IN SILICO PREDICTION OF THE STRUCTURE OF AN α-L-ARABINOFURANOSIDASE FROM *CHITINOPHAGA* sp

Moisés M. Chagas[1]*, Bárbara B. Buzzo[2], Gabriel Zazeri[3], Tatiane Iembo[4], Eliana G. Macedo Lemos[5], Eleni Gomes[6] & Gustavo O. Bonilla-Rodriguez[6]

[1] *Post-graduate Program in Microbiology, IBILCE-UNESP, São José do Rio Preto, Brazil.*
[2] *Post-graduate Program in Agricultural Microbiology, FCAV-UNESP, Jaboticabal, Brazil.*
[3] *Physics Department at the Federal University of Roraima (UFRR), Paricarana, Brazil.*
[4] *Medical course at College Ceres (Faceres), São José do Rio Preto, Brazil.*
[5] *São Paulo State University "Júlio de Mesquita Filho", FCAV-UNESP, Jaboticabal, Brazil.*
[6] *São Paulo State University "Júlio de Mesquita Filho", IBILCE-UNESP, São José do Rio Preto, Brazil.*
** Corresponding author's email address: moises.chagas@unesp.br*

ABSTRACT

Brazilian soil harbors a wealth of microorganisms possessing largely untapped biotechnological potential, fueling the quest for novel enzymes. Although arabinofuranosidases are typically viewed as supplementary enzymes, their interplay with other lignocellulose-degrading enzymes presents a promising avenue for biotechnological advancement, offering alternatives to conventional chemical technologies. In this study, we aimed to investigate the biochemical structural properties of an α-L-Arabinofuranosidase from *Chitinophaga* sp., isolated from a soil sample via metagenomic analysis. Thus far, structural characteristics have been assessed via (*in silico*), yielding physicochemical parameters, three-dimensional structural models, and stability evaluations both with and without the presence of the histidine-tag (His-tag). Our findings indicate that the His-tag does not induce structural alterations, with stable radii, but noticeable fluctuations at elevated temperatures. Notably, the protein segment spanning leucine 520 to alanine 630 was the most flexible and temperature sensitive region. We anticipate that our analytical results will align with predictions and that further transformation and expression analyses in *Komagataella phaffii* will contribute to elucidating kinetic parameters.

Keywords: Bioinformatics. Computer simulation. Molecular biology. Filamentous fungi. Biomass degradation.

1 INTRODUCTION

The cell wall of plants is a semi-rigid and complex structure; it functions in cellular maintenance and protection, playing a fundamental role in the exchange of chemical substances necessary for their metabolism. The plant cell wall and/or lignocellulosic biomass from agricultural and forest residues are the most abundant renewable resources in the world and are mainly composed of cellulose (35-50%), hemicellulose (20-50%), and lignin (10-35%), as well as proteins and water [1,2]. Although abundant, it has limitations regarding access, since its integrity confers resistance to attack by pathogens, pests, and enzymatic degradation. Therefore, there is an ongoing search for more efficient and low-cost enzymes for biomass hydrolysis. Among the investigated enzymes, α-L-arabinofuranosidases (Abfs) (EC.3.2.1.55) are hemicellulases that hydrolyze arabinofuranosyl residues in hemicellulose. Belonging to the family of glycoside hydrolases (GH) 3, 43, 51, 54, and 62, they act mainly on the side chains of hemicellulose, which have α-L-1,2-, α-L-1,3-, and/or residual non-reducing α-terminals α-L1,5-arabinofuranosyl of polysaccharides, to release arabinose[3]. The hydrolytic action of arabinofuranosidases applies to the production of biofuels and other products of industrial interest (food, cosmetics, medicines, paper, and cellulose) and can be used in the form of cocktails together with pectins, mannanases, and xylans to degrade polysaccharides linked to lignin. For this purpose, prior knowledge of the protein structure is necessary through Computational Biology, which refers to the use of computing techniques and tools to solve biological problems. Bioinformatics stands out as the most promising area because it results in the resolution of problems such as sequence comparison (DNA, RNA, and proteins), identification and analysis of gene expression, and determination of protein structure[4,5]. Therefore, the objective of this study was to predict the three-dimensional structure of α-l-arabinofuranosidase from *Chitinophaga* sp., with and without the His-tag, and to analyze how this structure behaves at different pH and temperature values.

2 MATERIAL & METHODS

The coding sequence for the enzyme α-L-Arabinofuranosidase was selected from the database (CB10) and constructed from the sequencing of a microbial consortium. This sequence was provided by Dr. Luciano Takeshi Kishi at the time linked to the Laboratory of Biochemistry of Microorganisms and Plants (LBMP) at UNESP/FCAV in Jaboticabal-SP, which has the genome of Chitinophaga sp.[6] This draft genome was predicted to have 80.3% similarity to an uncharacterized protein from a *Chitinophaga jiangningensis*[7] and was renamed 'Martins1' for subsequent vector construction and *in silico* analysis. However, some necessary changes were made to the sequence to avoid problems in processing heterologous proteins in *Komagataella phaffii*. These modifications included the elimination of the signal peptide in the 5' portion of the gene and replacement of the processing sites for Kex2 and Ste13, which have an XhoI site (CTCGAG) at their end. Furthermore, a site for SalI was added to the 3' region, in phase with the codons of the six histidines already present in the vector. Finally, the internal site of BspHI was eliminated without changing the amino acid (phenylalanine), which is useful for vector integration.

The initial predicted coordinates of the protein with and without His-Tag subjected to molecular dynamics were obtained by structural modeling using the AI program (RoseTTafold)[8]. Molecular dynamics simulations were performed using GROMACS/5.4.1, with a GROMOS54a6 force field[9]. The system was solvated using a Simple Point Charge (SPC) water model in a cubic box, where the protein was centered at least 1 nm from the edge. The system was neutralized with Na^+ and Cl^- ions. Energy minimization was performed using the steepest descent algorithm with 5000 steps and a tolerance of 10 kJ/mol. The first equilibrium step carried out in the NVT ensemble took 100 ps of simulation at constant temperatures of 30, 50, 70, and 90 °C (coupled to a modified Barendsen thermostat). Random velocities were generated using the Maxwell-Boltzmann distribution. The second equilibrium stage, carried out in the NPT ensemble, took 100 ps of simulation at constant temperatures of 30, 50, 70, and 90 °C and a pressure of 1 atm (coupled to the Parrinello-Rahman barostat). Molecular dynamics simulations were performed with 2 fs steps using the leapfrog algorithm to integrate the equations of motion. The results are expressed as the average of the three simulations.

3 RESULTS & DISCUSSION

In silico analysis of the protein occurred through the insertion of the primary sequence of the α-L-Arabinofuranosidase from *Chitinophaga* sp. into software available on the Internet to predict physicochemical parameters, conserved domains, and possible post-translational modifications (ProtParam)[10].

Table 1 Predicted parameters of a recombinant α-L-Arabinofuranosidase from *Chitinophaga* sp.

Item	Unit
Number of amino acids	775
Molecular weight (kDa)	87
Theoretical pI	7.82
Total number of negatively charged residues (Asp+Glu)	78
Total number of positively charged residues (Arg+Lys)	79
Formula	$C_{3911}H_{6062}N_{1108}O_{1135}S_{29}$
Total number of atoms	12245

Different microorganisms such as bacteria, fungi, and actinomycetes can produce and secrete arabinofuranosidases. However, large-scale production requires the best strains to have a high production capacity. The values of the predicted parameters found in this study are close to those found in *Thermotoga thermarum*[11], while they differ from other species of bacteria such as *Celullomonas fimi*[12], *Paenibacillus curdlanolyticud*[13], *Alicyclobacillus* sp[14], *Alicyclobacillus acidocaldarius*[15], *Caldicellulosiruptor saccharolyticus*[16], *and Geobacillus vulcani*[17], as their enzymes have Mw values between 50 and 60 kDa and their temperature tolerance varies between 35 and 80 °C.

Subsequently, the structure of α-L-Arabinofuranosidase (Figure 1) was predicted using artificial intelligence (AI), resulting in a complex structure composed of alpha helices corresponding to 24% of the total structure and 22% beta sheets, while the remaining 54% corresponded to random coil.

The α-L-Arabinofuranosidase was modelled with and without a His-tag. This analysis confirmed that the presence of the His-tag did not significantly alter the structure of the protein. However, it was verified that a small structural change may occur in a location close to the His-tag (Figure 1, red circle), which probably would not alter its function but could change its stability.

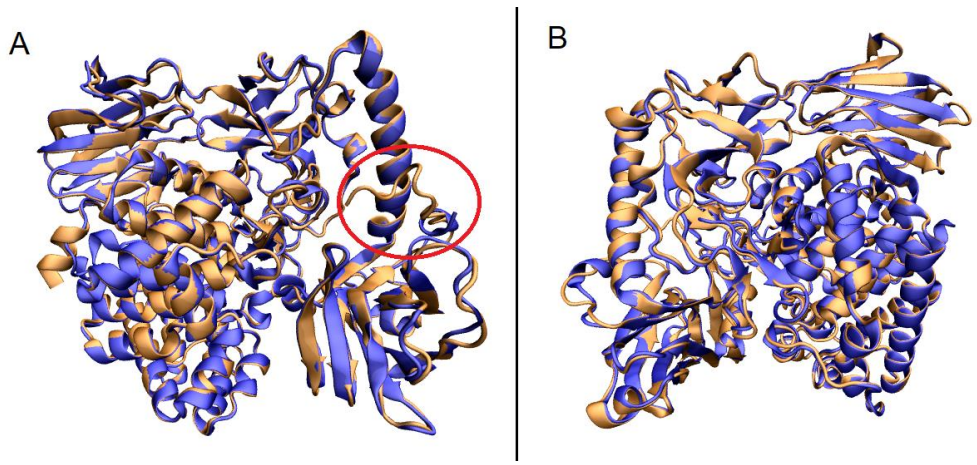

Figure 1 Three-dimensional (3D) predicted structure of α-L-arabinofuranosidase from *Chitinophaga* sp. The structure possessing a His-tag (in orange) and without (in blue) was modeled using the AI software RoseTTaFold (Washington University), which superimposed the VMD program (University of Illinois at Urbana Champaign).

The analysis of the radius of gyration or hydrodynamic radius with and without the His-tag was simulated at 180 nanoseconds (ns), showing that in the presence of the His-tag, there was a stability of more or less 2.75 nanometers (nm) with peaks above 2.77 nm at 70 °C, while in its absence, at a temperature of 30 °C, there was a decrease in Rg, meaning the protein would remain more stable without the His-tag. Regarding the analysis of the Mean Quadratic Deviation, in the presence of the histidine His-tag, the temperature range between 30 °C and 50 °C was more stable. However, when raising the temperature between 70 and 90 °C, there was a fluctuation; that is, the protein proved to be sensitive at higher temperatures. However, within the same temperature range, the proteins without the His-tag were stable. Analysis of the atomic mean square deviation verified the fluctuations in the protein of each of the 775 residues in different temperature ranges. In the presence of the His-tag, the backbone fluctuated at a temperature of 90 °C in the range of residues 520–630 with a peak above 0.8 nm, showing that this region of the protein is more flexible and sensitive to temperature changes than others, which was also observed when analyzed without the -His-tag.

4 CONCLUSION

Understanding the structure of an α-L-Arabinofuranosidase from *Chitinophaga* sp. through artificial intelligence and modeling tools will help in research on this and other proteins, as they serve as prediction models for real experiments and assist in the analysis of substrate specificity, stable protein rationality, and prediction of possible mutations. Such analyses are beneficial, as substrate specificity for arabinofuranosidase in a traditional model requires a large experimental determination, as well as high cost and time, since the substrate can include more than 30 types of various 4-Nitrophenyl (PNP)-glycosides, oligosaccharides, and polysaccharides.

REFERENCES

[1] BRETHAUER, S., STUDER, M. H. 2015. Biochemical Conversion Processes of Lignocellulosic Biomass to Fuels and Chemicals – A Review. Chimia (Aarau) 69 (10):572-581.

[2] RAZA, M., ABU-JDAYIL, B. 2022. Cellulose nanocrystals from lignocellulosic feedstock: a review of production technology and surface chemistry modification. Cellulose. 29 (2):685-722.

[3] AMORE, A., AMORESANO, A., BIROLO, L., HENRISSAT, B., LEO, G., PALMESE, A., FARACO, V. 2012. A family GH51 α-L-arabinofuranosidase from *Pleurotus ostreatus*: identification, recombinant expression, and characterization. Appl Microbiol Biotechnol. 4 (4):995-1006.

[4] SETUBAL, C., MEIDANIS, J. 1997. Introduction to Computational Molecular Biology. (ed). 1nd ed. PWS Publishing Co.

[5] BRUNAK, S., BALDI, P. 2001. Bioinformatics: The Machine learning Approach Bradford Book. (ed). 2nd ed. The MIT Press.

[6] KISHI, L. T., LOPES, E. M., FERNANDES, C. C., FERNANDES, G. C., SACCO, L. P., ALVES, L. M. C., LEMOS, E. G. M. 2017. Draft Genome Sequence of a *Chitinophaga* Strain isolated from a Lignocellulose Biomass-Degrading Consortium. Genome Announc. 5 (3): 01056-16.

[7] ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., MILLER, W., LIPMAN, D. J. 1997. “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

[8] BAEK, M., DIMAIO, F., ANISHCHENKO, I., DAUPARAS, J., OVCHINNIKOV, S., LEE, G. R., WANG, J., CONG, Q., KINCH, L. N., SCHAEFFER, R. D., MILLÁN, C., PARK, H., ADAMS, C., GLASSMAN, C. R., DEGIOVANNI, A., PEREIRA, J.H., RODRIGUES, A.V., VAN DIJK, A. A., EBRECHT, A. C., OPPERMAN, D. J., SAGMEISTER, T., BUHLHELLER, C., PAVKOV-KELLER, T., RATHINASWAMY, M. K., DALWADI, U., YIP, C. K., BURKE, J. E., GARCIA, K. C., GRISHIN, N. V., ADAMS, P. D., READ, R. J., BAKER, D. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 373 (6557):871-876.

[9] VAN DER SPOEL, D., LINDAHL, E., HESS, B., GROENHOF, G., MARK, A. E., BERENDSEN, H. J. 2005. GROMACS: Fast, flexible, and free. J Comput Chem. 26 (16): 1701-18.

[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: Walker, J. M., Ed., The Proteomics Protocols Handbook, Humana Press, New York. 571-607.

[11] SHI, H., ZHANG, Y., XU, B., TU, M., WANG, F. 2014. Characterization of a novel GH2 family α-l-arabinofuranosidase from hyperthermophilic bacterium *Thermotoga thermarum*. Biotechnol Lett. 36 (6):1321-1328.

[12] YANG, Y., ZHANG, L., GUO, M., SUN, J., MATSUKAWA, S., XIE, J., WEI, D. 2015. Novel α-L-Arabinofuranosidase from *Cellulomonas fimi* ATCC 484 and Its Substrate-Specificity Analysis with the Aid of Computer. J. Agric. Food Chem. 63 (14):3725-3733.

[13] LEE, S. H., LEE, Y. E. 2014. Cloning, expression, and characterization of a thermostable GH51 alpha-L-arabinofuranosidase from *Paenibacillus* sp. J Microbiol Biotech. 24 (2):236-244.

[14] YANG, W., BAY, Y., YANG, P., LOU, H., HANG, H., MENG, K., SHI, P., WANG, Y., YAO, B. 2015. A novel biofunctional GH51 exo-α-L-arabinofuranosidase/endo-xylanase from *Alicyclobacillus* sp. A4 with significant biomass-degrading capacity. Biotechnol Biofuels (8):197.

[15] AKKAYA, A., ENSARI, Y., OZSEKER, E. E., BATUR, O. O., BUYURAN, G., EVRAN, S. 2023. Recombinant Production and Biochemical Characterization of Thermostable N-arabinofuranosidase from Acidothermophilic Alicyclobacillus acidocaldarius. Protein J. 42 (4):437-450.

[16] LIM, Y. R., YOON, R. Y., SEO, E. S., KIM, Y. S., PARK, C. S., OH, D. K. 2010. Hydrolitic properties of a thermostable α-l-arabinofuranosidase from *Caldicellulosiruptor saccharolyticus*. J Appl Microbiol. 109 (4):1188-1197.

[17] ILGÜ, H., SÜRMELI, Y., SANLI-MOHAMED, G. 2018. A thermophilic α-L-Arabinofuranosidase from *Geobacillus vulcani* GS90: heterologous expression, biochemical characterization, and its synergistic action in fruit juice enrichment. Eur Food Res Technol. 244 (9):1624-1636.

ACKNOWLEDGEMENTS

This work was supported by the Amazonas State Research Support Foundation (FAPEAM), financing code 012/21, and São Paulo Research Foundation (FAPESP).