032. Green dashed lines indicate reported conventional hydrogen bonds, pale green dashed lines indicate carbon–hydrogen contacts, and red dashed lines indicate unfavorable contacts. The ligand label "DEAE-cellulose" in the interaction diagrams reproduces the PubChem record name and does not confirm the chemistry of authentic DEAE-substituted cellulose. Exported residue labels require chain-specific verification. The poses do not demonstrate a catalytic turnover.*

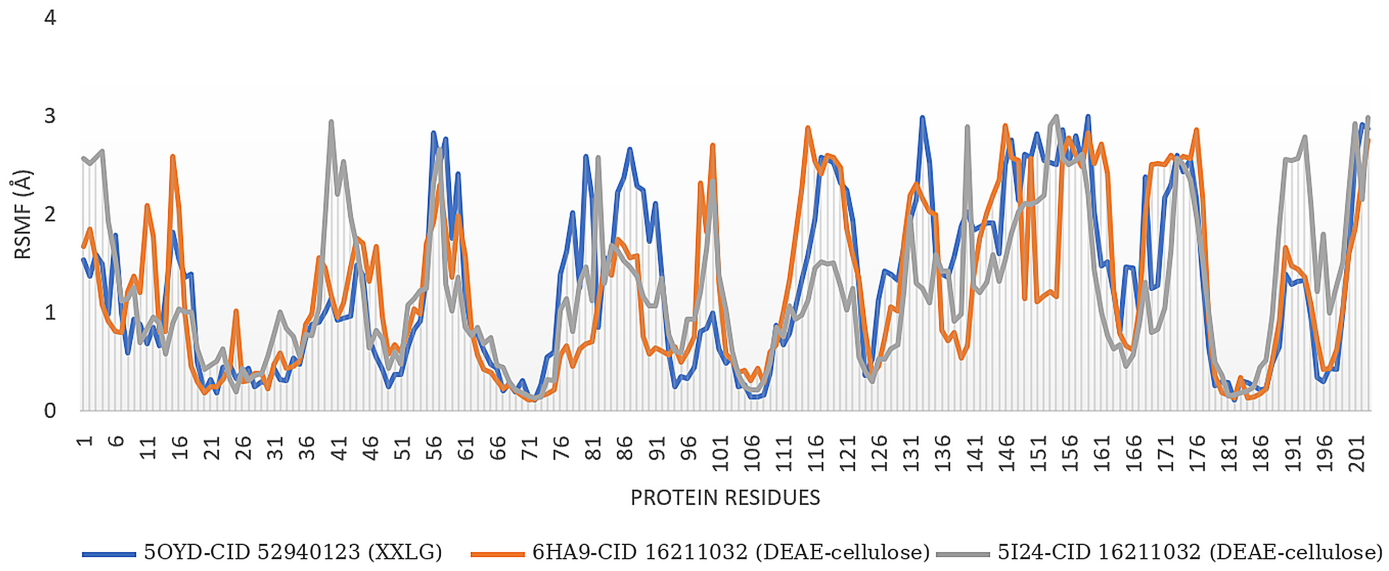


Figure 3: *CABS-flex 2.0 root mean square fluctuation (RMSF) profiles for the coordinate sets associated with 5OYD–XXLG, 6HA9–CID 16211032, and 5I24–CID 16211032. The combined graph displays indices 1–201. These indices are not aligned between the two enzyme identities and do not represent full-length mobility comparisons or evidence of ligand-bound complex stability. The plotted legend label "DEAE-cellulose" refers to the PubChem name for CID 16211032 and does not verify the chemical identity of a DEAE-substituted cellulose polymer.*

(Trott and Olson, 2010; Eberhardt *et al.*, 2021; Agboola *et al.*, 2026). The biological meaning of each result must therefore be assessed in relation to the enzyme's established function and chemical identity of the ligand.

The two Cel5D templates represent different structural states of the same GH5_4 endo-xyloglucanase, rather than independent enzyme classes. The unmutated 5OYD structure is associated with xyloglucan-related crystallographic components, whereas 6HA9 is a mutated structure deposited with a mechanism-based xyloglucan-derived inhibitor (Attia *et al.*, 2018; Jain *et al.*, 2018). These differences may affect pocket geometry and side-chain positions. Because 5OYD was docked with XXLG and 6HA9 with a chemically uncertain compound, the scores cannot isolate the effects of mutation, conformation, or ligand identity. A controlled comparison would require the same chemically verified ligand, harmonized preparation, and docking settings.

The 5OYD–XXLG pairing has the clearest enzyme–carbohydrate rationale among the three systems. Cel5D is an endo-xyloglucanase that cleaves the xyloglucan backbone, while XXLG is a defined xyloglucan oligosaccharide rather than bulk cellulose (DeBoy *et al.*, 2008; Attia *et al.*, 2018; Drula *et al.*, 2022). Its branched carbohydrate structure therefore provides a biologically relevant probe of possible recognition by a xyloglucan-active enzyme. However, the pose alone cannot determine whether XXLG is positioned as a substrate fragment or product-like ligand. Catalytic relevance depends on the orientation of the glycosidic bond relative to the catalytic machinery, which cannot be established from the docking score alone.

The 5OYD contact map provides a local geometric information consistent with a polar carbohydrate-binding environment. Seven favorable annotations are assigned to protein residues, including two contacts involving ASN216, where the shortest listed protein contact is ASN216–XXLG at 2.20548 Å. The eighth annotation has involved SO4 B:511, a crystallographic sulfate component, which was not counted as a protein contact. These contacts may guide further pose inspection, but software-defined distances and contact counts do not establish a productive catalytic geometry. Docking predictions require independent structural validation,

particularly in blind-docking settings (Trott and Olson, 2010; Eberhardt *et al.*, 2021; Agboola *et al.*, 2026).

Residue numbering is a further constraint on mechanistic interpretation. The Cel5D chains are reported as 396 residues long, yet the interaction exports include labels such as PHE440 and ILE740. These labels may reflect author numbering, chain handling, or renumbering during file preparation, but their correspondence to the deposited coordinates has not been established. The Protein Data Bank distinguishes chain identifiers and sequential versus author-defined residue numbering; thus, the prepared receptor files should be mapped to the deposited coordinates before any contact is assigned to a catalytic or substrate-binding position (Burley *et al.*, 2021). The remaining systems require particular caution because ligand chemistry, residue mapping, and the available flexibility output are not fully verified.

CID 16211032 is labelled "DEAE-cellulose" in PubChem, but its displayed formula ($C_{12}H_{22}O_{11}$) contains no nitrogen. Because authentic diethylaminoethyl substitution necessarily introduces nitrogen, the database structure cannot be assumed to represent the modified cellulose polymer. The scores and contacts for 6HA9 and 5I24 must therefore be attributed to the coordinates actually docked, not generalized to authentic DEAE-cellulose. Verification of the archived ligand coordinates and atom connectivity is required before assigning a chemical or a mechanistic meaning to the database label (Kim *et al.*, 2021).

The 6HA9–CID 16211032 map has contained eight favorable annotations, with SER510 showing the shortest listed contact (1.95772 Å); the 5I24–CID 16211032 map contained five, with SER198 showing the shortest contact (1.96980 Å). These values describe distances in the exported poses. They do not establish binding specificity, and the 6HA9 diagram has also shown unfavorable contacts. Without verified ligand chemistry and pose validation against an appropriate reference structure, neither a system can support claims about affinity for DEAE-cellulose nor degradation of native cellulose.

Agd31B/CJA_3248 is a GH31 α-transglucosylase, not a classical β-1,4-cellulose hydrolase (Larsbrink *et al.*, 2012). The 5I24 structure has been also studied in an activity-based-probe context (Jiang *et al.*, 2016), underscoring the importance of using chemically and mechanistically appropriate ligands. A favorable geometric fit to CID 16211032 does not change the established enzyme identity or demonstrate cellulose degradation.

The three scores are not directly comparable because the systems differ in ligand size and composition, protein identity, and structural state. XXLG is a large, branched oligosaccharide ($C_{45}H_{76}O_{38}$), whereas CID 16211032 has the smallest displayed formula $C_{12}H_{22}O_{11}$, where its relationship to the database label remains uncertain. Ligand size, conformational sampling, receptor preparation, and scoring-function assumptions can all affect the docking scores. The sequence of −11.5, −8.5, and −6.3 kcal/mol should therefore be treated as a descriptive result of this workflow, not as a ranking of substrate affinity or enzyme performance (Agboola *et al.*, 2026).

The contact annotations also require cautious interpretation. Interaction diagrams classify the close approaches using software-defined geometric criteria, and the five-decimal distances in Table 1 reflect exported numerical precision rather than experimental accuracy. Multiple contacts to one residue are not independent demonstrations of binding determinants, and crystallographic components such as sulfate may appear in the interaction environment. Before assigning mechanistic roles, the docked atoms, stereochemistry, protonation, donor–acceptor geometry, and residue mapping should be verified. For carbohydrate-active enzymes, the orientation of the relevant glycosidic linkage is more informative than the shortest distance alone.

The CABS-flex 2.0 profiles provide a separate, coarse-grained description of predicted protein mobility. The available graph covers indices 1–201 and shows RMSF values of approximately 0.1–3.0 Å, but it represents only a part of the deposited Cel5D and Agd31B sequences (Kuriata *et al.*, 2018). The indices are not aligned between the two enzyme identities; peaks at the same index therefore cannot be interpreted as differences at homologous functional sites. The graph supports a descriptive

reading of local fluctuations only, not a comparison of full-length rigidity or folding stability.

Protein RMSF should not also be confused with protein–ligand complex stability. The available output does not provide ligand RMSD, ligand-contact occupancy, or trajectory-derived binding energies and does not establish that a ligand remained bound during an all-atoms simulation. Chain-resolved profiles, structural alignment, and appropriately designed apo-versus-ligand bound simulations could inform future hypotheses, but these additional analyses were not performed in the present study.

The peatland context provides environmental motivation, not direct evidence of ecological function. *Cellvibrio* has been reported among the bacterial species associated with cellulose decomposition in peat soil from West Kalimantan (Khairiah *et al.*, 2013), but the proteins analyzed here came from the *C. japonicus* Ueda107 reference strain, which was not linked to a taxonomically verified local isolate. Docking therefore cannot establish the occurrence, expression, or activity of Cel5D or Agd31B in the Indonesian peatlands. Peat decomposition also depends on oxygen, moisture, redox conditions, temperature, and substrate accessibility (Dehaen *et al.*, 2025).

The next steps were to verify the atom-level identity of CID 16211032, reconcile chain and residue numbering against the deposited coordinates, and validate poses using experimentally characterized carbohydrate or inhibitor structures. Docking should then be repeated with documented search settings and suitable controls. Finally, purified or recombinant enzymes should be tested with chemically defined, enzyme-appropriate substrates, with a product formation and kinetic parameters measured where justified. Only after a substrate-specific activity is being established, experiments with peatland-derived isolates should assess environmental relevance under defined moisture and oxygen conditions.

In summary, 5OYD–XXLG provides the strongest biologically motivated hypothesis, because XXLG is a xyloglucan oligosaccharide and Cel5D is an established endo-xyloglucanase. The CID 16211032 systems remain conditional on chemical verification, while the partial RMSF traces cannot establish a complex stability or catalytic performance. Taken together, the docking scores, contact maps, and flexibility profiles can guide follow-up experiments but do not prove cellulose degradation or peatland function.

Conclusions and Recommendations

The assessment of three *Cellvibrio japonicus* protein–ligand combinations has identified the lowest AutoDock Vina score for 5OYD–XXLG (−11.5 kcal/mol), supporting a testable structural hypothesis involving the established xyloglucan-active enzyme Cel5D. The two combinations involving CID 16211032 have yielded scores of −8.5 and −6.3 kcal/mol, but the displayed PubChem formula does not represent an authentic nitrogen-containing DEAE-cellulose structure. Therefore, these results cannot demonstrate binding to or degradation of the modified polymer. Contact analyses provide pose-level geometric information, while the available indices of 1–201 offer only a partial description of the predicted protein flexibility. The findings of this study do not establish a comparative enzyme efficiency or peatland-remediation potential. Further studies are recommended to verify ligand chemistry and residue mapping, validate poses against experimentally characterized structures, and measure substrate-specific activity using several biochemical assays before environmental applications are considered.

Novelty Statement

The novelty of this study lies in its critical, database-based assessment that integrates docking scores, pose-derived contact annotations, and available partial CABS-flex RMSF profiles, while checking whether protein and compound identities support the proposed biological interpretations. In particular, it distinguishes the plausible Cel5D–XXLG hypothesis from the chemically unverified CID 16211032 systems and defines the evidence needed before catalytic or peatland-function claims can be made.

Acknowledgments

The authors acknowledge the Faculty of Agriculture, Universitas Tanjungpura, West

Kalimantan, Indonesia, for institutional support and scientific collaboration.

Conflict of Interests

The authors declare that they have no known competing financial interests or personal conflicts that could have influenced the work reported in this manuscript.

Authors' Contributions

Ahmad Mulyadi. S. Munir: Conceptualization, funding acquisition, supervision, validation, writing - original draft, and writing - review and editing. **Grin T. Panggabean**: Conceptualization and writing - original draft. **Yuli A. Tribudi**: Conceptualization and writing - original draft. **Ahmad F. Husen**: Conceptualization, project administration, validation, and writing - review and editing. **Silvi Z. Ilmiyah**: Conceptualization, software, validation, writing - original draft, and writing - review and editing. **Ja'far Umar**: Conceptualization, software, validation, writing - original draft, and writing - review and editing.

Ethical approval

Not applicable.

Funding source

This study was supported by the Faculty of Agriculture, Universitas Tanjungpura, Indonesia, through the 2025 institutional DIPA allocation (SP DIPA-139.03.2.693380/2025, dated 2 December 2024).

References

Agboola OE, Agboola SS, Aribigbola TC, Shaleye AB, Adedeji OS, Ayinla ZA, Omotuyi IO. 2026. Ranking metrics conceal pose errors in blind docking with AutoDock Vina, a cross-target benchmark. Results in Engineering. 32:112651. https://doi.org/10.1016/j.rineng.2026.112651.

Attia, M.A., Nelson, C.E., Offen, W.A., Jain, N., Davies, G.J., Gardner, J.G. and Brumer, H., 2018. In vitro and in vivo characterization of three Cellvibrio japonicus glycoside hydrolase family 5 members reveals potent xyloglucan backbone-cleaving functions. Biotechnol. Biofuels, 11: 45. https://doi.org/10.1186/s13068-018-1039-6.

Bugnon, M., Röhrig, U.F., Goullieux, M., Perez, M.A.S., Daina, A., Michielin, O. and Zoete, V., 2024. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. Nucleic Acids Res., 52(W1): W324-W332. https://doi.org/10.1093/nar/gkae300.

Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G.V., Christie, C.H., Dalenberg, K., Di Costanzo, L., Duarte, J.M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D.S., Ghosh, S., Green, R.K., Guranović, V., Henry, L., Hudson, B.P., Khokhriakov, I., Lawson, C.L., Liang, Y., Lowe, R., Peisach, E., Persikova, I., Randle, C., Rose, A., Rose, Y., Sali, A., Segura, J., Sekharan, M.R., Shao, C., Tao, Y.P., Voigt, M., Westbrook, J.D., Young, J.Y., Zardecki, C. and Zhuravleva, M., 2021. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res., 49(D1): D437–D451. https://doi.org/10.1093/nar/gkaa1038.

Dallakyan, S. and Olson, A.J., 2015. Small-molecule library screening by docking with PyRx. Methods Mol. Biol., 1263: 243-250. https://doi.org/10.1007/978-1-4939-2269-7_19.

DeBoy, R.T., Mongodin, E.F., Fouts, D.E., Tailford, L.E., Khouri, H., Emerson, J.B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H.J. and Nelson, K.E., 2008. Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus. J. Bacteriol., 190(15): 5455-5463. https://doi.org/10.1128/JB.01701-07.

Dehaen, E.M., Burke, E.J., Chadburn, S.E., Kaduk, J., Sitch, S., Smith, N.D. and Gallego-Sala, A.V., 2025. Drivers of soil heterotrophic respiration in tropical peatlands: a review to inform peat carbon accumulation modelling. Front. Geochem., 3: 1492386. https://doi.org/10.3389/fgeoc.2025.1492386.

Drula, E., Garron, M.L., Dogan, S., Lombard, V., Henrissat, B. and Terrapon, N., 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res., 50(D1): D571-D577. https://doi.org/10.1093/nar/gkab1045.

Eberhardt, J., Santos-Martins, D., Tillack, A.F. and Forli, S., 2021. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. J. Chem. Inf. Model., 61(8): 3891-3898. https://doi.org/10.1021/acs.jcim.1c00203.

Jain, N., Attia, M.A., Offen, W.A., Davies, G.J. and Brumer, H., 2018. Synthesis and application of a highly branched, mechanism-based 2-deoxy-2-fluoro-oligosaccharide inhibitor of endo-xyloglucanases. Org. Biomol. Chem., 16(45): 8732-8741. https://doi.org/10.1039/C8OB02250J.

Jiang, J., Kuo, C.L., Wu, L., Franke, C., Kallemeijn, W.W., Florea, B.I., van Meel, E., van der Marel, G.A., Codee, J.D.C., Boot, R.G., Davies, G.J., Overkleeft, H.S. and Aerts, J.M.F.G., 2016. Detection of active mammalian GH31 alpha-glucosidases in health and disease using in-class, broad-spectrum activity-based probes. ACS Cent. Sci., 2(5): 351-358. https://doi.org/10.1021/acscentsci.6b00057.

Kakacek, B.E., Liang, J., Dickerson, K.A. and Gardner, J.G., 2025. The physiological functions of the Cbp2D and Cbp2E proteins are important for insoluble cellulose-dependent growth in Cellvibrio japonicus. Appl. Environ. Microbiol., 91(10): e00818-25. https://doi.org/10.1128/aem.00818-25.

Khairiah, E., Khotimah, S. and Mulyadi, A., 2013. Karakterisasi dan kepadatan bakteri pendegradasi selulosa pada tanah gambut di Desa Parit Banjar Kabupaten Pontianak. Protobiont, 2(2): 87-92.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., Zaslavsky, L., Zhang, J. and Bolton, E.E., 2021. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res., 49(D1): D1388-D1395. https://doi.org/10.1093/nar/gkaa971.

Kuriata, A., Gierut, A.M., Oleniecki, T., Ciemny, M.P., Kolinski, A., Kurcinski, M. and Kmiecik, S., 2018. CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. Nucleic Acids Res., 46(W1): W338-W343. https://doi.org/10.1093/nar/gky356.

Larsbrink, J., Izumi, A., Hemsworth, G.R., Davies, G.J. and Brumer, H., 2012. Structural enzymology of Cellvibrio japonicus Agd31B protein reveals α-transglucosylase activity in glycoside hydrolase family 31. J. Biol. Chem., 287(52): 43288-43299. https://doi.org/10.1074/jbc.M112.416511.

Liang, J., Szczyrbak, J.J., Novak, J.K. and Gardner, J.G., 2026. Transcriptome data set of Cellvibrio japonicus grown using β-glucan media. Microbiol. Resour. Announc., 15(3): e01491-25. https://doi.org/10.1128/mra.01491-25.

Miettinen, J. and Liew, S.C., 2010. Status of peatland degradation and development in Sumatra and Kalimantan. Ambio, 39(5-6): 394-401. https://doi.org/10.1007/s13280-010-0051-2.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R., 2011. Open Babel: an open chemical toolbox. J. Cheminform., 3: 33. https://doi.org/10.1186/1758-2946-3-33.

Page, S.E., Rieley, J.O. and Banks, C.J., 2011. Global and regional importance of the tropical peatland carbon pool. Glob. Change Biol., 17(2): 798-818. https://doi.org/10.1111/j.1365-2486.2010.02279.x.

Sarkar, A., Concilio, S., Sessa, L., Marrafino, F. and Piotto, S., 2024. Advancements and novel approaches in modified AutoDock Vina algorithms for enhanced molecular docking. Results Chem., 7: 101319. https://doi.org/10.1016/j.rechem.2024.101319.

Trott, O. and Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem., 31(2): 455-461. https://doi.org/10.1002/jcc.21334.

Wróblewski, K., Zalewski, M., Kuriata, A. and Kmiecik, S., 2025. CABS-flex 3.0: an online tool for simulating protein structural flexibility and peptide modeling. Nucleic Acids Res., 53(W1): W95-W101. https://doi.org/10.1093/nar/gkaf412.