sciendo

ADVANCED RESEARCH IN LIFE SCIENCES
7, 2023, 135 – 141

www.degruyter.com/view/j/arls

DOI: 10.2478/arls-2023-0015
Research Article

In silico Structural and Functional Characterization of an Endoglucanase from *Actinoalloteichus hoggarensis*

Mahfoud Bakli [*1,2], Noureddine Bouras [2,3], Raul Paşcalău [4], and Laura Şmuleac [4]

[1]Université de Ghardaïa, Faculté des Sciences de la Nature et de la Vie et Sciences de la Terre, Département de Biologie, Ghardaïa, Algeria
[2]Université de Ghardaïa, Faculté des Sciences de la Nature et de la Vie et Sciences de la Terre, Laboratoire de Valorisation et Conservation des Écosystèmes Arides (LVCEA), Ghardaïa, Algeria
[3]Ecole Normale Supérieure de Kouba, Laboratoire de Biologie des Systèmes Microbiens (LBSM), Alger, Algeria
[4]University of Life Sciences "King Mihai I of Romania", Faculty of Agriculture, 119 Calea Aradului, 300645, Timişoara, Romania

Received May, 2023; Revised September, 2023; Accepted October, 2023

Abstract
Actinoalloteichus hoggarensis is a rare bacterial species that was isolated from the Algerian Saharan desert and is known for producing biologically active compounds. Despite its potential, little is understood about the enzymes it produces, including endoglucanases. These cellulase enzymes break down cellulose, the primary structural component of plant cell walls that provides strength and rigidity. The breakdown of cellulose by endoglucanases has numerous biotechnological applications, such as the production of biofuels, bioplastics, and paper. This study involves an *in silico* characterization of an endoglucanase from *A. hoggarensis* to gain insight into its structural and functional properties, with the goal of informing the development of novel biotechnological applications. Our study represents a major milestone in understanding the potential of this rare bacterial species and its enzymes, opening up exciting new avenues for further research and development.

Keywords: Endoglucanase, *Actinoalloteichus hoggarensis*, structural and functional properties, *in silico* characterization.

Introduction

Cellulose is the primary structural component of plant cell walls that provides strength and rigidity and is the most abundant biopolymer on earth [1]. Cellulase enzymes encompass endoglucanases (EC 3.2.1.4), β-glucosidase (EC 3.2.1.21), and exoglucanases (EC 3.2.1.91). Endoglucanase (EG) is the essential component for hydrolyzing β-1,4-glycosidic bonds of the carbon skeleton in cellulose to produce celloligosaccharides [2]. Endoglucanases are enzymes that break down cellulose by hydrolyzing the internal bonds between glucose units, and have numerous biotechnological applications, such as the production of biofuels, bioplastics, and paper [3]. Despite their potential, the industrial-scale production of endoglucanases using traditional methods, such as bacterial or fungal fermentation, is costly and inefficient [4].

Actinoalloteichus hoggarensis is a rare actinobacterial species that was isolated from the Algerian Saharan desert and is known for producing bioactive secondary metabolites [5]. Little is understood about the enzymes it produces, including endoglucanases. However, its potential for biotechnological applications is promising due to its unique properties, such as thermotolerance, halotolerance, and the ability to

* Corresponding author: **M. Bakli, Email:** mahfoud.bakli@gmail.com

grow in extreme conditions [6]. The study of endoglucanases from rare bacterial species like *A. hoggarensis* is still in its early stages, leaving a significant knowledge gap in understanding their properties. Further research is needed to fill this gap and explore the potential applications of these enzymes. However, The experimental expression and purification of proteins can pose challenges and require significant time and effort [7].

Computational approaches, such as homology modeling, molecular dynamics simulation, mutational analysis, etc., provide a promising alternative to design endoglucanase enzymes with desirable characteristics for industrial applications [8]. In this study, we use *in silico* methods to perform a structural and functional characterization of an endoglucanase from *A. hoggarensis*, with the goal of gaining insights into its properties and potential for biotechnological applications.

Materials and methods

Amino acid sequence retrieval

The FASTA format of the endoglucanase protein sequence from *Actinoalloteichus hoggarensis* DSM 45943^T (accession no. A0A221VZV6) was retrieved using UniProt (Universal Protein Resource, https://www.uniprot.org/), and this sequence served as the input for all bioinformatics analyses.

Physicochemical characterization and solubility prediction

ExPASy-ProtParam tools (http://expasy.org/tools/protparam.html) were employed to conduct an analysis of various physicochemical properties, including molecular weight, isoelectric point (pI), aliphatic index (AI), instability index (II), the number of positive and negative charged residues, extinction coefficient (EC), half-life, and Grand Average Hydropathy (GRAVY) [9]. Furthermore, to predict the hydropathy of the endoglucanase amino acid sequence, the ProtScale analysis tool based on the Kyte and Doolittle scale (https://web.expasy.org/protscale/) was utilized [10]. In addition, the protein's transmembrane tendency was predicted using the SOSUI server (http://harrier.nagahama-i-bio.ac.jp/sosui/) [11].

Secondary structure prediction

Two approaches were utilized to predict secondary structure of the protein, improved self-optimized prediction (SOPMA) from the NPS@ server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) [12] and PSI-blast-based secondary structure PREDiction, PSIPRED 4.0 workbench server (http://bioinf.cs.ucl.ac.uk/psipred/) [13]. These methods predicted the location of alpha-helix, beta-sheet, and coil regions, providing information on the secondary structure.

Tertiary structure prediction, structure validation, and quality prediction

Seven different protein structure prediction tools, namely; the SWISS-MODEL server (http://swissmodel.expasy.org/) [14], the Raptor X protein structure property prediction server (http://raptorx.uchicago.edu/StructurePropertyPred/predict/) [15], the $(PS)^2$-v2 Server (http://ps2.life.nctu.edu.tw/) [16], Phyre^2 server (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) [17], the Local Meta-Threading-Server, LOMETS (https://zhanggroup.org/LOMETS/) [18], and the Iterative Threading Assembly Refinement server, I-TASSER (https://zhanggroup.org/I-TASSER/) [19] were utilized to build the 3D structure models of *A. hoggarensis* endoglucanase protein query sequence based on homology modeling method. Every best predicted three-dimensional structure initial model, thus obtained, was refined in order to improve its physical quality by using a high-resolution protein structure refinement tool server; The ModRefiner (http://zhanglab.ccmb.med.umich.edu/ModRefiner/) [20]. Both initial and refined models were then assessed by Ramachandran's plot using PROCHECK tool on SAVES v6.0 online server (https://saves.mbi.ucla.edu/) [21]. The PyMOL Molecular Graphics System (Schrödinger LLC, version 2.3) was utilized to generate and optimize images of various 3D structure models [22].

Ligand binding site prediction and protein localization

The best refined structural model was submitted to the COACH algorithm available at The meta-server, COACH (https://zhanglab.ccmb.med.umich.edu/COACH/) [23]. This approach matches target the submitted models with proteins in the BioLiP database to predict the location of ligand-binding sites [24]. The protein subcellular localization prediction tool, PSORT-B v3.0.3 (https://www.psort.org/psortb/) was used to predict the endoglucanase protein localization [25], Moreover, SignalP-6.0 server (https://services.healthtech.dtu.dk/services/SignalP-6.0/) was employed to predict signal peptide cleavage sites in the sequence, indicating the protein localization [26]. Information about the transmembrane protein was determined using the TMHMM server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [27].

Functional analysis

The query sequence of the target endoglucanase from *A. hoggarensis* was analyzed to determine the protein functional families using MotifFinder

server (http://www.genome.jp/tools/motif/) by selecting Pfam annotation database.

Results and discussion

Physicochemical characterization and solubility prediction

The protein sequence of endoglucanase from *A. hoggarensis*, retrieved from UNIPROT (ID: A0A221VZV6) is 875 amino acids in length. The physiochemical properties of endoglucanase protein were calculated using ProtParam (Table 1). The endoglucanase protein exhibited a molecular weight of 92.88 kDa. The isoelectric point (pI) value of 4.08 defines the acidic behavior of the protein in nature and instability index (II) 33.03 which is less than 40 defined its stable nature. There are more predicted negative residues in the target protein than positive residues. The endoglucanase protein showed a high extinction coefficient value, suggesting a high concentration of Trp and Tyr that facilitates the quantitative analysis of protein-protein and protein-ligand interactions in solution. The half-life of the endoglucanase is 30 hours for mammalian reticulocytes *in vitro.* The aliphatic index (AI) of a protein is a measure of the relative volume occupied by aliphatic side chains (such as alanine, valine, isoleucine, and leucine) within the protein sequence, and it reflects the thermostability of the protein [28]. The AI value of the endoglucanase was determined to be 78.22, indicating that the protein possesses thermostable properties that enable it to maintain its structural integrity over a wide range of temperatures. This observation was consistent with numerous studies on thermostable cellulase and endoglucanase enzymes, which are present in various organisms, including bacteria and fungi [29]. The GRAVY (the Grand average of hydropathicity) score measures a protein's overall hydrophobicity or hydrophilicity by summing the hydropathy values of its amino acids and dividing it by the total number of amino acids [10]. GRAVY index of the endoglucanase had negative value. Accordingly, this result revealed that the enzyme had good interactions with water (hydrophilicity). Overall, acidic and thermostable valuable properties of the target endoglucanase enzyme could be exploited in the industry of bioethanol production [30]. In addition, According to the SOSUI server prediction, the endoglucanase protein was determined to be soluble.

Secondary structure prediction

The secondary structure was predicted using SOPMA and PSIPRED, revealing that the *A. hoggarensis* endoglucanase consisted of 22.97% α-helix, 20.00% β-sheets, and 57.03% random coil. The protein did not exhibit any disordered protein binding sites. No disordered protein binding sites were detected in the protein (Figure 1).

Table1. The physiochemical properties of the predicted endoglucanase

Property	Value
Number of amino acids (AA)	875
Molecular weight (Da)	92883.92
Theoretical pI	4.08
Total number of negatively charged residues (Asp + Glu)	124
Total number of positively charged residues (Arg + Lys)	33
Extinction coefficient (EC)	172855
half-life	30 hr
Instability index (II)	33.03 (Stable)
Aliphatic index (AI)	78.22
Grand average of hydropathicity (GRAVY)	-0.176

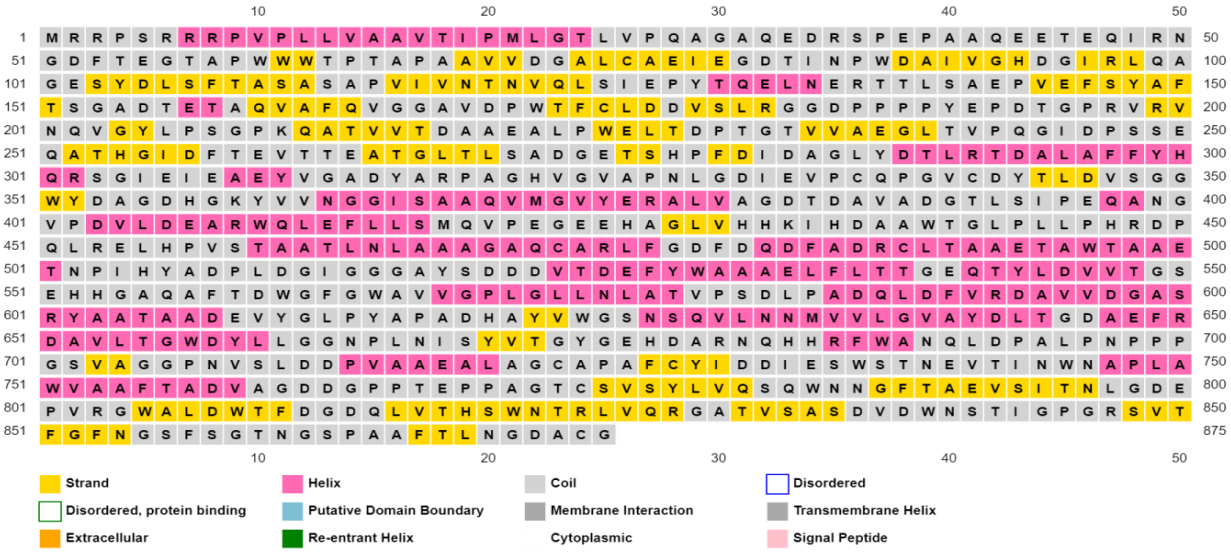


Figure 1. Sequence annotation plot analysing the secondary structure prediction of endoglucanase from *A. hoggarensis* by PSIPRED (yellow for β-strands, pink for α-helix and grey for coil structures).

Tertiary structure prediction and validation

The tertiary structure of *A. hoggrarensis* endoglucanase protein was constructed by using different structure prediction server, which are Modeller, Phyre^2, (PS)^2-v2, Raptor X, I-TASSER, LOMETS, and SWISS-MODEL. Then, 3D builded models were refined by ModRefiner server and the geometry of all initial and refined models was validated with Ramachandran plot using PROCHECK tool. Table 2 displays a comparison of the Ramachandran plots for each structural model. The best refined model was initially generated by (PS)^2-v2, with PDB ID: 1ut9A (chain A, cellulose 1,4-beta-cellobiosidase from *Clostridium thermocellum*), as a template. In addition, the percentage of Ramachandran plot highest favored regions was 91.8% and with the minimum percent of the disallowed region (0) suggesting the traits of a high-quality model (Figure 2).

Table 2. Ramachandran plot calculation of the endoglucanase models computed with PROCHECK tool

Model	Type of model	Favored region (%)	Additional allowed region (%)	Generously allowed region (%)	Disallowed region (%)
SWISS-MODEL	Initial	88.0	11.4	0.2	0.4
	Refined	89.6	9.4	0.4	0.6
Raptor X	Initial	86.8	10.3	1.9	1.0
	Refined	86.8	10.3	1.9	1.0
(PS)^2-V2	Initial	81.0	17.7	1.4	0.0
	Refined	91.8	8.2	0.0	0.0
Phyre^2	Initial	84.5	13.6	1.3	0.6
	Refined	89.2	9.1	0.8	0.8
MODELLER	Initial	90.0	8.1	1.5	0.4
	Refined	90.5	8.3	0.8	0.4
LOMETS	Initial	73.0	19.3	5.5	2.2
	Refined	78.9	14.5	4.1	2.5
I-TASSER	Initial	59.6	29.7	7.4	3.3
	Refined	78.6	15.4	3.9	2.1

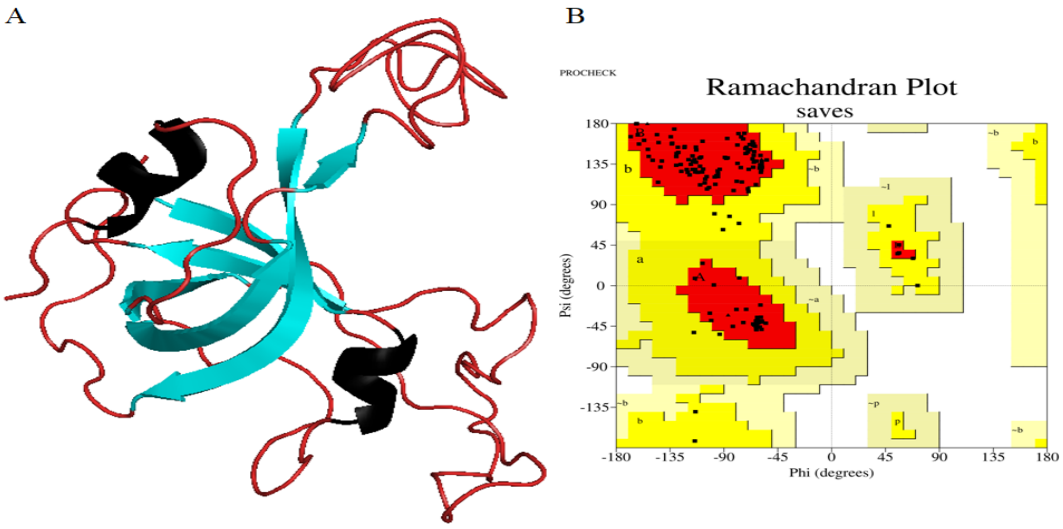


Figure 2. Tertiary structure prediction and validation for the endoglucanase from *A. hoggarensis*. (A) Tertiary structure prediction of the protein was done by homology modeling using $(PS)^2$-v2**.** This structural model was refined by ModRefiner and visualised by PyMOL. α-helix (black), β-sheet (cyan), and loop (red). (B) The predicted and refined structure's Ramachandran plot was validated by the PROCHECK program, with residues classified based on their conformational preferences as per Ramachandran plot analysis. Regions with favoured (A, B, and L), additional allowed (a, b, l, and p), generously allowed (~a, ~b, ~l, and ~p), and disallowed conformations are highlighted in red, yellow, beige, and white, respectively. Non-glycine and non-proline residues are represented by filled black squares, while glycines (excluding those at the ends of the polypeptide chains) are depicted as filled black triangles.

Ligand binding site prediction and protein localization

COACH meta server was utilized to predict the consensus ligand binding residues located within the predicted active site of the endoglucanase enzyme. The result showed that the enzyme interacted with ligand (PDB CCD ID : 1XT) in the following consensus binding residues; Leu 29, Lys 63, Tyr 65, Leu 76, Glu 78, Ser 80, Gln 81, Tyr 113, Thr 115, Val 136, Val 138, Asn 140, and Asn 142 with a higher confidence score (C-score) indicating a more reliable prediction of 0.15 (Figure 3). Furthermore, PSORTb v3.0.3 identified the endoglucanase protein to be extracellular. SignalP-6.0 server indicated that this protein has a signal peptide sequence with a cleavage site between position 31 and 32. Moreover, according to the TMHMM-2.0 server, the protein does not contain any transmembrane helices. Based on the protein localization results, it can be inferred that the *A. hoggarensis* endoglucanase is likely to be functional in the extracellular space.

Functional analysis.

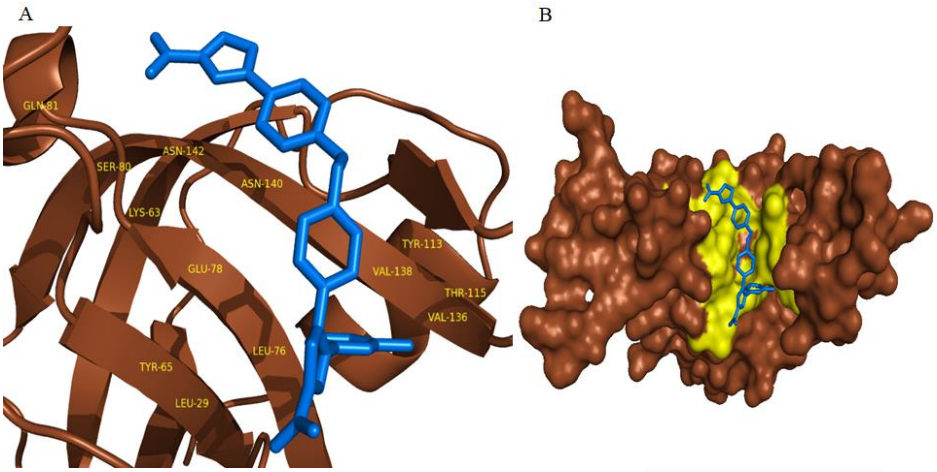
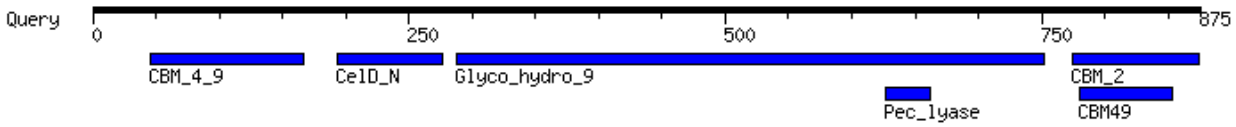


Figure 3. Ligand binding sites of endoglucanase from *A. hoggarensis* predicted by COACH. (A) Ligand binding sites (Leu 29, Lys 63, Tyr 65, Leu 76, Glu 78, Ser 80, Gln 81, Tyr 113, Thr 115, Val 136, Val 138, Asn 140, and Asn 142). (B) Surface view of the protein with ligand in pocket exposed active site. Sticks represent the ligand (blue). Ligand binding sites (yellow) within the protein structure (brown).



Pfam (6 motifs)

Pfam	Position(Independent E-value)		Description
Glyco_hydro_9	287..752(1.9e-102)	Detail	PF00759, Glycosyl hydrolase family 9
CBM_2	774..874(2.8e-32)	Detail	PF00553, Cellulose binding domain
CelD_N	194..277(1.8e-20)	Detail	PF02927, Cellulase N-terminal ig-like domain
CBM_4_9	46..167(2.9e-15)	Detail	PF02018, Carbohydrate binding domain
Pec_lyase	627..662(0.14)	Detail	PF09492, Pectic acid lyase
CBM49	780..853(0.2)	Detail	PF09478, Carbohydrate binding domain CBM49

Figure 4. The motif analysis of *A. hoggarensis* endoglucanase determined by MotifFinder.

MotifFinder server suggested six functional motifs through the entire glucananase protein sequence, including Glycosyl hydrolase family 9 (GH9), Cellulose binding domain (CBM2), Cellulase N-terminal ig-like domain (CelD_N), Carbohydrate binding domain (CBM_4_9), Pectic acid lyase (Pec_lyase), Carbohydrate binding domain (CBM49) (Figure 4).

Conclusions

In silico analysis of the endoglucanase from *A. hoggarensis* revealed several key properties, including its acidic nature, thermostability, hydrophilicity, and extracellular location. The protein's secondary structure was found to be predominantly random coil, with some alpha-helix and beta-sheet elements. Additionally, we conducted both structural homology modeling and functional analyses. Nevertheless, further experimental studies, such as enzyme kinetics, substrate specificity, and protein expression, are needed to fully understand the potential of this endoglucanase and its applications in biotechnology.

Acknowledgements

This work was supported by PRFU “Projet de Recherche et Formation Universitaire” Project Research, D01N01UN460120220001.

References

1. **Sharma, A., Thakur, M., Bhattacharya, M., Mandal, T., & Goswami, S.** (2019). Commercial application of cellulose nano-composites–A review. *Biotechnology Reports, 21*, e00316. DOI: 10.1016/j.btre.2019.e00316

2. **Li, Y., Song, W., Han, X., Wang, Y., Rao, S., Zhang, Q., Zhou, J., Li, J., Liu, S., Du, G.** (2022). Recent progress in key lignocellulosic enzymes: Enzyme discovery, molecular modifications, production, and enzymatic biomass saccharification. *Bioresource Technology*, 127986. DOI: 10.1016/j.biortech.2022.127986

3. **Gupta, G. K., Dixit, M., Kapoor, R. K., & Shukla, P.** (2022). Xylanolytic enzymes in pulp and paper industry: new technologies and perspectives. *Molecular biotechnology*, 1-14. DOI: 10.1007/s12033-021-00396-7

4. **Ejaz, U., Sohail, M., & Ghanemi, A.** (2021). Cellulases: from bioactivity to a variety of industrial applications. *Biomimetics, 6*(3), 44. DOI: 10.3390/biomimetics6030044

5. **Boudjelal, F., Zitouni, A., Bouras, N., Schumann, P., Spröer, C., Sabaou, N., & Klenk, H.-P.** (2015). *Actinoalloteichus hoggarensis* sp. nov., an actinomycete isolated from Saharan soil. *International journal of systematic and evolutionary microbiology, 65*(6), 2006-2010. DOI: 10.1099/ijs.0.000216

6. **Xie, F., Rangseekaew, P., & Pathom-aree, W.** (2022). Actinobacteria from Arid Environments and Their Biotechnological Applications. *Natural Products from Actinomycetes: Diversity, Ecology and Drug Discovery* (pp. 91-118): Springer. DOI: 10.1007/978-981-16-6132-7_4

7. **Bakli, M., Karim, L., Mokhtari-Soulimane, N., Merzouk, H., & Vincent, F.** (2020). Biochemical characterization of a glycosyltransferase Gtf3 from *Mycobacterium smegmatis*: a case study of improved protein solubilization. *3 Biotech, 10*(10), 1-13. DOI: 10.1007/s13205-020-02431-x

8. **Mandeep, Liu, H., & Shukla, P.** (2021). Synthetic biology and biocomputational approaches for improving microbial endoglucanases toward their innovative applications. *ACS omega, 6*(9), 6055-6063. DOI: 10.1021/acsomega.0c05744

9. **Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A.** (2005). Protein identification and analysis tools on the ExPASy server.

The proteomics protocols handbook, 571-607. DOI: 10.1385/1-59259-890-0:571
10. **Kyte, J., & Doolittle, R. F.** (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology, 157*(1), 105-132. DOI: 10.1016/0022-2836(82)90515-0
11. **Hirokawa, T., Boon-Chieng, S., & Mitaku, S.** (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics (Oxford, England), 14*(4), 378-379. DOI: 10.1093/bioinformatics/14.4.378
12. **Geourjon, C., & Deleage, G.** (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics, 11*(6), 681-684. DOI: 10.1093/bioinformatics/11.6.681
13. **Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., & Jones, D. T.** (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research, 41*(W1), W349-W357. DOI: 10.1093/nar/gkt381
14. **Arnold, K., Bordoli, L., Kopp, J., & Schwede, T.** (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics, 22*(2), 195-201. DOI: 10.1093/bioinformatics/bti770
15. **Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J.** (2012). Template-based protein structure modeling using the RaptorX web server. *Nature protocols, 7*(8), 1511-1522. DOI: 10.1038/nprot.2012.085
16. **Chen, C.-C., Hwang, J.-K., & Yang, J.-M.** (2009). $(PS)^2$-v2: template-based protein structure prediction server. *Bmc Bioinformatics, 10*(1), 1-13. DOI: 10.1186/1471-2105-10-366
17. **Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J.** (2015). The Phyre^2 web portal for protein modeling, prediction and analysis. *Nature protocols, 10*(6), 845-858. DOI: 10.1038/nprot.2015.053
18. **Wu, S., & Zhang, Y.** (2007). LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic acids research, 35*(10), 3375-3382. DOI: 10.1093/nar/gkm251
19. **Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y.** (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods, 12*(1), 7-8. DOI: 10.1038/nmeth.3213
20. **Xu, D., & Zhang, Y.** (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical journal, 101*(10), 2525-2534. DOI: 10.1016/j.bpj.2011.10.024
21. **Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M.** (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography, 26*(2), 283-291. DOI: 10.1107/S0021889892009944
22. **DeLano, W.** (2019). The PyMOL Molecular Graphics System, version 2.3. 1. *Schrodinger LLC: New York, NY, USA*.
23. **Yang, J., Roy, A., & Zhang, Y.** (2013). Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics, 29*(20), 2588-2595. DOI: 10.1093/bioinformatics/btt447
24. **Yang, J., Roy, A., & Zhang, Y.** (2012). BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic acids research, 41*(D1), D1096-D1103. DOI: 10.1093/nar/gks966
25. **Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J.** (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics, 26*(13), 1608-1615. DOI: 10.1093/bioinformatics/btq249
26. **Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., Heijne, G. O., & Nielsen, H.** (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature biotechnology, 40*(7), 1023-1025. DOI: 10.1038/s41587-021-01156-3
27. **Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L.** (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology, 305*(3), 567-580. DOI: 10.1006/jmbi.2000.4315
28. **Ikai, A.** (1980). Thermostability and aliphatic index of globular proteins. *The Journal of Biochemistry, 88*(6), 1895-1898. DOI: 10.1093/bioinformatics/11.6.681
29. **Cai, L.-N., Xu, S.-N., Lu, T., Lin, D.-Q., & Yao, S.-J.** (2022). Salt-tolerant and thermostable mechanisms of an endoglucanase from marine *Aspergillus niger*. *Bioresources and Bioprocessing, 9*(1), 1-15. DOI: 10.1186/s40643-022-00533-3
30. **Akram, F., ul Haq, I., Imran, W., & Mukhtar, H.** (2018). Insight perspectives of thermostable endoglucanases for bioethanol production: a review. *Renewable Energy, 122*, 225-238. DOI: 10.1016/j.renene.2018.01.095