**BIODIVERSITAS**
Volume 25, Number 6, June 2024
Pages: 2705-2718

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d250641

# Production and diversity analysis of cellulases from *Anoxybacillus* genus

**RAHMAT EKO SANJAYA[1,2,4] , NI NYOMAN TRI PUSPANINGSIH[2,3,♥], ALI ROHMAN[2,3], HAKIKI RAHMASARI[2,3], ROSLI MD. ILLIAS[5,6], NAWEE JANTARIT[7,8], KAZUHITO FUJIYAMA[9], ANDRE PRATAMA[10], FATIHA KHAIRUNNISA[2,3]**

[1]Program of Mathematics and Natural Science, Faculty of Science and Technology, Universitas Airlangga. Jl. Dr. Ir. H. Soekarno, Surabaya 60115, East Java, Indonesia
[2]Proteomic Laboratory, University-CoE-Research Center for Bio-Molecule Engineering, Universitas Airlangga. Jl. Dr. Ir. H. Soekarno, Surabaya 60115, East Java, Indonesia. Tel.: +62-31-5922427, Fax.: +62-31-5936502, ♥email: ni-nyoman-t-p@fst.unair.ac.id
[3]Department of Chemistry, Faculty of Science and Technology, Universitas Airlangga. Jl. Dr. Ir. H. Soekarno, Surabaya 60115, East Java, Indonesia
[4]Department of Chemistry, Faculty of Mathematics and Natural Science, Universitas Lambung Mangkurat. Jl. A. Yani Km. 36.4, Banjarbaru 70714, South Kalimantan, Indonesia
[5]Department of Bioprocess and Polymer Engineering, Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia. 81310 Skudai, Johor, Malaysia
[6]Institute of Bioproduct Development, Universiti Teknologi Malaysia. 81310 Skudai, Johor, Malaysia
[7]Laboratory of Protein Crystallography, Institute for Protein Research, Osaka University. 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[8]Department of Macromolecular Science, Graduate School of Science, Osaka University. 1-1 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan
[9]International Center for Biotechnology, Osaka University. 2-1 Yamadaoka, Suita-Shi, Osaka 565-0871, Japan
[10]Department of Bacterial Infections, Research Institute for Microbial Diseases, Osaka University. 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Manuscript received: 17 February 2024. Revision accepted: 25 June 2024.

**Abstract.** *Sanjaya RE, Puspaningsih NNT, Rohman A, Rahmasari H, Illias RM, Jantarit N, Fujiyama K, Pratama A, Khairunnisa F. 2024. Production and diversity analysis of cellulases from* Anoxybacillus *genus. Biodiversitas 25: 2705-2718.* Bioconversion of cellulose into sustainable renewable resources of oligosaccharides and glucose requires cellulase action. Genomic sequences of *Anoxybacillus* cellulases are available in GenBank databases, but no information about their biochemical properties. *Anoxybacillus flavithermus* TP-01 is a thermophilic, facultatively anaerobic, cellulase-producing bacterium isolated from the mountain of Gunung Pancar hot springs, Bogor, Indonesia. This study aimed to characterize the cellulase produced by *Anoxybacillus flavithermus* TP-01, a local Indonesian isolate. This cellulase was characterized and the diversity analysis on biochemical properties and structural characteristics was done by in silico approaches. SOPMA, SWISS-MODEL, I-TASSER, ProSA, and SAVES were used for computational analysis on physicochemical characterization, phylogenetic construction, functional analysis, multiple sequence alignment, and secondary-tertiary structure prediction. Physicochemical analysis showed pI<7, indicating acidic property for these proteins. Furthermore, the proteins were thermostable and hydrophilic, as proved by their relatively high aliphatic index and negative GRAVY values. The five sequence motifs harbored a conserved domain of the M42 peptidase/endoglucanase family. Alpha helix was the predominant secondary structure, and the tertiary structure fulfilled the structural quality criteria of QMEAN4, ERRAT, Ramachandran plot, and Z-score. Structural comparison between the template structure and the models revealed significant differences in the loop sections. This study identified a potentially thermostable cellulase of *Anoxybacillus flavithermus* TP-01, and provided new insights of the physicochemical, functional, and explores the structure-function relationship of cellulase from *Anoxybacillus* genus based on in silico data.

**Keywords:** *Anoxybacillus* sp., cellulase, in silico, renewable resource, thermophilic

**Abbreviations**: Mol. Wt.: Molecular Weight; GRAVY: Grand average of hydropathicity; pI: Isoelectric point; *II*: Instability index; AI: Aliphatic index; –R: Number of negative residues (aspartic acid & glutamic acid); +R: Number of positive residues (arginine & lysine); MEME: multiple EM for motif elicitation; SOPMA: Self Optimized Prediction Method with Alignment; CDD: Conserved Domain Database; QMEAN: Qualitative Model Energy Analysis; MEGA: Molecular Evolutionary Genetics Analysis; NCBI: National Center for Biotechnology Information; PDB: Protein Data Bank; I-TASSER: Iterative Threading ASSEmbly Refinement

## INTRODUCTION

Cellulose is the most abundant organic polymer found in nature. The annual amount of pure cellulose generated from photosynthesis is estimated to be $10^{11}$-$10^{12}$ tonnes of total biomass production. Cellulase as a hydrolytic enzyme, plays the central role in the bioconversion of cellulose into renewable products (Loow et al. 2017). Enzymatic hydrolysis of cellulose breaks β (1 → 4) glycosidic bonds and results in the formation of oligosaccharides and glucose. Demand for cellulases has increased since hazardous chemicals in the industrial sector are being actively replaced by green industrial approaches (Ejaz et al. 2021). Cellulases are commonly used in the feedstock, textile, wine and beer, paper and pulp, cotton, and detergent industries (Juturu and Wu 2014; Jaramillo et al. 2015; Tiwari et al. 2018). Cellulase can also be used to convert cellulose into sugar, which can be further used for producing bioethanol (Patel et al. 2019). Owing to its wide applications, cellulase is the third most widely used for industrial enzyme in the world (Bajaj and Mahajan 2019).

Cellulase is naturally produced by fungi, actinomycetes, and bacteria (Khoshnevisan et al. 2019). Fungi tend to produce more cellulases than bacteria; however, cellulases from bacteria exhibit better catalytic activity. Even in extreme environments, such as high temperatures or extreme pH, bacteria can produce enzymes that are stable under these conditions (Merino et al. 2019). These microorganisms are found on plants, soil, and water surfaces. *Acidothermus cellulolyticus, Pectobacterium chrysanthami, Thermobifida fusca, Thermotoga maritima, Paenibacillus* sp., *Bacillus* sp., and *Aeromonas* sp. are examples of cellulolytic producing bacteria (Islam and Roy 2018).

Cellulases from thermophilic bacteria are useful in various industries. One of the advantages offered by enzyme biotechnology is enabling the reduction of protein contamination. *Anoxybacillus* sp. is a facultative anaerobic bacterium inhabiting high-temperature environments, such as hot springs. *Anoxybacillus flavithermus* appears dark yellow because of the accumulation of carotenoid pigments in its cell membranes. *Anoxybacillus* sp. forms dark yellow colonies and was first discovered in hot springs in New Zealand. These bacteria are non-motile, Gram-positive, straight, and rod-shaped with round ends. *Anoxybacillus* spp. produce inducible enzymes, such as cellulase, with thermophilic properties (Genc et al. 2015; Hussain et al. 2017). The amino acid sequences of cellulases can be used to determine their structure, properties, and characteristics. Identifying properties and characteristics of cellulases may be the preliminary step for further developing properties and features, as well as understanding the diversity of cellulases produced by *Anoxybacillus* sp. Software and internet tools are available for phylogenetic analysis of cellulases from *Anoxybacillus* sp. as well as determining overall of physicochemical characteristics; primary, secondary, and tertiary structures; and conserved motifs.

Gunung Pancar hot springs in Bogor, West Java, Indonesia, is a prime location for exploring potential cellulase-producing thermophilic bacteria. Apart from having high temperatures, these hot springs are surrounded by biomass containing lignocellulose, making it a potential source for exploring bacteria that produce lignocellulose-degrading enzymes, especially cellulase. In a recent study, four isolates, P1, P2, P3, and P4 were identified from the Gunung Pancar hot springs. These four isolates exhibited various types of hydrolytic activities including pectinolytic, amylolytic, xylanolytic, and cellulolytic activities. The P2 isolate showed potential cellulolytic activity compared to the other isolates. Biochemical and molecular identification of P2 isolate using 16S rDNA sequences revealed it to be *Anoxybacillus flavithermus* (Laras et al. 2017), hereinafter referred to as *Anoxybacillus flavithermus* TP-01.

Cellulases from *Anoxybacillus* have not been extensively studied; therefore, it is important to study their diversity based on in silico approaches. This study aimed to characterize the cellulase produced by *Anoxybacillus flavithermus* TP-01, a local Indonesian isolate. In addition, this study used in silico tools to characterize the physicochemical, functional, and structural properties of cellulases isolated from members of *Anoxybacillus* genus. Since biochemical characterization data remains unavailable for cellulases from *Anoxybacillus*, the in silico characterization data presented here add valuable insights to the available scientific literature.

## MATERIALS AND METHODS

### Isolation and screening of cellulolytic activity

*Anoxybacillus flavithermus* TP-01 a local Indonesian isolate from Gunung Pancar hot spring, Bogor, West Java, Indonesia, was obtained from The Proteomic Laboratory University-CoE Research Center for Bio-Molecule Engineering Universitas Airlangga. It was grown on agar plates containing Carboxymethyl Cellulose (CMC) and incubated at 60°C for 48 h. Following incubation, the plate was cooled to room temperature and then dripped with 0.1% Congo red solution. The cellulolytic halo test was used to characterize cellulase activity, which showed a clear area around the colony, whereas the other areas retained a red or orange color. The cellulolytic index was obtained as follows:

$$\text{Cellulolytic Index} = \frac{\text{Diameter of halo zone} - \text{Diameter of bacterial colony}}{\text{Diameter of bacterial colony}}$$

### Optimization of cellulase production

The cellulase production time for *Anoxybacillus flavithermus* TP-01 was optimized by determining cellulase activity and bacterial growth by collecting samples from active culture from 0 to 40 h at 2-4 h intervals.

Culture duration leading to the highest cellulase activity was considered optimum production time. Briefly, 1% inoculum was added to the media and incubated at 60°C at 150 rpm. Culture samples were retrieved at predetermined times for determining enzyme activity and optical density ($OD_{600}$). The samples were centrifuged at 7000 rpm for 10 minutes, and the supernatant was collected and treated with crude cellulase extract to test cellulase activity. Cellulase activity was determined using the dinitro salicylic method that measures the amount of reducing sugars in the reaction mixture (Miller 1959). Briefly, 100 µL crude extract was mixed with 100 µL CMC substrate in 0.1 M phosphate citrate (pH 7), and incubated at 70°C for 30 minutes. Next, 600 µL of 3,5-dinitrosalicylic acid was added to the samples, and the mixture was first heated for 15 minutes and then cooled for 20 minutes. The absorbance was measured at 550 nm using a UV-Vis spectrophotometer. The control was prepared using the same process without the incubation step. One unit activity of cellulase (U) was defined as the amount of enzyme required to produce 1 µmol glucose per minute. The optical density of the samples was determined using a UV-Vis spectrophotometer at 600 nm, and sterile media without isolates was used as a control.

### Ammonium sulfate precipitation

Ammonium sulfate precipitation was a widely applied method for protein purification and fractionation and separates proteins based on their solubility in presence of a high salt concentration. Ammonium sulfate concentration was optimized by precipitating the cellulase crude extract

in a solution containing 30%-90% ammonium sulfate and homogenized. The mixture of crude extract and ammonium sulfate was incubated in a refrigerator for approximately one day to optimize precipitation and then centrifuged at 7000 rpm 5 minutes at 4°C. The precipitate was dissolved in 0.1 M phosphate citrate (pH 7) and cellulase activity was tested. The percentage of ammonium sulfate with the highest cellulase activity was selected for further experiments. The same experiment was repeated using 1000 mL crude cellulase extract saturated with optimized concentration of ammonium sulfate.

In addition to determining cellulase activity, the precipitate crude extract was analyzed for determining protein concentration using Bradford method (Bradford 1976). Briefly, 20 µL sample was mixed with 1 mL Bradford solution and the absorbance was measured at 595 nm. The concentrations were determined using the standard curve generated from the absorbance values of bovine serum albumin solutions prepared at different concentrations. A solution of 0.1 M phosphate citrate, pH 7 was used as a control.

**In silico sequence and physiochemical property analysis**

The sequences of cellulase amino acid obtained from *Anoxybacillus* sp. was obtained from the GenBank (http://www.ncbi.nlm.nih.gov/). Eleven full sequences were saved in the FASTA format. ExPASy-ProtParam tools (Duvaud et al. 2021) (http://web.expasy.org/protparam/) were used to analyze the physicochemical properties of the selected sequences, including molecular weight (Mol. Wt.), theoretical isoelectric point (pI), Instability Index (*II*), Aliphatic Index (AI), Grand Average Of Hydropathicity (GRAVY), negative (–R) and positive (+R) charges, Extinction Coefficient (EC), and amino acid composition. Motifs or consensus sequence was identified using the Multiple EM for Motif Elicitation (MEME) server (https://meme-suite.org/meme/tools/meme) (Bailey et al. 2015). The maximum number of motifs was set to five, with a motif width ranging from 6-50 amino acids, along with other default values.

**Multiple sequence alignment and phylogenetic analysis**

The full-length amino acid sequences of the cellulase were aligned using ClustalW using MEGA 11 (Kumar et al. 2018) and visualized using the ESPript 3.0 server (Robert and Gouet 2014) (https://espript.ibcp.fr/ESPript/ESPript/). A cladogram of cellulase sequences from *Anoxybacillus* sp. was constructed using the Maximum Likelihood method with the LG+G4 model (Trifinopoulos et al. 2016; Minh et al. 2020) using MEGA 11 (Kumar et al. 2018).

**Secondary and tertiary structure prediction**

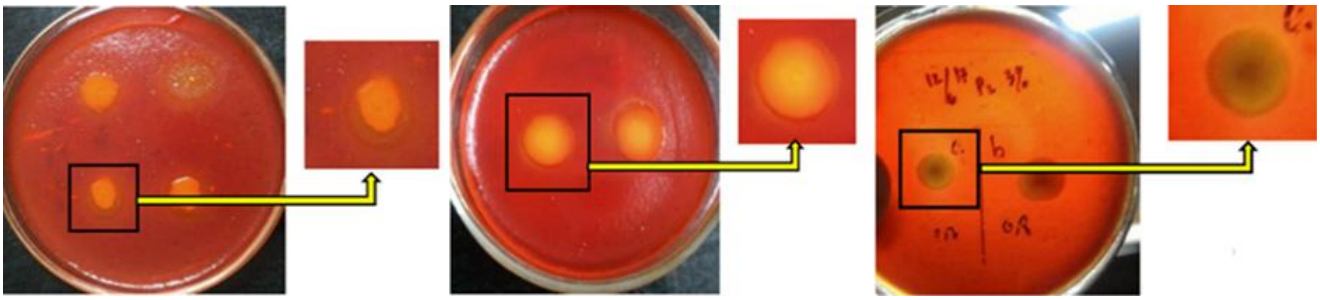
Self-Optimized Prediction Method with Alignment (SOPMA) was used to calculate the secondary structure-related features of the selected protein sequences (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). The in silico analysis of secondary structures included identification of alpha helices, extended strands, beta turns, and random coils. All sequences used in this study were searched for a template three-dimensional structure using comparative homology modeling with SWISS-MODEL (https://swissmodel.expasy.org/) (Biasini et al. 2014; Bienert et al. 2017; Waterhouse et al. 2018). Sequences in clusters A, C, D, and E had a protein structure template PDB ID of 1vhe.1. A, with Global Model Quality Estimate (GMQE) values ranging from 0.89 to 0.96, and sequence identity ranging between 50% and 77.72%. However, the sequences in cluster B had very low GMQE and sequence identity values, so the three-dimensional structure of one of the three sequences from cluster B (ANB65284.1) was constructed using the I-TASSER server (https://zhanggroup.org/I-TASSER/) (Yang et al. 2014; Yang and Zhang 2015; Zheng et al. 2021; Zhou et al. 2022).

The quality of the structural models was assessed using the PROCHECK and ERRAT servers (https://saves.mbi.ucla.edu/). PROCHECK was used to assess the stereochemical quality of protein structures by analyzing the residue-by-residue and overall structural geometries, and the results were presented as a Ramachandran plot. ERRAT values were used to assess the resolution of protein structures, with values approximately or higher than 95% representing high-resolution structures and values approximately 91% representing lower resolutions (2.5 to 3) (Dutta et al. 2018). ProSA-web was used to assess the Z-score and energy plots (https://prosa.services.came.sbg.ac.at/prosa.php). The desirable Z-score should be <1 compared to non-redundant set of PDB structures. QMEAN4 was used to fit cumulative QMEAN values on a global scale (range: 0 to 1) (Pramanik et al. 2017). The models were visualized using PyMOL (PyMOL Molecular Graphics System, Version 1.8; Schrödinger, LLC).

## RESULTS AND DISCUSSION

**Isolation and screening of cellulolytic activity**

The results showed that *Anoxybacillus flavithermus* TP-01 exhibited maximum cellulase activity with the largest halo index of 0.2222 when grown in the presence of 2% CMC compared to halo indices of 0.0833 and 0.1428 when grown in presence of 1% and 3% CMC, respectively (Figure 1). *Anoxybacillus flavithermus* TP-01 was spotted on a medium plate containing CMC (1%, 2%, and 3%) substrate, a polymer that induces bacteria to produce cellulase (Genc et al. 2015). Cellulase activity was indicated by a clear area around the isolate after adding Congo red dye. The clear zone formed due to Congo red strongly binding polysaccharides containing β-(1-4)-D-glucopyranosyl bonds and indicated the presence of cellulase, which breaks the β-1,4-glycosidic bond (Sazci et al. 1986).



**A** **B** **C**

**Figure 1.** Cellulase halo or clear zone *Anoxybacillus flavithermus* TP-01 colony. Agar plat containing: A. 1% CMC; B. 2% CMC; C. 3% CMC

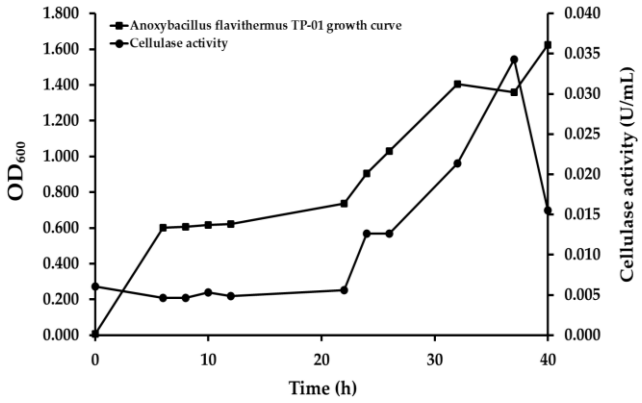



**Figure 2.** The growth curve and cellulase expression of *Anoxybacillus flavithermus* TP-01

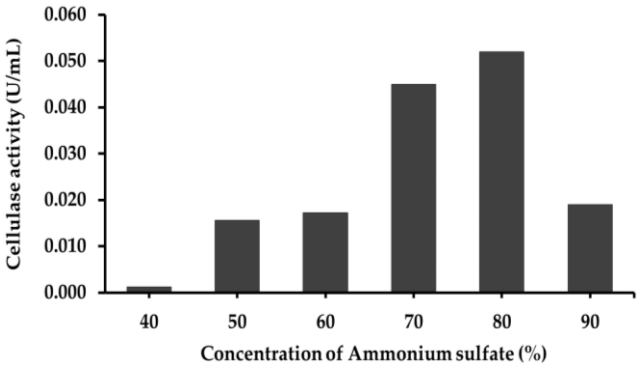



**Figure 3.** Optimization of maximum cellulase activity following ammonium sulfate precipitation

## Optimization of cellulase production

*Anoxybacillus flavithermus* TP-01 produced cellulase by 37 hours, with the highest glucose concentration of 0.103 µmol and activity of 0.034 U/mL (Figure 2). The cellulase production decreased after 40 hours of growth. This result is in accordance with the findings of Ibrahim and El-diwany (2007) who reported cellulase production at the end of the stationary phase at 36 hours. The optimum conditions for cellulase production are achieved after 24 hours induction, and decreased after 48 hours (Hussain et al. 2017). The decrease in enzymatic activity with increasing incubation duration can be attributed to nutrient depletion and the production of by-products in the medium, which causes stress to growing bacteria and inactivates enzyme production. In addition to being an inducer, CMC also acts as a carbon source, as LB media is nitrogen-rich and has minimal carbon sources.

## Ammonium sulfate precipitation

Ammonium sulfate precipitation is a popular method for large- and laboratory-scale purification and fractionation of proteins. This method is used to separate proteins based on their differential solubility in varying salt concentrations. Ammonium sulfate attracts water molecules from the protein surface to induce their aggregation and precipitation (Wingfield 2016). The two concepts involved in ammonium sulfate precipitation are *salting in* and *salting out*; *salting in* means that a low concentration of salt ions will still protect the protein molecules, so that the protein is still dissolved, whereas *salting out* means that high concentrations of salt ions will increase the electric charge around the protein, thereby decreasing protein solubility (Duong-Ly and Gabelli 2014; Wingfield 2016).

Optimum cellulase activity of 0.052 U $mL^{-1}$ was achieved when precipitated using 80% ammonium sulfate (Figure 3). Generally, cellulases show optimum activity when precipitated using 60%-80% ammonium sulfate (Genc et al. 2015; Islam and Roy 2018; Pachauri et al. 2020). The precipitation should be performed at 4°C to maintain enzyme activity and prevent protein denaturation. While stirring, the ammonium sulfate was sprinkled gradually until it reached 80% concentration. The mixture of crude extract and ammonium sulfate was incubated in a refrigerator for approximately one day to optimize protein precipitation. These pellets obtained after centrifuging the precipitated solution were dissolved in 0.1 M phosphate citrate pH 7. Table 1 shows that the specific activity of cellulase increased by 1.40-fold after its partial purification using 80% ammonium sulfate. Earlier, cellulase from *Trichoderma longibrachiatum* showed a 1.351-fold increase in specific activity after purification with ammonium sulfate (Pachauri et al. 2020).

**In silico sequence and physiochemical property analysis**

The amino acid sequences of cellulases from the genus *Anoxybacillus* were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/) and saved in FASTA format. Eleven full amino acid sequences were found in GenBank; two sequences each from *Anoxybacillus flavithermus* AK1, *Anoxybacillus gonensis, Anoxybacillus flavithermus* NBRC 109594, and *Anoxybacillus ayderensis*, and one sequence each from *Anoxybacillus rupiensis, Anoxybacillus* sp. P3H1B, and *Anoxybacillus* sp. B7M1 (Table 2). *Anoxybacillus* sp. is a thermophilic microorganism inhabiting hot springs and geothermal areas. The results showed that cellulase-producing *Anoxybacillus* shows more potential for exploration than other genera because of the limited number of sequences in the GenBank database. Primary, secondary, and tertiary structures; physicochemical properties; alignment; phylogeny; and motifs were determined for retrieved cellulase sequences using various online tools and servers.

The physicochemical properties of proteins provide preliminary information for determining their uniqueness. The results of analysis of pI, *II*, AI, GRAVY, number of positive (+R) and negative (-R) residues, and Extinction Coefficient (EC) for each cellulase sequence using the ExPASy-ProtParam tool and the results are shown in Table 3. The pI of the cellulases from the genus *Anoxybacillus* were less than 7, indicating their acidic nature. The pI value corresponds to the pH at which the mobility of a protein becomes zero, with a more compact and stable conformation (Pergande and Cologna 2017). pI estimation is useful in protein crystallization trials as well as in analytical biochemistry and proteomics techniques, such as two-dimensional polyacrylamide gel electrophoresis or capillary isoelectric focusing, which utilize pI information based on the amino acid sequence.

The cellulases from *Anoxybacillus* sp. were stable with an instability index of less than 40. The instability index (*II*) assesses the stability of a protein based on its amino acid sequence; values less than 40 indicate stability, whereas values greater than 40 indicate instability. A high Aliphatic Index (AI), which measures the volume occupied by aliphatic side chains, such as isoleucine, leucine, alanine, and valine, indicates thermostability of proteins. This can positively enhance the thermostability of globular proteins. The amino acids cellulase sequences (accession number EMI10668.1) had the highest AI value among the analyzed sequences, indicating that it was the most thermostable. The positive (+R) and negative charges (–R) of the sequences were determined by counting the number of arginine, lysine residues and aspartic acid, glutamic acid residues, respectively. The extinction coefficient (EC) of the sequences was computed using the ExPASy-ProtParam tool and ranged from 28670 $M^{-1}s^{-1}$ to 101760 $M^{-1}s^{-1}$.

**Table 1.** Purification of cellulase-producing *Anoxybacillus flavithermus* TP-01

| Purification steps | Total activity (U) | Total protein (mg) | Enzyme activity (U $mL^{-1}$) | Specific activity (U $mg^{-1}$) | Yields (%) | Purification (fold) |
|---|---|---|---|---|---|---|
| Crude enzyme | 28.985 | 951.539 | 0.032 | 0.030 | 100.00 | 1.00 |
| Ammonium sulfate fractional precipitation (80%) | 13.316 | 319.968 | 0.034 | 0.042 | 45.94 | 1.40 |

**Table 2.** Protein sequences of selected cellulase from *Anoxybacillus* sp.

| Source organisms | Protein sequence accession number | Number of sequences |
|---|---|---|
| *Anoxybacillus flavithermus* AK1 | EMT45536.1, EMT45196.1 | 2 |
| *Anoxybacillus gonensis* | EMI10759.1, EMI10668.1 | 2 |
| *Anoxybacillus flavithermus* NBRC 109594 | GAC90965.1, GAC92240.1 | 2 |
| *Anoxybacillus ayderensis* | EPZ39606.1, EPZ37744.1 | 2 |
| *Anoxybacillus rupiensis* | WP_221211947.1 | 1 |
| *Anoxybacillus* sp. P3H1B | WP_066147649.1 | 1 |
| *Anoxybacillus* sp. B7M1 | ANB65284.1 | 1 |

**Table 3.** Physicochemical properties of cellulase from *Anoxybacillus* sp. analyzed using ExPASy-ProtParam tool

| Accession numbers | Numbers of amino acids | Mol. Wt. (Dalton) | pI | *II* | AI | GRAVY | –R (Asp + Glu) | +R (Arg + Lys) | EC ($M^{-1}cm^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|
| EMT45536.1 | 361 | 39544.68 | 5.58 | 26.49 | 95.57 | –0.074 | 52 | 43 | 33920 |
| EMT45196.1 | 355 | 38980.84 | 6.05 | 26.93 | 89.58 | –0.139 | 43 | 36 | 28670 |
| EMI10759.1 | 355 | 39199.12 | 5.61 | 28.72 | 92.87 | –0.136 | 46 | 37 | 28670 |
| EMI10668.1 | 361 | 39634.89 | 5.75 | 25.65 | 96.40 | –0.072 | 51 | 43 | 33920 |
| GAC90965.1 | 355 | 38993.78 | 5.50 | 29.81 | 90.42 | –0.150 | 45 | 35 | 28670 |
| GAC92240.1 | 364 | 39913.14 | 5.75 | 24.91 | 96.10 | –0.073 | 52 | 44 | 35410 |
| EPZ39606.1 | 361 | 39657.79 | 5.58 | 27.99 | 95.01 | –0.097 | 52 | 43 | 33920 |
| EPZ37744.1 | 355 | 39280.20 | 5.76 | 29.20 | 90.96 | –0.165 | 45 | 38 | 28670 |
| WP_221211947.1 | 492 | 57152.51 | 5.91 | 27.41 | 77.91 | –0.478 | 67 | 52 | 101760 |
| WP_066147649.1 | 492 | 57166.53 | 5.91 | 27.56 | 77.91 | –0.478 | 67 | 52 | 101760 |
| ANB65284.1 | 492 | 57117.33 | 5.81 | 28.54 | 76.52 | –0.490 | 68 | 50 | 101760 |

**Table 4.** The best matched amino acid sequences of five motifs with respective conserved domains observed among 11 sequences of cellulases from *Anoxybacillus* sp.

| Motifs | Width | Best possible amino acid sequence | Conserved domain |
|---|---|---|---|
| 1 | 21 | GHPBVVYYDATVIEEKGLREF | M42 peptidase/endoglucanase |
| 2 | 25 | PYAEFTPMNNEKKLLAKAWDNRIGC | M42 peptidase/endoglucanase |
| 3 | 21 | ZPBJAYRVDISPAGGTDAGVI | M42 peptidase/endoglucanase |
| 4 | 21 | IGIPARYIHTHAAILHKDDYE | M42 peptidase/endoglucanase |
| 5 | 21 | SKPPHJLPEEARRKPVEIKDM | M42 peptidase/endoglucanase |

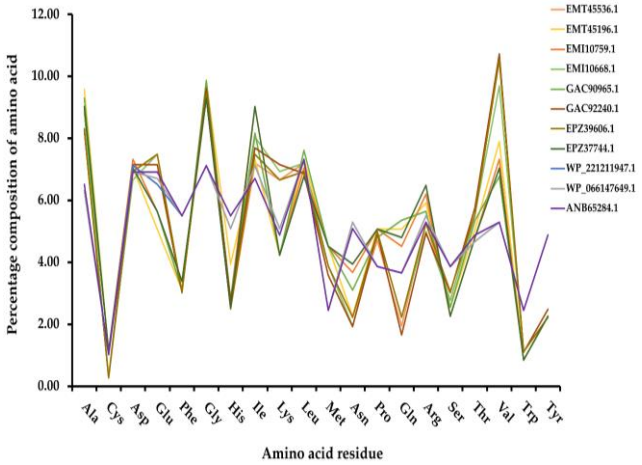



**Figure 4.** Graphical representation of amino acid composition of cellulase from *Anoxybacillus* sp. using ExPASy-ProtParam tool.

All cellulase sequences obtained from *Anoxybacillus* sp. exhibited negative GRAVY values, indicating good water-protein interactions. Some of the sequences (accession numbers EMT45536.1, EMI10668.1, GAC92240.1, and EPZ39606.1) had GRAVY values close to zero. Results showed overall similar cellulase amino acid composition from *Anoxybacillus* spp. (Figure 4). The most abundant amino acid residues were alanine, glycine, valine, aspartate, leucine, and isoleucine. This amino acid residue profile was similar to that of cellulases from uncultured microorganisms, which also contain high concentrations of alanine, glycine, and valine (Sanjaya et al. 2021). As alanine and glycine tend to form alpha helix secondary structures, cellulases from *Anoxybacillus* sp. have a dominant alpha helix structure.

Table 4 and Figure 5 presents the five conserved motifs identified through MEME analysis. These motifs can serve as targets for enzyme engineering and provide insights into the functional identities of enzymes. Sequence motifs often represent the catalytic sites, binding sites, and structural motifs of proteins. Although a protein may not be identical to any known protein, the presence of sequence motifs can provide crucial clues regarding its function. The results of the BLAST analysis of the motif sequences revealed that the five motifs had a conserved domain of M42 peptidase/endoglucanase. The M42 peptidase/endoglucanase family of enzymes exhibits dual functionality as peptidases and cellulase/endoglucanases. There are two different perspectives regarding to M42 peptidase/endoglucanase family; (1) according to the MEROPS database (http://merops.sanger.ac.uk/), a subfamily of non-peptidases are homolog with CelM, an endoglucanase of *Clostridium thermocellum*; (2) according to the NCBI Conserved Domain Database (CDD) defines that the M42 aminopeptidase family as one consisting of aminopeptidases, endoglucanases, and proteins of the Frv operon, which are involved in the biosynthesis and degradation of polysaccharides (Sharma et al. 2019). Although CelM has previously been assigned to this family, several of its homologs have recently been identified as aminopeptidases (Ma et al. 2020).

**Multiple sequence alignment and phylogenetic analysis**

Eleven cellulase sequences from *Anoxybacillus* spp. were subjected to multiple sequence alignment using ClustalW. The alignment revealed the presence of conserved residues, which were identical or similar residues (Figure 6). This analysis was important as conserved residues serve as markers of important residues regarding cellulase structure and function. Changes in conserved residues can result in variations in protein structure and function. Alignment analysis identified glycine as the most conserved residue, whereas proline, leucine, tryptophan, asparagine, valine, and tyrosine were the least commonly observed residues, with only one appearing as a conserved or identical residue. The gap observed in the alignment was attributed to residues in the sequences WP_221211947.1, WP_066147649.1, and ANB65284.1. If these residues were excluded from analysis, the other sequences would have a higher number of identical residues. Notably, eight sequences were from the same species but arranged differently in the phylogenetic tree. For instance, cellulases from *Anoxybacillus ayderensis* (EPZ39606.1) were classified in cluster E, whereas those from *Anoxybacillus ayderensis* (EPZ37744.1) were classified in cluster A. Similarly, cellulases from *Anoxybacillus gonensis* and *Anoxybacillus flavithermus* were found in clusters C and D, respectively. Interestingly, *Anoxybacillus flavithermus* AK1 also had its cellulases in cluster A, suggesting that cellulases from the same species and strain can exhibit different levels of evolution.

The cladogram in Figure 7 shows the results of phylogenetic analysis, which grouped the cellulase sequences into five clusters based on their ancestors. Cluster A contained the most members compared to the other clusters, with four sequences, two of which were from the same species but different strains. Clusters C and D contained only one sequence each derived from *Anoxybacillus gonensis* and *Anoxybacillus flavithermus* AK1. Notably, *Anoxybacillus flavithermus* AK1 also contained cellulases in cluster A, indicating that cellulases from the same species and strain

may have different evolutionary levels. Similarly, cellulases from *Anoxybacillus ayderensis* were observed in clusters A and B, whereas sequences with accession numbers WP_221211947.1, WP_066147649.1, and ANB65284.1 were present only in cluster B, confirming their unique nature compared to the other cellulase sequences investigated in this study.

**Secondary and tertiary structure prediction**

Secondary structure prediction revealed that alpha helix was the predominant secondary structure in all sequences (Figure 8). As detailed earlier, this can be attributed to the abundance of alanine and glycine, which have a tendency to form alpha helix secondary structures (Sanjaya et al. 2021). Random coil secondary structures were the second-most common structures in all sequences. These secondary structures are more flexible and dynamic than other secondary conformational structures (Dutta et al. 2021). The least common secondary structure was the beta turn. All sequences exhibited similar profiles, with the alpha helix being the most common structure, followed by random coil, extended strand, and beta turn in cellulases from *Anoxybacillus* sp..

The amino acid query was inputted into the SWISS-MODEL server and modeling predictions were evaluated using computational tools. Table 5 shows the template, GMQE score, and identity of each sequence in each cluster. Except for the sequence in cluster B, all sequences in each cluster had the best template of 1vhe.1. A, with GMQE values ranging from 0.89 to 0.96, and sequence identity between 50.00% and 77.72%. The GMQE estimates the overall quality of the model and considers properties from both the target-template alignment and template structure. Sequences with accession numbers GAC90965.1 (cluster A), EMI10668.1 (cluster C), EMT45536.1 (cluster D), and EPZ39606.1 (cluster E) were selected for SWISS-MODEL homology modeling using template 1vhe.1. A.

**Table 5.** Cellulase structure modeling using SWISS-MODEL

| Clusters | Accession numbers | Best template | Residues | GMQE | Sequence identity (%) |
|---|---|---|---|---|---|
| A | EMI10759.1 | 1vhe.1.A | 2-353 | 0.89 | 50.28 |
| | EPZ37744.1 | 1vhe.1.A | 2-353 | 0.89 | 50.00 |
| | EMT45196.1 | 1vhe.1.A | 2-353 | 0.89 | 50.57 |
| | GAC90965.1 | 1vhe.1.A | 2-353 | 0.89 | 51.42 |
| B | ANB65284.1 | 1cec.1.A | 19-371 | 0.39 | 27.09 |
| | WP_221211947.1 | 1cec.1.A | 19-371 | 0.39 | 27.09 |
| | WP_066147649.1 | 1cec.1.A | 19-371 | 0.39 | 27.09 |
| C | EMI10668.1 | 1vhe.1.A | 2-360 | 0.96 | 77.72 |
| D | EMT45536.1 | 1vhe.1.A | 2-360 | 0.96 | 77.72 |
| E | GAC92240.1 | 1vhe.1.A | 5-363 | 0.95 | 77.72 |
| | EPZ39606.1 | 1vhe.1.A | 2-360 | 0.96 | 77.72 |

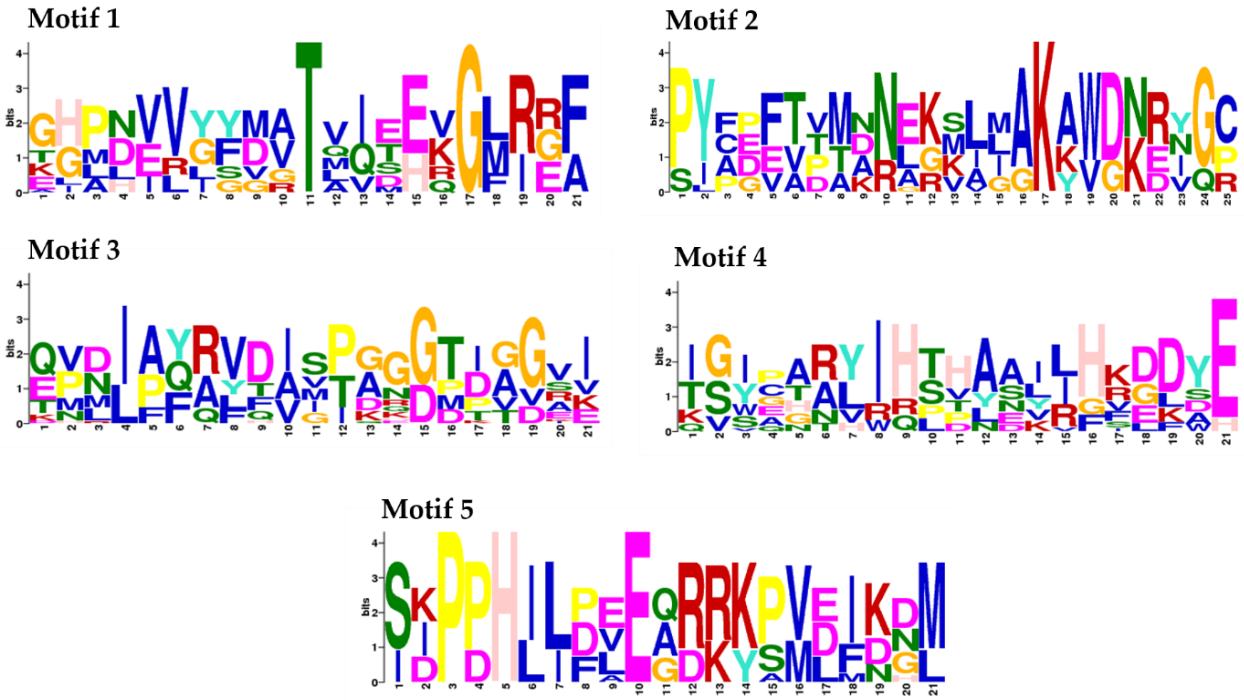



**Figure 5.** Logo of conserved sequence motifs of 11 sequences of cellulases from *Anoxybacillus* sp. elicited by MEME

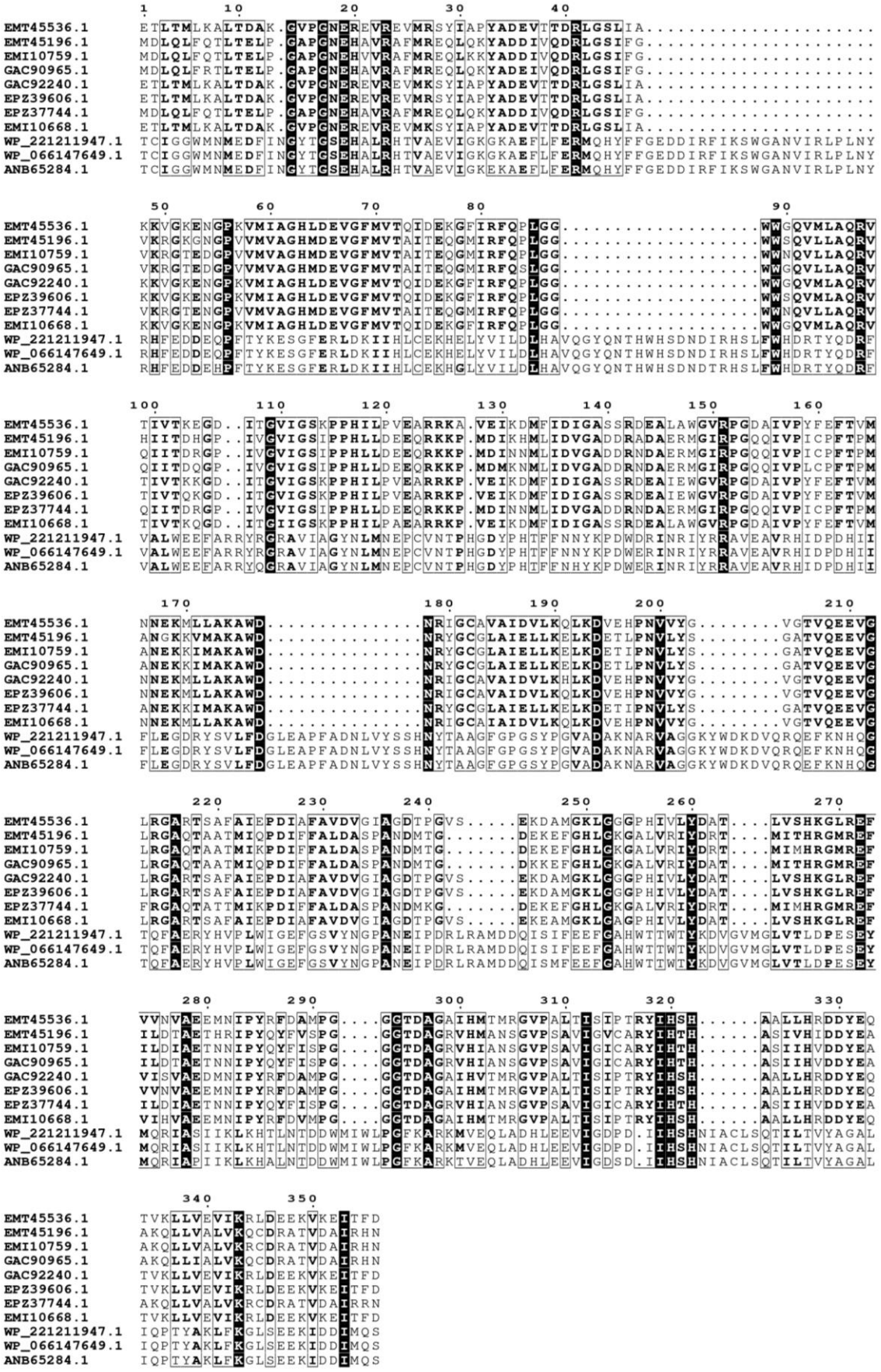



**Figure 6.** Multiple sequence alignment of cellulase from *Anoxybacillus* sp. Black bold characters and boxed showed similar residues, white bold characters boxed in black indicated identical residues

Quality parameters, including QMEAN4, ERRAT, Ramachandran plot, and Z-score, were used to evaluate the tertiary model structures generated using the SWISS-MODEL. Table 6 lists the quality parameter values of the resulting models. The generated models were of good quality, as evidenced by the residue in the favored region with value of over 90%. The ERRAT value was used to assess the resolution of the models, with higher three-dimensional structural resolutions yielding values of approximately 95%. The average overall quality factor for the cellulase structure models was approximately 91%, with all models from the SWISS-MODEL displaying ERRAT values between 94.97% and 95.61%, indicating a sufficiently high structural resolution. Data anomalies were observed during the analysis of the QMEAN4 values against the structural model from the GAC90965.1 sequence. A good QMEAN score for a high-quality protein structure ranges between 0 and 1 (Santhoshkumar and Yusuf 2020). However, the model structure of GAC90965.1 had a QMEAN score of -0.01, which was not desirable. Nonetheless, other parameters indicated that the cellulase structure model of the selected sequences had overall good quality.

For sequences in cluster B with GMQE and identity values below 50%, three-dimensional structure was predicted using Iterative Threading ASSEmbly Refinement (I-TASSER). I-TASSER is a web-based tool that uses multiple threading alignment approach to recognize structural templates from the Protein Data Bank (PDB) (Yang and Zhang 2015). Full-length structural models were constructed by performing iterative fragment assembly simulations. Finally, functional insights were obtained by matching the predicted structural models with known proteins from functional databases.

The sequence with accession number ANB65284.1 (cluster B) using I-TASSER was modeled. This sequence was chosen based on the phylogenetic tree, where it was closer to the direct node than the sequences WP_221211947.1 and WP_066147649.1. Modeling of ANB65284.1 using I-TASSER produced five models, which were evaluated using parameters such as QMEAN4 score, ERRAT value, Ramachandran plot, and Z-score to assess their quality. Table 7 presents the QMEAN4, ERRAT, and Ramachandran plots for the five models and the results indicated that the quality of the models did not meet the desired criteria. The ideal QMEAN4 value should be between 0-1, whereas the models should have a value between -16.38 and -13.91. The Z-score of each model was below 0, indicating that the resulting models were far from the experimental protein structure range (Santhoshkumar and Yusuf 2020). The Z-score was used to validate the predicted three-dimensional structure of the protein; if the Z-score of the model was within the experimental protein structure range, it was considered accurate. Protein Structure Analysis (ProSA) tool was used to calculate the Z-score and assess the accuracy of the model structures. Experimental protein structures based on X-ray crystallography or NMR spectroscopy magnetic resonance spectroscopy were used for the analysis.

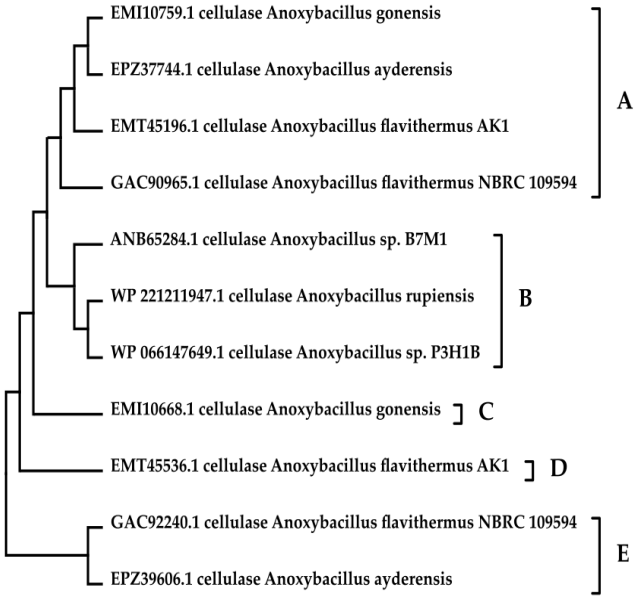



**Figure 7.** Cladogram of cellulase sequences from *Anoxybacillus* sp. using the maximum likelihood method based on LG+G4 model using MEGA 11 with 1000 bootstrap replicates. A-E: Clusters formed on the construction of cladogram

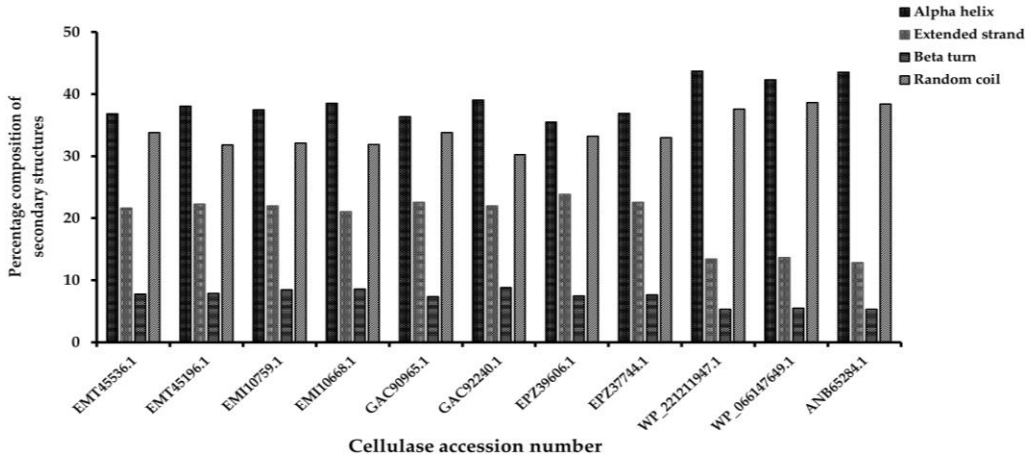



**Figure 8.** Percentage of secondary structures prevalent in cellulase from *Anoxybacillus* sp. as determined using SOPMA

A homology model of the cellulase structure generated using SWISS-MODEL that met the requirements of a high-quality model, is shown in Figure 9. The figure also presents a Ramachandran plot that displays the distribution of the amino acid φ/ψ angles. The structural models were compared with 1vhe.1 template, which was the crystal structure of the aminopeptidase/glucanase homolog, obtained by X-ray diffraction. A comparison between the 1vhe.1 template structure and the four models revealed significant differences in loop structure. Specifically, the model structure of GAC90965.1 had a longer and different loop than the other model structures and template (Figure 10). According to several studies, extended loops affect thermal stability (Zheng et al. 2018). However, extended loops also promote additional intramolecular contacts, which may be associated with higher structural stability. From the results it was suggested that cellulase with the accession number GAC90965.1 had more stable properties than its homologs. Further experimental research is needed to validate the differences in the loop structure resulting from GAC90965.1.

**Table 6.** Comparison of QMEAN4, ERRAT, Ramachandran plot, and Z-score for the quality assessment of selected three-dimensional cellulase structures model from Clusters A, C, D, and E using SWISS-MODEL

| Accession numbers | QMEAN4 score | ERRAT quality factor (%) | Ramachandran plot | | | | Z-score |
|---|---|---|---|---|---|---|---|
| | | | Residues in favored region (%) | Residues in additional allowed region (%) | Residues in generously allowed region (%) | Residues in disallowed region (%) | |
| GAC90965.1 | -0.01 | 94.97 | 92.6 | 7.0 | 0.0 | 0.3 | -8.79 |
| EMI10668.1 | 0.32 | 95.32 | 93.8 | 5.9 | 0.0 | 0.3 | -9.64 |
| EMT45536.1 | 0.18 | 95.31 | 94.1 | 5.6 | 0.0 | 0.3 | -9.2 |
| EPZ39606.1 | 0.27 | 95.61 | 94.1 | 5.6 | 0.0 | 0.3 | -9.55 |

**Table 7.** Comparison of QMEAN4, ERRAT, Ramachandran plot, and Z score for the quality assessment of top five final three-dimensional ANB65284.1 (Cluster B) structures using I-TASSER program

| ANB65284.1 | QMEAN4 score | ERRAT quality factor (%) | Ramachandran plot | | | | Z-score |
|---|---|---|---|---|---|---|---|
| | | | Residues in favored region (%) | Residues in additional allowed region (%) | Residues in generously allowed region (%) | Residues in disallowed region (%) | |
| Model 1 | –15.18 | 62.40 | 62.2 | 28.2 | 5.7 | 3.9 | –5.09 |
| Model 2 | –15.08 | 70.87 | 57.8 | 31.7 | 8.3 | 2.3 | –7.4 |
| Model 3 | –16.38 | 60.12 | 60.8 | 29.1 | 6.9 | 3.2 | –4.47 |
| Model 4 | –13.91 | 78.72 | 58.7 | 29.4 | 8.7 | 3.2 | –7.01 |
| Model 5 | –14.88 | 72.48 | 60.3 | 27.8 | 8.3 | 3.7 | –2.41 |

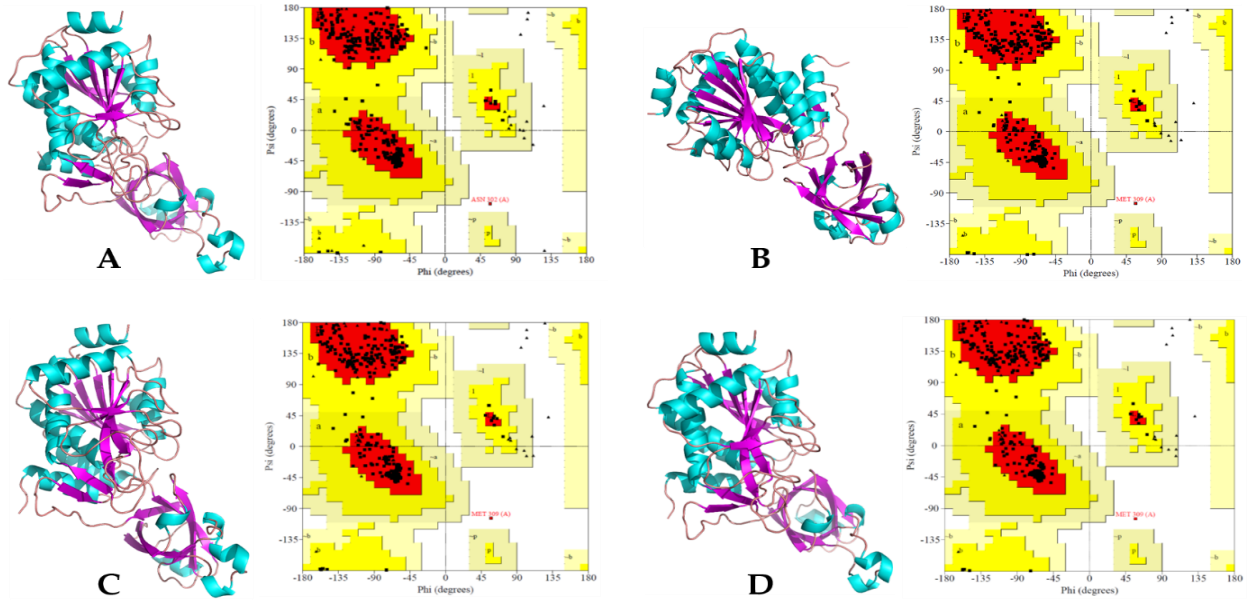



**Figure 9.** Three-dimensional structure models of cellulase using SWISS-MODEL homology modeling and Ramachandran plot of each model. Cellulase from: A. *Anoxybacillus flavithermus* NBRC 109594 (GAC90965.1); B. *Anoxybacillus gonensis* (EMI10668.1); C. *Anoxybacillus flavithermus* AK1 (EMT45536.1); D. *Anoxybacillus ayderensis* (EPZ39606.1)

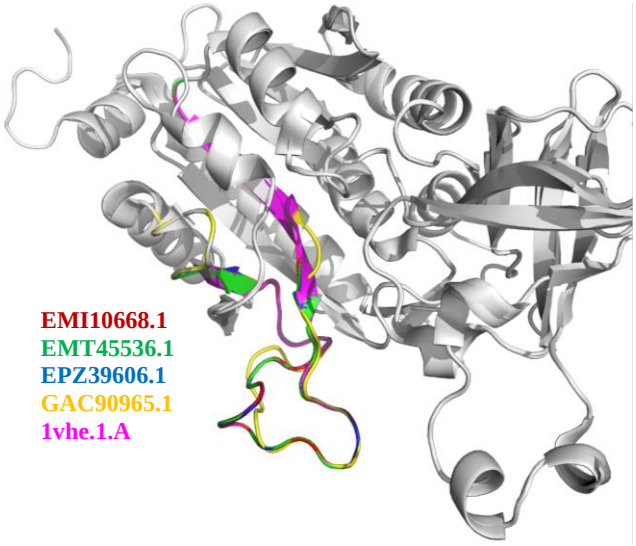



**Figure 10.** Comparative structural analysis of cellulase predicted with template (1VHE, aminopeptidase/glucanase from *Bacillus subtilis*)

## Discussion

*Anoxybacillus* is a facultative anaerobic bacterium that is generally found in high-temperature environments. Therefore, *Anoxybacillus* is a potential resource for obtaining thermostable enzymes. Bacteria from this genus produce unique biocatalysts that function under extreme conditions comparable to those used in various industrial processes. Thus, enzymes produced by thermophilic bacteria show promise for industrial applications. *Anoxybacillus flavithermus* TP-01, isolated from Gunung Pancar hot spring, is one *Anoxybacillus* species that possesses cellulase activity (Laras et al. 2017). This bacterium produces extracellular cellulases at 60°C, and this enzyme exhibits catalytic activity at 70°C and pH 7, showing its thermostability. Previous studies have shown that cellulases produced by *Anoxybacillus* are catalytically active at temperatures and pH values above 65°C and 6, respectively. *Anoxybacillus flavithermus* MBT002 exhibits optimal activity at 70°C and pH 7.0 and remains active up to 70°C (Sharif et al. 2023). *Anoxybacillus gonensis* O9 isolated from an Agri Diyadin hot spring, Turkey also exhibited thermophilic properties, with the optimum temperature and pH for crude enzyme activity being 50°C and 3.0 and 10.0, respectively (Genc et al. 2015). These findings indicate that *Anoxybacillus* genus is a promising source of thermostable cellulase-producing bacteria.

An important characteristic of these enzymes is their ability to be purified. In this study, cellulase derived from *Anoxybacillus flavithermus* TP-01 was partially purified using ammonium sulfate to determine the relationship between its purity and activity. Cellulase precipitated using 80% ammonium sulfate showed the highest enzymatic activity compared to enzymes precipitated at other concentrations. Partial purification with ammonium sulfate increases the activity and purity of cellulase. For example, the purity of cellulase obtained from *Bacillus tequilensis* G9 increased 3.6-fold when precipitated using 70% ammonium sulfate (Dar et al. 2019). Similar increases were observed for cellulases obtained from *Pleurotus ostreatus* (Okereke et al. 2017) and *Trichoderma longibrachiatum* (Pachauri et al. 2020). Ammonium sulfate increases cellulase activity by salting-out, a process that decreases protein solubility by increasing salt ion concentrations. A precise concentration of salt allows for the separation and precipitation of different proteins (Septiani et al. 2019). The 80% ammonium sulfate fraction showed the highest cellulase activity, which is consistent with the findings of Nisar et al. (2022), who reported the highest cellulase activity of *Thermomyces dupontii* when precipitated with 80% ammonium sulfate. Generally, 80% ammonium sulfate is the optimal concentration for partial purification of cellulases from various species (Gaur and Tiwari 2015; Bhardwaj et al. 2021).

In this study, amino acid sequences of cellulases from *Anoxybacillus* genus were subjected to in silico analysis. Computational approaches have become valuable tools to complement our understanding of proteins and their biochemical properties. In silico analysis is one of the most useful tools in computational biology for exploring the structural and functional properties of proteins (Nene et al. 2022). This study aimed to analyze the structural and functional properties of cellulases from *Anoxybacillus* using various programs and servers. The ExPASy-ProtParam tool was used to analyze the physicochemical properties and distribution of amino acid residues in each sequence. The GenBank, NCBI database was used to retrieve amino acid sequence; this database has 22 cellulase sequences originating from *Anoxybacillus* sp., however, only 11 sequences were complete, with no truncated or partial sequences.

The pI value provides information on the acidic or basic properties of protein sequences. Our analysis showed that the pI value of all sequences was below 7.0, indicating that all cellulases had acidic properties. In silico analysis of cellulase amino acids sequences from uncultured microorganisms (Sanjaya et al. 2021) and *Ruminococcus albus* (Hoda et al. 2021) also indicated their pI values being below 7, suggesting their acidic nature. The pI value is important in determining buffer systems for protein purification and crystallization (Kantardjieff and Rupp 2004). The isoelectric point is the pH at which the net protein charge is zero; thus, at their isoelectric point, the enzymes are compact and stable.

The Instability Index (*II*) of all sequences predicted that the cellulase would be stable in the test tube. This is in line with the relatively large AI values of all sequences (range: 76.52-96.40). The heat stability of proteins depends on their aliphatic index. A higher aliphatic index indicates that proteins can better withstand high temperatures (Bhattacharya et al. 2018; Vidhya et al. 2018). Although cellulase from *Anoxybacillus* sp. is predicted to have good stability based on *II* and AI values, experimental research is needed to validate this. For instance, a protein was deemed unstable based on the in silico analysis (Sanjaya et al. 2021), but experimental results indicates good stability for the protein. Gamage et al. (2019) also observed a similar phenomenon,

wherein they reported stability differences between the results of the computational and experimental analyses. This indicates that the Instability Index (*II*) prediction does not accommodate all the relevant information for the determination of protein stability under in vitro conditions. The application of *II* in the prediction of protein stability depends on the intrinsic nature of the protein and the conditions of the protein milieu.

The GRAVY index is an important parameter for studying the physicochemical properties of proteins. The GRAVY index represents the solubility of proteins and their interactions with water (Pramanik et al. 2018). Increasing positive scores indicate greater hydrophobicity, whereas a low GRAVY value indicates good interaction between water and protein. Proteins with hydrophobicity scores below 0 are more likely to be globular (hydrophilic proteins), whereas those with scores above 0 are more likely to be membranous (hydrophobic proteins). Four of the eleven sequences analyzed showed unusual GRAVY values. The GRAVY indexes of EMT45536.1, EMI10668.1, GAC92240.1, and EPZ39606.1 were –0.074, –0.072, –0.073, and –0.097, respectively. These values are relatively small compared to the GRAVY values of cellulases in general. Therefore, compared to other cellulases, these four cellulases may interact poorly with water.

Protein secondary structure can be predicted from their primary sequences (Zhang et al. 2018). The alpha helix structure is the most abundant secondary structure in cellulase compared to other secondary structures, which is related to the amino acid composition. Alanine, glycine, valine, aspartate, leucine, and isoleucine were the most prevalent amino acid residues in all cellulase sequences, whereas cysteine was the least common residue. Alanine, glycine, and leucine tend to form alpha helix, while threonine and valine typically form beta sheets. Aspartic acid bonds with the solvent, which is supported by hydrogen bonds. All analyzed cellulase sequences were hydrophobic in nature because most of the amino acids present in cellulase sequences namely, alanine, glycine, leucine, valine, proline, isoleucine, tryptophan, phenylalanine, and methionine have hydrophobic side chains.

Random coils are irregular secondary structures present in the N– and C–terminal arms and loops of protein structures and form due to electrostatic repulsion and steric hindrance of bulky adjacent residues, such as isoleucine, or charged residues, such as glutamic acid or aspartic acid. Thus, the presence of large number of random coils may be attributed to a high number of glycine and proline residues in protein (Vidhya et al. 2018). A high percentage of glycine in the sequence provides flexibility and structural rigidity to the polypeptide chain. Proline, on the other hand, promotes a coiling structure because of the crinkled polypeptide chains that interfere with the secondary structures. In the random-coil state, the average conformation of each amino acid residue was independent of the conformation of all the residues other than those in its immediate proximity in the primary structure. These five sequences are likely present in cellulases from *Bacillus thuringiensis* and *Bacillus pumilus* (Lugani and Sooch 2017), which have a higher percentage of alpha-helical structures than other secondary structures and lowest percentage of beta turns or reverse turns. They usually connect via antiparallel beta strands and are stabilized by hydrogen bonds connecting the carbonyl oxygen and amide hydrogen. Glycine and proline are commonly found in beta turns.

Four sequences, GAC90965.1, EMI10668.1, EMT45536.1, and EPZ39606.1, were selected for homology modeling to represent each cluster in the phylogenetic tree. The homology modeling results showed that the predicted structures for four of these sequences met the criteria for high-quality protein structures. These structural models were aligned to the template structure (PDB ID IVHE). In the resulting loop, the model of the GAC90965.1 sequence, which is a cellulase sequence from *Anoxybacillus flavithermus* NBRC 109594 and belongs to cluster A, differed notably from the template and other structural models. The loop conformation was different and an extended loop was noted, which contrasted with the beta sheet structure in other structures. This is interesting because loops play important roles in protein function, including enzyme activity and ligand binding (Corbella et al. 2023). Loops connect two secondary structural elements that are shorter than beta turns. Beta turns usually involve glycine and/or proline and are characterized by a hydrogen bond between the C=O of residue *i* and the N-H of residue *i+3* (i.e., between the first and fourth residues of the turn). The AI value of GAC90965.1 was relatively high, indicating that longer loops with a different orientation may have a stabilizing effect on the predicted structure; although the AI value of EMI10668.1 was the highest, the difference between the two sequences was not significant.

## ACKNOWLEDGEMENTS

The authors would like to appreciate to the Directorate General of Higher Education, Research and Technology, Ministry of Education, Culture, Research and Technology, Indonesia for financing this research.

## REFERENCES

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. Nucleic Acids Res 43 (W1): W39-W49. DOI: 10.1093/nar/gkv416.

Bajaj P, Mahajan R. 2019. Cellulase and xylanase synergism in industrial biotechnology. Appl Microbiol Biotechnol 103 (21-22): 8711-8724. DOI: 10.1007/s00253-019-10146-0.

Bhardwaj N, Kumar B, Agrawal K, Verma P. 2021. Current perspective on production and applications of microbial cellulases: A review. Bioresour Bioprocess 8: 95. DOI: 10.1186/s40643-021-00447-6.

Bhattacharya M, Hota A, Kar A, Chini DS, Malick RC, Patra BC, Das BK. 2018. In silico structural and functional modelling of Antifreeze protein (AFP) sequences of Ocean pout (*Zoarces americanus*, Bloch & Schneider 1801). J Genet Eng Biotechnol 16 (2): 721-730. DOI: 10.1016/j.jgeb.2018.08.004.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. 2014. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42 (W1): W252-W258. DOI: 10.1093/nar/gku340.

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. 2017. The SWISS-MODEL Repository-new features and

functionality. Nucleic Acids Res 45 (D1): D313-D319. DOI: 10.1093/nar/gkw1132.
Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248-254. DOI: 10.1016/0003-2697(76)90527-3.
Corbella M, Pinto GP, Kamerlin SCL. 2023. Loop dynamics and the evolution of enzyme activity. Nat Rev Chem 7 (8): 536-547. DOI: 10.1038/s41570-023-00495-w.
Dar MA, Pawar KD, Rajput BP, Rahi P, Pandit RS. 2019. Purification of a cellulase from cellulolytic gut bacterium, *Bacillus tequilensis* G9 and its evaluation for valorization of agro-wastes into added value byproducts. Biocatal Agric Biotechnol 20: 101219. DOI: 10.1016/j.bcab.2019.101219.
Duong-Ly KC, Gabelli SB. 2014. Salting out of proteins using ammonium sulfate precipitation. Methods Enzymol 541: 85-94. DOI: 10.1016/B978-0-12-420119-4.00007-0.
Dutta B, Banerjee A, Chakraborty P, Bandopadhyay R. 2018. In silico studies on bacterial xylanase enzyme: Structural and functional insight. J Genet Eng Biotechnol 16 (2): 749-756. DOI: 10.1016/j.jgeb.2018.05.003.
Dutta B, Deska J, Bandopadhyay R, Shamekh S. 2021. In silico characterization of bacterial chitinase: Illuminating its relationship with archaeal and eukaryotic cousins. J Genet Eng Biotechnol 19 (1): 19. DOI: 10.1186/s43141-021-00121-6.
Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. 2021. Expasy, the Swiss bioinformatics resource portal, as designed by its users. Nucleic Acids Res 49 (W1): W216-W227. DOI: 10.1093/nar/gkab225.
Ejaz U, Sohail M, Ghanemi A. 2021. Cellulases: From bioactivity to a variety of industrial applications. Biomimetics 6 (3): 44. DOI: 10.3390/biomimetics6030044.
Gamage DG, Gunaratne A, Periyannan GR, Russell TG. 2019. Applicability of instability index for in vitro protein stability prediction. Protein Pept Lett 26: 339-347. DOI: 10.2174/0929866526666190228144219.
Gaur R, Tiwari S. 2015. Isolation, production, purification and characterization of an organic-solvent-thermostable alkalophilic cellulase from *Bacillus vallismortis* RG-07. BMC Biotechnol 15: 19. DOI: 10.1186/S12896-015-0129-9.
Genc B, Nadaroglu H, Adiguzel A, Baltaci O. 2015. Purification and characterization of an extracellular cellulase from *Anoxybacillus gonensis* O9 isolated from geothermal area in Turkey. J Environ Biol 36: 1319-1324.
Hoda A, Tafaj M, Sallaku E. 2021. In silico structural, functional and phylogenetic analyses of cellulase from *Ruminococcus albus*. J Genet Eng Biotechnol 19 (1): 58. DOI: 10.1186/s43141-021-00162-x.
Hussain AA, Abdel-Salam MS, Abo-Ghalia HH, Hegazy WK, Hafez SS. 2017. Optimization and molecular identification of novel cellulose degrading bacteria isolated from Egyptian environment. J Genet Eng Biotechnol 15 (1): 77-85. DOI: 10.1016/j.jgeb.2017.02.007.
Ibrahim ASS, El-diwany AI. 2007. Isolation and identification of new cellulases producing thermophilic bacteria from an Egyptian hot spring and some properties of the crude enzyme. Aust J Basic Appl Sci 1 (4): 473-478.
Islam F, Roy N. 2018. Screening, purification and characterization of cellulase from cellulase producing bacteria in molasses. BMC Res Notes 11 (1): 445. DOI: 10.1186/s13104-018-3558-4.
Jaramillo PMD, Gomes HAR, Monclaro AV, Silva COG, Filho EXF. 2015. Lignocellulose-degrading enzymes. In: Gupta VK, Mach RL, Sreenivasaprasad S (eds). Fungal Biomolecules: Sources, Applications and Recent Developments. John Wiley & Sons, Ltd, Chichester, UK. DOI: 10.1002/9781118958308.ch6.
Juturu V, Wu JC. 2014. Microbial cellulases: Engineering, production and applications. Renew Sustain Energy Rev 33: 188-203. DOI: 10.1016/j.rser.2014.01.077.
Kantardjieff KA, Rupp B. 2004. Protein isoelectric point as a predictor for increased crystallization screening efficiency. Bioinformatics 20 (14): 2162-2168. DOI: 10.1093/bioinformatics/bth066.
Khoshnevisan K, Poorakbar E, Baharifar H, Barkhi M. 2019. Recent advances of cellulase immobilization onto magnetic nanoparticles: An update review. Magnetochemistry 5 (2): 36. DOI: 10.3390/magnetochemistry5020036.
Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35 (6): 1547-1549. DOI: 10.1093/molbev/msy096.
Laras GA, Handayani MN, Yamani LN, Purwani NN, Purkan, Puspaningsih NNT. 2017. Identification and characterization of thermophilic and pectinolytic bacteria isolated from gunung pancar hot spring bogor, Indonesia. Asian J Microbiol Biotechnol Environ Sci 19 (3): 496-508.
Loow Y-L, New EK, Yang GH, Ang LY, Foo LYW, Wu TY. 2017. Potential use of deep eutectic solvents to facilitate lignocellulosic biomass utilization and conversion. Cellulose 24: 3591-3618. DOI: 10.1007/s10570-017-1358-y.
Lugani Y, Sooch BS. 2017. In silico characterization of cellulases from genus *Bacillus*. Intl J Curr Res Rev 9 (13): 30-37. DOI: 10.7324/ijcrr.2017.9136.
Ma L, Zhao Y, Meng L, Wang X, Yi Y, Shan Y, Liu B, Zhou Y, Lü X. 2020. Isolation of thermostable lignocellulosic bacteria from chicken manure compost and a M42 Family Endocellulase cloning from *Geobacillus thermodenitrificans* Y7. Front Microbiol 11: 281. DOI: 10.3389/fmicb.2020.00281.
Merino N, Aronson HS, Bojanova DP, Feyhl-Buska J, Wong ML, Zhang S, Giovannelli D. 2019. Living at the extremes: Extremophiles and the limits of life in a planetary context. Front Microbiol 10: 780. DOI: 10.3389/fmicb.2019.00780.
Miller GL. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31: 426-428. DOI: 10.1021/ac60147a030.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37 (5): 1530-1534. DOI: 10.1093/molbev/msaa015.
Nene T, Yadav M, Yadav HS. 2022. Plant catalase in silico characterization and phylogenetic analysis with structural modeling. J Genet Eng Biotechnol 20: 125. DOI: 10.1186/s43141-022-00404-6.
Nisar K, Abdullah R, Kaleem A, Iqtedar M, Aftab M, Saleem F. 2022. Purification, characterization and thermodynamic analysis of cellulases produced from *Thermomyces dupontii* and its industrial applications. Saudi J Biol Sci 29 (12): 103483. DOI: 10.1016/j.sjbs.2022.103483.
Okereke OE, Akanya HO, Egwim EC. 2017. Purification and characterization of an acidophilic cellulase from *Pleurotus ostreatus* and its potential for agrowastes valorization. Biocatal Agric Biotechnol 12: 253-259. DOI: 10.1016/j.bcab.2017.10.018.
Pachauri P, Aranganathan V, More S, Sullia SB, Deshmukh S. 2020. Purification and characterization of cellulase from a novel isolate of *Trichoderma longibrachiatum*. Biofuels 11 (1): 85-91. DOI: 10.1080/17597269.2017.1345357.
Patel AK, Singhania RR, Sim SJ, Pandey A. 2019. Thermostable cellulases: Current status and perspectives. Bioresour Technol 279: 385-392. DOI: 10.1016/j.biortech.2019.01.049.
Pergande MR, Cologna SM. 2017. Isoelectric point separations of peptides and proteins. Proteomes 5 (1): 4. DOI: 10.3390/proteomes5010004.
Pramanik K, Ghosh PK, Ray S, Sarkar A, Mitra S, Maiti TK. 2017. An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of *Pseudomonas aeruginosa*. J Genet Eng Biotechnol 15 (2): 527-537. DOI: 10.1016/j.jgeb.2017.05.003.
Pramanik K, Kundu S, Banerjee S, Ghosh PK, Maiti TK. 2018. Computational-based structural, functional and phylogenetic analysis of *Enterobacter* phytases. 3 Biotech 8 (6): 262. DOI: 10.1007/s13205-018-1287-y.
Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42: W320-W324. DOI: 10.1093/nar/gku316.
Sanjaya RE, Putri KDA, Kurniati A, Rohman A, Puspaningsih NNT. 2021. In silico characterization of the GH5-cellulase family from uncultured microorganisms: Physicochemical and structural studies. J Genet Eng Biotechnol 19 (1): 143. DOI: 10.1186/s43141-021-00236-w.
Santhoshkumar R, Yusuf A. 2020. In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2, and CURS3) proteins of *Curcuma longa*. J Genet Eng Biotechnol 18 (1): 24. DOI: 10.1186/s43141-020-00041-x.
Sazci A, Erenler K, Radford A. 1986. Detection of cellulolytic fungi by using Congo red as an indicator: A comparative study with the dinitrosalicyclic acid reagent method. J Appl Bacteriol 61 (6): 559-562. DOI: 10.1111/j.1365-2672.1986.tb01729.x.
Septiani D, Suryadi H, Mun'im A, Mangunwardoyo W. 2019. Production of cellulase from *Aspergillus niger* and *Trichoderma reesei* mixed culture in carboxymethylcellulose medium as sole carbon. Biodiversitas 20 (12): 3539-3544. DOI: 10.13057/biodiv/d201211.
Sharif S, Shah AH, Fariq A, Jannat S, Rasheed S, Yasmin A. 2023. Optimization of amylase production using response surface methodology

from newly isolated thermophilic bacteria. Heliyon 9 (1): e12901. DOI: 10.1016/j.heliyon.2023.e12901.

Sharma D, Sharma P, Dev K, Sourirajan A. 2019. Endoglucanase gene of M42 aminopeptidase/endoglucanase family from thermophilic *Bacillus* sp. PW1 and PW2 isolated from Tattapani hot spring, Himachal Pradesh, India. J Genet Eng Biotechnol 17 (1): 4. DOI: 10.1186/s43141-019-0001-8.

Tiwari R, Nain L, Labrou NE, Shukla P. 2018. Bioprospecting of functional cellulases from metagenome for second generation biofuel production: A review. Crit Rev Microbiol 44 (2): 244-257. DOI: 10.1080/1040841X.2017.1337713.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44 (W1): W232-W235. DOI: 10.1093/nar/gkw256.

Vidhya VG, Swarnalatha Y, Bhaskar A. 2018. An in silico analysis of physicochemical characterization and protein-protein interaction network analysis of human anti-apoptotic proteins. Asian J Pharm 12 (4): S1397-S1407. DOI: 10.22377/ajp.v12i04.2941.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res 46 (W1): W296-W303. DOI: 10.1093/nar/gky427.

Wingfield PT. 2016. Protein precipitation using ammonium sulfate. Curr Protoc Protein Sci 84: A.3F.1-A.3F.9. DOI: 10.1002/0471140864.psa03fs84.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2014. The I-TASSER suite: Protein structure and function prediction. Nat Methods 12 (1): 7-8. DOI: 10.1038/nmeth.3213.

Yang J, Zhang Y. 2015. I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Res 43 (W1): W174-W181. DOI: 10.1093/nar/gkv342.

Zhang B, Li J, Lü Q. 2018. Prediction of 8-state protein secondary structures by a novel deep learning architecture. BMC Bioinformatics 19 (1): 293. DOI: 10.1186/s12859-018-2280-5.

Zheng F, Tu T, Wang X, Wang Y, Ma R, Su X, Xie X, Yao B, Luo H. 2018. Enhancing the catalytic activity of a novel GH5 cellulase GtCel5 from *Gloeophyllum trabeum* CBS 900.73 by site-directed mutagenesis on loop 6. Biotechnol Biofuels 11: 76. DOI: 10.1186/s13068-018-1080-5.

Zheng W, Zhang C, Li Y, Pearce R, Bell EW, Zhang Y. 2021. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. Cell Rep Methods 1 (3): 100014. DOI: 10.1016/j.crmeth.2021.100014.

Zhou X, Zheng W, Li Y, Pearce R, Zhang C, Bell EW, Zhang G, Zhang Y. 2022. I-TASSER-MTD: A deep-learning-based platform for multi-domain protein structure and function prediction. Nat Protoc 17 (10): 2326-2353. DOI: 10.1038/s41596-022-00728-0.