GMR
The Original

INTEGRATED IN SILICO STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF CELLULASE ENZYMES FROM CELLULOLYTIC YEASTS

Bilwashri H[1], Seema J Patel[2*], Gajendrakumar Pujar [2], Ganesh G Tilve [3], Shivalingsarj Desai[4]

[1]Department of Biotechnology, BIET, Davangere, Karnataka, India
[2*]Department of Studies of Biotechnology, Davangere University, Davangere, Karnataka, India.
[3]Department of Biotechnology, FET, SE, G M University, Davangere, Karnataka, India.
[4]Department of Biotechnology, KLE Technological University, B V Bhoomaraddi campus, Hubballi, Karnataka, India.
[*]Corresponding author: - Seema J Patel, ssjp176@gmail.com

ABSTRACT

Yeasts are microscopically eukaryotic fungi that are distributed everywhere in nature, but are most abundant in environments with rich sugar content, such as fruits and flower nectar; over 1,500 species of yeasts have been found, and one of them, *Saccharomyces cerevisiae*, is a model organism that has been most intensely studied. Few yeasts are now found to produce cellulases. An integrated in silico approach was employed to comparatively characterize cellulase enzymes from the isolates *Vanrija humicola*, *Geotrichum candidum*, and *Aureobasidium pullulans*, focusing on their structural, functional, and regulatory attributes. Protein sequences were retrieved from UniProt and assessed for homology by BLAST. Amino acid composition, molecular weight, and basic physicochemical properties, instability index, molar extinction coefficient, and the prediction of secondary structural elements were performed. Functional domains, conserved motifs, phosphorylation sites, transmembrane regions, and salt-bridge compositions were identified to evaluate protein functionality and stability. Physicochemical parameters were assessed by ProtParam, while signal peptides and transmembrane helices were predicted using the DeepTMHMM programs, respectively. Three-dimensional modeling of structures was performed using SWISS-MODEL and trRosetta, and model quality was further validated by QMEAN, and InterPro analyses. Other analyses included motif scanning, phosphorylation site prediction by means of NetPhos. The in-silico characterization described here is useful in understanding structure-function relationships in yeast proteins, it therefore supports further applications in biotechnology and industrial research. Sequence homology and evolutionary analyses revealed strong conservation of catalytic residues and motifs typical of endo-β-1,4-glucanases, indicating a shared mechanism of cellulose hydrolysis. Overall, these findings underscore the functional complementarity of fungal cellulases, integrating structural robustness with catalytic flexibility. The distinct yet synergistic properties of the analyzed enzymes support their relevance as promising candidates for targeted applications in lignocellulosic biomass conversion, bioethanol production, and industrial biotechnology, while providing a rational framework for future experimental validation and enzyme engineering strategies.

KEYWORDS: Fungal Cellulase, cellulolytic yeast, Insilico studies, Sequence similarity, Structure prediction.

INTRODUCTION

Cellulases are very specific biocatalysts that work together to break down sugars, especially glucose, which is very useful in industry because it can be turned into bioethanol. Many different kinds of microorganisms in nature make cellulases. Some of these microorganisms are called "truly cellulolytic" because they can break down natural cellulose quickly and easily [1]. Cellulase is a key enzyme in biorefineries based on lignocellulose and plays a central role in a variety of industrial applications. Due to its broad utilization, a large number of prokaryotic and eukaryotic microorganisms have been evaluated for effective cellulase production. However, yeasts have only rarely been investigated for their ability to degrade plant cell walls, despite their potential as strong and industrially viable producers of enzymes [2]. By expressing cellulolytic enzymes in bacteria and yeast, respectively, the cost of cellulase production can be reduced and some of the steps required for the pretreatment of making ethanol from cellulosic biomass can be simplified. Recently, attention has been focused on constructing ethanologenic bacterial and yeast strains capable of directly converting cellulosic materials to ethanol. For instance, using genetic engineering techniques, recombinant *Klebsiella oxytoca* SZ21 was constructed that was capable of producing ethanol from amorphous cellulose, although the yield was not high. Even with such progress, in most reports involving recombinant ethanologenic microbes for cellulose fermentation, supplementation with commercial cellulase is still required to achieve effective ethanol production [3].

Microbial enzymes have various advantages over enzymes of plant and animal origin. They are inexpensive and

can be mass-produced in a short period and using limited space. Besides, microbial enzymes are readily extractable and purifiable, and microorganisms can produce a broad range of enzymes in varying environmental conditions, thus showing extreme genetic flexibility. Therefore, the identification and isolation of yeast species with cellulolytic activity indicates a key development in enzyme technology, offering an outline to improve the yield and versatility of cellulase production [4].

The term "proteome" describes the complete set of proteins that are produced by a genome, cell, tissue, or organism. Despite revolutionary advances in sequencing technologies, most protein sequences predicted remain functionally uncharacterized. The major challenge, therefore, involves the generation of statistical, comparative, structural, and functional information for these sequences. Such analyses shall be collectively essential to accomplish an integrated understanding of biological systems at the gene, transcript, protein, and functional level [5]. Protein analysis forms the basis of most biological data being used for drug development and healthcare. Proteins are large, complex molecules that form the very basis of cellular function. They drive biochemical processes through enzymatic action and modulate a variety of metabolic reactions, hence the need to study and comprehend proteins to understand their functions and their mode of action within the human body [6].

This study presents an integrated in silico structural and functional characterization of cellulase enzymes sourced from cellulolytic yeasts. Using a robust bioinformatics pipeline, the study sets out to provide information on the primary sequence features, physicochemical properties, and secondary and tertiary structures, including conserved domains, functional motifs, and evolutionary relationships of yeast-derived cellulases. These results therefore provide important insights into enzyme stability, catalytic potential, and suitability for industrial applications and thus would inform the rational design of effective cellulase systems for sustainable biotechnological processes.

MATERIALS AND METHODS

Three isolates *Vanrija humicola*, *Geotrichum candidum* and *Aureobasidium pullulans* were selected for the work as previous lab work results showed them to be cellulolytic isolates from fruit wastes.

1. Protein Sequence Retrieval

Protein sequences were retrieved from the UniProt Knowledgebase (UniProtKB), a comprehensive, curated database that provides high-quality protein sequence data along with functional annotations. FASTA-format sequences of cellulase proteins from *Vanrija humicola* (UniProt accession: A0A7D8Z3G1), *Geotrichum candidum* (UniProt accession: A0A9P5KPQ9), and *Aureobasidium pullulans* (UniProt accession: A0A4V4K0C7) were retrieved from the UniProt Knowledgebase and used for subsequent in silico analyses.

2. Sequence Similarity and Homology Analysis

Protein sequence similarity searches were performed using BLASTP (Basic Local Alignment Search Tool for Proteins) against the non-redundant (nr) protein database to identify homologous sequences of cellulase proteins (7). Additionally, PSI-BLAST (Position-Specific Iterative BLAST) was employed to detect distant homologs of these cellulase proteins by generating position-specific scoring matrices through iterative database searches. This approach enhanced sensitivity for identifying evolutionarily conserved regions and functionally important residues across diverse taxa (8).

3. Secondary Structure Prediction

Secondary structure prediction of the cellulase protein sequences was performed using SOPMA (Self-Optimized Prediction Method with Alignment). SOPMA predicts secondary structural elements, including α-helices, β-sheets, β-turns, and random coils, with an overall prediction accuracy of approximately 69.5%. The method is based on multiple sequence alignments derived from a reference dataset comprising
126 non-homologous protein chains sharing less than 25% sequence identity, allowing reliable identification of conserved structural features (9).

4. Physicochemical Characterization

The physicochemical properties of the cellulase protein sequences were analysed using the ProtParam tool available on the ExPASy server. ProtParam was employed to calculate key parameters, including amino acid composition, molecular weight, theoretical isoelectric point (pI), atomic composition, extinction coefficient, instability
index, aliphatic index, and the grand average of hydropathicity (GRAVY). These parameters provide insights into protein stability, solubility, thermostability, and overall biochemical behavior (10).

5. Homology Modeling and Tertiary Structure Prediction

Three-dimensional (3D) structure prediction of the cellulase protein sequences was carried out using SWISS-MODEL, a web-based automated homology modeling server that generates 3D protein structures based on

experimentally resolved template structures. Additionally, trRosetta (Template-based Rosetta) was employed for protein structure prediction by integrating deep learning–based inter-residue distance and orientation predictions with Rosetta modeling protocols. trRosetta enables accurate structure prediction for protein sequences ranging from 10 to 1000 amino acid residues, particularly in cases where suitable homologous templates are limited (11).

6. Model Quality Assessment

The quality of the predicted three-dimensional protein models of cellulases from was *Vanrija humicola, Geotrichum candidum*, and *Aureobasidium pullulans* evaluated using QMEAN (Qualitative Model Energy Analysis). QMEAN is a composite scoring function that provides both global and local quality estimates by comparing structural features of the predicted models with those of high-resolution experimentally determined protein structures, thereby enabling reliable assessment of model accuracy and structural reliability (12).

7. Motif and Domain Analysis

Protein motifs and functional domains of the cellulase sequences from were identified using MotifScan, which scans protein sequences such as *Vanrija humicola, Geotrichum candidum*, and *Aureobasidium pullulans* against curated motif and pattern databases to detect conserved functional signatures. Functional annotation was further supported using InterPro, an integrated resource that combines multiple protein family, domain, and functional site databases to provide comprehensive protein characterization and classification.

8. Transmembrane Helix Prediction

Prediction of transmembrane helices and membrane topology of the cellulase protein sequences from *Vanrija humicola, Geotrichum candidum*, and *Aureobasidium pullulans* was performed using TMHMM (Topology Prediction using Hidden Markov Models). TMHMM is a widely used computational tool that identifies transmembrane regions and predicts membrane topology in protein sequences based on probabilistic hidden Markov model–based approaches.

9. Phosphorylation Site Prediction

Potential phosphorylation sites in the cellulase protein sequences from *Vanrija humicola, Geotrichum candidum*, and *Aureobasidium pullulans* were predicted using NetPhos, a neural network–based tool that identifies serine, threonine, and tyrosine phosphorylation sites in eukaryotic proteins. This analysis provides insight into potential post-translational modifications that may influence protein regulation, stability, and function.

RESULTS AND DISCUSSIONS

1. Sequence Retrieval and Homology Analysis

Cellulase protein sequences from *Vanrija humicola* (A0A7D8Z3G1), *Geotrichum candidum* (A0A9P5KPQ9), and *Aureobasidium pullulans* (A0A4V4K0C7) were retrieved from UniProtKB and analyzed for sequence similarity. BLASTP results showed complete or near-complete query coverage with extremely low E-values, confirming strong homology and evolutionary conservation among these cellulases. *V. humicola* exhibited the highest sequence similarity, while *G. candidum* and *A. pullulans* also retained conserved catalytic motifs and functional domains characteristic of endo-β-1,4-glucanases (Tables 1–3). This conserved structural framework highlights their shared role in cellulose degradation and supports their potential as efficient enzymes for industrial and bioethanol-related applications.

2. Secondary Structure Prediction:

Secondary structure analysis revealed marked differences among the three fungal cellulases. The *Geotrichum candidum* cellulase (~527 aa) was dominated by random coils (~85%), with minimal α-helices (~3%) and β-strands (~10%), indicating high structural flexibility typical of linker-rich and surface-exposed regions that may enhance substrate accessibility. In contrast, the *Vanrija humicola* cellulase (~592 aa) exhibited substantially higher α-helical content (~34%) and reduced disorder, suggesting a compact and stable architecture often associated with enhanced thermostability and industrial robustness. The *Aureobasidium pullulans* cellulase (~341 aa) displayed an intermediate profile, balancing ordered (α-helices ~31%, β-strands ~16%) and disordered regions. Overall, these structural differences likely contribute to variations in stability and functional performance, positioning *V. humicola* cellulase as a robust candidate for stability-demanding applications, while *G. candidum* and *A. pullulans* cellulases may be advantageous where conformational flexibility is beneficial as shown in Fig 1.

3. Physicochemical Properties:

Among the three enzymes, cellulase from *Vanrija humicola* showed the largest protein size (592 amino acids), the highest extinction coefficient (~144,410 M^{-1} cm^{-1}), and a stable instability index (38.18), indicating greater aromatic residue content, structural stability, and improved suitability for spectrophotometric analysis and

purification. Its higher aliphatic index (75.10) also suggests enhanced thermostability.
In contrast, cellulases from *Geotrichum candidum* and *Aureobasidium pullulans* exhibited similar physicochemical properties, including higher instability indices (44.23), lower aliphatic indices (47.83), and strongly negative GRAVY values (– 0.400), reflecting lower predicted stability but greater hydrophilicity typical of extracellular fungal cellulases. Overall, *V. humicola* cellulase appears more favorable for industrial applications, while the other two enzymes may be better suited for efficient cellulose degradation under mild aqueous conditions (Table 4).

4. Homology Modeling and Tertiary Structure Prediction:

Homology modeling using trRosetta revealed a conserved cellulase fold with distinct structural features among *Vanrija humicola*, *Geotrichum candidum*, and *Aureobasidium pullulans*. The *V. humicola* cellulase exhibited a compact and highly ordered tertiary structure with excellent model reliability (TM-score ~0.93), indicating strong structural stability. In contrast, the *G. candidum* cellulase showed greater conformational flexibility due to extended linker regions (TM-score ~0.64), which may enhance substrate accessibility, while *A. pullulans* displayed an intermediate architecture combining a well-defined catalytic core with moderate flexibility (TM-score ~0.92) as shown in Fig 2. Collectively, these structural differences likely underlie species-specific variations in stability and catalytic behavior, highlighting their complementary potential for diverse industrial and biotechnological applications.

5. Model Quality Assessment

Structural quality assessment using SWISS-MODEL revealed species-specific differences among the modeled cellulases. The *Geotrichum candidum* model showed the highest reliability, with a favorable QMEAN4 Z-score (–0.45) and uniformly high local quality across the β-sheet–rich catalytic core. The *Vanrija humicola* cellulase exhibited moderate quality (QMEANDisCo = 0.46), characterized by a well-defined globular core and increased flexibility in terminal regions. In contrast, the *Aureobasidium pullulans* model displayed lower overall quality (QMEAN4 = –3.67), reflecting deviations mainly within extended loop regions, although the α-helical core remained structurally coherent as shown in Fig 3. Collectively, all models retained structurally plausible catalytic domains, supporting their functional relevance despite variations in conformational flexibility.

6. Motif and Domain Analysis

InterProScan analysis revealed that the cellulase proteins from *Geotrichum candidum, Vanrija humicola, and Aureobasidium pullulans* belong to the endoglucanase 1 family within glycoside hydrolase family 45 (GH45). All sequences contained a conserved GH45 catalytic domain spanning the majority of the protein length, confirming their role in endo-β-1,4- glucan hydrolysis. In addition, one or more carbohydrate-binding module family 1 (CBM1) domains were identified, primarily in the N-terminal region, which are characteristic of fungal cellulases and facilitate efficient binding to crystalline cellulose. Short intrinsically disordered regions were also detected, likely acting as flexible linkers between functional domains. The conserved modular architecture observed across all three species supports their functional role in lignocellulosic biomass degradation and highlights their potential applicability in bioethanol and industrial biotechnology processes.
MotifFinder analysis highlights distinct glycosyl hydrolase families and domain organizations among the three cellulolytic yeasts. *G. candidum* and *A. pullulans* encode soluble extracellular cellulases, whereas *V. humicola* possesses a membrane-associated cellulase with additional surface-related motifs. These variations likely reflect different ecological strategies for cellulose degradation and may influence enzyme efficiency, substrate interaction, and industrial applicability as shown in Fig 4.

7. Transmembrane Helix Prediction

TMHMM analysis reveals that cellulases from *G. candidum* and *A. pullulans* are likely secreted extracellular enzymes, whereas the cellulase from *V. humicola* contains a C-terminal transmembrane helix, suggesting membrane association. These differences in membrane topology may influence enzyme localization, substrate accessibility, and functional roles in lignocellulosic biomass degradation.

8. Phosphorylation Site Prediction

NetPhos 3.1a analysis predicted multiple potential phosphorylation sites in the cellulase proteins from *Geotrichum candidum, Vanrija humicola,* and *Aureobasidium pullulans*. Phosphorylation was predominantly observed on serine and threonine residues, with tyrosine phosphorylation occurring less frequently. The predicted sites were distributed throughout the protein sequences, including within the catalytic regions, suggesting a role in post- translational regulation of enzyme activity, stability, and secretion.
Notably, *V. humicola* and *A. pullulans* displayed a higher density of predicted phosphorylation sites, indicating a greater potential for regulatory control and environmental adaptability.

CONCLUSION

The comparative bioinformatics evaluation of cellulase enzymes from *Vanrija humicola*, *Geotrichum candidum*, and *Aureobasidium pullulans* demonstrates a conserved catalytic framework underlying cellulose hydrolysis, accompanied by distinct structural and regulatory adaptations. Preservation of GH45 catalytic domains and CBM1 carbohydrate-binding modules across all three enzymes confirms a shared functional core essential for efficient interaction with crystalline cellulose.

Despite this conservation, pronounced species-specific differences were evident in structural organization, conformational dynamics, and regulatory features. The compact and α-helix- enriched architecture of the *V. humicola* cellulase suggests enhanced structural stability, supporting its potential suitability for industrial applications requiring tolerance to challenging physicochemical conditions. In contrast, the increased flexibility observed in *G. candidum* and *A. pullulans* cellulases, driven by disordered regions and linker segments, may confer adaptive advantages for substrate accessibility and catalytic performance in heterogeneous lignocellulosic substrates.

Variations in domain organization, membrane association, and transmembrane topology further indicate divergent ecological strategies for enzyme localization and substrate utilization. The presence of multiple predicted phosphorylation sites, particularly in *V. humicola* and *A. pullulans*, highlights the role of post-translational regulation in modulating cellulase activity, stability, and secretion in response to environmental cues. Overall, these findings underscore the functional complementarity of fungal cellulases, integrating structural robustness with catalytic flexibility. The distinct yet synergistic properties of the analyzed enzymes support their relevance as promising candidates for targeted applications in lignocellulosic biomass conversion, bioethanol production, and industrial biotechnology, while providing a rational framework for future experimental validation and enzyme engineering strategies.

REFERENCES

1. Goldbeck, R., Andrade, C. C. P., Pereira, G. A. G., & Maugeri Filho, F. (2010). *Screening and identification of cellulase-producing yeast-like microorganisms from Brazilian biomes*. University of Campinas (UNICAMP).
2. Sohail, M., Barzkar, N., Michaud, P., Tamadoni Jahromi, S., Babich, O., Sukhikh, S., Das, R., & Nahavandi, R. (2022). Cellulolytic and xylanolytic enzymes from yeasts: Properties and industrial applications. *Molecules, 27*(12), 3783. https://doi.org/10.3390/molecules27123783.
3. Fujita, Y., Ito, J., Ueda, M., Fukuda, H., & Kondo, A. (2004). Synergistic saccharification and direct fermentation to ethanol of amorphous cellulose by an engineered yeast strain codisplaying three types of cellulolytic enzymes. *Applied and Environmental Microbiology, 70*(2), 1207–1212.
4. Kim, J.-Y., Jung, H.-Y., Park, J.-S., Cho, S.-J., Lee, H. B., Sung, G.-H., Subramani, G., & Kim, M. K. (2019). Isolation and characterization of cellulolytic yeast belonging to *Moesziomyces* sp. from the gut of grasshopper. *Korean Journal of Microbiology, 55*(3), 234–241. https://doi.org/10.7845/kjm.2019.9056.
5. Pruess, M., & Apweiler, R. (2002). Bioinformatics resources for *in silico* proteome analysis. European Bioinformatics Institute, EMBL Outstation, Hinxton, Cambridge, UK.
6. Singh, N., Upadhyay, S., Jaiswar, A., & Mishra, N. (2016). *In silico analysis of protein*. Journal of Bioinformatics, Genomics and Proteomics, 1(2), 1007.
7. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
8. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389.
9. Geourjon, C., & Deléage, G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer Applications in the Biosciences, 11*(6), 681–684. https://doi.org/10.1093/bioinformatics/11.6.681
10. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In
11. J. M. Walker (Ed.), The proteomics protocols handbook (pp. 571–607). Humana Press. https://doi.org/10.1385/1-59259-890-0:571
12. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and
13. complexes. *Nucleic Acids Research, 46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427
14. Benkert, P., Tosatto, S. C. E., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics, 71*(1), 261–277. https://doi.org/10.1002/prot.21715
15. Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version

7.0 for bigger datasets. *Molecular Biology and Evolution, 33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Table 1. Top 2 clusters showing significant BLAST alignments of *Vanrija humicola* cellulase

Cluster No.	Representative Sequence	Organism	Accession	Max Score	E-value	Identity (%)	Query Coverage (%)
1	Hypothetical protein VHum_00002	*Vanrija humicola*	TXT15499.1	1191	0.0	100	100
2	Endoglucanase E-4	*Vanrija pseudolonga*	XP_062622428.1	983	0.0	82.89	100

Table 2. Top 2 clusters producing significant B L A S T alignments for *Geotrichum candidum* cellulase

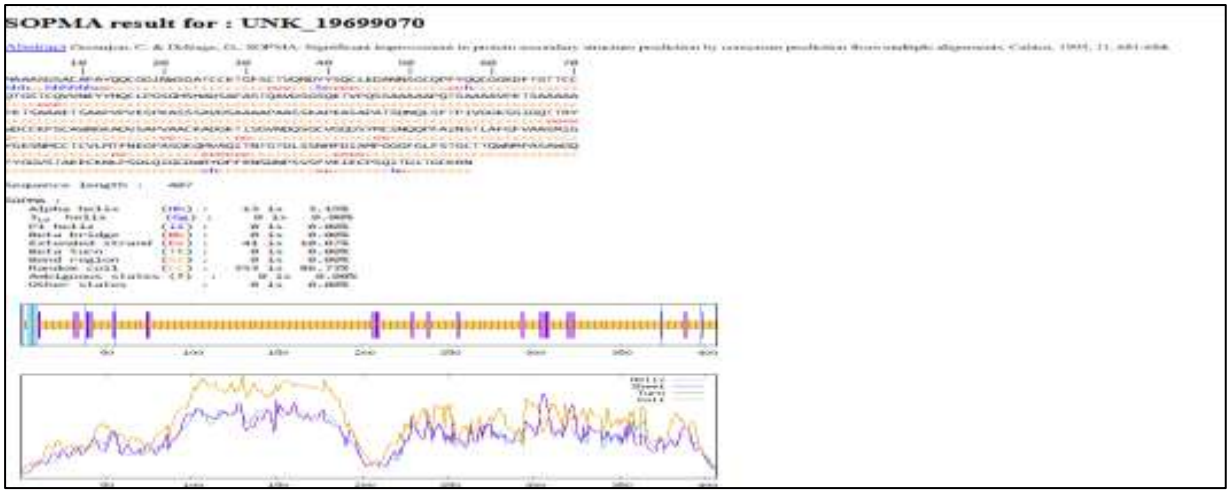
Cluster No.	Representative Protein	Organism	Accession No.	Max Score	Query Coverage (%)	Identity (%)	E-value
1	Hypothetical protein DV451_004422	*Geotrichum candidum*	KAF5096033.1	1058	100	100.00	0.0
2	Hypothetical protein DZ00_001766	*Geotrichum galactomycetum*	KAF5099171.1	748	100	72.02	0.0

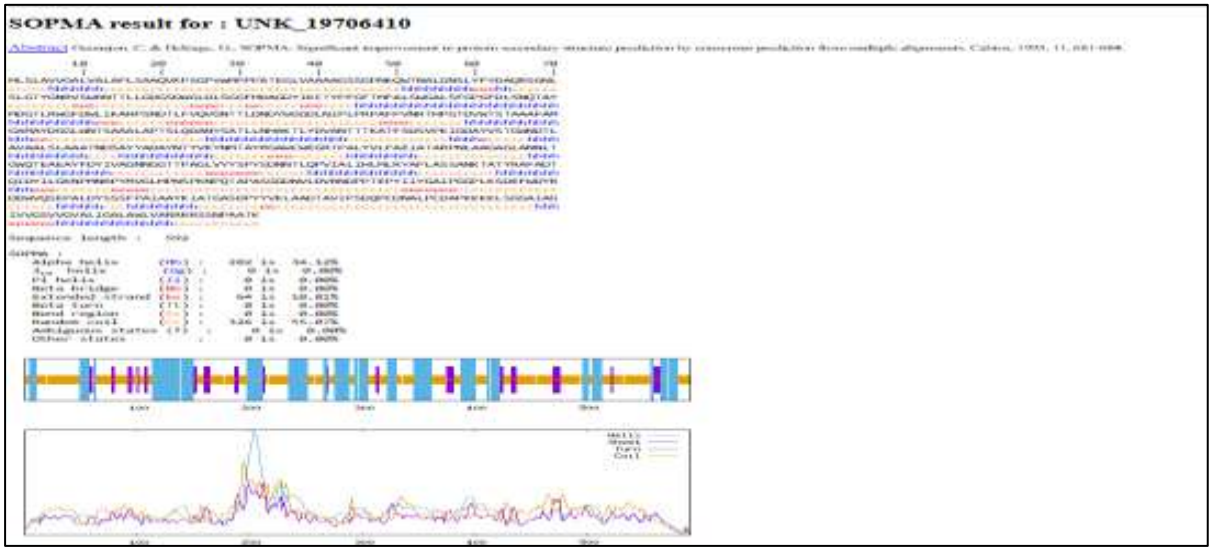
Table 3. Top 2 clusters producing significant BLAST alignments for *Aureobasidium pullulans* cellulase

Cluster No.	Representative Protein	Organism	Accession No.	Max Score	Query Coverage (%)	Identity (%)	E-value
1	Unnamed protein product	*Aureobasidium pullulans*	CAD0037178.1	686	100	98.25	0.0
2	Glycoside hydrolase	*Aureobasidium subglaciale*	KAI5202486.1	634	100	88.99	0.0

Table 4: The physicochemical properties of cellulases from *Vanrija humicola, Geotrichum candidum,* and *Aureobasidium pullulans* predicted using ProtParam.

Parameter	***Vanrija humicola*** **(A0A7D8Z3G1)**	***Geotrichum candidum*** **(A0A9P5KPQ9)**	***Aureobasidium pullulans*** **(A0A4V4K0C7)**
Protein length (aa)	592	526	~530*
Total number of atoms	8,464	7,118	7,118
Extinction coefficient (M^{-1} cm^{-1}, 280 nm)	~144,410	79,475	79,475
Estimated half-life (mammalian reticulocytes, in vitro)	30 h	4.4 h	4.4 h
Estimated half-life (yeast, in vivo)	10 min	>20 h	>20 h
Estimated half-life (E. coli, in vivo)	10 h	>10 h	>10 h
Instability index	38.18 (Stable)	44.23 (Unstable)	>10 h
Aliphatic index	75.10	47.83	44.23 (Unstable)
GRAVY value	–0.175	–0.400	47.83
Overall hydropathicity	Moderately hydrophilic	Strongly hydrophilic	Strongly hydrophilic





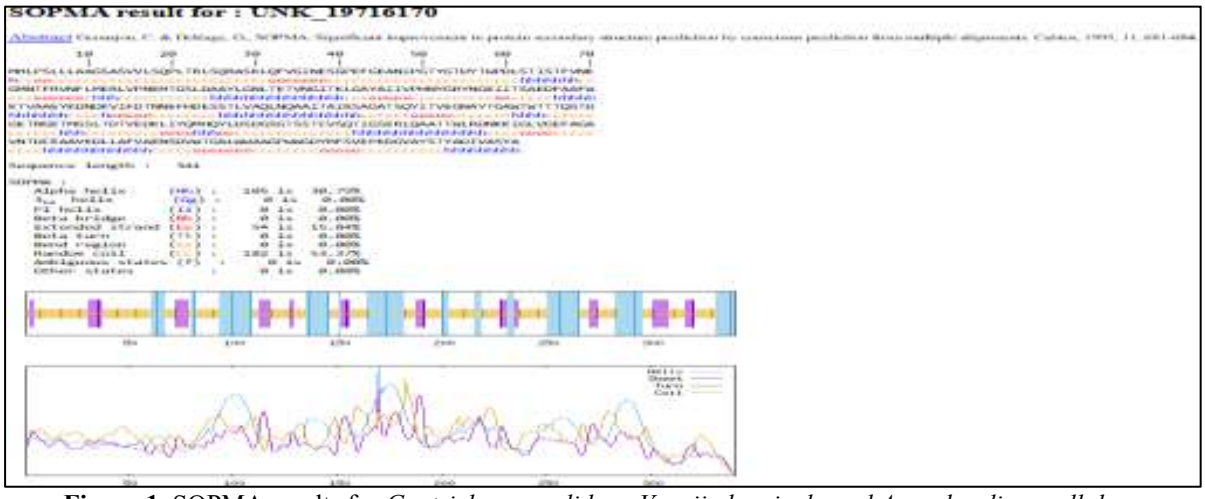
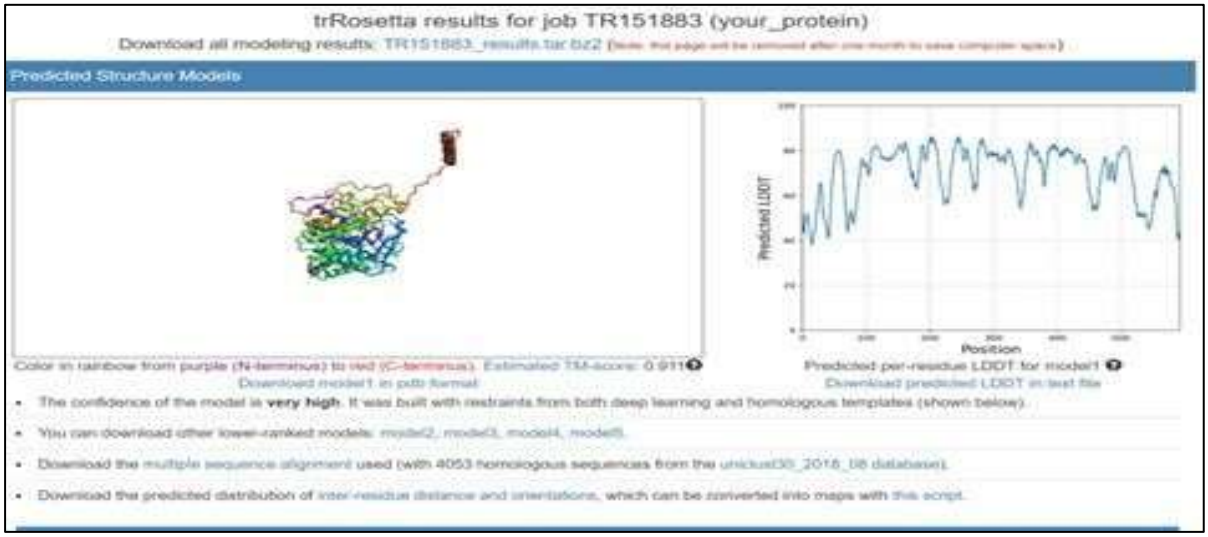


Figure 1. SOPMA results for *Geotrichum candidum, Vanrija humicola and Aureobasdium pullulans*

A total of 548 templates were found to match the target sequence. This list was filtered by a heuristic down to 50. The top templates are:

Template	Sequence Identity	Biounit Oligo State	Description
A0ATD8Z3G1.1	100.00	monomer	cellulase AlphaFold DB model of A0ATD8Z3G1_IARHU (gene: A0ATD8Z3G1_IARHU, organism: Iarrija humicola (Yeast) (Cryptococcus humicola))
1f94.2	37.50	monomer	T. FUSCA ENDOEXO-CELLULASE E4 CATALYTIC DOMAIN AND CELLULOSE-BINDING DOMAIN ENDOEXOCELLULASE FROM THERMOMONOSPORA
1f94.2	38.07	monomer	T. FUSCA ENDOEXO-CELLULASE E4 CATALYTIC DOMAIN AND CELLULOSE-BINDING DOMAIN ENDOEXOCELLULASE FROM THERMOMONOSPORA
8v4f.1	32.33	monomer	Endoglucanase Crystal Structure of BCel9A from Glycoside Hydrolase Family 9 in Complex with Cellohexaose
1ga2.1	31.87	monomer	ENDOGLUCANASE 9G THE CRYSTAL STRUCTURE OF ENDOGLUCANASE 9G FROM CLOSTRIDIUM CELLULOLYTICUM COMPLEXED WITH CELLOBIOSE

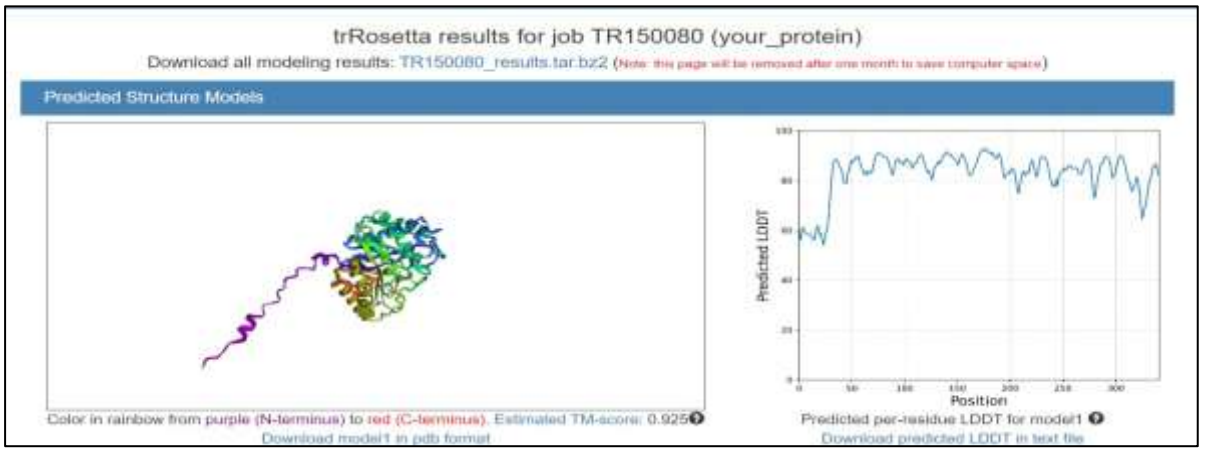
Show full template details



Template Results

A total of 1732 templates were found to match the target sequence. This list was filtered by a heuristic down to 50. The top templates are:

Template	Sequence Identity	Biounit Oligo State	Description
A0A4S9NV51.1	98.83	monomer	UniProtKB entry unknown, most likely obsolete AlphaFold DB model of A0A4S9NV51 (gene: unknown, organism: unknown)
6zlg.1	65.11	monomer	Endo-beta-1, 4-glucanase Crystal structure of a beta-1,4-endoglucanase from Bispora sp. MEY-1
6zlg.1	65.82	monomer	Endo-beta-1, 4-glucanase Crystal structure of a beta-1,4-endoglucanase from Bispora sp. MEY-1
6zlk.1	64.69	homo-tetramer	cellulase Crystal structure of a beta-1,4-endoglucanase from Bispora sp. MEY-1 in complex with cellotetraose
6zlk.1	64.49	homo-tetramer	cellulase Crystal structure of a beta-1,4-endoglucanase from Bispora sp. MEY-1 in complex with cellotetraose



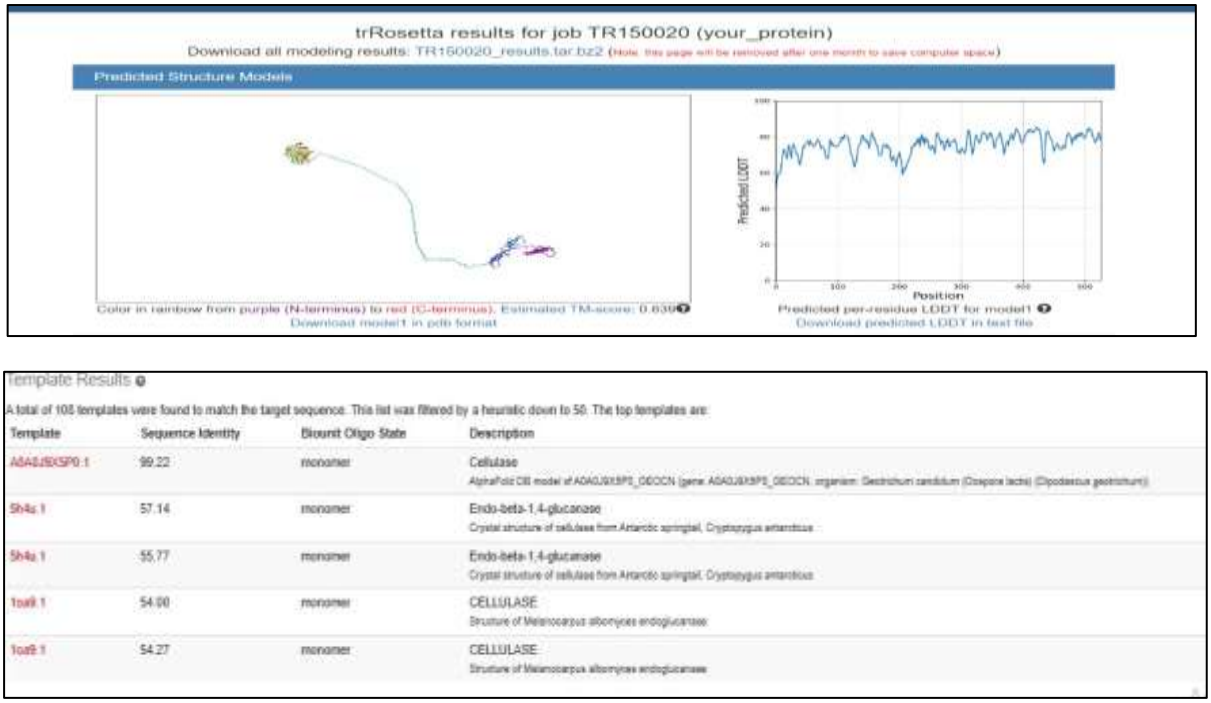
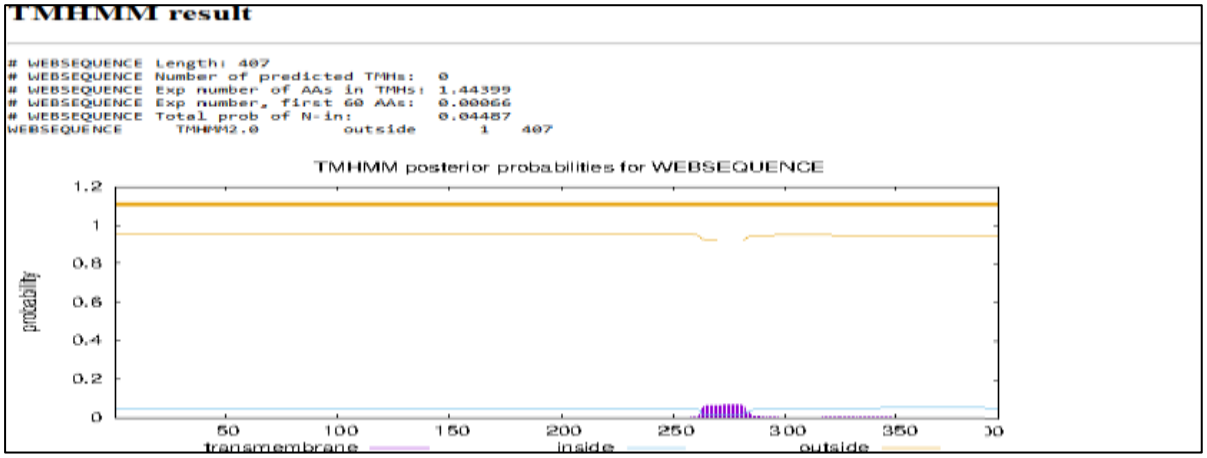
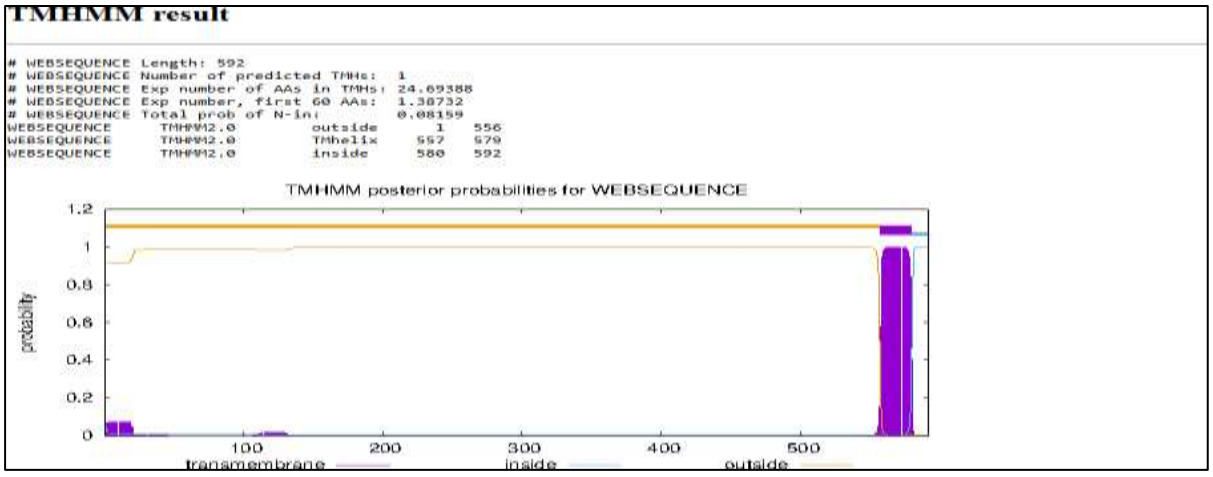


Figure 2. trRossetta result *Vanrija humicola, Aureobasidium pullulans and Geotrichum candidum*,





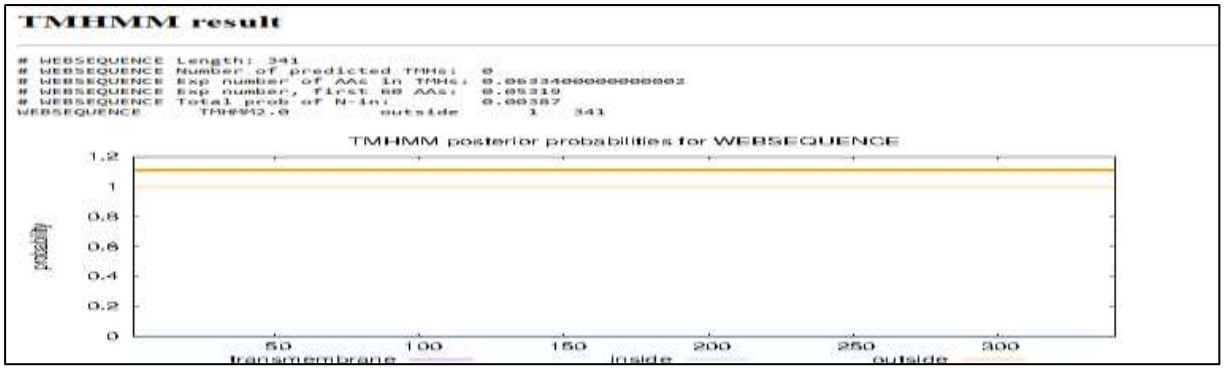
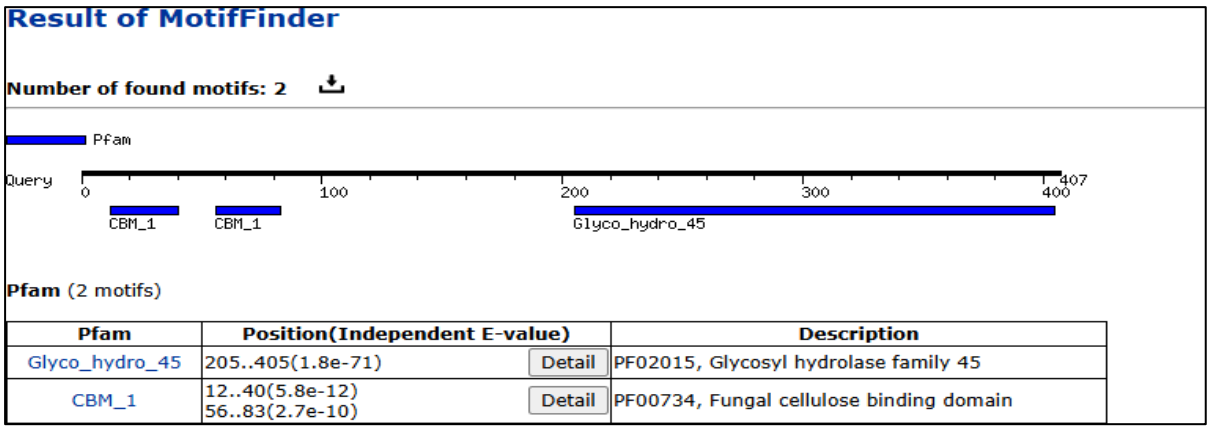
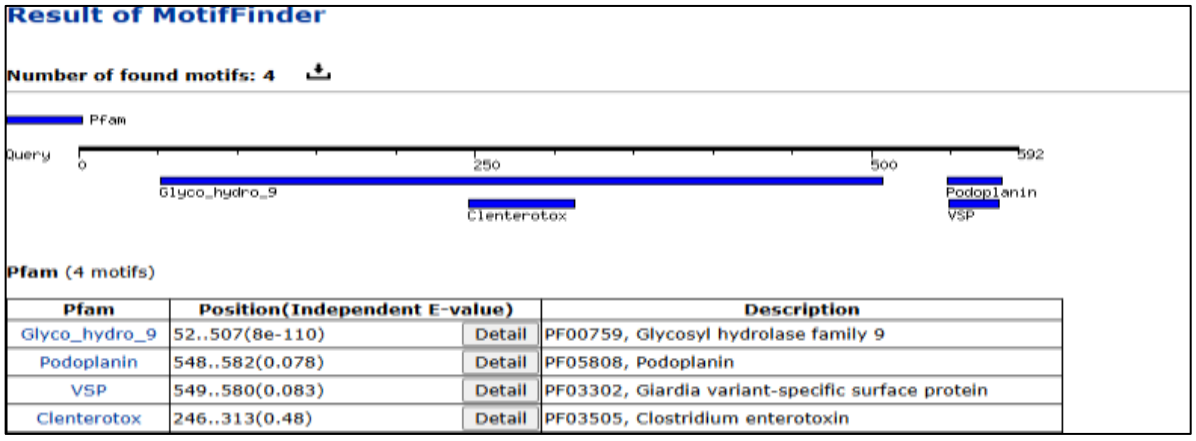


Figure 3. TMHMM result for *Geotrichum candidum, Vanrija humicola, and Aureobasidium pullulans*





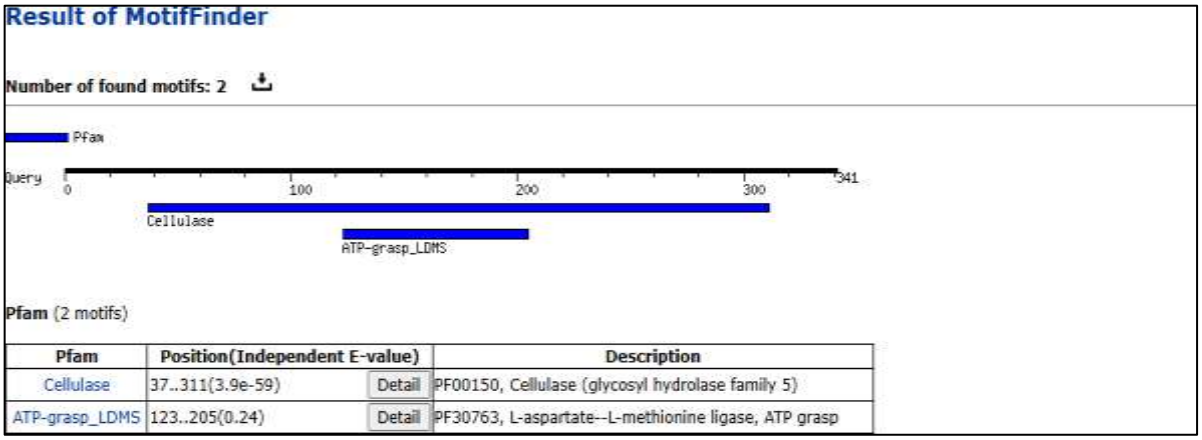
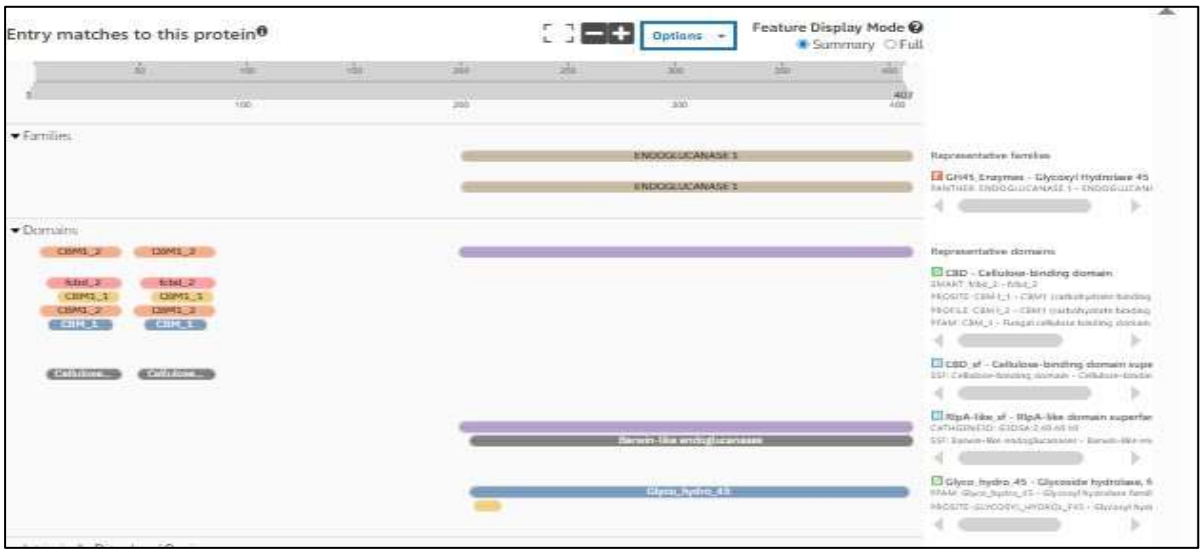


Figure 4. MotifFinder result for *Geotrichum candidum, Vanrija humicola, and Aureobasidium pullulans*

Vanrija humicola

Geotrichum candidum



Aureobasidium pullulans

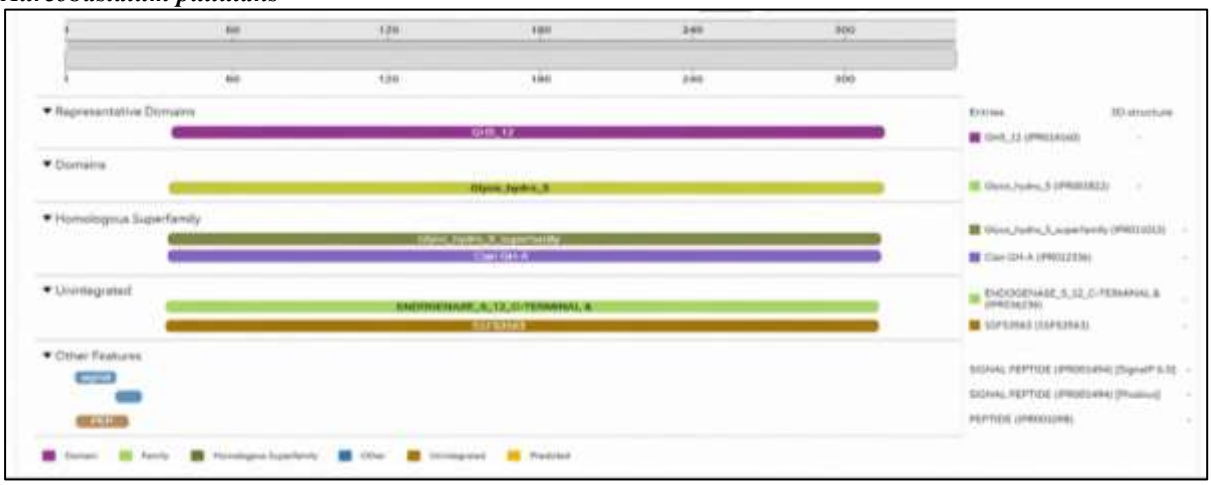
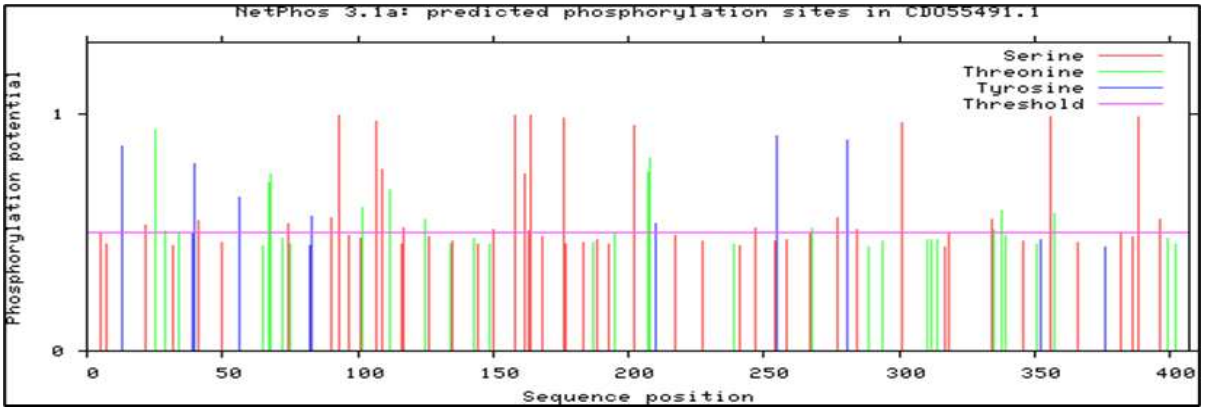
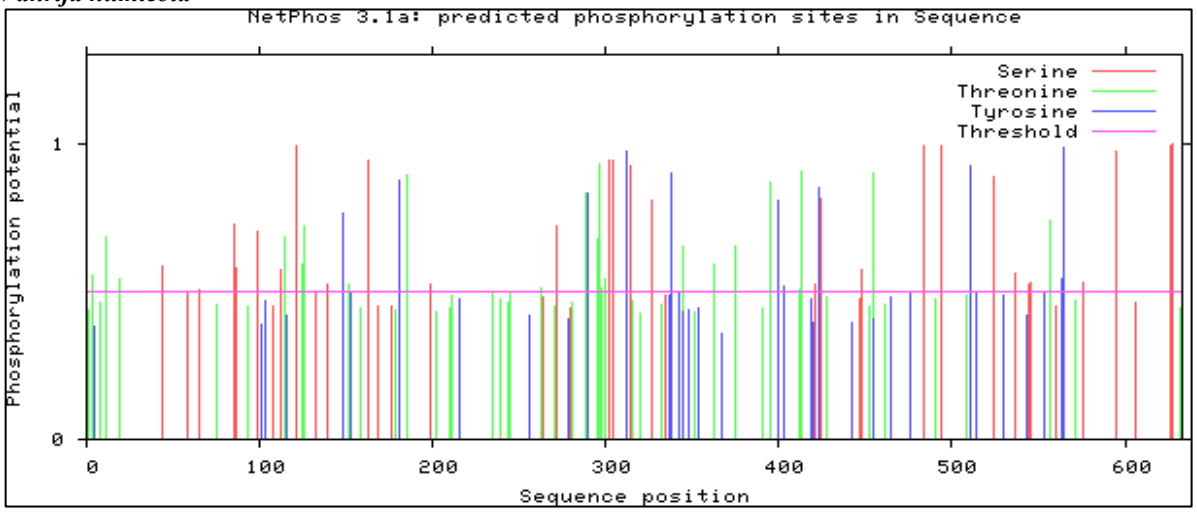


Figure 5. Interpro result for *Geotrichum candidum, Vanrija humicola, and Aureobasidium pullulans*

Geotrichum candidum



Vanrija humicola



Aureobasidium pullulans

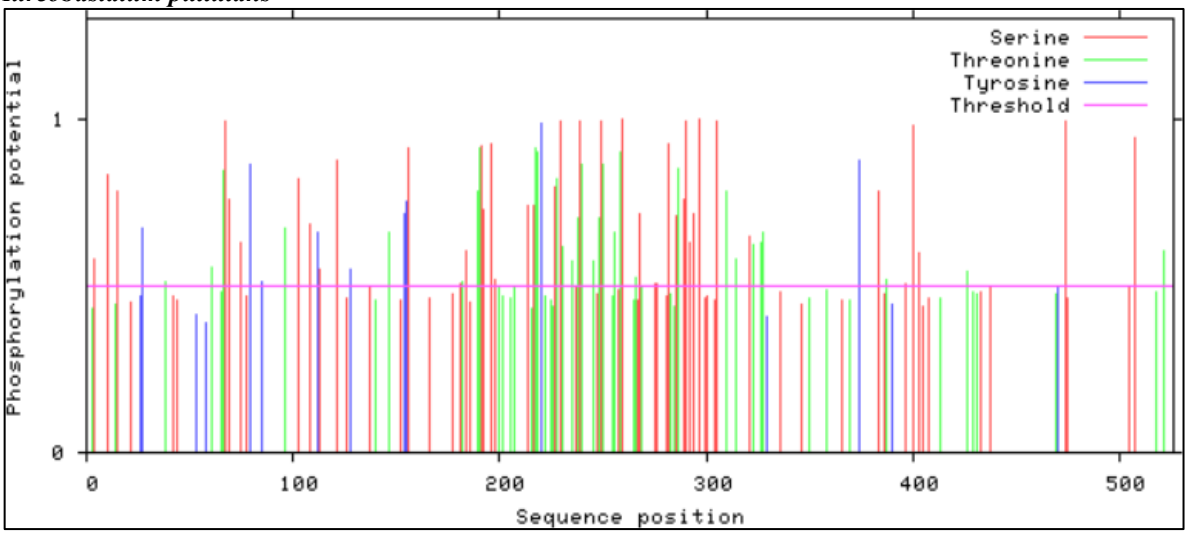


Figure 6. Netphos results for *Geotrichum candidum, Vanrija humicola, and Aureobasidium pullulans*