

RESEARCH

Open Access



Optimization, gene cloning, expression, and molecular docking insights for enhanced cellulase enzyme production by *Bacillus amyloliquefaciens* strain eh1

Elham F. El-Khamisi¹, Effat A. M. Soliman¹, Ghada M. El-Sayed¹, Shaimaa A. Nour^{2*}, Mohamed O. Abdel-Monem³ and Mervat G. Hassan³

Abstract

Background In this study, we isolated a cellulase-producing bacterium, *Bacillus amyloliquefaciens* strain eh1, from rice peel. We employed two optimization methods to enhance the yield of cellulase. Firstly, we utilized a one-variable-at-a-time (OVAT) approach to evaluate the impact of individual physical and chemical parameters. Subsequently, we employed response surface methodology (RSM) to investigate the interactions among these factors. We heterologously expressed the cellulase encoding gene using a cloning vector in *E. coli* DH5 α . Moreover, we conducted in silico molecular docking analysis to analyze the interaction between cellulase and carboxymethyl cellulose as a substrate.

Results The bacterial isolate eh1 exhibited an initial cellulase activity of 0.141 ± 0.077 U/ml when cultured in a specific medium, namely Basic Liquid Media (BLM), with rice peel as a substrate. This strain was identified as *Bacillus amyloliquefaciens* strain eh1 through 16S rRNA sequencing, assigned the accession number OR920278 in GenBank. The optimal incubation time was found to be 72 h of fermentation. Urea was identified as the most suitable nitrogen source, and dextrose as the optimal sugar, resulting in a production increase to 5.04 ± 0.120 U/ml. The peak activity of cellulase reached 14.04 ± 0.42 U/ml utilizing statistical optimization using Response Surface Methodology (RSM). This process comprised an initial screening utilizing the Plackett–Burman design and further refinement employing the BOX–Behnken Design. The gene responsible for cellulase production, *egl*, was effectively cloned and expressed in *E. coli* DH5 α . The transformed cells exhibited a cellulase activity of 22.3 ± 0.24 U/ml. The *egl* gene sequence was deposited in GenBank with the accession number PP194445. In silico molecular docking revealed that the two hydroxyl groups of carboxymethyl cellulose bind to the residues of Glu169 inside the binding pocket of the CMCase. This interaction forms two hydrogen bonds, with an affinity score of -5.71 .

Conclusions Optimization of cultural conditions significantly enhances the yield of cellulase enzyme when compared to unoptimized culturing conditions. Additionally, heterologous expression of *egl* gene showed that the recombinant form of the cellulase is active and that a valid expression system can contribute to a better yield of the enzyme.

Keywords Cellulase, *Bacillus amyloliquefaciens*, Optimization, Gene cloning and expression, Molecular docking

*Correspondence:

Shaimaa A. Nour

shaimaafars83@gmail.com; shaimaa_nour@yahoo.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Agricultural and industrial wastes are major contributors to environmental pollution. However, if we transform them into valuable resources, we can reduce their negative impacts. Common examples of these wastes in many countries, such as leaves, straws, cereals, and corn-cobs, are often used as animal feed [1]. Unfortunately, a significant amount of these materials is left on farmlands to decompose naturally with the help of microorganisms like bacteria and fungi, which produce cellulase enzymes responsible for the biodegradation of cellulose, the major constituent of agricultural waste [2]. Both cultivated and uncultivated bacteria can yield cellulases. Microbial cellulases are obtained from cultivated microorganisms that can be isolated in a laboratory. Numerous microorganisms produce cellulases, including bacteria, fungi, and actinomycetes. Conversely, the cultivation-independent approach, which relies on uncultured microorganisms, faces limitations due to the inability to cultivate the majority of microorganisms, especially those inhabiting soil environments, in laboratory settings [3]. Bacterial cellulases have gained more attention than their fungal counterparts for several reasons. Bacterial cellulases are preferred due to their lower production costs compared to fungal cellulases [4, 5]. Moreover, bacteria have faster growth rates than fungi, enabling them to reach high cell densities more quickly. This, in turn, facilitates efficient enzyme production. Additionally, certain bacterial cellulases are expressed in multiple complexes, which enhance overall performance through synergistic effects [6]. Most researchers are currently focusing on scaling up the production of a vital enzyme for the industry by using diverse bacterial strains. However, there is significant variability among these enzymes in terms of molecular weight, stability, amino acid composition, classification into protein families and domains, and secondary and tertiary structures [6, 7]. To address this complexity, bioinformatics—an interdisciplinary field—is employed to analyze the structure and function of proteins using various computational tools and databases. The information obtained from these tools and databases can help select the most efficient bacterial strains for industrial enzyme production. Additionally, this information can be used to guide the development of new microbial strains with enhanced enzyme production capabilities through the application of recombinant DNA technology [8, 9]. A breakthrough in cellulase research occurred in 1982 when the first cellulase gene was successfully cloned [10]. Since then, numerous cellulase genes have been engineered for expression in different host organisms, including *Escherichia coli*, *Saccharomyces cerevisiae*, and *Pichia pastoris* [11–13]. Each organism exhibited varying levels of enzymatic activity. For example, the Cellobiohydrolase

I (CBHI) gene, obtained from *Trichoderma koningii*, was integrated into the pGAPZ A plasmid and introduced into *Pichia pastoris*, which exhibited significant activity, reaching 0.1276 U/ml in the supernatant [14]. In another example, the gene encoding cellobiohydrolase was isolated from the *Clostridium clariflavum* and was adequately heterologous expressed in *Escherichia coli* BL21 (DE3). Through careful optimization of parameters such as induction time, pH, isopropyl β -D-1-thiogalactopyranoside (IPTG) concentration, and temperature, the highest enzyme activity achieved was 2.78 U/ml [15].

The core components of cellulase, which include endoglucanases (EC 3.2.1.4), exoglucanases (EC 3.2.1.74), and β -glucosidase (EC 3.2.1.21), play vital roles in fully breaking down cellulose [15, 16]. Cellulase is the second most important enzyme after amylase and is crucial for environmentally friendly and economically viable biofuel development [17]. Since the 1960s, cellulase has been increasingly used in various industries such as food, paper, pulp, textiles, and pharmaceuticals [18]. Specifically, cellulase (E.C. 3.2.1.4) is essential for breaking down lignocellulosic materials into simpler sugars, demonstrating its versatility in fields such as alternative energy, textiles, detergents, livestock feed, pharmaceuticals, food production, nutrition, and agriculture [19]. While the conventional chemical breakdown of cellulose is simple, the enzymatic process is distinctive for its lack of pollution, cost-effectiveness, and economic viability. Given these advantages, industries are increasingly looking for cost-effective cellulase with versatile applications, leading to the exploration of microorganisms capable of efficiently producing cellulase [20]. This pursuit opens up new opportunities for the discovery of economically viable cellulase-producing microorganisms. Numerous factors can be adjusted to enhance enzyme productivity and yield, including media components such as carbon, nitrogen, mineral sources, additives, and inducers, as well as physical parameters like pH, aeration, and temperature. These factors play a crucial role in determining the cost of enzyme production, which is often considered a primary challenge in biotechnological processes [21]. Researchers typically use a one-variable-at-a-time (OVAT) approach to screen physical and chemical parameters. The aim is to identify optimal conditions and understand the specific impact of each variable. However, this method is time-consuming and may overlook interactions between variables, leading to suboptimal results [21, 22].

To address these challenges, investigational methods based on factorial designs coupled with statistics have been developed and utilized to achieve faster and more reliable outcomes. Among these methods, the two-level factorial model is particularly advantageous

as it facilitates the analysis of interactions between factors. The Plackett–Burman design (PBD) is one factorial-based statistical technique used to assess the significance of critical variables. Additionally, response surface methodology (RSM) is employed to study the interactions among independent process variables [23].

The computational methods rely on diverse structural and physicochemical properties of protein sequences to extensively analyze and characterize, providing a thorough understanding of the connections between function, structure, and interactions with substrates. To tackle the challenge posed by this complexity, the most reliable and precise approach is considered to be template-based modeling. By using computational techniques, numerous models have been developed and evaluated, specifically revealing the intricate protein structures of cellulase found in *Bacillus amyloliquefaciens* and various *Bacillus* species [24]. As a result, a semi-rational approach has emerged as a viable strategy for protein redesign. This understanding has resulted in significant advancements in catalytic activity, enzyme binding efficiency, and practical benefits for industries, agriculture, and medicine [25].

In this investigation, the strain of *Bacillus amyloliquefaciens* was isolated from rice peel and molecularly identified, revealing its potential for cellulase production. Our objective was to optimize the cultural conditions for maximization of cellulase production from *Bacillus amyloliquefaciens* using the methods of one-factor-at-a-time (OFAT) and BOX–Behnken Design (BBD). The data obtained from this research can provide insights into the specific conditions influencing cellulase yield for practical application. Furthermore, we cloned and heterologously expressed the cellulase-encoding gene, resulting in significant cellulase production. Moreover, our study highlights the value of *in silico* analyses of the cellulase encoded by the *egl* gene, which led to enhanced predictions for enzyme binding and catalytic activity. Cellulase production optimization was accomplished through the utilization of both the one-factor-at-a-time (OFAT) approach and the BOX–Behnken Design (BBD) method.

Results

Isolation, screening, selection and molecular identification of CMCase producing bacteria

Among forty-seven bacterial isolates obtained from agriculture wastes suspensions, eleven strains, each with distinct morphology, tested positive in the zone of clearance test. The strain that exhibited the largest zone of clearance (Fig. 1) was chosen for further evaluation of enzyme production. The highest production of carboxymethyl cellulose (CMC) hydrolase (CMCase), 0.141 ± 0.006 U/ml, was reported by the bacterium that was originally



Fig. 1 Zone of hydrolysis produced on nutrient agar plates supplemented with 1% (W/V) CMC by bacterial isolate elh1

isolated from rice peel and was designated as isolate elh1. Genotypic characterization based on nucleotide homology, a phylogenetic analysis of 16S rDNA sequence, revealed that the isolate elh1 has 99% similarity with *Bacillus amyloliquefaciens* strain NBRC 15535 (Sequence ID: NR041455). Phylogenetic trees were created using the neighbor-joining (NJ) method, as shown in Fig. 2. The analysis of the phylogenetic tree confirmed that the bacterial strain elh1 is classified under the *Bacillus* genus. The elh1 strain was submitted to the GenBank database and assigned the accession number OR920278. Based on the phylogenetic analysis, the bacterial isolate elh1 was identified as *Bacillus amyloliquefaciens* strain elh1.

Optimization of CMCase production conditions

One variable at a time (OVAT)

Choosing the appropriate strain is crucial for achieving effective enzyme production. However, to obtain the highest enzyme yield, it is necessary to meticulously optimize the production processes and culture conditions. This requires finely adjusting various parameters, inoculum size, incubation period, nitrogen source, sugars as a carbon source and substrate concentration, to maximize CMCase productivity. In this experiment, the rice peel—the efficient carbon source for CMCase production by *Bacillus amyloliquefaciens* strain elh1—was added in various concentrations; the maximum CMCase yield (0.141 ± 0.077 U/ml) was given at rice peel concentration of 1% (w/v) while the minimum production was detected at a concentration of 4% (Fig. 3a) after 48 h incubation. In terms of the temporal progression of CMCase production through submerged fermentation, Fig. 3b

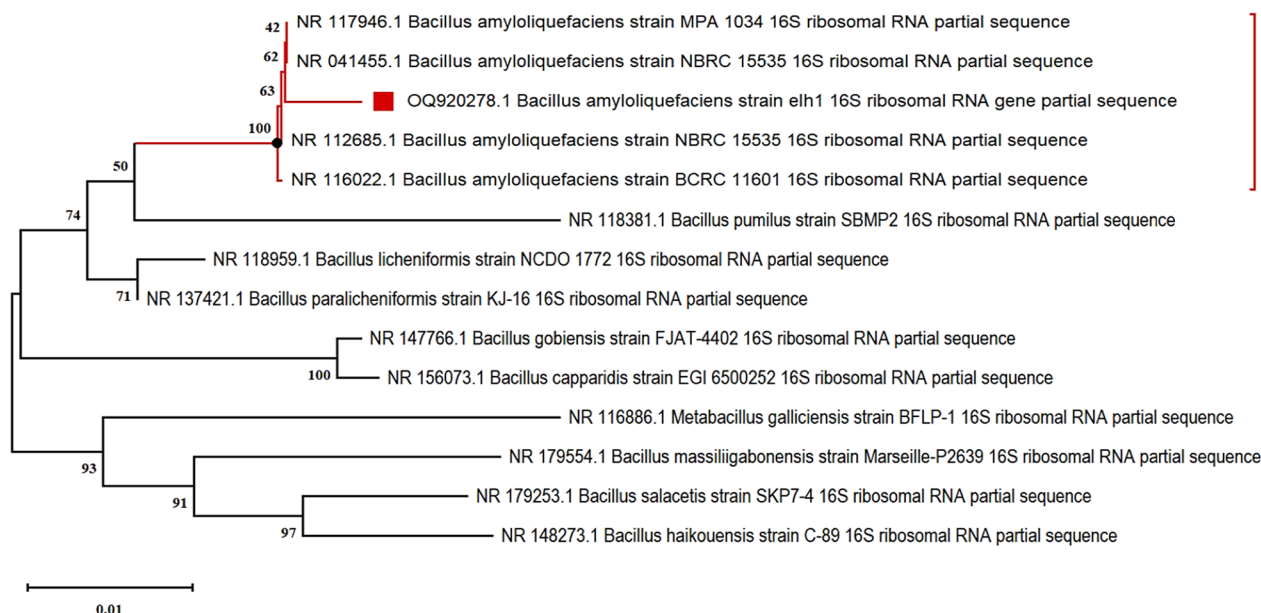


Fig. 2 Phylogenetic Tree Reveals Evolutionary Relationships of Bacterial Strain elh1, indicated by a red square, within the *Bacillus* Genus

demonstrates that CMCCase reached its peak production after 72 h of fermentation, with a yield of 0.173 ± 0.006 U/ml, and was reduced after 96 h. In this study, different nitrogen sources were applied to the production media; the component urea gave the highest CMCCase yield (2.22 ± 0.162 U/ml) (Fig. 3c). Considering this data, urea was used in different concentrations; the productivity increased to 3.03 ± 0.131 U/ml at urea concentration of 0.15 g/L (Fig. 3d). Evaluating the efficacy of the various sugar supplements tested in this study in terms of their impact on CMCCase production, dextrose was found to be the best sugar, where the yield reached 4.48 ± 0.028 U/ml (Fig. 3e) and increased to 5.04 ± 0.120 U/ml when dextrose concentration was 0.3% (w/v) (Fig. 3f).

Statistical optimization of rice peel based-medium for CMCASE production

The Plackett–Burman design (PBD)

For CMCCase production, various process parameters were optimized and nutritional conditions were screened using placket burman design of response surface methodology. The main nutritional components—seven variables—and eleven experiments were conducted for screening of various nutrients for CMCCase production and results are mentioned in Table 1. The findings revealed the substantial impact of various factors on the fermentation process, as evidenced by the highest CMCCase value (6.94 U/g) observed in trial 1, which included the following: rice peel (2 g/flask), Urea (0.05%), Dextrose (0.2 g/100), $MgSO_4 \cdot 7H_2O$ (0.25 g/l), KH_2PO_4

(4 g/l), $CaCl_2$ (0.3 g/l) and inoculum (4 ml). In contrast, the lowest CMCCase value (0.42 U/ml) was observed in trial 8, which consisted of rice peel (3 g/flask), Urea (0.25%), Dextrose (0.4 g/100), $MgSO_4 \cdot 7H_2O$ (0.75 g/l), KH_2PO_4 (4 g/l), $CaCl_2$ (0.5 g/l), and inoculum (4 ml). The primary effects of the investigated parameters on CMCCase production have been estimated and visually represented in Fig. 4. Analysis revealed that while the factors examined—rice peel, Urea, Dextrose, $MgSO_4 \cdot 7H_2O$, Inoculum, and $CaCl_2$ —exhibited negative effects, KH_2PO_4 showed a positive effect. The confidence level, P-effect, and t-test results of the statistical analysis of the Plackett–Burman Design (PBD) are detailed in Table 2. Consequently, due to the high significance level indicated by the P-value for the variables rice peel, Urea, and Dextrose, they were chosen for further optimization. The equation provided below represents a first-order model that elucidates the relationship between the seven components and CMCCase activity:

$$Y_{\text{activity}} = 10.18062 - 2.52951 X_1 - 10.3084 X_2 - 10.9023 X_3 - 0.28682 X_4 + 1.568245 X_5 - 0.52281 X_6 + 1.946743 X_7$$

box–behnken design (BBD)

To optimize the concentrations of rice peel, Urea and Dextrose, Box–Behnken design of response surface

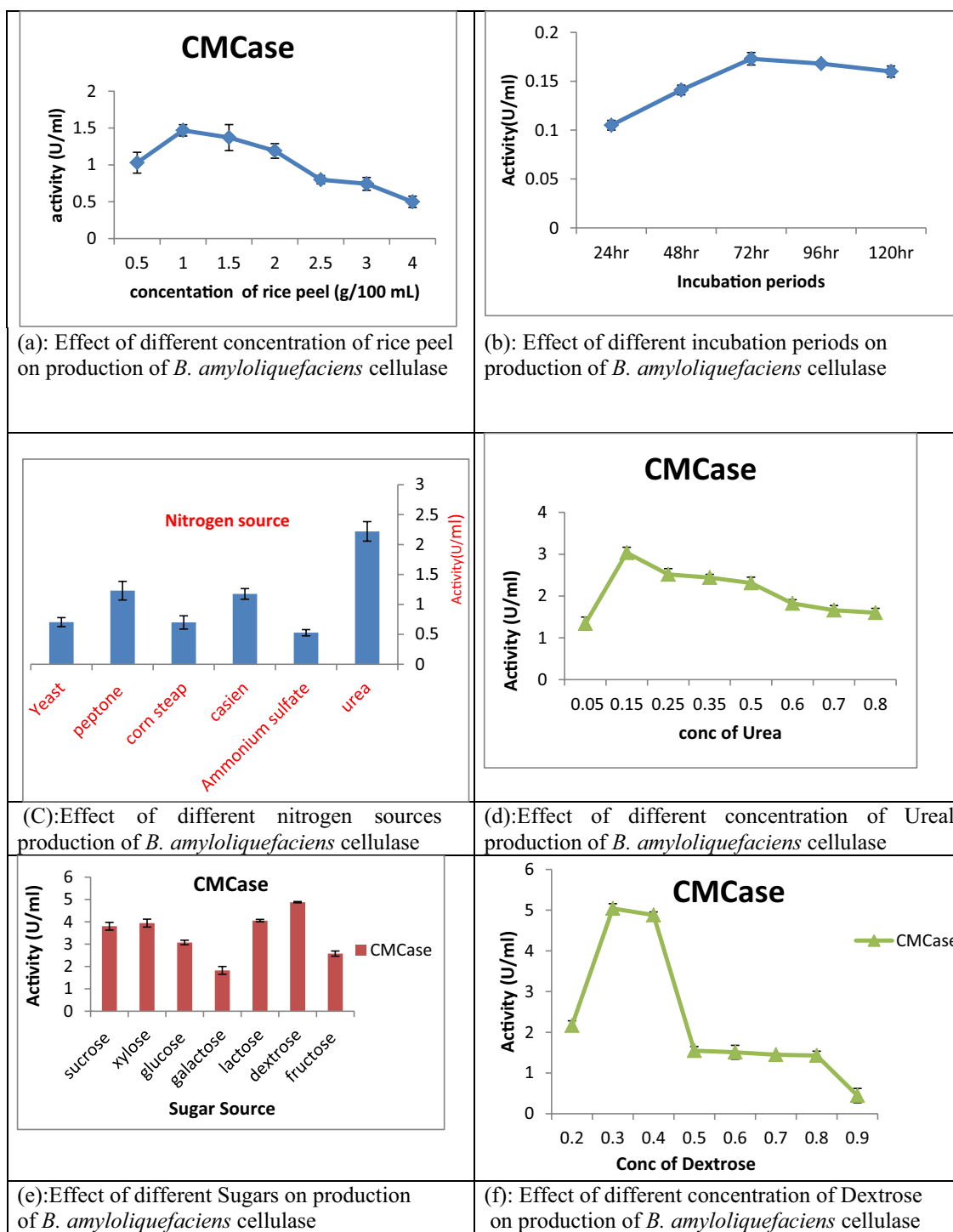


Fig. 3 Effect of different factors on production of *B. amyloliquefaciens* cellulase

methodology with three levels (-1, 0, and +1) obtained through 17 experimental runs. The independent variables were assessed at five discrete levels within the experimental design. Table 3 displays the theoretical and

observed increases in CMCCase derived from the statistical analysis of the test factors. The maximum CMCCase production (14.04 ± 0.424 U/ml) was observed using 1.5% rice peel, 0.2% Urea and 0.75% Dextrose (experimental

Table 1 Plackett–Berman experiment coded levels and real values

Trial	Rice peel	Urea	Dextrose	MgSO4 7H2O	KH2PO4	Inoculum	CaCl ₂	Activity (U/ml)
1	2(-)	0.05(-)	0.2(-)	0.25(-)	4(+)	4(+)	0.3(-)	6.94 ± 0.353
2	3(+)	0.05(-)	0.4(+)	0.25(-)	3(-)	2(-)	0.5(+)	2.10 ± 0.290
3	2(-)	0.25(+)	0.4(+)	0.25(-)	3(-)	4(+)	0.3(-)	1.48 ± 0.353
4	3(+)	0.25(+)	0.2(-)	0.25(-)	4(+)	2(-)	0.5(+)	4.13 ± 0.081
5	2(-)	0.05(-)	0.4(+)	0.75(+)	4(+)	2(-)	0.3(-)	6.01 ± 0.148
6	3(+)	0.05(-)	0.2(-)	0.75(+)	3(-)	4(+)	0.5(+)	3.44 ± 0.049
7	2(-)	0.25(+)	0.2(-)	0.75(+)	3(-)	2(-)	0.3(-)	4.22 ± 0.071
8	3(+)	0.25(+)	0.4(+)	0.75(+)	4(+)	4(+)	0.5(+)	0.42 ± 0.063
9	1(0)	0.15(0)	0.3(0)	0.5(0)	2(0)	3(0)	0.4(0)	5.03 ± 0.064
10	1(0)	0.15(0)	0.3(0)	0.5(0)	2(0)	3(0)	0.4(0)	5.03 ± 0.064
11	1(0)	0.15(0)	0.3(0)	0.5(0)	2(0)	3(0)	0.4(0)	5.03 ± 0.064

R-square, 0.993942; adjusted R-square, 0.979807. P-value: 0.002538

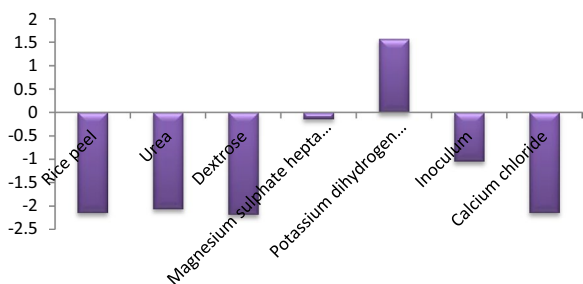


Fig. 4 Main effects of independent variables on cellulase production according to the results of the PBD

Table 2 A statistical analysis of the Plackett–Burman design shows coefficient values, effect, t- and P-values for each variable on the cellulase study

	Coefficients	Standard error	t stat	P-value
Intercept	10.18062	0.871588	11.68054	0.001348
Rice peel	-2.52951	0.235685	-10.7326	0.00173
Urea	-10.3084	0.993327	-10.3777	0.001909
Dextrose	-10.9023	0.993327	-10.9755	0.001619
MgSO ₄ 7H ₂ O	-0.28682	0.397331	-0.72186	0.522558
KH ₂ PO ₄	1.568245	0.198665	7.8939	0.004237
Inoculum	-0.52281	0.099333	-5.26321	0.013365
CaCl ₂	1.946743	1.541228	1.263111	0.295792

run#1). The predicted CMC_{Case} production under these conditions were almost near to the observed value depicting the accuracy of the model. The analysis of variance for the quadratic regression model revealed a highly significant F-value of 74.47 (Table 4). The adjusted coefficient of determination, “Adj R-Squared,” was calculated to be 0.9764, resulting in a predictive R Square value of

0.8346, which is deemed very high according to the F-test (Tables 4 and 5). The R-squared value indicates the proportion of variability in observed response values that can be explained by the experimental factors and their interactions. The effects of interactions and variable responses were examined using the Box-Behnken design (Fig. 5).

The model validation

To verify the model’s accuracy, a set of 17 random production combinations was employed to conduct experimental retests of CMC_{Case} production. The optimized conditions determined from the model projected a CMC_{Case} production of 14.04 U/ml. Impressively, the experimental value obtained precisely matched the predicted value, indicating the validity and reliability of the model. This close alignment between the experimental and predicted values is further supported by the data presented in Table 4. The Lack of Fit F-value of 1.34 suggests that the Lack of Fit is not significant compared to the pure error, with a 39.24% likelihood that a Lack of Fit F-value of this magnitude could occur due to noise. A non-significant Lack of Fit is desirable as it indicates a well-fitting model. The Predicted R² of 0.7798 closely corresponds to the Adjusted R² of 0.9063, with the difference being less than 0.2. The Adeq Precision, indicating the signal-to-noise ratio, demonstrates an adequate signal with a ratio of 14.967, suggesting that this model can effectively explore the design space. It is advisable to select the highest-order polynomial where the additional terms are significant and the model is not aliased. The Model F-value of 74.47 suggests the model’s significance. With only a 0.01% likelihood of such a large F-value arising from noise, it underscores the model’s robustness. P-values below 0.0500 signal significant model terms; in

Table 3 Examined concentration of the key variables and results of BBD experiment

Run	Factor 1 A:waste conc g/100 ml	Factor 2 B:urea g/L	Factor 3 C:dextrose g/L	Response 1 activity U/ml	Predicted value	Residual
1	1.5(0)	0.2(0)	0.75(0)	14.04±0.424	14.04	0.0000
2	1(-)	0.25(+)	0.75(0)	3.92±0.076	2.81	1.12
3	1.5(0)	0.15(-)	0.5(-)	0.414±0.027	-0.5032	0.9178
4	2(+)	0.2(0)	1(+)	2.672±0.126	1.93	0.7399
5	1.5(0)	0.2(0)	0.75(0)	14.04±0.424	14.04	0.0000
6	1.5(0)	0.2(0)	0.75(0)	14.04±0.423	14.04	0.0000
7	2(+)	0.2(0)	0.5(-)	2.723±0.118	2.53	0.1979
8	1.5(0)	0.25(+)	1(+)	0.677±0.066	1.59	-0.9179
9	2(+)	0.25(+)	0.75(0)	3.825±0.202	3.65	0.1780
10	1(-)	0.2(0)	1(+)	2.989±0.043	3.19	-0.1979
11	1.5(0)	0.2(0)	0.75(0)	14.04±0.424	14.04	0.0000
12	1(-)	0.2(0)	0.5(-)	0.662±0.033	1.40	-0.7399
13	1(-)	0.15(-)	0.75(0)	2.032±0.154	2.21	-0.1780
14	1.5(0)	0.2(0)	0.75(0)	14.04±0.424	14.04	0.0000
15	2(+)	0.15(-)	0.75(0)	0.129±0.001	14.04	0.0000
16	1.5(0)	0.25(+)	0.5(-)	0.799±0.014	2.81	1.12
17	1.5(0)	0.15(-)	1(+)	0.644±0.031	-0.5032	0.9178

Table 4 ANOVA for the quadratic response surface model (RSM) from the cellulase production

Source	Sum of squares	df	Mean square	F-value	p-value (prob > F)
Model	545.14	9	60.57	74.47	< 0.0001*
A-waste conc	0.0086	1	0.0086	0.0106	0.9208
B-urea	4.52	1	4.52	5.55	0.0506
C-dextrose	0.7089	1	0.7089	0.8715	0.3816
AB	0.8095	1	0.8095	0.9951	0.3517
AC	1.41	1	1.41	1.74	0.2289
BC	0.0309	1	0.0309	0.0380	0.8510
A ²	103.86	1	103.86	127.69	< 0.0001*
B ²	183.11	1	183.11	225.11	< 0.0001*
C ²	195.37	1	195.37	240.18	< 0.0001*
Residual	5.69	7	0.8134		
Lack of fit	5.69	3	1.90		
Pure error	0.0000	4	0.0000		
Std. Dev	0.9019		R ²	0.9897	
Mean	5.39		Adjusted R ²	0.9764	
C.V. %	16.72		Predicted R ²	0.8346	
			Adeq Precision	21.0246	

* Significant variable at 95% confidence. R²: 0.9897, R² Adj: 0.9764. Adequate precision ratio: 21.0246. DF, degree of freedom

this instance, A², B², and C² are notable. Conversely, values surpassing 0.1000 indicate insignificant model terms (Fig. 6).

Amplification and cloning of *egl* gene from *Bacillus amyloliquefaciens* strain elh1

Using the genomic DNA of *Bacillus amyloliquefaciens* strain elh1 as a template, the particular primers created for the *egl* gene successfully amplified a DNA fragment of about 1500 base pairs. Using colony PCR, a gene fragment of roughly 1500 base pairs was amplified using the

Table 5 Coefficients in terms of coded factors

Factor	Coefficient estimate	df	Standard error	95% CI low	95% CI high	VIF
Intercept	14.04	1	0.4033	13.09	14.99	
A-waste conc	-0.0329	1	0.3189	-0.7869	0.7212	1.0000
B-urea	0.7513	1	0.3189	-0.0027	1.51	1.0000
C-dextrose	0.2977	1	0.3189	-0.4563	1.05	1.0000
AB	0.4499	1	0.4509	-0.6165	1.52	1.0000
AC	-0.5945	1	0.4509	-1.66	0.4719	1.0000
BC	-0.0879	1	0.4509	-1.15	0.9784	1.0000
A ²	-4.97	1	0.4395	-6.01	-3.93	1.01
B ²	-6.59	1	0.4395	-7.63	-5.56	1.01
C ²	-6.81	1	0.4395	-7.85	-5.77	1.01

egl-F/egl-R primers, and the resulting construct was dubbed pGEM-egl. This confirmed the correct assembly of the construct (Fig. 7A). Subsequent double digestion assays confirmed the integrity of the pGEM-egl vector. The digestion yielded a fragment size of 3015 base pairs for the pGEM Teasy vector and approximately 1500 base pairs for the egl gene fragment (Fig. 7B). This dual confirmation through colony PCR and double digestion ensures the successful construction and proper assembly of the pGEM-egl vector, demonstrating the accurate incorporation of the egl gene fragment into the vector construct.

CMCase heterologous expression in *E. coli* DH5 α

The *E. coli* colonies containing the recombinant vector pGEM-egl were grown on nutrient agar plates supplemented with 1% carboxymethyl cellulose (CMC). Clear zones were observed around these cells, in contrast to the non-transformed *E. coli* cells (Fig. 7C, D), indicating the presence of cellulase activity. These colonies were then chosen for CMCase evaluation. The expression levels of the enzyme showed slight variation among the positive colonies tested, with the highest production reaching 22.3 ± 0.24 U/ml.

DNA sequence and phylogenetic analysis of CMCase encoding gene (egl)

When comparing the predicted amino acid sequence of CMCase with homologous sequences in the UniProt database, a remarkable 98.8% similarity to the *Bacillus subtilis* endoglucanase entry P07983 was discovered. After submission to GenBank, the nucleotide sequence was assigned the entry number PP194445. This 1500 base pair open reading frame (ORF) encodes a pre-enzyme consisting of 499 amino acids. The analysis indicated that the CMCase belongs to the glycosyl hydrolase family 5. SignalP analysis identified residues 1–29 as

the signal peptide, responsible for the enzyme's excretion (Fig. 8). The evolutionary relationship between the inferred amino acid sequence and the closest relatives from the UniProt database is shown in Fig. 9. The ESPript server was used to align the sequence of CMCase with its homologs, and the results, along with the secondary structure prediction, are shown in Fig. 10. The analysis of the CMCase model reveals a secondary structure consisting of 10 α -helices and 10 β -sheets. Additionally, two glutamic acid residues, Glu169 and Glu257, highlighted in black boxes, are identified as the active sites involved in the CMCase binding to the substrates.

Homology modelling and structure validation of modelled CMCase

Using SWISS-MODEL, a 3D model of CMCase was created; *Bacillus subtilis* 168 endo-1,4-beta-glucanase (PDB ID: 3PZT.1; resolution: 1.97 Å) was the best match. With a zero e-value, a GMQE value of 0.80, a QMEAN of 0.08, and a noteworthy 97.6% identity, the model demonstrates its high reliability and quality. The majority of residues, as shown in Fig. 11A, have values that are near to 1, which indicates good estimations of local quality. Low quality residues were defined as those having values less than 0.6. To further support its reliability, the modeled CMCase structure aligns perfectly with other protein structures in the PDB (Fig. 11B). The Ramachandran plot (Fig. 12A) and accompanying statistics (Fig. 12B) reveal that 86.7% of the modeled cellulase residues are in the most favored regions, 12.9% are in additional allowed regions, and 0.4% are in generously allowed regions, confirming the model's high quality. The model's validity was confirmed by the Verify3D plot (Fig. 12C), which showed a PASS for structural validation with a mean 3D-1D score of >0.1 for 89.4% of residues. The remarkable quality of the anticipated model was indicated by the Z-score of -9.05 obtained from ProSA-web analysis (Fig. 12D). The 3D

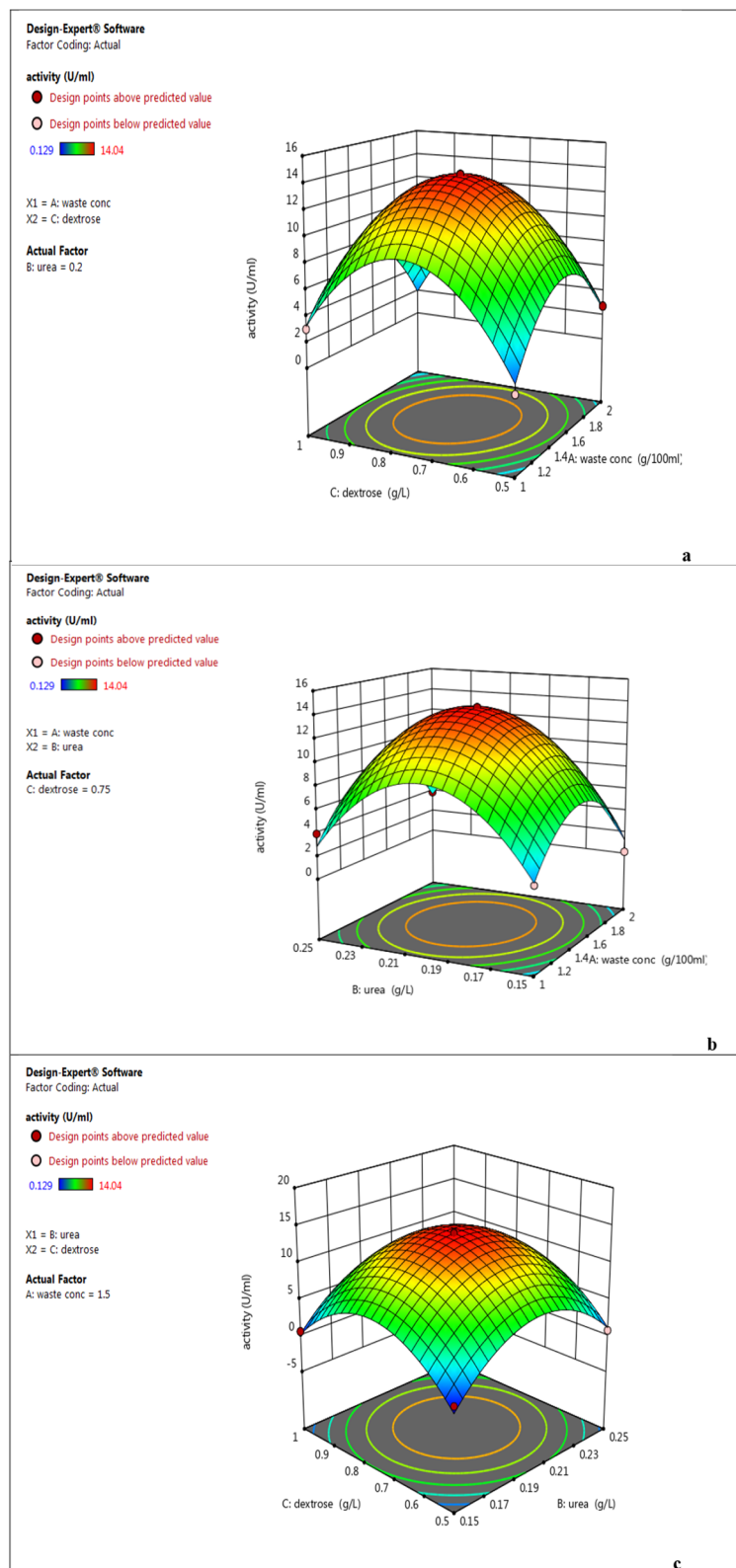


Fig. 5 Three-dimensional response surface plots showing the effect of different variables on cellulase production: waste concentration, Dextrose (a); waste conc, Urea (b); Urea, Dextrose (c). Green, yellow, and red color showed low, medium, and high cellulase activity (U/mL), respectively

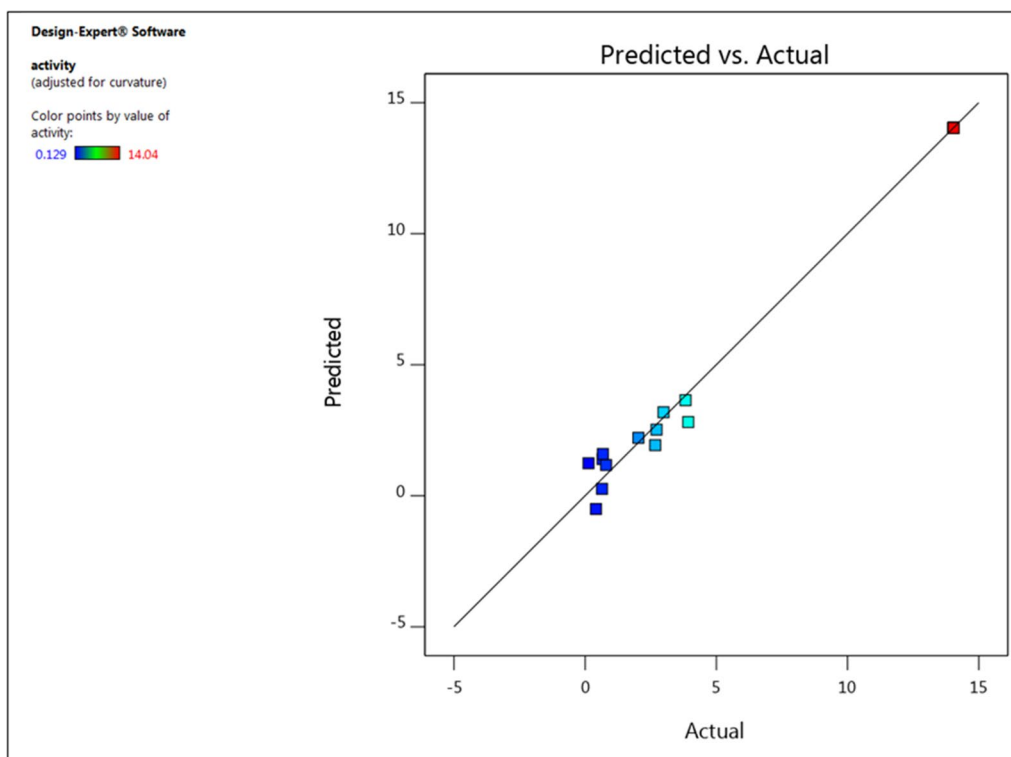


Fig. 6 Residual plot of the observed-predicted values (residuals) versus the response (optimization process) of *B. amyloliquefaciens* CMCase

model's correct positioning of the protein backbone dihedral angles phi (ϕ) and psi (ψ) is verified by this investigation. Additionally, the ERRAT Complete Overall Quality Factor, which plots error function values against the location of a 9-residue sliding window, was 96.8085. This factor evaluates the statistics of non-bonded interactions among different types of atoms.

Alignment of the CMCase model and template (3PZT) structure

Figure 13A, B show the 3D representations of CMCase and its template (3PZT), respectively. Alignment calculated using PyMOL Molecular Viewer revealed an RMSD value of 0.062 (285 to 285 atoms), indicating a close structural similarity (Fig. 13C). In the figures, the template structure is represented by cyan helices, while the protein homology model is represented by green helices. The template is homodimers and has two identical chains, A and B. The alignment of chain A in the template (3PZT) with the homology model demonstrates the high quality of the model.

Investigations of docking and molecular interactions

As demonstrated by Fig. 14, the two hydroxyl groups on the ligand, carboxymethyl cellulose, create two hydrogen bonds with the CMCase protein within the binding

pocket. The glutamic acid residues Glu169 and Glu257 are involved in these bonds, with corresponding distances of 2.2 and 2.6 Å. Glu169 and Glu257's oxygen atoms accept hydrogen bonds from both hydrogen bonds, acting as hydrogen bond donors (HBD), resulting in an interaction affinity score of -5.71 kcal/mol. Additionally, a hydrogen bond with a distance of 2.4 Å is formed with the residue Trp219 (tryptophan). Visualization using Chimera revealed three hydrogen bonds and clearly showed the 3D structural interactions (Fig. 15). The angles of the hydrogen bonds were measured at 170.9° , indicating that they are reliable and effective.

Discussion

This research comprised two main investigations: firstly, the production and optimization of a cellulase derived from *Bacillus amyloliquefaciens* strain elh1, in the presence of rice straw as a substrate; secondly, the cloning of egl gene, the characterization and in silico analysis of the protein. The accumulation of agricultural waste presents a significant environmental challenge as it contributes to pollution and requires a large amount of land for disposal. Addressing this issue can be costly, so it is crucial to explore sustainable solutions [2]. These wastes are primarily broken down by specific microorganisms that naturally reside in various environments. These

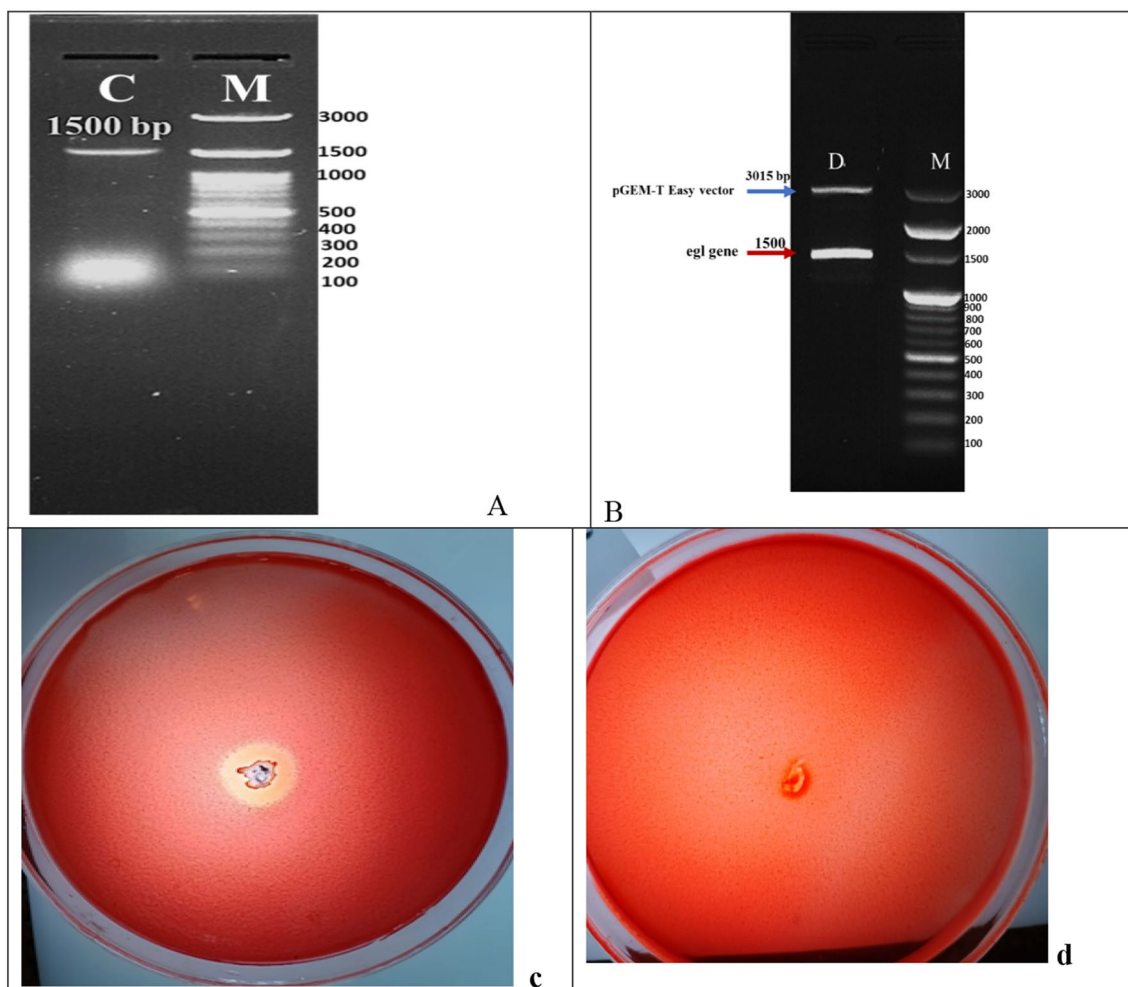


Fig. 7 Validation of pGEM-egl Recombinant Construct Assembly through Multiple Techniques: **A** Colony PCR of Positive Transformants: The amplified fragment of the *egl* gene was visualized on an agarose gel, with the expected size of approximately 1500 base pairs (**C**). The DNA ladder used for size reference was the 100 bp DNA ladder H3 RTU (GeneDirex, Inc.)(**M**). **B** Double Digestion of the Recombinant Construct: the *egl* gene fragment (indicated by a red arrow) and the linearized pGEM Teasy vector (indicated by a blue arrow). The DNA ladder used for size reference was the 100 bp DNA Ladder SolisFAST[®](**M**). **C** Cellulase Activity Assay: The appearance of a clear zone of hydrolysis around the transformant colonies confirmed the expression and activity of the *egl* gene (**c**), absence of hydrolysis zone in case of non-transformant *E.coli* DH5 α (**d**)

microorganisms produce enzymes that break down lignocellulosic materials into simpler sugars [26]. Utilizing these waste materials and transforming them into valuable products offers economic benefits and reduces environmental impact. This represents a shift away from traditional harmful chemical treatments towards safe and sustainable alternatives. Cellulose-degrading microbes can be found in diverse habitats such as animal waste, gastrointestinal tracts, soil, and aquatic ecosystems. However, current population densities of these microbes are not sufficient to meet industrial demands. The predominant microorganisms of significance include bacteria and fungi, with bacteria being particularly favored over fungi due to their adaptability to industrial

environments. Ongoing research efforts are focused on exploring different ecological niches to isolate bacterial strains that exhibit promising cellulolytic activity.

In this study, a robust bacterial strain called *Bacillus amyloliquefaciens* elh1 was isolated. This strain was chosen for its significant cellulase production when cultured on a 1% carboxymethyl cellulose (CMC) at 37 °C for 48 h (0.141 ± 0.077 U/ml). Similar studies have also demonstrated the effectiveness of these approaches in utilizing bacterial strains for cellulase production. In a similar study, Islam and Roy [27] discovered that the bacterial species *Paenibacillus* sp. exhibited considerable promise for achieving maximum cellulase production ($0.9 \mu\text{mol ml}^{-1} \text{min}^{-1}$) at pH 7.0 following a 24 h incubation

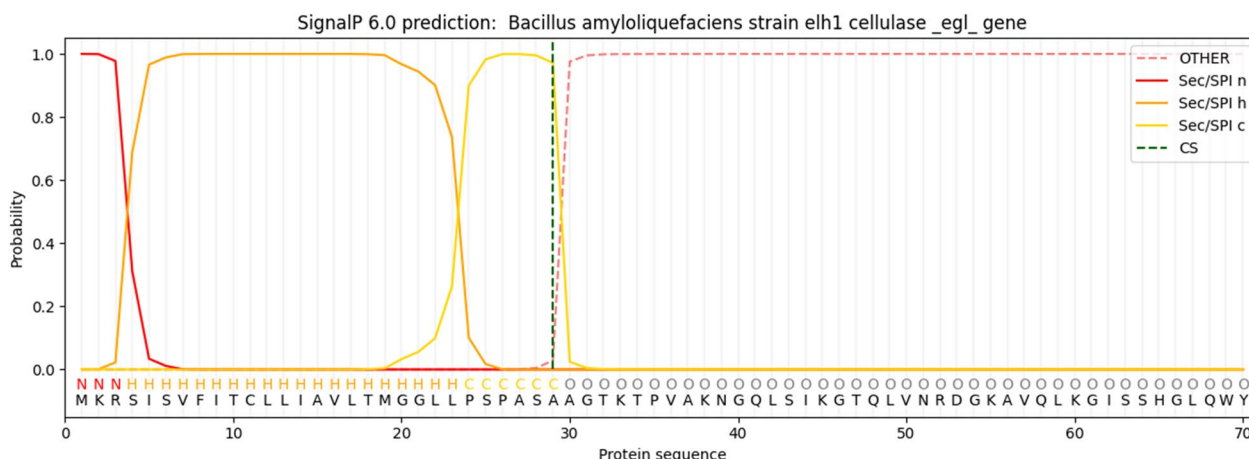


Fig. 8 Analysis of cellulase signal peptide and pro-peptide using signalp version 6.0

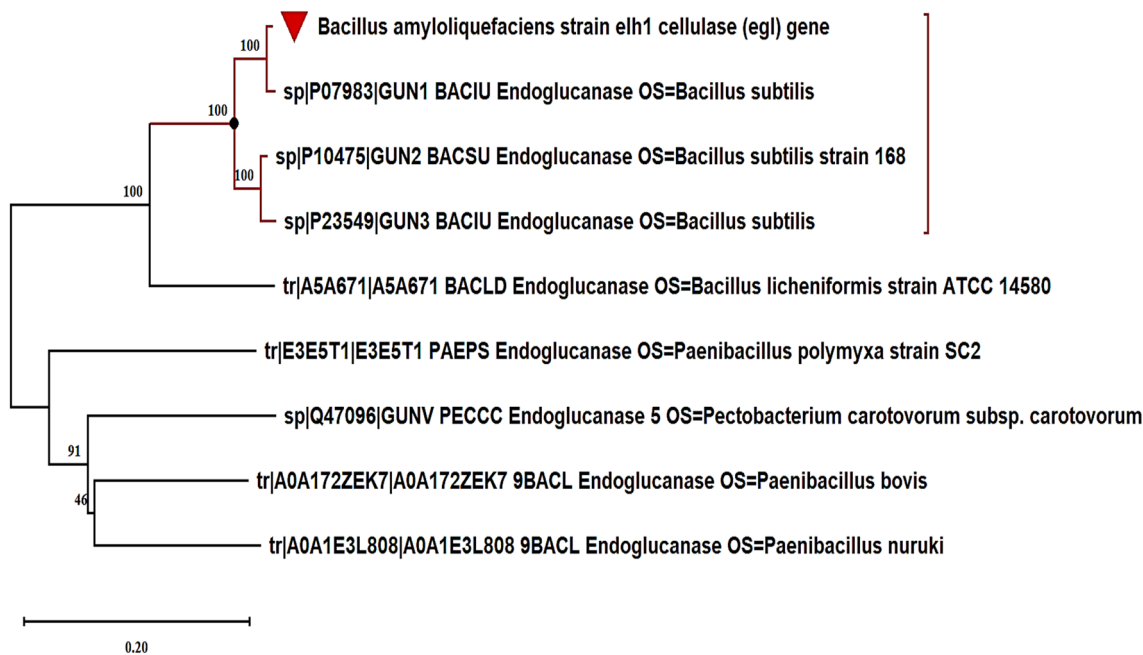


Fig. 9 Phylogenetic tree of the egl inferred amino acid sequence (red triangle) and its nearest relatives, reconstructed in mega 11 through the neighbor-joining method from the uniprot database. bootstrap percentages are shown from 1000 replications

(See figure on next page.)

Fig. 10 Multiple structure alignment of deduced amino acids sequence of egl gene and relative proteins retrieved from UniProt and rendering of secondary structure information regarding the prediction of α -helice and β -sheet regions. Residues invariable among sequences are typed in red on a white background; residues conserved within each group are displayed as white letters on a red background, and the residues representing active sites are shown in a black box. Secondary structure elements from known endoglucanase structure are indicated at the top of the alignment. TT letters represent strict beta turns

period at 40 °C. This observation was made in a medium comprising 1.0% CMC. In a study conducted by Fouda et al. [28], it was revealed that the dominant cellulase

activity of *Bacillus amyloliquefaciens* M7 was measured at 11.6 ± 0.4 U/ml. In another investigation by Singh et al. [14], the CMCcase activity observed in the cell-free

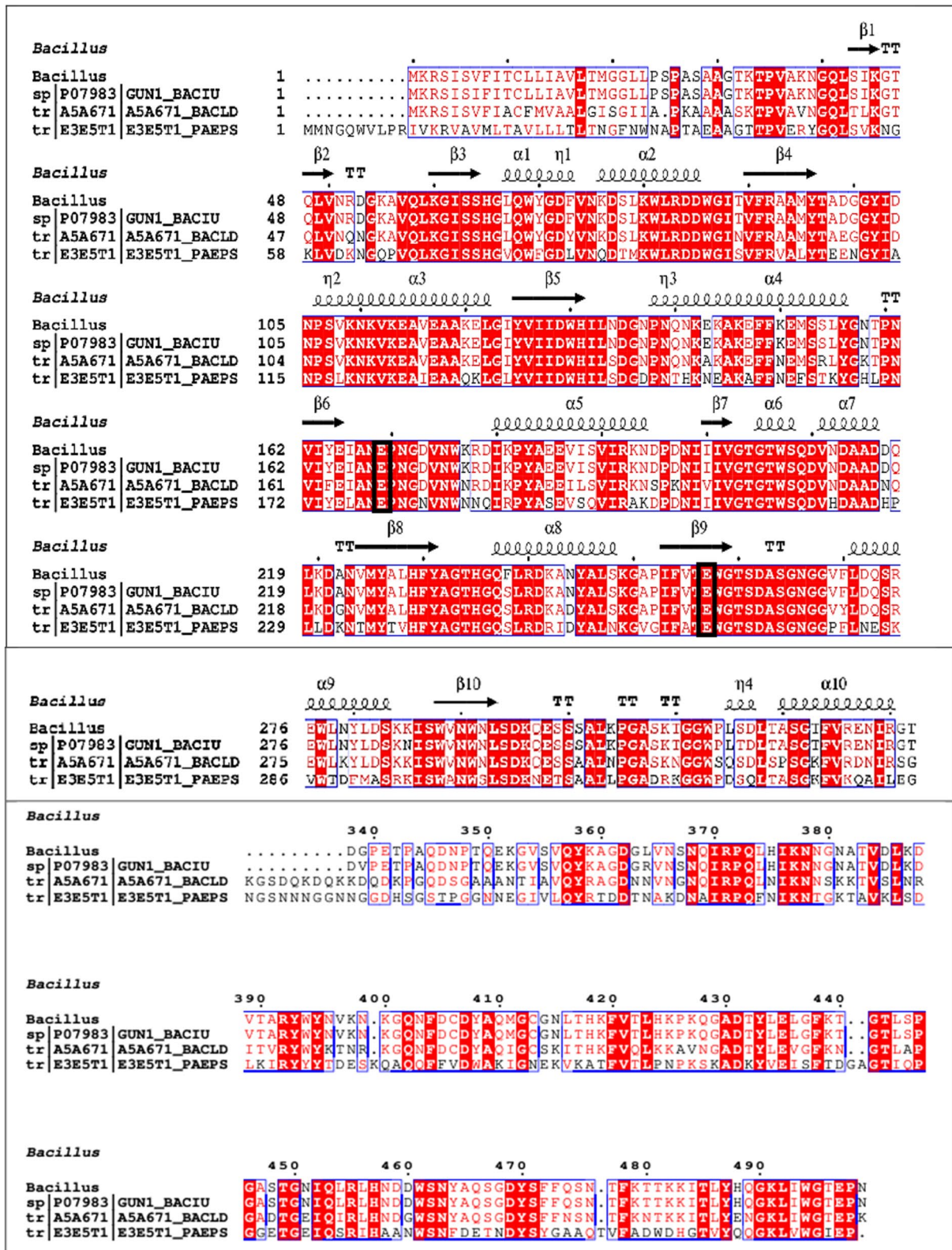


Fig. 10 (See legend on previous page.)

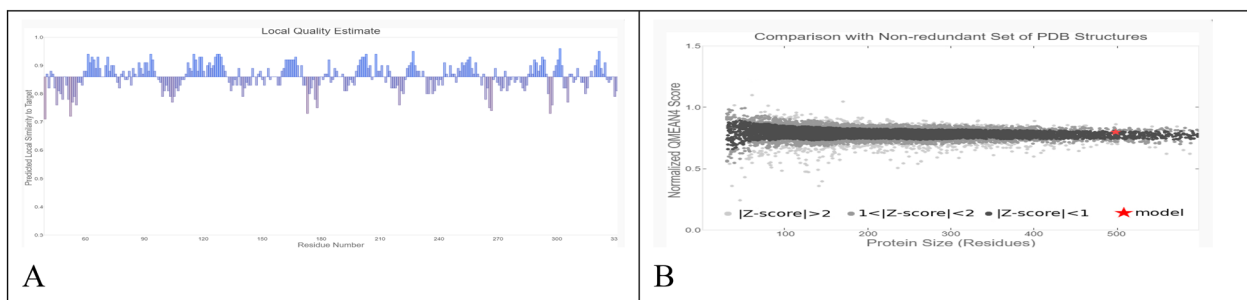


Fig. 11 Validation of the modeled cellulase structure involves: **A** The evaluation of local quality of the residues in the anticipated cellulase model; **B** the comparison of the predicted cellulase structure to a nonredundant set of PDB structures

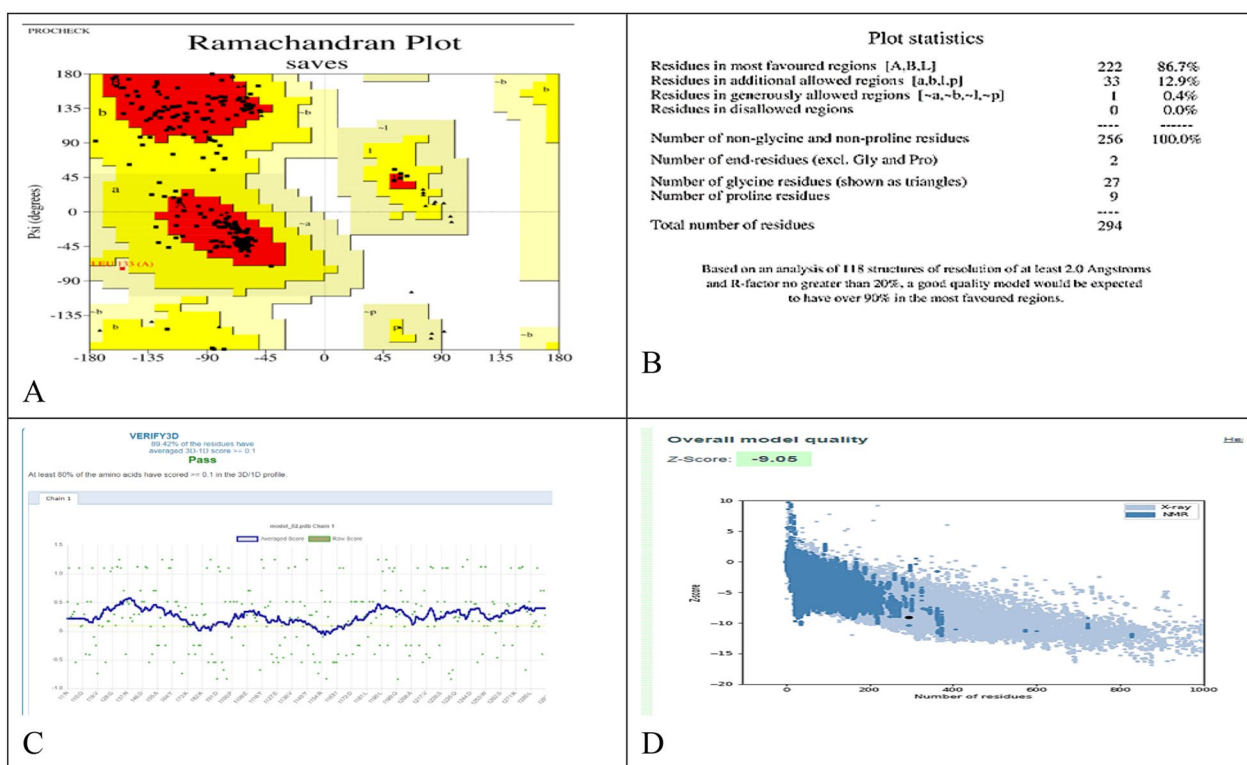


Fig. 12 Estimated Model Assessment: **A** Ramachandran Plot Analysis; **B** Ramachandran Plot Statistics; **C** Verify3D Results; **D** ProSA-web Z-Score

supernatant of *B. amyloliquefaciens* SS35 ranged from 0.132 to 0.528 U/ml. Additionally, Ahmad et al. [29] isolated the bacterial strain *Aneurinibacillus aneurinilyticus* from urban Himalayan freshwater. This strain exhibited cellulolytic activity on 0.5% CMC agar and showed a zone of hydrolysis. The utilization of agro-wastes as substrates for cellulase production offers a comprehensive approach with dual benefits. Firstly, it helps to eliminate accumulated agro-wastes. Secondly, it facilitates the production of cellulase from a cost-effective source. In this investigation, rice straw was the best substrate for *Bacillus amyloliquefaciens* strain elh1 to produce cellulase

enzyme. The study by Pham et al.[1] investigated the ability of *Bacillus* sp. to produce cellulase (140 U/ml) from coconut-mesocarp. Also, Bala et al. [5] exploited the cheap agro-waste, sugarcane bagasse as a substrate for cellulase enzyme production using *Bacillus licheniformis* MTCC 429. In this study, rice straw is identified as the optimal substrate for *Bacillus amyloliquefaciens* strain elh1 in the production of cellulase enzymes. The total quantity of cellulase is significantly influenced by nutritional and physical culture conditions [30]. Numerous researchers have explored cellulase yield in various media and production environments [29–31]. Initially,

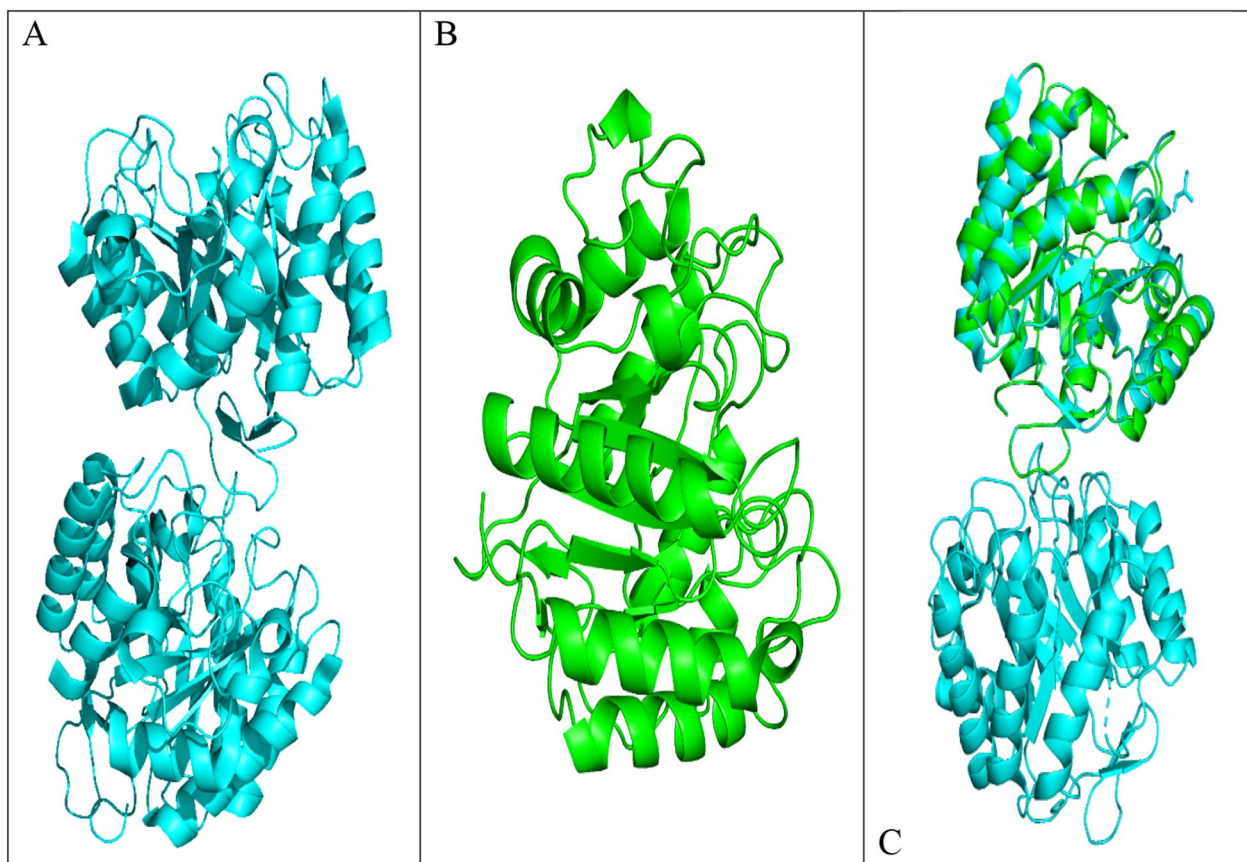


Fig. 13 **A** The three-dimensional structures of 3PZT (template model); **B** the cellulase homology model; **C** Structure Alignment Between the Cellulase Target Protein and the Template Protein (3PZT)

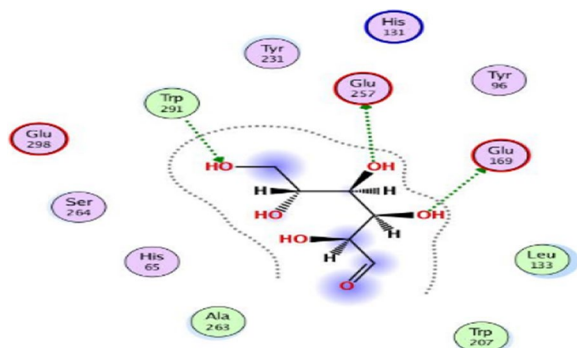


Fig. 14 Molecular docking interaction of carboxymethyl cellulose with cellulase from *Bacillus amyloliquefaciens* Strain elh1

cellulase production in bacterial isolates was conducted at 37 °C and a pH of 7, consistent with findings from multiple published sources [20, 31]. In this investigation, we conducted an optimization of production medium supplements, focusing on carbon sources (sugars), organic nitrogen, incubation durations, and substrate

concentration (rice straw) using a one-factor analysis approach. Among the carbon sources assessed, dextrose demonstrated the highest production yield compared to the other carbon sources examined in this study. These findings diverge from previous research, which identified glucose as the optimal carbon source for cellulase production [29, 32]. Another crucial determinant impacting cellulase activity involves the incorporation of supplementary organic and inorganic nitrogen sources into the production media. The greatest cellulase activity ($30.9 \pm 0.1 \text{ Uml}^{-1}$) was noted when urea was employed as the nitrogen source, while utilization of alternative nitrogen sources resulted in a reduction in enzyme activity. However, Singh et al. [18] proposed yeast extract and peptone as the most effective organic nitrogen sources for cellulase production in *Bacillus amyloliquefaciens* SS35, whereas Fouda et al. [28] suggested peptone as the preferable choice over yeast extract for optimizing cellulase production in *Bacillus amyloliquefaciens* M7. Upon examining the impact of the incubation period, we observed that cellulase production commenced 24 h after incubation and experienced a significant increase,

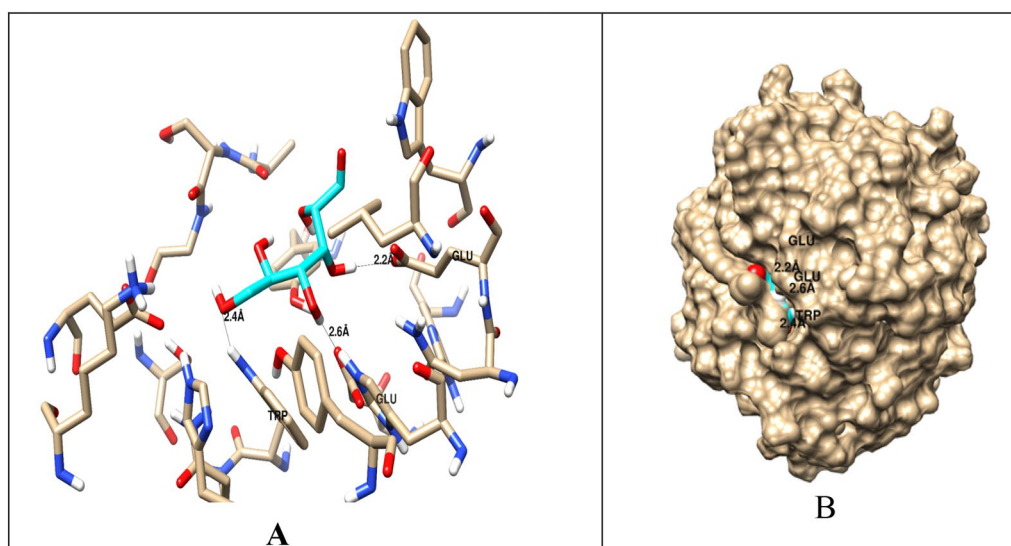


Fig. 15 **A** Chimera software was used to visualise the binding disposition and receptor-ligand interactions of carboxymethyl cellulose (the carbon skeleton is displayed in cyan) at the cellulase binding site. **B** 3D Complex Interaction

reaching its peak after 72 h. At the 72-h time point, the maximum cellulase production of $54.3 \pm 0.1 \text{ Uml}^{-1}$ was attained, after which it gradually declined. This decline can be elucidated by the bacterial strain's progression through different growth phases. Initially, during the initial 24 h period, the bacterial strain enters a lag phase, wherein it adapts to the new environmental conditions and metabolizes the available nutrients. Subsequently, a phase of exponential growth follows, characterized by a rapid proliferation of bacterial strains. However, after reaching an optimal incubation period, cellulase activity diminishes due to a decline in the metabolic activity of bacterial species, possibly due to nutrient depletion or the accumulation of detrimental metabolites in the growth medium. Our findings align with the results reported by Ye et al. [19] but differ from those of Fouada et al. [28]. They concluded that the optimal incubation period for maximizing cellulase production is 24 h. The concentration of enzyme–substrate within fermentative media, such as rice peel, plays a crucial role in augmenting enzymatic activity. Employing the *B. amyloliquefaciens* strain elh1, the highest recorded cellulase activity ($64.7 \pm 0.3 \text{ Uml}^{-1}$) was attained at a concentration of 1 g/L of rice peel. Beyond this threshold, cellulase activity gradually decreased due to substrate inhibition [33]. Inadequate availability of rice peel at lower concentrations (0.5 gL^{-1}) resulted in reduced cellulase activity due to insufficient substrate for the enzymes to act upon. Conversely, substrate inhibition was observed at higher concentrations of rice peel. Increased substrate concentrations led to the accumulation of end products

or intermediate metabolites, which hindered cellulase enzymes by disrupting active sites or catalytic activity, consequently diminishing enzyme efficacy. Hence, optimizing the equilibrium between enzyme activity and substrate concentration is imperative. Deviating from the optimal concentration range, whether lower or higher, negatively impacts enzyme activity either due to substrate insufficiency or substrate inhibition. It is important to note that when the concentration of rice peel increases, it absorbs the available liquid. This absorption then mitigates the agitation required for optimal aeration conditions. As a result, these conditions become unsuitable for ideal bacterial growth and consequently reduce cellulase production. The impact of inorganic salts on CMCase production is surprisingly minimal. As indicated by the negative Plackett–Burman model coefficient, all the factors exhibited negative effect on CMCase production except the potassium phosphate, which posed positive effect. It can be explained the well-known roles of potassium phosphate in cell growth-enhancing buffer solutions [32, 34, 35]

The detection of significant activity on carboxymethyl cellulose (CMC) suggests that the CMCase produced by *Bacillus amyloliquefaciens* strain elh1 may meet the criteria outlined by Coughlan and Mayer [36] for classification as an endoglucanase. Therefore, our investigation indicates that the enzyme produced by *Bacillus amyloliquefaciens* strain elh1 is likely functioning as an endo- β -1,4-glucanase, based on its specificity towards the substrate. The gene responsible for encoding an endo- β -1,4-glucanase was screened and

characterized. The designed primers, egl-F and egl-R successfully amplified a 1500 base pairs fragment. This fragment was then cloned into the pGEM Teasy vector. The amplified fragment corresponds to the egl gene found in *Bacillus amyloliquefaciens* strain elh. It was aligned with sequences in GenBank and translated into deduced amino acids to validate the sequence. In silico analysis confirmed that the amplified fragment corresponds to the egl gene responsible for CMCase production. Our findings are consistent with the results reported by Sun et al. [37]. In their study, Sun and his colleagues successfully isolated a gene that is 1500 base pairs long and encodes a specific type of cellulase found in *B. amyloliquefaciens* S1. Thakkar and Saraf [38] identified a gene responsible for encoding cellulase, composed of 1300 base pairs in length. Subsequently, they conducted cloning of the gene, that was cloned into the pCR4-TOPOR vector. Moreover, cellulase-encoding genes of 1500 base pairs in length were also discovered and sequenced in various *Bacillus* species, such as *B. subtilis* (natto strain) [39] and *Bacillus subtilis* IARI-SP-1 [40]. The use of heterologous expression is a valuable strategy for enhancing cellulase production. Although primarily a cloning vector, the pGEM-Teasy vector effectively facilitates heterologous gene expression, particularly when the gene product is non-toxic to the host organism, *E. coli* DH5 α . Numerous studies have employed the pGEM-Teasy vector for both cloning and heterologous gene expression. For example, Abdel-Salam et al. [41] and their colleagues successfully cloned and expressed the avicelase Gene from *Bacillus subtilis* subsp. *subtilis* 168 in *E. coli* DH5 α using the pGEM-Teasy vector. Similarly, from *Rhodotorula mucilaginosa*, Abd El-Aziz et al. [42] proficiently cloned and expressed the gene that encodes Endo-polygalacturonase into *E. coli* DH5 α via the pGEM-Teasy vector. Analysis of the inferred amino acid sequence of cellulase indicated its categorization within the Glycoside hydrolase family 5 (GH5) subfamily 1 (EC 3.2.1.4). Within GH5, Various conserved residues of amino acids were identified, namely histidine (H), asparagine (N), glycine (G), arginine (R), tyrosine (Y), glutamic acid (E), and tryptophan (W). These residues play a crucial task in the catalytic machinery, particularly glycine, arginine, and tryptophan, facilitating substrate binding and influencing hydrolytic activities [43]. Among these residues, two glutamic acids (E) were the most significant in the GH5 family. These residues of glutamic acid serve as proton donors and nucleophiles, respectively, thereby playing a pivotal role in the catalytic activity [8]. In our investigations the glutamic acid residues stand for Glu 169 and Glu 257.

Computational modelling and docking investigations are useful tools for exploring the relationship between bacterial cellulases and CMC. Investigating the

structure–function connection and substrate-protein interactions of proteins is important, despite their limitations in comparison to experimental methods, especially given their affordability [44]. Web-based servers validated the 3D structural model of cellulase by employing evaluations and quality assessments through QMEAN4, Z-score, Ramachandran plot analyses and ERRAT. The QMEAN score helps us to understand the geometric aspects of protein structures and the arrangement of variable residues. A higher QMEAN4 score indicates a better structure, while negative scores indicate instability [46]. QMEAN4 predicts the overall quality of model structures by combining four descriptors: local geometry, distance-dependent interaction, ensure consistency between the predicted secondary structure and solvent accessibility, as well as solvation potential. In this study, the QMEAN4 score for the 3D structure of cellulase was 0.08, indicating proper folding into a compact three-dimensional entity. Notably, desirable QMEAN scores ranged from 0 to 1 [45]. Validation of the 3D structures was further confirmed through crystallography, as represented by the ERRAT values that are associated with structural resolution, assessing protein structures based on the distances between pairs of atoms. Higher 3D structure resolutions typically yield rates of around 95% or higher, while lower resolutions indicate an average overall quality factor of approximately 91% [46]. In this study, interestingly, the cellulase structure exhibited an overall quality factor, as indicated by the ERRAT value of 94.96%, suggesting satisfactory structural resolution. Furthermore, a Ramachandran plot was established to visualize the position of each amino acid residue. As a result, 86.7% of the residues were located in the most favored regions. According to this resulted percentage, the constructed model is considered to be of good quality [47]. Molecular docking analysis revealed that carboxymethyl cellulase formed favorable interactions with *Bacillus amyloliquefaciens* strain elh1-cellulase within the active site of the enzyme. Among the amino acids involved, two glutamic acid residues were identified. Specifically, Glu 169 acts as a proton donor, facilitating the protonation of the glycosidic bond and subsequent bond fission. Additionally, Glu 257 acted as a nucleophile, aiding the reaction by stabilizing the resulting carbonium ion intermediate. These residues form hydrogen bonds with the substrate, carboxymethyl cellulase, within the active site.

Two main criteria were used to validate the docking protocol. These were the binding score and root mean square deviation (RMSD). The docking protocol involved the cellulase substrate-binding site and CMC ligand. With an RMSD of 1.5 Å and a significant interaction affinity score of -5.71 kcal/mol, the chosen pose was found to be reliable. It should be noted that an RMSD

below 2.0 Å indicates a favorable docking solution [48], along with the best scoring energy. Several other poses also met these criteria. To ensure accuracy in the selection process, it is highly advisable to select the best docking solution based on other structural considerations indicated for related ligands, in addition to the scoring function [49]. These docking orientations are consistent with previous research conducted by Maryanty et al. [50] who performed molecular docking of cellulose with cellulase as a ligand and identified the active site composed of Glu169, Glu257, and Trp207. The discrepancy lies in the amino acid tryptophan, where our results identified the tryptophan site as 219. Moreover, data from the web-based server <https://www.uniprot.org/uniprotkb/P07983/entry> verified the representation of the active site by the Glu169 and Glu257 residues. Additionally, these findings support previous studies by Santos et al. [52] who also identified Glu169 and Glu257 as critical residues involved in interactions with carboxymethyl cellulose (CMC).

Conclusion

In this investigation, we conducted the production of the cellulase enzyme using the bacterial strain *Bacillus amyloliquefaciens* strain elh1. We achieved optimal cellulase production from *B. amyloliquefaciens* after a fermentation period of 72 h. We found urea was the most suitable nitrogen source, while dextrose was the optimal sugar. This resulted in a production increase to 5.04 ± 0.120 U/ml. By employing statistical optimization through Response Surface Methodology (RSM), we reached a peak cellulase activity of 14.04 ± 0.42 U/ml.

Furthermore, we successfully cloned and expressed the cellulase encoding gene, *egl*, in *E. coli* DH5 α . The transformed cells exhibited a cellulase activity of 22.3 ± 0.24 U/ml. In silico analysis of the protein sequence provided insight into the physicochemical properties of the cellulase produced by *Bacillus amyloliquefaciens* strain elh1. Molecular docking helped elucidate the amino acids within the active sites that are involved in substrate binding, particularly carboxymethyl cellulose. Our future research plans to modify these specific amino acids through site-directed mutation to enhance cellulase functionality.

Materials and methods

Bacterial strains, growth media, and chemicals

The endophytic bacterial strain elh1 was isolated from rice straw. Non-transformed and transformed *E. coli* DH5 α cells were cultured at 37 °C in Luria–Bertani (LB) without and with the appropriate antibiotics (ampicillin

50 $\mu\text{g ml}^{-1}$), respectively. Kits for GeneJET Genomic DNA Purification, GeneJET Plasmid Miniprep, and T4 DNA Ligase, were purchased from Thermo Scientific™ (Waltham, Massachusetts, United States). Primers were manufactured by Macrogen Inc. (Amsterdam, the Netherlands). Takara Bio was the source of the EmeraldAmp® GT PCR 2 \times master mix (Shiga, Japan). Promega Co. provided the pGEM-T easy-cloning vector (Madison, WI, USA).

Primary screening and selection of the cellulase producing bacterial isolates

Nine different agricultural wastes were chosen for the purpose of isolation of potential cellulase producing bacteria: rice straw (RS), palm kernel (PK), wheat bran (WB), saw dust (SD), rice peel (RP), corn stalks (CS), pomegranate peel (PP), olive kernel (OK) and garlic peel (GP). After grinding of these dry wastes, one gram of powder for each was dissolved in 50 ml sterile saline solution under aseptic conditions in 250 ml conical flask, which then were permitted to shake at 150 rpm for 1 h. The suspensions were serially diluted until the concentration of 10^{-9} ; 100 μL from each dilution were plated onto the nutrient agar medium and subjected to incubation at a temperature of 37 °C until appearance of bacterial colonies. Colonies displaying distinct morphologies were isolated, purified, and preserved on nutrient agar slants at 4 °C; this procedure was executed following the methodology described by Waghmare et al. [52]. The cellulolytic/CMCase activity of the bacterial isolates was assessed on nutrient agar plates supplemented with 1% (W/V) carboxymethyl cellulose (CMC) following the method described by Vu et al. [53]. Briefly, the plates underwent incubation at 37 °C for 48 h, followed by flooding with a 0.1% Congo red solution and subsequently treated with 1 M NaCl. The presence of a transparent zone surrounding the bacterial growth indicated the hydrolysis of CMC. The colony exhibiting the most extensive clearance zone was chosen for molecular characterization through 16S rDNA sequencing.

Production and activity assay of extracellular CMCase enzyme

A single colony of the bacterium showing the largest zone of clearance was first cultured on nutrient broth overnight, until the optical density reached 0.8 at 600 nm—it was prepared as bacteria for inoculation. Then 0.5 ml of bacteria for inoculation were added to a 250 ml conical flask containing 50 ml of specific medium, namely Basic Liquid Media (BLM) for cellulase production; this media consisted of (g/l): glucose 0.5, peptone 0.75, FeSO $_4$ 0.01, KH $_2$ PO $_4$ 0.5, MgSO $_4$ 0.5 and 0.25 g from each agriculture

wastes added individually. The substrate served as the favorable source for higher cellulase enzyme activity was considered as suitable substrate for further studies. The inoculation medium was prepared and shaken at 180 rpm for 48 h at 37 °C. The culture medium was then centrifuged for 20 min at 4 °C and 8000 rpm. The supernatants were recovered as a source of the crude enzyme to determine the cellulase activity, the method was implemented properly as described by Lingouangou et al. [54]. Cellulase activity was measured following standardized procedure of Smogyi by estimating reducing sugar content [55]. In brief, 0.5 ml of cell free supernatant was taken as crude enzyme to form a reaction mixture with 1 ml of 1% (w/v) carboxymethyl cellulose (CMC) in 50 mM sodium phosphate buffer, pH 6.0 for 10 min at 50 °C. The reaction was stopped by adding 1 ml Somogyi copper reagent and boiling for 20 min in a water bath. The reaction mixture was first cooled down and then received 1 ml of Nelson reagent and completed to 25 ml with distilled water [56]. The density of the developed color was measured spectrophotometrically at 660 nm against blank—containing all the reagents except crude enzyme. One unit of enzyme activity is defined as the amount of enzyme that liberates 1 μ mole of reducing sugar, with mannose as a standard per min per ml of culture filtrate. The blank coincided with the experiment in which the culture filtrate was boiled before the reaction.

Molecular identification of the most efficient bacterium in CMCase production

Bacterial DNA purification and PCR amplification

The most cellulase-productive strain was designated as elh1. A single colony of elh1 was incubated overnight in nutrient broth (NB) at 30 °C. The bacterial culture was then centrifuged to obtain the pellet, which served as the source of genomic DNA. The GeneJET Genomic DNA Purification Kit (Thermo Scientific™, USA) was used to extract the genomic DNA. The 16S rRNA gene was amplified using universal bacterial primers: 8f forward and 1429R reverse primers. The sequences of these primers are as follows: 5'-AGAGTTTGGATCMTGGCTCAG' and 5'-TACGGYTACCTTGTTACGACTT'. The PCR reaction mixture (50 μ L) was prepared as follows: 2 μ L each of the forward and reverse primers (at a concentration of 10 pmol), 25 μ L of 2 \times EmeraldAmp® GT PCR Master Mix, 4 μ L of bacterial DNA template, and PCR-grade water adjusted to 50 μ L. The PCR conditions were as follows: a three-minute initial denaturation at 95 °C, followed by 35 cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 50 °C, and elongation for one minute at 72 °C. A BioRad T100 Thermal Cycler was used for the last extension phase, which was carried out for ten more minutes at 72 °C. The target fragment was visualized on

an ethidium bromide-stained 1.2% agarose gel. Purification of the gel was performed using the GeneJET Gel Extraction Kit (Thermo Scientific™, USA). The sequencing of the target gene was carried out by MacroGen Inc. (Amsterdam, the Netherlands) using the Applied Biosystems 3730XL sequencer.

Phylogenetic analysis

BioEdit 7.1.10 was used to assemble the sequences [57]. The sequences were compared with those in the GenBank database (<http://www.ncbi.nlm.nih.gov/blast>) to identify comparable species. A collection of 16S rRNA genes from 13 related species available in the GenBank database was employed to construct a phylogenetic tree. Multiple sequence alignments were carried out using the MUSCLE algorithm [58] in MEGA11 [59]. The evolutionary history was inferred using the neighbor-joining method [60] with a 1000-bootstrap runs [61], and evolutionary distances were computed using the Jukes and Cantor method [62].

Optimization of CMCase production conditions

One-variable-at-a-time approach (OVAT)

In this experiment, BLM medium components were changed to enhance the cellulase activity by identification of the key components of the medium. Each experiment was replicated three times to minimize variation. Results were expressed as mean \pm standard deviation. Experimental factors included varying incubation durations (24, 48, 72, 96, and 120 h); rice peel, identified as the most effective waste for cellulase production, was incorporated at different concentrations (g/100 ml: 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0); nitrogen sources (yeast, peptone, corn steep, casein, (NH₄)₂SO₄, and urea) were added at a concentration equivalent to that in the original culture medium; sugar sources (sucrose, xylose, glucose, galactose, lactose, dextrose, and fructose) were included at a concentration of two grams per liter. Also, the resulting most efficient organic nitrogen source, urea, was then used at diverse concentrations (0.05, 0.15, 0.25, 0.35, 0.5, 0.6, 0.7 and 0.8 gm g/L) in the culture medium to determine the optimal concentration for cellulase production. Also, dextrose as the most efficient sugar for cellulase production was utilized in different concentrations (g/100 ml): 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8. All experiments were supplied with inoculum size of 1% (v/v) and established at 37 °C, pH 7, for 48 h. Each parameter was examined individually while maintaining all other conditions at their optimal levels. At the conclusion of the specified incubation period for each parameter, approximately 5 ml of the incubated medium was collected and subjected to centrifugation at 4 °C for 10 min at 7000 rpm. The resulting supernatants were then utilized for the determination of cellulase activity via the aforementioned assay method. Each parameter was tested

in triplicate. This method was implemented following that of Fouada et al. [28]. Results were expressed as mean ± S.D.; preliminary screening results will serve as a basis for the RSM experiment.

Statistical optimization of rice peel based-medium for CMCase production

The Plackett–burman design (PBD)

The relative significance of the different components within the medium was assessed and evaluated using the Plackett–Burman experimental design. Seven key factors with the greatest impact on enzyme production were selected for optimization. In order to pinpoint the critical factors that promote increased cellulase production, a total of seven variables were examined carefully, encompassing six components of the media (rice peel (X1), Urea (X2), Dextrose (X3), MgSO₄·7H₂O (X4), KH₂PO₄ (X5), and CaCl₂ (X6)), and one cultivation parameter (inoculum size (X7)). Eleven experiments were conducted, with each variable designated and utilized at either high (+) or low (–) concentration. [30]. As indicated in Table 1, Each row denotes a test iteration, while each column signifies the concentration of a distinct independent variable. The experimental data were fitted in the following linear regression Eq. (1):

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i \tag{1}$$

where Y represents the response for cellulase enzyme activity (U/ml), β₀ is the model intercept, β_i is the linear coefficient, and X_i is the level of the independent variable.

The effect of each variable was determined by the following Eq. (2)

$$E(X_i) = 2 \left(\sum Mi+ - \sum Mi- \right) / N \tag{2}$$

where Mi+ and Mi– represent the response for cellulase enzyme activity (U/ml) from trials in which the Mi+ and Mi– are the activity percentage in trials. The independent variable (X_i) was varied at both high and low concentrations, denoted as such, with N representing the total number of trials. The standard error (SE) was computed using the square root of each effect’s variance. The significance of each concentration effect was determined through Student’s t-test, with a significance level set at p < 0.05, as per Eq. (3).

$$t(X_i) = E(X_i) / SE \tag{3}$$

The effect of variable X_i is denoted by E(X_i). Regression analysis of the experimental data was conducted using SPSS Version 15.0.

Box–Behnken design (BBD)

Based on the findings from the initial one-factor-at-a-time (OFAT) experiments, the experimental setup and statistical study were executed using Design-Expert software, trial version 11.0 (Stat-Ease Inc., Minneapolis, USA). The BBD (Box-Behnken design) three-level, three-factor design was used, comprising of 17 experimental runs. In this study, three variables were included: (X₁) rice peel, (X₂) Urea, and (X₃) Dextrose. Each variable varied over three coded levels of –1, 0, +1; the designation of “high” was represented as (+1), “medium” as (0), and “low” as (–1) [63]. The generalized second-order polynomial model employed in the RSM is presented in Eq. (4):

$$Y_{Activity} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 \tag{4}$$

where Y represents the predicted response [i.e. CMCase activity (U/ml)]; β₀ is the model constant; X₁, X₂, and X₃ are the coded input variables which influence the response variable; β₁, β₂, and β₃ are the linear coefficients; β₁₂, β₁₃, and β₂₃ are the cross-product coefficients; β₁₁, β₂₂, and β₃₃ are the quadratic coefficients. All experiments were performed three times independently. All experimental outcomes and the corresponding standard deviations were derived from the mean of triplicate trials. A paired t-test was performed using the Statistical Package for Social Sciences (SPSS) for Microsoft Windows Version 15.0 to determine the actual and anticipated responses.

Statistical evaluation of the model

The model underwent statistical analysis to evaluate the analysis of variance (ANOVA). The significance of the model equation was determined by Fisher’s test value, while the proportion of variance explained by the model was indicated by the estimation of multiple coefficients for each variable. Quadratic models were visualized using contour plots (3D), and response surface curves were generated using Design-Expert software, trial version 11.0 (Stat-Ease Inc., Minneapolis, USA). The predictive accuracy of the polynomial model equation was assessed using the coefficient of determination R² and adjusted R².

Primer Design and egl Gene amplification

The full-length egl gene sequence, annotated as an endo-1,4-β-glucanase, was retrieved from the genomic DNA of *Bacillus amyloliquefaciens*, Reference genome ASM1939692v1. We have performed an automated UniProt Blast using this sequence as query, which recorded

98.8% identity with uniprotkb/P07983/entry (Endoglucanase from *Bacillus subtilis* as a reviewed entry). Additionally, the *egl* gene sequence retrieved from EMBL (<https://www.ebi.ac.uk/ena/browser/api/embl/MK675502>) under the MK675502.1 entry was utilized for the primer design. Primer3 program (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) was utilized for this purpose. The sense primer, *egl*-F, has a nucleotide sequence of 5'-GAGGCTCATGAAACGGTC AATCTCTATT-3', emphasizing a *Sac*I recognition site, and the antisense primer, *egl*-R, has a sequence of 5'-GCA TGCCTAATTTGGTTCTGTTCCCCAA-3', emphasizing an *Sph*I recognition site. PCR-based amplification of the *egl* gene was carried out using Phusion High-Fidelity DNA Polymerase from Thermo Scientific™ according to its provided protocol, with an annealing temperature of 50 °C. Phusion DNA Polymerases exhibit robust performance, short protocol times, tolerance to PCR inhibitors, and produce higher yields with lower enzyme amounts compared to other DNA polymerases.

Gene cloning and transformation

A 1.2% agarose gel was used to electrophorese the PCR product, and the band believed to include the desired gene was removed. The GeneJET Gel Extraction Kit from Thermo Scientific (Waltham, Massachusetts, United States) was used for processing. A 1500 bp fragment compatible with the ORF of the *egl* gene was isolated from the agarose gel using the extraction kit. The purified product was then ligated into the pGEM-T Easy-cloning vector using T4 DNA Ligase (Thermo Scientific™) following the recommended procedures. The ligated product was transformed into competent *E. coli* DH5 α cells (100 μ L) prepared according to the method outlined by Chung et al. [64]. The recombinant plasmid was introduced into *E. coli* DH5 α cells via the heat shock method: the mixture was immediately added to 500 μ L of LB media in a sterilized 2 ml Eppendorf, which was shaken for two hours at 37 °C to permit the expression of antibiotic ampicillin gene. Subsequently, an appropriate volume was plated on LB agar plates containing 0.5 mM IPTG, 50 μ g/ml ampicillin, and 40 μ g X-gal, followed by overnight incubation at 37 °C. White colonies were selected and examined for the presence of the recombinant vector. Purification of recombinant vector was accomplished using Plasmid Miniprep Kit. Standard protocols for restriction endonuclease digestions, agarose gel electrophoresis, DNA purification from agarose gels, DNA ligation, and other cloning-related techniques were employed as described by Sambrook and Russell [65]. Two techniques were used to evaluate the integrity of the recombinant plasmids: colony PCR and double digestion of the recombinant vector using *Sac*I and *Sph*I restriction enzymes. The

ORF of *egl* gene was subjected to DNA sequencing and the recombinant vector was designated as pGEM-*egl*. The positive transformant cells were subsequently transferred to LB agar media supplemented with 1% CMC (w/v) to observe clear zone. Then approximately 0.5 ml of overnight culture derived from positive transformant cells, exhibiting an optical density of 0.8 at 600 nm, was introduced into a 250 ml conical flask containing 50 ml of Basic Liquid Media (BLM) supplemented with 0.25 g of rice peel. The flasks were incubated for 72 h at 37 °C with continuous shaking at 180 rpm. After incubation, the culture medium was centrifuged for 20 min at 8000 rpm and 4 °C. The resulting supernatant was collected as the crude enzyme source, which was utilized for cellulase activity determination. The same procedure was followed for the not-transformed parent *E. coli* strain.

DNA sequencing and in silico analysis of *egl* gene

The nucleotide sequence was determined using the dideoxynucleotide chain termination method with the specific primers, *egl*-F and *egl*-R [66]. Following sequencing, editing was conducted to correct inaccuracies and trim unreadable portions at the 3' and 5' ends using BioEdit version 7.0.2 software. The edited sequence was compared against the NCBI nucleotide database (<https://blast.ncbi.nlm.nih.gov>) to determine its taxonomic identity, and a unique accession number was assigned to the sequence. To identify homologous proteins, the deduced amino acid sequence of the *egl* gene was aligned with the UniProt protein database (<https://legacy.uniprot.org/align/>). The ESPript 3.0 program was used to incorporate superimposed predicted secondary structures into the alignment, providing insights into the structural features of the *egl* gene product [67]. The Conserved Domain Database, accessible via the NCBI website (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), was searched to investigate the conserved domains. This analysis helped to determine the functional domains and potential catalytic regions of the endoglucanase enzyme. Additionally, version 6.0 of the SignalP software (<http://www.cbs.dtu.dk/services/SignalP/>) was employed to analyze the N-terminal signal peptide, which is crucial for the secretion of the enzyme [68]. To investigate the evolutionary relationships between the deduced *egl* amino acid sequence and other homologous proteins, the MEGA11 software was utilized [59]. The MUSCLE algorithm was used to conduct multiple sequence alignment [58], and the neighbor-joining method was applied to infer the phylogenetic relationships [60]. A 1000-replicate bootstrap test was conducted to assess the statistical support for the branching patterns in the phylogenetic tree [61].

Template search, comparative modeling and model confirmation

The ExPasy translate tools, accessible at web.expasy.org/translate/, were used to translate nucleotide sequences and align deduced amino acid sequences. A homology model was constructed using SWISS-MODEL [69–72]. The translation of the cellulase encoded by the *egl* gene provided the protein sequence input for the model. In order to guarantee adequate coverage of the query sequence and sequence identity within the template library, a template search was carried out on the SWISS-MODEL web server using Blast and HHBlits in order to construct the model. The target sequence was compared to the primary amino acid sequence in the SMTL database using the BLAST algorithm [73]. To choose the most trustworthy 3D structure, it is essential to evaluate the values of both the Qualitative Model Energy Analysis (QMEAN) and Global Model Quality Estimation (GMQE) [75]. more numbers, which usually lie between 0 and 1 for GMQE values, suggest greater reliability of the projected structure. Higher reliability is indicated by a QMEAN value below 4.0 in the model's quality assessment [76].

Structure validation of modelled protein

To validate the structure of the modeled protein, we ran the following tools: the ProSA server (<https://prosa.services.came.sbg.ac.at/prosa.php>) and SAVES version 6.0 (Structure Analysis and Verification Server). SAVES v6.0 includes five programs: PROCHECK [53], VERIFY-3D [77], and ERRAT [78]. All these tools were utilized to assess the 3D protein models, evaluate the quality of the model by examining the allowed and disallowed regions on the plot, and determine the similarity of the model to native nuclear magnetic resonance/X-ray crystal structures [79]. This comprehensive validation process ensures the accuracy and reliability of the modeled protein structure by assessing its compliance with established structural standards and principles. In particular, the Ramachandran plot and statistics were used to examine the permitted and prohibited areas on the plot in order to assess the quality of the model. A Z score value was produced by the ProSA web service [79], which provided information about the model's general quality and similarity to native nuclear magnetic resonance/X-ray crystal structures. This comprehensive validation process ensures the accuracy and reliability of the modeled protein structure by assessing its conformity to established structural standards and principles.

Alignment of the CMCase model and the template structure

The alignment between the modelled CMCase structure and the template structure was implanted by the PyMOL molecular viewer [80] and proximity of the carbon atoms also illustrated. The root mean square deviation (RMSD) is the main metric used to measure the difference in carbon atom locations between the template and model. The degree of structural similarity between two entities is larger when the RMSD is lower, almost nil [81].

Molecular docking

A computational analysis was conducted using the Molecular Operating Environment (MOE) software Version 2015.10 [82]. Preparation of the modeled protein, CMCase, and ligand Carboxymethyl cellulose (PubChem CID 24748) involved removing water molecules and adding hydrogen atoms. Subsequently, a molecular database (MDB) file containing the protein's 3D structure attached to the ligand for docking simulations was created [83]. Docking poses were chosen based on the ratings and root mean square deviation values. Amino acids in the active site were determined. Chimaera was used to visualize the receptor-binding site and analyze ligand-receptor interactions [84], with a focus on key amino acid residues involved in hydrogen bonding. To evaluate the interactions, established criteria for molecular interactions were followed [85].

Data analysis

Data processing was done using Design-Expert software, trial version 11.0 (Stat-Ease Inc., Minneapolis, USA).

Acknowledgements

We extend our thanks to Academy of Scientific Research and Technology (ASRT) and National Research Centre for providing us with access to facilities, necessary chemicals, and instruments essential for the advancement of our research

Author contributions

E.A.; Conceptualization, methodology E.S.; Conceptualization, visualization and supervision. G.E.; Conceptualization, methodology, software, validation, writing—original draft preparation, writing—review and editing. S.N.; Conceptualization, methodology, software, validation, writing—original draft preparation, writing—review and editing. M.A.; validation, investigation and supervision. M.H.; Conceptualization, validation, investigation and supervision. All authors have read and agreed to the published version of the manuscript.

Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Availability of data and materials

All data generated or analyzed during this study are included in this article.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of National Research Centre, Cairo, Egypt.

Consent for publication

Approved by all named authors.

Competing interests

The authors declare no competing interests.

Author details

¹Microbial Genetics Department, Biotechnology Research Institute, National Research Centre, 33 El-Bohouth St., (Former El-Tahrir St.) Dokki, P.O. 12622, Giza, Egypt. ²Chemistry of Natural and Microbial Products Department, Pharmaceutical and Drug Industries Research Institute, National Research Centre, 33 El-Bohouth St., Dokki, P.O. 12622, Giza, Egypt. ³Botany and Microbiology Department, Faculty of Science, Benha University, Benha 13511, Egypt.

Received: 21 March 2024 Accepted: 7 June 2024

Published online: 02 July 2024

References

- Pham VHT, Kim J, Shim J, Chang S, Chung W. Coconut mesocarp-based lignocellulosic waste as a substrate for cellulase production from high promising multienzyme producing *Bacillus amyloliquefaciens* FW2 without pretreatments. *Microorganisms*. 2022;10:327. <https://doi.org/10.3390/microorganisms10020327>.
- Hassan O, Khalil MS, Elnaway AS. Biodegradation of agricultural Wastes by cellulase producing bacteria. *IJSET Int J Innov Sci Eng Technol*. 2022;9(1):218–32.
- Weiping Y, Fanxu M, Jiayin P, Peng H, Fang F, Li M, Binyun C. Isolation and identification of a cellulolytic bacterium from the tibetan pig's intestine and investigation of its cellulase production. *Electron J Biotechnol*. 2014;17(6):262–7. <https://doi.org/10.1016/j.ejbt.2014.08.002>.
- Dumorné K, Córdova DC, Astorga-Eló M, Renganathan P. Extremozymes: a potential source for industrial applications. *J Microbiol Biotechnol*. 2017;27:649–59.
- Bala Kumaran MD, Kalachelvan PT, Santhi R. Exploitation of agro-industrial wastes as substrates for cellulase production by *Bacillus licheniformis* MTCC 429. *Microbiol J*. 2015;5(2):36–42.
- Shweta S, Arun D, Arun G. Enhanced catalytic efficiency of *Bacillus amyloliquefaciens* SS35 endoglucanase by ultraviolet directed evolution and mutation analysis. *Renew Energy*. 2020;151:1124–33. <https://doi.org/10.1016/j.renene.2019.11.105>.
- Maryanty Y, Sumitro SB. Interaction of enzyme-substrate from indigenous cellulolytic bacteria by bioinformatics. *IOP Conf Ser: Mater Sci Eng*. 2020;854:012068. <https://doi.org/10.1088/1757-899X/854/1/012068>.
- Sanjaya RE, Putri KDA, Kurniati A, et al. In silico characterization of the GH5-cellulase family from uncultured microorganisms: physicochemical and structural studies. *J Genet Eng Biotechnol*. 2021;19:143. <https://doi.org/10.1186/s43141-021-00236-w>.
- Arsov A, Petrov K, Petrova P. Enhanced activity by genetic complementarity: heterologous secretion of clostridial cellulases by *Bacillus licheniformis* and *Bacillus velezensis*. *Molecules*. 2021;26(18):5625.
- Whittle DJ, Kilburn DG, Warren RAJ, Miller JRC. Molecular cloning of a cellulomonas firmi cellulase gene in *Escherichia coli*: recombinant DNA; plasmid pBR322; immunoassay. *Gene*. 1982;17:139–45.
- Laymon RA, Adney WS, Mohagheghi A, Himmel ME, Thomas SR. Cloning and expression of full-length *Trichoderma reesei* cellobiohydrolase I cDNA in *Escherichia coli*. *Appl Biochem Biotechnol*. 1996;57–58:389–97.
- Penttilä M, Andre L, Lehtovaara P, Bailey M, Teeri TT, Knowles J. Efficient secretion of two cellobiohydrolases by *Saccharomyces cerevisiae*. *Gene*. 1988;63:103–12.
- Zahri S, Zamani MR, Motallebi M, Sadeghi M. Cloning and characterization of CBH II gene from *Trichoderma parceramosum* and Its expression in *Pichia pastoris*. *Iran J Biotechnol*. 2005;3:204–15.
- Lu M, Chang J, Wang P, Yin Q, Dang X, Gao T, Lu F. Study on cellulase gene expressed in pichia pastoris and analyses of its biochemical characters. *J Mater Sci Chem Eng*. 2018;6:43–52.
- Zafar A, Aftab MN, Asif A, Karadag A, Peng L, Celebioglu HU, Afzal MS, Hamid A, Iqbal I. Efficient biomass saccharification using a novel cellobiohydrolase from *Clostridium clariflavum* for utilization in biofuel industry. *RSC Adv*. 2021;11:9246–61.
- Sidar A, Albuquerque ED, Voshol GP, Ram AFJ, Vijgenboom E, Punt PJ. Carbohydrate binding modules: diversity of domain architecture in amylases and cellulases from filamentous microorganisms. *Front Bioeng Biotechnol*. 2020;8:871.
- Coughlan MP. *Cellulose degradation by fungi*. Enzyme Microb Technol. 2nd ed. Dordrecht: Springer; 1990. p. 1–36.
- Singh S, Moholkar VS, Goyal A. Optimization of carboxymethylcellulase production from *Bacillus amyloliquefaciens* SS35. *Biotech*. 2014;4(4):411–24. <https://doi.org/10.1007/s13205-013-0169-6>.
- Ye M, Sun L, Yang R, Wang Z, Qi KZ. The optimization of fermentation conditions for producing cellulase of *Bacillus amyloliquefaciens* and its application to goose feed. *R Soc*. 2017;4: 171012. <https://doi.org/10.1098/rsos.171012>.
- Aslam S, Hussain A, Qazi JI. Production of cellulase by *Bacillus amyloliquefaciens*-ASK11 under high chromium stress. *Waste Biomass Valor*. 2019;10:53–61. <https://doi.org/10.1007/s12649-017-0046-3>.
- Bouzaiane T, Ziadi M, Enneifer M, Sellami A, Aydi A, Cherif A, Hamdi M. Cellulolytic bacillus strain: production optimization using wheat bran under solid-state fermentation and investigation of its probiotic potential. *Sustainability*. 2023;15:8394. <https://doi.org/10.3390/su15108394>.
- Tabssum F, Irfan M, Shakir HA, Qazi JI. RSM based optimization of nutritional conditions for cellulase mediated Saccharification by *Bacillus cereus*. *J Biol Eng*. 2018;3(12):7. <https://doi.org/10.1186/s13036-018-0097-4>.
- Aanchal NA, Kanika DG, Goyal A. Response surface methodology for optimization of microbial cellulase production. *Rom Biotechnol Lett*. 2016;21(5):11832–41.
- Maryanty Y, adryati and Suharjo and Sumitro. TSB. Interaction of enzyme-substrate from indigenous cellulolytic bacteria by bioinformatics. *IOP Conf Ser Mater Sci Eng*. 2020;854(1):012068. <https://doi.org/10.1088/1757-899X/854/1/012068>.
- Selvam K, Senbagam D, Selvankumar T, Sudhakar C, Kamala-Kannan S, Senthilkumar B, Govarthanan M. Cellulase enzyme: homology modeling, binding site identification and molecular docking. *J Mol Struct*. 2017;1150:61–7. <https://doi.org/10.1016/j.molstruc.2017.08.067>.
- Leo VV, Lalthafala Z, Ramesh N, Singh BP. Microorganisms as an efficient tool for cellulase production: availability, diversity, and efficiency. In: *New and Future Developments in Microbial Biotechnology and Bioengineering*. Amsterdam: Elsevier; 2019. p. 45–61.
- Islam F, Roy N. Screening, purification and characterization of cellulase from cellulase producing bacteria in molasses. *BMC Res Notes*. 2018;11:445. <https://doi.org/10.1186/s13104-018-3558-4>.
- Fouda A, Alshallah KS, Atta HM, El-Gamal MS, Bakry MM, Salem AMI, SS. A thermo-tolerant cellulase enzyme produced by *Bacillus amyloliquefaciens* M7, an insight into synthesis, optimization, characterization, and biopolishing activity. *Green Process Synth*. 2023;12(1):20230127. <https://doi.org/10.1515/gpps-2023-0127>.
- Ahmad T, Sharma A, Gupta G, Mansoor S, Jan S, Kaur B, Paray BA, Ahmad A. Response surface optimization of cellulase production from *Aneurini-bacillus aneurinilyticus* BKT-9: an isolation of urban himalayan freshwater. *Saudi J Biol Sci*. 2020;27(9):2333–43. <https://doi.org/10.1016/j.sjbs.2020.04.036>.
- Shajahan S, Ganesh Moorthy I, Sivakumar N, Selvakumar G. Statistical modeling and optimization of cellulase production by *Bacillus licheniformis* NCIM 5556 isolated from the hot spring Maharashtra, India., *J King Saud Univ-Sci*. 2017;29(3):302–10. <https://doi.org/10.1016/j.jksus.2016.08.001>.
- Lee YJ, Kim BK, Lee BH, Jo KI, Lee NK, Chung CH, et al. Purification and characterization of cellulase produced by *Bacillus amyloliquefaciens* DL-3 utilizing rice hull. *Bioresour Technol*. 2008;99:378–86. <https://doi.org/10.1016/j.biortech.2006.12.013>.
- Jin IH, Jing DY, Son CW, Kim SK, Gao W, Chung CH, Lee JW. Enhanced production of heteropolysaccharide-7 by *beijerinikia indica* HS-2001 in repeated batch culture with optimized substitution of culture medium. *Biotechnol Bioprocess Eng*. 2011;16:245–55.

33. Kokkonen P, Beier A, Mazurenko S, Damborsky J, Bednar D, Prokop Z. Substrate inhibition by the blockage of product release and its control by tunnel engineering. *RSC Chem Biol*. 2021;2:645–55. <https://doi.org/10.1039/DOCB00171F>.
34. Gao W, Lee EJ, Lee SU, Li J, Chung CH, Lee JW. Enhanced carboxymethyl-cellulase production by a newly isolated marine bacterium *cellulophaga lytica* LBH-14 using rice bran. *J Microbiol Biotechnol*. 2012;22:1412–22.
35. Singh S, Moholkar VS, Goyal A. Isolation, identification and characterization of a cellulolytic *Bacillus amyloliquefaciens* strain SS35 from rhinoceros dung. *ISRN Microbiol*. 2013. <https://doi.org/10.1155/2013/728134>.
36. Coughlan MP, Mayer F. Cellulose-decomposing bacteria and their enzyme systems. In: Balows A, Truper HG, Dworkin M, Harder W, Schleifer KH, editors. *The prokaryotes*. New York: Springer; 1992. p. 460–516.
37. Sun L, Cao J, Liu Y, Wang J, Guo P, Wang Z. Gene cloning and expression of cellulase of *Bacillus amyloliquefaciens* Isolated from the cecum of goose. *Anim Biotechnol*. 2017;28(1):74–82. <https://doi.org/10.1080/10495398.2016.1205594>.
38. Thakkar A, Saraf M. Molecular cloning and in silico analysis of cellulase gene from *Bacillus amyloliquefaciens* MBAA 3. *Indian J Biotechnol*. 2017;16:487–94.
39. Vadala BS, Deshpande S, Apte-Deshpande A. Soluble expression of recombinant active cellulase in *E. coli* using *B. subtilis* (natto strain) cellulase gene. *J Genet Eng Biotechnol*. 2021;19(1):7. <https://doi.org/10.1186/s43141-020-00103-0>.
40. Pandey S, Kushwah J, Tiwari R, Kumar R, Somvanshi VS, Nain L, Saxena AK. Cloning and expression of β -1, 4-endoglucanase gene from *Bacillus subtilis* isolated from soil long term irrigated with effluents of paper and pulp mill. *Microbiol Res*. 2014;169(9–10):693–8. <https://doi.org/10.1016/j.micres.2014.02.006>.
41. Abdel-Salam MS, Hafez SS, Fadel M, Mohamed SAH, Hegazy WK, Khalil BE. Bio ethanol production from rice straw saccharification via avicelase gene in *E. coli* recombinant strain coli recombinant strain. *Clean Technol*. 2023;5:451–65. <https://doi.org/10.3390/cleantechnol5020023>.
42. Abd El-Aziz NM, Moharam ME, El-Gamal NN, et al. Enhancement of novel endo-polygalacturonase expression in *Rhodotorula mucilaginosa* PY18: insights from mutagenesis and molecular docking. *Microb Cell Fact*. 2023;22:252. <https://doi.org/10.1186/s12934-023-02253-5>.
43. Yan J, Liu W, Li Y, Lai H-L, Zheng Y, Huang J-W, Chen C-C, Chen Y, Jin J, Li H, Guo R-T. Functional and structural analysis of *Pichia pastoris* expression *Aspergillus niger* 1,4- β -endoglucanase. *Biochem Biophys Res Commun*. 2016;475:8–12. <https://doi.org/10.1016/j.bbrc.2016.05.012>.
44. Sehar U, et al. Three-dimensional (3D) structure prediction and substrate-protein interaction study of the chitin binding protein CBP24 from *B. thuringiensis*. *Bioinformation*. 2013;9(14):725–9. <https://doi.org/10.6026/97320630009725>.
45. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011;27:343–50. <https://doi.org/10.1093/bioinformatics/btq662>.
46. Bhattacharya M, Hota A, Kar A, Sankar Chini D, Chandra Mallick R, Chandra Patra B, Kumar DB. In silico structural and functional modelling of anti-freeze protein (AFP) sequences of ocean pout (*Zoarces americanus*, bloch & schneider 1801). *J Genet Eng Biotechnol*. 2018;16:721–30. <https://doi.org/10.1016/j.jgeb.2018.08.004>.
47. Saleem A, Rajput S. Insights from the in silico structural, functional and phylogenetic characterization of canine lysyl oxidase protein. *J Genet Eng Biotechnol*. 2020;18:20. <https://doi.org/10.1186/s43141-020-00034-w>.
48. Mena-Ulecia K, Tiznado W, Caballero J. Study of the differential activity of thrombin inhibitors using docking, qsar, molecular dynamics, and MM-GBSA. *PLoS ONE*. 2015;10:e0142774. <https://doi.org/10.1371/journal.pone.0142774>.
49. Ramírez D, Caballero J. Is it reliable to take the molecular docking top scoring position as the best solution without considering available structural data? *Molecules*. 2018;23(5):1038. <https://doi.org/10.3390/molecules23051038>.
50. Maryanty Y, Adryati T, Suharjo Sumitro SB. Interaction of enzyme-substrate from indigenous cellulolytic bacteria by bioinformatics. *IOP Conf Ser: Mater Sci Eng*. 2020. <https://doi.org/10.1088/1757-899X/854/1/012068>.
51. Santos CR, Paiva JH, Sforca ML, Neves JL, Navarro RZ, Cota J, Akao PK, Hoffmam ZB, Meza AN, Smetana JH, Nogueira ML. Dissecting structure–function–stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochem J*. 2012;441(1):95–104. <https://doi.org/10.1042/BJ20110869>.
52. Waghmare PR, Patil SM, Jadhav SL, Jeon BH, Govindwar SP. Utilization of agricultural waste biomass by cellulolytic isolate *Enterobacter* sp. SUK-Bio Agr Nat Resour. 2018;52:399–406.
53. Vu V, Farkas C, Riyad O, Bujna E, Kilin A, Sipiczki G, Sharma M, Usmani Z, Gupta VK, Nguyen QD. Enhancement of the enzymatic hydrolysis efficiency of wheat bran using the bacillus strains and their consortium. *Bioresour Technol*. 2022. <https://doi.org/10.1016/j.biortech.2021.126092>.
54. Lingoungou TM, Ampa R, Nguimbi E, Bissombolo PM, Ngo-Itsouhou MFAS, Ngoulou TB, Nzaou SAE. Optimization of cellulase production conditions in bacteria isolated from soils in brazzaville. *J Biosci Med*. 2022;10(8):14–28.
55. Smogyi M. Notes on sugar determination. *J Biol Chem*. 1952;195:19–23.
56. Nelson N. A photometric adaptation of the somogyi method for the determination of glucose. *J Biol Chem*. 1944;153:375–80.
57. Hall TA. Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/ NT. *Nucleic Acids Symp Ser*. 1999. https://doi.org/10.14601/phytopathol_mediterr-14998u1.29.
58. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340>.
59. Tamura K, Stecher G, Kumar S. Mega11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–7. <https://doi.org/10.1093/molbev/msab120>.
60. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4(4):406–25. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>.
61. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evol Int J Org Evol*. 1985;39(4):783–91. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>.
62. Jukes TH, Cantor CR. Evolution of protein molecules, in mammalian protein metabolism. Amsterdam: Elsevier; 1969. p. 21–132. <https://doi.org/10.1016/B978-1-4832-3211-9.50009-7>.
63. Zambare V, Jacob S, Din MFM, Ponraj M. Box-behnken design-based optimization of the saccharification of primary paper-mill sludge as a renewable raw material for bioethanol production. *Sustainability*. 2023;15(13):10740. <https://doi.org/10.3390/su151310740>.
64. Chung CT, Niemela SL, Miller RH. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A*. 1989;86(7):2172–5. <https://doi.org/10.1073/pnas.86.7.2172>.
65. JSambrook and D Russell. *Molecular cloning : a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press. 2001, 2024;10. <https://ci.nii.ac.jp/ncid/BA4950703>.
66. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463–7.
67. Robert X, Guet P. Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res*. 2014;42:W320–324. <https://doi.org/10.1093/nar/gku316>.
68. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Brunak S, Heijne GV, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37(4):420–3. <https://doi.org/10.1038/s41587-019-0036-z>.
69. Biasini M, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, nucleic acids Res; Web Server issue. 2014;42:W252–258. <https://doi.org/10.1093/nar/gku340>.
70. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinforma Oxf Engl*. 2006;22(2):195–201. <https://doi.org/10.1093/bioinformatics/bti770>.
71. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res*. 2009;37:D387–392. <https://doi.org/10.1093/nar/gkn750>.
72. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*. 2009;30(1):162–73. <https://doi.org/10.1002/elps.200900140>.
73. Altschuls F, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402. <https://doi.org/10.1093/nar/25.17.3389>.

74. Cardoso JM, Fonseca L, Egas C, Abrantes I. Cysteine proteases secreted by the pinewood nematode, *bursaphelenchus xylophilus*: in silico analysis. *Comput Biol Chem*. 2018;77:291–6.
75. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res*. 2009;37:510–4. <https://doi.org/10.1093/nar/gkp322>.
76. Odusele GO, Ajani OO, Ajamma YU, Brors B, Adebijoyi E. Homology modeling and molecular docking studies of selected substituted benzo[d]imidazol-1-yl) methyl) benzimidamide scaffolds on plasmodium falciparum adenylosuccinate lyase receptor. *Bioinforma Biol Insights*. 2019;13:1177932219865533. <https://doi.org/10.1177/1177932219865533>.
77. Colovos C, Yeates TO. Verification of protein structures: patterns of non-bonded atomic interactions. *Protein Sci Publ Protein Soc*. 1993;2(9):1511–9. <https://doi.org/10.1002/pro.5560020916>.
78. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992;356(6364):83–5. <https://doi.org/10.1038/356083a0>.
79. Rishad RK, Sherin V, Jisha MS. Sequence analysis and docking performance of extracellular chitinase from *bacillus pumilus* MCB-7, a novel mangrove isolate. *Enzyme Microb Technol*. 2020;140: 109624. <https://doi.org/10.1016/j.enzmictec.2020.109624>.
80. PyMOL | pymol.org." 2024; 10: <https://pymol.org/2/>
81. Bramucci E, Paiardini A, Bossa F, Pascarella S. PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinform*. 2012;13(4):S2. <https://doi.org/10.1186/1471-2105-13-S4-S2>.
82. Iftikhar F, et al. Design, synthesis, in-vitro thymidine phosphorylase inhibition, in-vivo antiangiogenic and in-silico studies of C-6 substituted dihydropyrimidines. *Bioorganic Chem*. 2018;80:99–111. <https://doi.org/10.1016/j.bioorg.2018.05.026>.
83. Silvestri A, Barone G, Ruisi G, Lo Giudice MT, Tumminello S. The interaction of native DNA with iron (III)-N, N'-ethylene-bis(salicylideneiminato)-chloride. *J Inorg Biochem*. 2004;98(4):589–94. <https://doi.org/10.1016/j.jinorgbio.2004.01.010>.
84. Pettersen EF, et al. UCSF chimera-a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605–12. <https://doi.org/10.1002/jcc.20084>.
85. Bissantz C, Kuhn B, Stahl M. A medicinal chemist's guide to molecular interactions. *J Med Chem*. 2010;53(14):5061–84. <https://doi.org/10.1021/jm100112j>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.