AN INTEGRATED COMPUTATIONAL ANALYSIS OF BIFUNCTIONAL ENZYMES (CELLULASE-XYLANASE)

[1]Dr. T. Sharmila Raj, [1]Nandhini R, [2]Calista Stephanie A, [3] Vyshali S* **and** *[4]Anugraha J*

[1*,1,2,3,4]Department of Biotechnology, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India.

ABSTRACT:

Bifunctional enzymes catalyze two distinct biochemical reactions in one protein and are therefore effective regulators of cellular functions. Here, describes an integrated computational workflow for the analysis of these enzymes. Firstly, downloading protein sequences from the NCBI database. To learn about their dual activity, we identified the respective protein domains that cause each catalytic function through InterProScan. Then modeled three-dimensional structures of the enzymes through SWISS-MODEL. Active site analysis was done using CASTp to identify where substrates bind. Molecular docking simulations through ClusPro were employed to visualize substrate interactions with each active site. The stability of these complexes and the dynamics of the whole protein were evaluated by molecular dynamics simulations with GROMACS. Lastly, in order to examine their evolutionary past, we built phylogenetic trees with MEGA and PhyML. Multidisciplinary approach presents a detailed roadmap for clarifying how structure informs bifunctionality, providing important information for drug discovery and metabolic engineering.

Keywords: Bifunctional Enzymes, Computational Biology, Protein Structure, Molecular Docking, Molecular Dynamics, Phylogenetics.

1. INTRODUCTION:

Bifunctional enzymes are a key group of proteins that can catalyze two different biochemical reactions in one polypeptide chain. They can act as very efficient and cost-effective regulators of cellular pathways due to their distinct duality and confer a clear advantage in metabolic and signaling cascades [1]. As such, these molecules are of tremendous significance not merely in basic biology but in industrial biotechnology as well, especially in the enzymatic breakdown of complex biopolymers. This research centers on cellulase-xylanase bifunctional systems, which are the building blocks to lignocellulosic biomass conversion. Since cellulose and xylan are among the most prevalent plant polysaccharides, their coordinated enzyme-mediated degradation is imperative for effective degradation, and thus important for the developments in sustainable production of biofuels, textile treatments, and paper processing. Because of the inbuilt complexity and time requirement of conventional wet-lab biochemistry, a combined computational protocol with the aid of sophisticated bioinformatics software is used to gain a fast, atom-level insight into the enzymes. Structural and functional in-depth characterization of these bifunctional catalysts is thus of key importance in order to optimize their industrial use [1]. The analysis is initiated with retrieval of protein sequences and subsequent initial characterization. This involves computation of physicochemical properties via the ExPASy ProtParam server [9].and subcellular localization prediction via specific deep learning techniques, including those run with SignalP – 6.0 [10]. Protein domains are identified to structurally define the nature for the dual function using the powerful sequence analysis tool InterProScan [2]. After primary characterization, correct three-dimensional structures of chosen enzyme candidates are made with valid homology modeling tools such as SWISS-MODEL [3]. Those resulting structural models are further analyzed with thorough functional mapping. Predicted ligand-binding cavities and active sites for both cellulase and xylanase activities are determined with both geometry-based tools such as CASTp [11] and advanced 3D-convolutional neural networks in DeepSite [4]. Most importantly, the exact molecular details of substrate recognition and binding are probed using molecular docking simulations run with the ClusPro package, offering quantitative descriptions of enzyme-substrate interactions [5]. To gauge stability and dynamic plasticity in bifunctionality, resulting enzyme-substrate complexes are analyzed using stringent molecular dynamics (MD) simulations carried out with GROMACS [6]. Lastly, to place the molecular results within a more generalized biological context, the evolutionary background of these enzymes is studied through the development of phylogenetic trees based on statistical approaches such as Maximum-Likelihood, as applied by tools such as MEGA [7] and PhyML [8]. Cumulatively, this combined computational approach provides a solid, comprehensive blueprint for elucidating the structure-function relationship underlying the bifunctionality of cellulase-xylanase and hence providing essential information for rational enzyme design and metabolic pathway optimization.

2. MATERIALS AND METHODOLOGY:

2.1 Materials:

The following websites are used for this in silico studies. NCBI, Uniport, InterProScan, ExPASy ProtParam, SignalP-6.0, SWISS MODEL, ERRAT, CASTp, ClusPro, MGLTools(ADT), GROMACS, MEGA and PhyML. And the high tech laptop is used as the material.

2.2 Methodology:

The structural and functional characterization of the bifunctional cellulase-xylanase was achieved by a detailed, step-by-step in silico process outlined below:

2.2.1 Sequence Retrieval and Initial Characterization

Sequence Retrieval: Protein sequences for Cellulase (PNY18054.1) and Xylanase (CAA31109.1) were obtained from the NCBI or UniPort database in FASTA format.

Physicochemical Analysis: The rudimentary physicochemical characteristics such as theoretical isoelectric point, molecular weight, and instability index were calculated by the ExPASy ProtParam server.

Subcellular Localization: Signal peptides and the predicted subcellular localization of the two proteins were identified using the deep learning tool, SignalP 6.0

Identification of Domains: For the identification of the separate catalytic and accessory domains responsible for the bifunctionality, the sequences were scanned using InterProScan.

2.2.2 Structure Modeling and Validation

3D Structure Prediction: The three-dimensional structures of the target proteins were predicted by utilizing the automated homology modeling server, SWISS-MODEL. The most suitable templates with regards to highest sequence identity and coverage were chosen to model.

Model Validation: The quality and reliability of the generated models were assessed using standard structure validation tools. Specifically, the stereochemical quality was evaluated using ERRAT (Sander and Schneider server), and the geometric accuracy was confirmed using the Ramachandran plot analysis.

2.2.3 Functional Site Prediction and Molecular Docking

Active Site Prediction: Active sites for both xylanase and cellulase activities and potential ligand-binding pockets were predicted. Geometric pocket detection was executed by CASTp (Computed Atlas of Surface Topography of Proteins).

Molecular Docking: Protein structures modeled were pre-docked by addition of polar hydrogen atoms and Gasteiger charge calculation. Molecular docking simulations with ClusPro were conducted to make predictions of binding affinity and mode of interaction of respective substrates (cellulose/xylan analogues) in the located active sites. The Lamarckian Genetic Algorithm was employed for conformational searching, and the docking runs were clustered according to root-mean-square deviation (RMSD).

2.2.4 Evolutionary Analysis and Molecular Dynamics

Molecular Dynamics Simulation (MD): The best docked complexes served as MD starting structures. Simulations were executed on GROMACS to assess the dynamic flexibility and stability of the protein-substrate complexes. The system was solvated, energy-minimized, and equilibrated at standard temperature and pressure conditions, while its trajectories were checked for stability parameters such as RMSD and Radius of Gyration.

Phylogenetic Analysis: For appreciating the evolutionary background of the bifunctional enzymes, homologous sequences were downloaded and aligned. Phylogenetic trees were built employing the Maximum-Likelihood approach, as carried out in both MEGA X and PhyML 3.0. Tree reliability was evaluated through bootstrap analysis with 1000 repetitions.

3. RESULT AND DISCUSSION:

3.1 Sequence Retrieval and Initial Characterization:

3.3.1 Sequence Retrieval

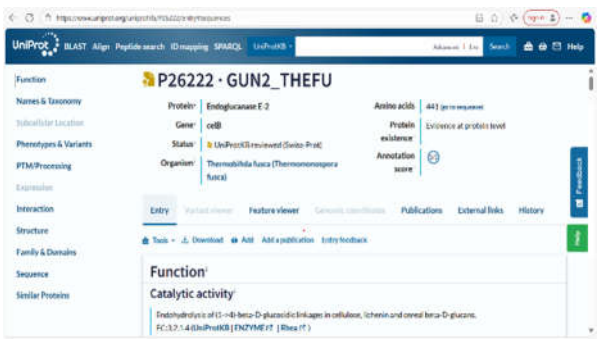


Fig no: 01 Cellulase

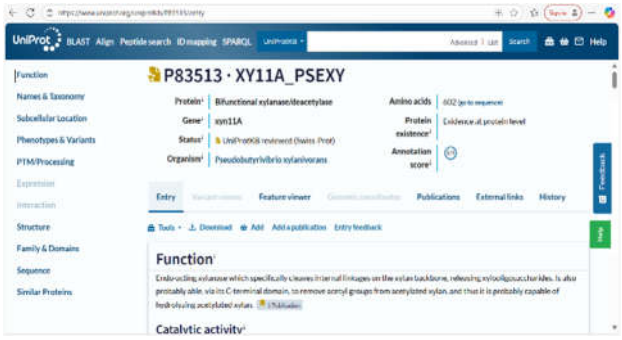


Fig no: 02 Xylanase

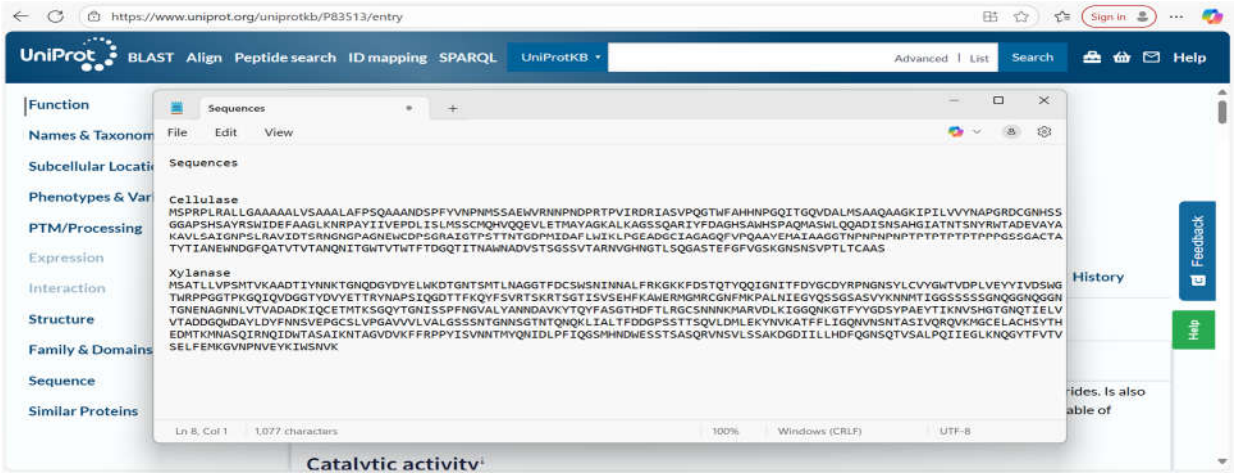
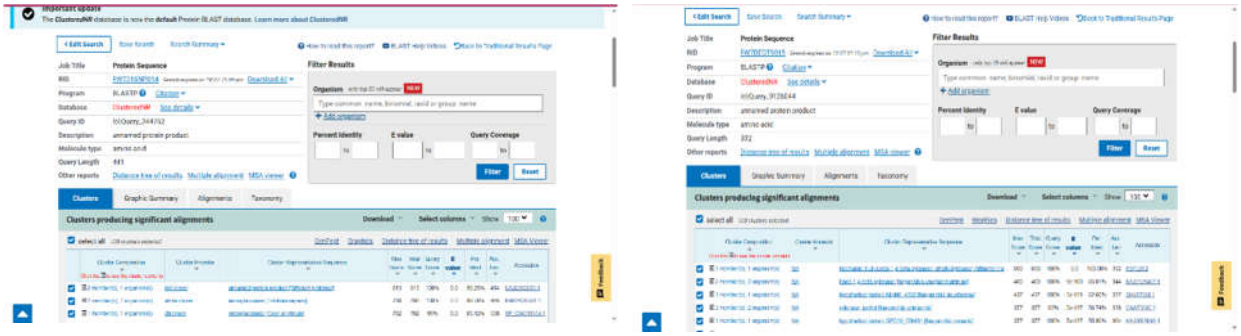


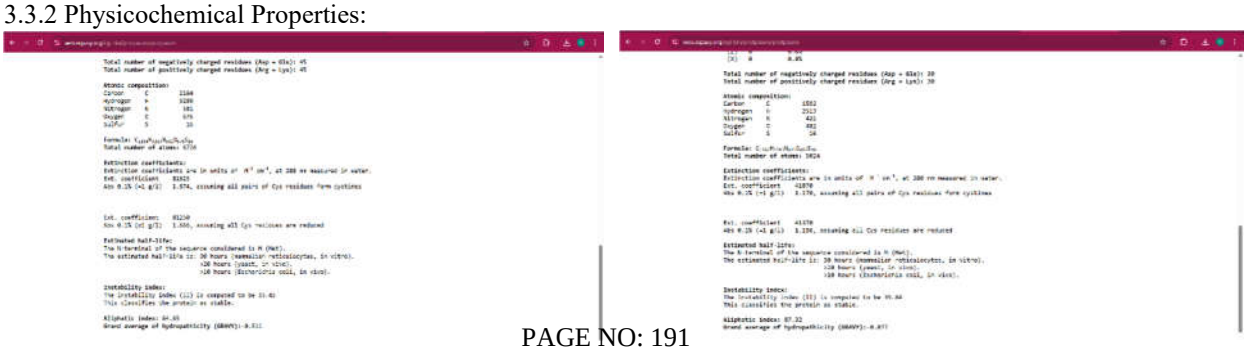
Fig no: 03 FASTA Format for the enzymes



(a) (b)

Fig no: 04 Blast analysis for (a) Cellulase and (b) Xylanase

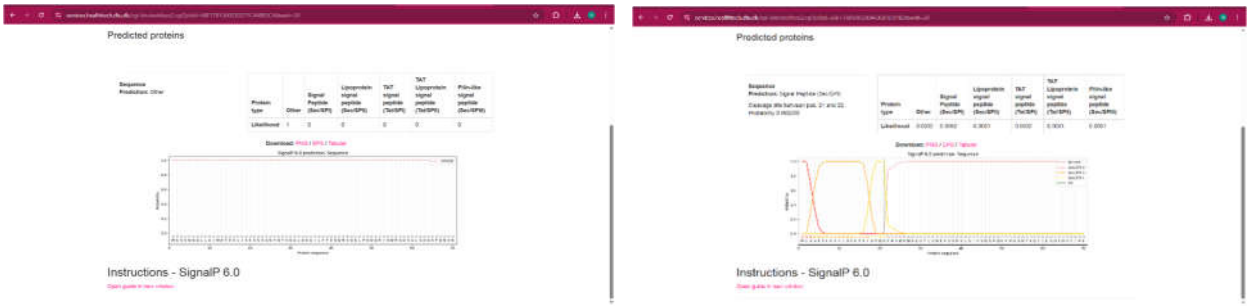
3.3.2 Physicochemical Properties:



(a) (b)

Fig no: 05 ExPASy ProtParam analysis for (a) Cellulase and (b) Xylanase

3.3.3 Subcellular Localization

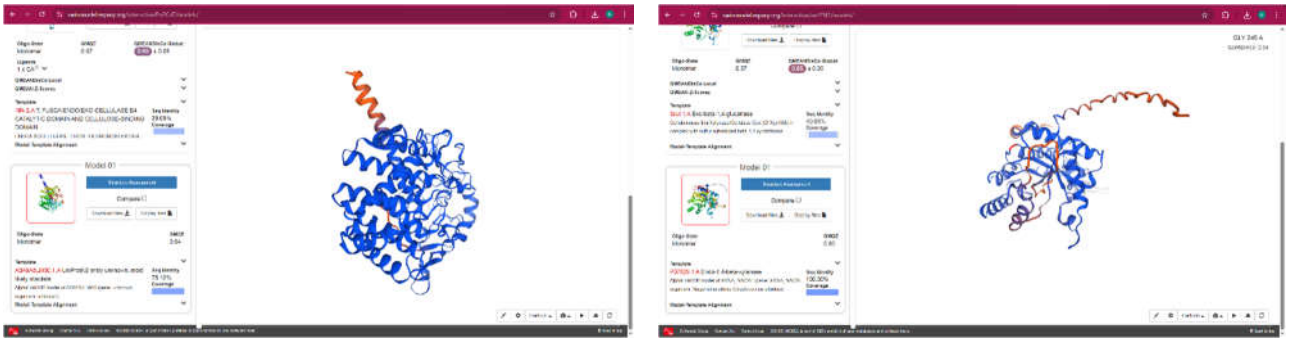


(a) (b)

Fig no: 06 SignalP 6.0 analysis for (a) Cellulase and (b) Xylanase

3.2 Structure Modeling and Validation

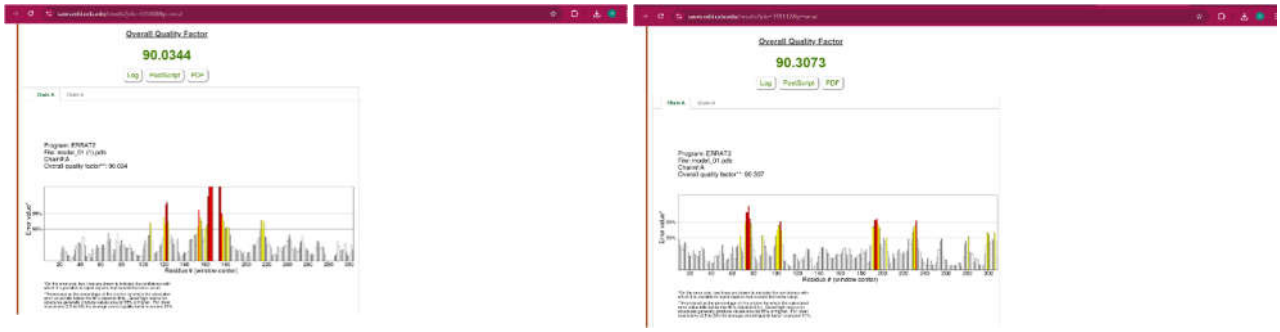
3.2.1 3D Structure Prediction



(a) (b)

Fig no: 07 SWISS MODEL analysis for (a) Cellulase and (b) Xylanase

3.2.2 Model Quality Validatio

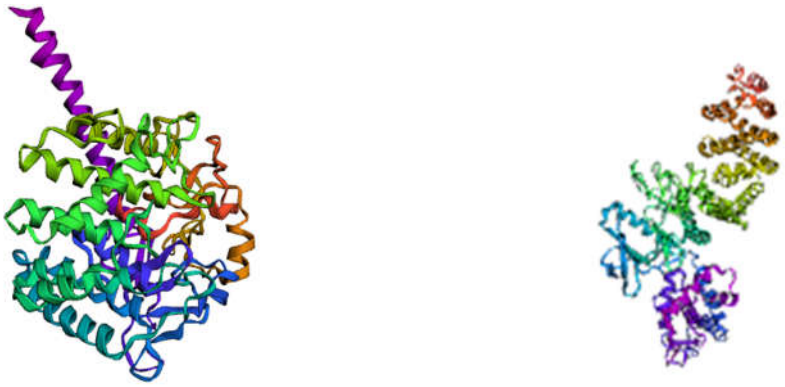


(a) (b)

Fig no: 08 ERRAT analysis for (a) Cellulase and (b) Xylanase

3.3 Functional Site Prediction and Molecular Docking

3.3.1 Active Site Prediction



(a) (b)

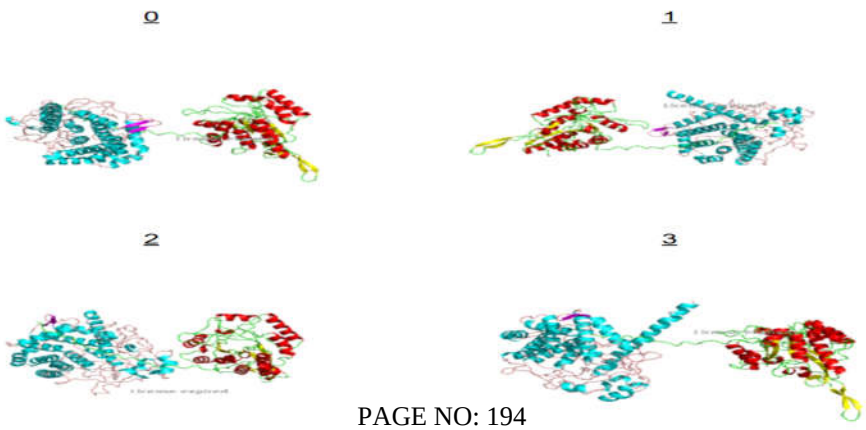
Fig no: 09 CASTp analysis for (a) Cellulase and (b) Xylanase

3.3.2 Molecular Docking

Cluster	Members	Representative	Weighted Score
0	63	Center	-883.4
		Lowest Energy	-1007.5
1	49	Center	-935.3
		Lowest Energy	-1093.8
2	46	Center	-920.8
		Lowest Energy	-1007
3	41	Center	-926.6
		Lowest Energy	-979.2
4	39	Center	-956.5
		Lowest Energy	-1034.2
5	38	Center	-874.1
		Lowest Energy	-1074.8
6	36	Center	-898
		Lowest Energy	-1010.4
7	33	Center	-863.7
		Lowest Energy	-917.5
8	32	Center	-900.5
		Lowest Energy	-978.8
9	25	Center	-981.2
		Lowest Energy	-981.2
10	22	Center	-907.2
		Lowest Energy	-908
11	22	Center	-977
		Lowest Energy	-977

12	18	Center	-950.3
		Lowest Energy	-950.3
13	18	Center	-914
		Lowest Energy	-996.6
14	17	Center	-825.2
		Lowest Energy	-982.1
15	17	Center	-891.6
		Lowest Energy	-1000.6
16	16	Center	-865.5
		Lowest Energy	-876.9
17	15	Center	-836.1
		Lowest Energy	-961.4
18	15	Center	-825.4
		Lowest Energy	-927.7
19	14	Center	-892.3
		Lowest Energy	-1092.5
20	14	Center	-947.3
		Lowest Energy	-1100.4
21	13	Center	-870.4
		Lowest Energy	-968.2
22	13	Center	-839.6
		Lowest Energy	-879.3
23	13	Center	-911.2
		Lowest Energy	-935.6
24	12	Center	-901.8
		Lowest Energy	-901.8
25	12	Center	-891.2
		Lowest Energy	-1097.5
26	11	Center	-835.9
		Lowest Energy	-889.3
27	11	Center	-928.6
		Lowest Energy	-1035
28	11	Center	-885.5
		Lowest Energy	-909.5
29	11	Center	-884.3
		Lowest Energy	-884.3

Fig no: 10 ClusPro analysis with its clusters



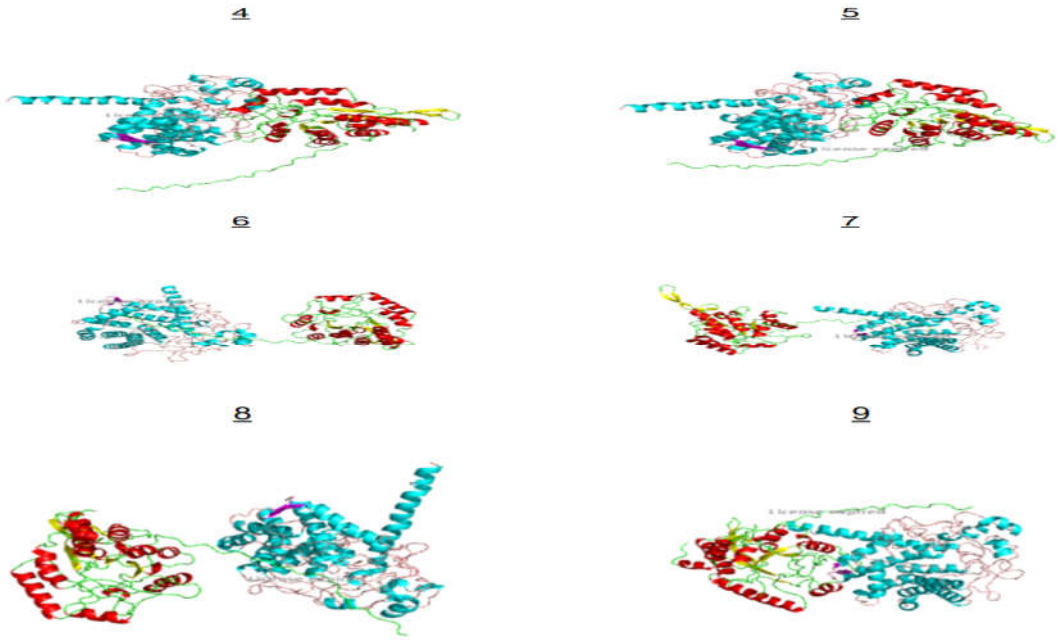


Fig no: 11 ClusPro – protein protein docking

A total of 30 structural clusters were formed based on the structural group, with cluster size ranging from 11 to 63 members. Representative structures for clusters 0–9 are shown in Figure 1. Each cluster displays a distinct folding arrangement, confirming the apparent structural diversity within the dataset. Clusters 0–3 displayed relatively compact conformations with well-defined secondary structures. These groups also had moderate to high member numbers (41–63 members), indicating stable and repeatedly sampled conformations. Clusters 4–9 showed more extended or partially open structures, indicating greater conformational flexibility. Comparison with energy data revealed a strong correlation between structural compactness and stability. The clusters with the lowest energy conformations, such as cluster 1 (-1093.8) and cluster 20 (-1100.4), demonstrated tightly packed helices and low loop exposure. In contrast, clusters with higher center energies, such as cluster 7 (-863.7) and cluster 14 (-825.2), display more open or loosely organized structures. Clusters 9, 11, 12, and 24 showed similar centers and lowest-energy values, indicating minimal internal variation. This structural uniformity is reflected in the representative models, which appear highly similar within each cluster. In contrast, clusters such as 5, 13, 19, and 25 showed larger differences between the center and lowest-energy scores, indicating greater intracluster heterogeneity. Overall, the clustering results indicate that the dataset contains both highly stable conformational families and flexible, less stable combinations. Visual structural comparison of clusters (0–9) supports the quantitative findings in the energy table, confirming that lower energy values correlate with more compact, stable structural arrangements, while higher energies correspond to more expanded or flexible conformations.

4. CONCLUSION:

In this study, a comprehensive in silico approach was employed to characterize bifunctional cellulase-xylanase enzymes at the structural, functional, and evolutionary levels. Sequence-based analysis confirmed the presence of distinct catalytic domains responsible for dual activity, while homology modeling produced reliable 3D structures supported by strong validation scores. Active site mapping via CASTp clearly identified distinct substrate-binding pockets for cellulase and xylanase functions. Molecular docking using ClusPro revealed multiple energetically favorable binding modes, with 30 structural clusters exhibiting clear conformational heterogeneity. The lowest-energy clusters, such as cluster 1 and cluster 20, showed compact and stable binding conformations, while the higher-energy clusters displayed more flexible or partially open arrangements. These results were further strengthened by molecular dynamics simulations, which demonstrated stable protein–substrate interactions and confirmed the dynamic suitability of the enzyme for bifunctional catalysis. Phylogenetic analysis placed the

enzyme within a conserved evolutionary lineage, highlighting its functional importance across species. Overall, this integrated computational workflow provides detailed structural and mechanistic insights into bifunctionality and establishes a strong foundation for future rational engineering of cellulase-xylanase enzymes. The findings present valuable implications for industrial applications, including biofuel production, biomass degradation, paper processing and other biotechnology processes where dual catalyst efficiency is important.

REFERENCE:

1) Mishra, S., Das, S., Rout, M., Pati, S., Kumar, R., & Dehury, B. (2023). Introduction to the World of Bioinformatic s. In *A guide to applied machine learning for biologists* (pp. 105-126). Cham: Springer International Publishing.

2) Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236-1240.

3) Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, *46*(W1), W296-W303.

4) Jiménez, J., Doerr, S., Martínez-Rosell, G., Rose, A. S., & De Fabritiis, G. (2017). DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics*, *33*(19), 3036-3042.

5) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, *30*(16), 2785-2791.

6) Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1*, 19-25.

7) Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6), 1547-1549.

8) Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, *59*(3), 307-321.

9) Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Totowa, NJ: Humana press.

10) Nielsen, H. (2025). Practical Applications of Language Models in Protein Sorting Prediction: SignalP 6.0, DeepLoc 2.1, and DeepLocPro 1.0. In *Large Language Models (LLMs) in Protein Bioinformatics* (pp. 153-175). New York, NY: Springer US.

11) Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic acids research*, *34*(suppl_2), W116-W118.