

## Structure Flexibility of Alpha-galactosidase from a Marine Psychrophilic Yeast, *Glaciozyma antarctica* PI12

Shuhaila Mat-Sharani<sup>1\*</sup>, Reyad Al-Moheer<sup>2</sup>, Farah-Diba Abu-Bakar<sup>3</sup>, Nor-Muhammad Mahadi<sup>4</sup>, and Abdul-Munir Abdul-Murad<sup>3</sup>



OPEN ACCESS

<sup>1</sup> School of Health Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

<sup>2</sup> Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

<sup>3</sup> Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM-Bangi, Selangor, Malaysia

<sup>4</sup> Malaysia Genome Institute, National Institutes of Biotechnology Malaysia, Jalan Bangi, Kajang 43000, Selangor, Malaysia. Current Affiliation: Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM-Bangi, Selangor, Malaysia

\*Corresponding Author: shuhailamsr@usm.my

**Received:** 20 July 2022

**Accepted:** 30 September 2022

**Published:** 30 December 2022

**Academic Editor:** Dr. Fazilah Ariffin

<sup>©</sup>Squalen Bulletin of Marine and Fisheries Postharvest and Biotechnology, 2021. Accreditation Number: 148/M/KPT/2020. ISSN: 2089-5690, e-ISSN: 2406-9272. <https://doi.org/10.15578/squalen.680>

### Abstract

Factors that contribute to maintaining the flexibility or stability of an enzyme structure may depend on the composition of each amino acid with different characteristics, providing a purpose and bonding features within the structure. Based on this assumption, a study using homology modeling and a comparative study to observe different structure behaviors of an enzyme at an extremely low temperature (psychrophile) against temperate (mesophile) and high temperature (thermophile) was performed. The subject,  $\alpha$ -galactosidase from *Glaciozyma antarctica* as a marine psychrophilic candidate was chosen against  $\alpha$ -galactosidase from *Trichoderma reesei* (mesophile) and *Ramsonia emersonii* (thermophile). This enzyme catalyzes the hydrolysis of  $\alpha$ -1-6 linked terminal galactosyl residues which can be found in a wide range of the organism. The ability of *G. antarctica* to grow in extremely cold temperatures rendered the question that the enzyme must have special characteristics to adapt to the cold condition. Based on the homology modeling and molecular dynamics study, a comparison of the structure of *G. antarctica*  $\alpha$ -galactosidase enzymes with its homolog from the mesophilic and thermophilic fungi showed that *G. antarctica*  $\alpha$ -galactosidase enzyme confers its flexibility by the increased number of small amino acids with reduced charges, more loops, a fewer number of hydrogen and disulfide bonds in its structure. Furthermore,  $\alpha$ -galactosidase has potential for commercialization in bleach paper and the baking industry also a treatment for bloating and Fabry disease.

**Keywords:** *Glaciozyma antarctica*, psychrophilic yeast, cold adaptation, structure flexibility, molecular dynamics

### Introduction

Psychrophilic microorganisms growing at or below 15 °C are available in various permanently cold environments, including Antarctica. For a long time, Antarctica has been known as a geographical area that is characterized and investigated by microbiologists to harbor a diversity of cold-adapted microorganisms (Bajaj & Singh, 2015). Psychrophilic microorganisms that can survive under such extreme conditions require complex physiological, morphological and metabolic adaptations (Yusof et al., 2021). Enzymes, also defined as proteins that can catalyze all biochemical reactions within an organism that render them compatible with life, are described as essential targets for the adaptation of an organism to a cold environment (Parvizpour et

al., 2021). Cold-adapted enzymes are known for their high catalytic efficiency at low temperatures resulting from an inherent increase flexibility. This will allow the molecular motions necessary for activity in low energy environments leading to the observed low stability of these enzymes (Collins et al., 2003). Several factors can impact the flexibility of psychrophilic enzymes including reduction of the density of a protein structure, the number of hydrophobic side chains that are exposed to the solvent, the high number of glycine and lysine, the low number of arginine and proline and the weak intramolecular bond (Metpally & Reddy, 2009; Mohammadi et al., 2018).

The interest in cold-adapted enzymes, both as models for thermal stability studies and molecular

protein adaptation, and as potential candidates for biotechnological applications has been long been discussed (Adapa et al., 2022; Parvizpour et al., 2021; Yusof et al., 2021).  $\alpha$ -Galactosidase (EC 3.2.1.22) belonging to glycosyl hydrolase family 27 is an enzyme involves in the hydrolysis of the terminal  $\alpha$ -galactosyl moieties from oligosaccharides such as melibiose, raffinose, stachyose and verbascose, galactopolysaccharides and glycoconjugates such as glycolipids and glycoproteins (Álvarez-Cao et al., 2019; Bhatia et al., 2020).  $\alpha$ -Galactosidase has been reported to be found in a variety of organisms including bacteria (Fei et al., 2020; Xu et al., 2014), fungi (Xie et al., 2020), plants (Fujimoto et al., 2003; Tsaniklidis et al., 2016) and human (Dwyer et al., 2021; Modrego et al., 2021). This enzyme is widely used and mostly found in the markets as intestinal gas prevention treatment (Fabris et al., 2021), increasing crystallized sugar yield (Belcourt & Labuza, 2007), improving bleaching in the pulp and paper industry (Clarke et al., 2000), improving the nutritional value of feed additives (Ghazi et al., 2003), in the baking industry (Liljestrom-Suominen et al., 1988) and in the medicinal treatment of Fabry disease (Germain et al., 2016).

In elucidating the cold adaptation of psychrophilic  $\alpha$ -galactosidase from *G. antarctica*, we modeled the structure and compared it to the homologous  $\alpha$ -galactosidase from mesophilic and thermophilic enzyme of *Trichoderma reesei* and *Ramsonia emersonii*, respectively. Molecular dynamic (MD) was also utilized in this study to assess the flexibility of  $\alpha$ -galactosidase of a marine psychrophilic yeast, *G. antarctica* with mesophilic and thermophilic  $\alpha$ -galactosidase from fungi.

## Materials and Methods

### cDNA Amplification of $\alpha$ -galactosidase

First-strand cDNA was synthesized using a kit from Super Script™ III First-Strand Synthesis System for RT-PCR (Invitrogen, USA). A cDNA encoding  $\alpha$ -galactosidase was amplified using specific primers Lan-08-216F 5' ATGCTTCTCACTCGCCTCGGATCAC for sense primer and Lan-08-216R 5' TCATCTCCACCGTCCCCGCCTCACC for antisense primer. PCR products were inserted into a pGEM-T Easy cloning vector, transformed into *E. coli* DH5 $\alpha$  and verified as  $\alpha$ -galactosidase cDNA via sequencing.

### $\alpha$ -Galactosidase Data Mining

*Glaciozyma antarctica* PI12 is a marine psychrophilic yeast obtained from Casey Research Institute, Antarctica (Firdaus-Raih et al., 2018).  $\alpha$ -galactosidase sequence of *G. antarctica* encoded by

LAN\_08\_216 was downloaded from *G. antarctica* database [http://mfrlab.org/glacier/]. The crystal structure of mesophilic fungi, *T. reesei* 1SZN was obtained from Protein Data Bank [http://www.rcsb.org] while the sequence of the thermophilic enzyme from *R. emersonii* (NCBIID:XP013331263.1) was retrieved from National Center for Biotechnology Information (NCBI) [http://www.ncbi.nlm.nih.gov/]. Amino acid sequences from these three organisms were subjected to various sequence analyses with BLASTP (Johnson et al., 2008), ProtParam (Gasteiger et al., 2005), InterProScan (Mulder & Apweiler, 2007) and SignalPv4.0 (Petersen et al., 2011).

### Sequence and Structure Analysis of $\alpha$ -galactosidases

The construction of predicted structures for both  $\alpha$ -galactosidases from *G. antarctica* and *R. emersonii* was done using MODELLER9v10 (Sali, 2008). Firstly, both sequences were blasted against the PDB database and a template was chosen based on the identity higher than 30% and e-value lower than  $10 \times e^{-05}$ . The optimization and energy minimization of the resulting model was performed using force field CHARM-22 and algorithm steepest descent from Accelrys Discovery Studio 2.5. The structure quality was evaluated using PROCHECK (Laskowski et al., 1996), VERIFY3D (Eisenberg et al., 1997), ERRAT (Colovos & Yeates, 1993) and ProSA-web (Wiederstein & Sippl, 2007). Structure superposition was performed with SuperPose (Maiti et al., 2004). Both the hydrogen bonds and the disulfide bond were defined using DSSP (Kabsch & Sander, 1983), STRIDE (Heinig & Frishman, 2004), and PDB file. Figures of all molecular structures were generated using CHIMERA (Pettersen et al., 2004).

### Molecular Dynamic (MD) Simulation of $\alpha$ -galactosidases

MD simulations of  $\alpha$ -galactosidases were simulated using GROMACS 4.5.5 and gromos 54A7 force field. To examine the structure stability, MD simulation was carried out at room temperature for *G. antarctica*, *T. reesei*, and *R. emersonii*.  $\alpha$ -Galactosidases were solvated in a 150 box of explicit simple point charge (SPC) water molecules and stimulated using periodic boundary condition (PBC) with a minimum of 0.1 nm between solute and box. The structures were further minimized by several steps for minimization using the steepest descent method. Simulations at certain temperatures were conducted with applied isotropic pressure. The result showed that after 200 ps, all systems were equilibrated using solute position-restrained MD. LINCS algorithm was used to constrain

bond length with a time step of 2 fs for all calculations. The particle mesh Ewald (PME) summation was applied to improve the electrostatic interactions. Van der Waals and Coulomb interactions were truncated at 0.1 nm and conformation was stored every 2 ps. The systems were equilibrated then subjected to MD simulations for 20 ns (20000 ps). Protein structures were further visualized using Visual Molecular Dynamics (VMD). The stability and flexibility were counted in backbone root mean square deviations (RMSD) for 20 ns while in carbon-alpha and protein-h root mean square fluctuation (RMSF) from the last 5 ns.

## Result and Discussion

### cDNA Amplification of $\alpha$ -galactosidase

The sequence of  $\alpha$ -galactosidase *G. antarctica* was predicted from genome annotation analysis using computational analysis. To validate the sequence, cDNA was amplified and sequenced with an open reading frame of 1404 bp. The protein consists of 467 amino acid residues with a predicted molecular weight of 48.59 kDa and isoelectric PI of 4.7. Signal peptide analysis of  $\alpha$ -galactosidase *G. antarctica* using SignalP 4.1 (Petersen et al., 2011) indicated the presence of a signal peptide with 21 amino acids residues at the N-terminal of the protein (Figure 1). The presence of a signal peptide suggested that the protein is a secreted protein. Furthermore, the presence of transmembrane protein sequences was predicted using TMHMM software. It was found that  $\alpha$ -galactosidase did not have any transmembrane domain hence this suggests that the protein was not located in the membrane.

### $\alpha$ -galactosidase Data Mining

The sequence of  $\alpha$ -galactosidase from *G. antarctica* and thermophilic fungi of, *R. emersonii* were submitted to BLASTP against the Protein Data Bank (PDB) database to obtain a suitable template. Prior to that, the signal peptides were removed from *G. antarctica* and *R. emersonii* at a cleavage position of 21 and 24, respectively. The result of the BLASTP against the PDB suggested that  $\alpha$ -galactosidase from *G. antarctica* and *R. emersonii* was built using a homology modeling approach since the identity to the template was approximately 50% with all known structures in the PDB database. The *G. antarctica*  $\alpha$ -galactosidase sequence matched with 3A5V chain A from mesophilic fungi, *Mortierella vinacea* with 49% identity and the resolution at 0.2 Å, while *R. emersonii*  $\alpha$ -galactosidase matched with  $\alpha$ -galactosidase (1SZN chain A) from mesophilic fungi, *T. reesei* with 59% identity and resolution 1.54Å. The PDB structure from *T. reesei* (1SZN) was thus selected as the highly probable

template for *G. antarctica*  $\alpha$ -galactosidase due to its higher structural identity. It is noted that the sequence alignment of *G. antarctica*  $\alpha$ -galactosidase against *T. reesei* and *R. emersonii*'s counterpart showed low sequence identity, which is 32% and 39%, respectively. Although the sequence similarity is low, all the sequences have the same domain of  $\alpha$ -galactosidase.

### Sequence Analysis of $\alpha$ -galactosidase

The cold-adapted enzyme is capable of functioning efficiently at low temperatures and this represents one of the adaptive mechanisms for cold environments. Some characteristics of the cold-adapted enzymes were suggested based on the comparisons between the primary structures of the respective enzymes from different organisms with different growth temperature ranges (Sheridan et al., 2000). Furthermore, analyses of the crystal structures of cold-adapted enzymes revealed the diversity of their cold adaptation mechanisms (Bentahir et al., 2000; Russell et al., 1998). Based on the amino acid composition,  $\alpha$ -galactosidase from *G. antarctica* has a higher non-polar group than  $\alpha$ -galactosidase from *T. reesei* and *R. emersonii*.

```

1 atgctctcactcgcctcggatcaactcgcgctctctgctcgcctctgagcgcctgcttc 60
1 M L L T R L G S L A V F A R L A T P A F 20
61 gcccctcga caa tgggttggcctctccgcctcctggctggagctcgtggaacaacttc 120
21 A L D N G V G V L P A L G W N S W N T F 40
121 gctaccgacatcagcagggagatcatcaatcaccgctcgcagagagatggttaactctggg 180
41 A T D I S E E I I I T A A K E M V N S G 60
181 ctgctcgcctaggatcaagatcatctggtcgcagcagcagcagcagcagcagcagcagc 240
61 L A A L G Y K Y I L V D D G A T S H E R 80
241 acgcccgaaggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
81 T P E G A I Q A N T T K F P S G V K N L 100
301 accgacagatattcaatcaataaacctcaaggtcggcagcagcagcagcagcagcagcagc 360
101 T D S I H S I N L K V G M Y S D A G E F 120
361 accctcgcctcctaccctcctcctcctgggacaagtcctgacgagcagcagcagcagcagc 420
121 T C G S Y P G S L G H E V S D A Q T W A 140
421 accctgggagtgattaccctcaagtaacgacacctcctcaacaaagcagcagcagcagcagc 480
141 A W G V D Y L K Y D N C F T N Q G Q S G N 160
481 gccaacattctctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540
161 A N I S Y E R Y E R M S R A L N A T D R 180
541 gcaattctcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 600
181 P I L Y S L C T W G Q D N V W D W A S T 200
601 atcgcgaacagcctaccctctcctcctcctcctcctcctcctcctcctcctcctcctcctc 660
201 I A N S Y R I S G D I Y P F D R F D A 220
661 agatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 720
221 R C P C D T Y D C L L P G Y H C S V T N 240
721 atcgtcga caa gggcgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 780
241 I V D K G A P L G Q K V R A G S Y A G D W 260
781 gacagctcgaagtaggctcctggagctcctgagcagcagcagcagcagcagcagcagcagc 840
261 D H L E V G V G L H T P A E Q V A H F 280
841 tccctcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 900
281 S L W S M A K S P L I L G N D L G N M S 300
901 accgagcagccttctattacgca caa agtctgatttccatca caa cagcagcagcagcagc 960
301 T E T L S I I S N K V M I S I N Q A D S L 320
961 tccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1020
321 S S P A H R V F K K E V A A E E D S S L 340
1021 gcaaccctggcagcctcagcctcaca cgcctcctcctcctcctcctcctcctcctcctcctc 1080
341 A T W Q L Q P T T A G S L Q L W S G R L 360
1081 gaagatggatctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1140
361 E D G F V V A L F N T S P R N L T Y D L 380
1141 gagctcagcagggcctcgaacatttctgactaccgcccagggcgaacactcctcctgatat 1200
381 E L S E A L D I F D Y G P E A Q T L L Y 400
1201 aaactatcagcagcctcggcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1260
401 N V Y D V W A P A A D P T A P K N A S R 420
1261 gcgagagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1320
421 A E A A S R L C G A D A G V V S G S I P 1340
1321 gcctctggttgtaggagcagtgaggatagggctctcctcctcctcctcctcctcctcctcct 1380
441 A V V V E A H G I R V F K L T A A M E G 460
1381 gtgagcgggagcaggtggagatga 1404
461 V R R G R W R * 467

```

Figure 1. The cDNA and amino acid sequences of  $\alpha$ -galactosidase from *G. antarctica*. The nucleotide sequences are represented by lowercase, while amino acid sequences are represented by uppercase letters. Amino acid sequences which are bold and italic represent a signal peptide sequence. The stop codon was indicated by asterisks. This cDNA has 1404 base pairs which encode 467 amino acids.

Besides that, several charges and the aromatic amino acid group is lower in *G. antarctica*'s  $\alpha$ -galactosidase compared to *T. reesei* and *R. emersonii*. The number of small amino acids in *G. antarctica*'s  $\alpha$ -galactosidase is higher compared to *T. reesei* and *R. emersonii*. Furthermore, the psychrophilic enzyme has a greater number of glycine, which may contribute to structure flexibility. Editing to that, a lower number of charged amino acids present in  $\alpha$ -galactosidase of *G. antarctica* also contributes to increased flexibility due to the interaction with ion pairs that are important in stabilizing protein surface conformation (Table 1).

Many prolines were substituted in *G. antarctica*'s  $\alpha$ -galactosidase at positions 2, 67, 136, 306, 320, 329, 338, and 365 with either hydrophobic (Alanine and Valine), polar (glutamine and asparagine) or charged (Aspartic acid) residue while proline is conserved throughout *T. reesei* and *R. emersonii* (Figure 2). Proline is inherently absent in a psychrophilic enzyme such as *G. antarctica*'s  $\alpha$ -galactosidase. This is because the amino acid side chain is covalently linked to the N-C $\alpha$  bond of the backbone (Davail et al., 1994), its pyrrolidine ring imposes severe restrictions on the conformational freedom, causing structural rigidity. The strict dependence between main- and side-chain proline conformations and the lack of a backbone hydrogen-bonding donor. This structure limits proline capability to form main-chain hydrogen bonds. Because of these intrinsic properties, the presence of proline strongly influences the protein structure by increasing protein stability and local rigidity by reducing the configurational entropy of unfolding and reducing the flexibility of the structure.

Another amino acid group that contributes to the stability of a protein is aromatic properties (tryptophan

and phenylalanine). An aromatic amino acid may contribute to ion pair electrostatic interactions which are a critical binding force to maintain conformational

Table 1. Amino acid composition of  $\alpha$ -galactosidase from *G. antarctica*, *T. reesei* and *R. emersonii* analysis

	<i>G. antarctica</i> Psychrophilic 446aa	<i>T. reesei</i> Mesophilic 417aa	<i>R. emersonii</i> Thermophilic 428aa
Non-polar (A, G, F, I, L, M, P, V, W)	48.1%	47.0%	45.3%
Polar (C, N, Q, S, T, Y)	28.6%	27.6%	29.4%
Small (A, C, V, G, N, P, S, D, T)	45.0%	46.3%	43.5%
Gly (G)	9.1%	8.4%	9.6%
Pro (P)	4.1%	4.6%	6.1%
Charged(D, E, H, K, R)	20.2%	21.8%	21.0%
Glu (E)	4.3%	2.9%	4.7%
Asp (D)	6.4%	8.4%	7.7%
Aromatic (F, H,W,Y)	11.7%	13.2%	21.0%
Isoelectric point (pI)	5.1	5.0	4.5
GRAVY	-0.17	-0.23	-0.42
Arg/ (Arg +lys) ratio	0.54	0.45	0.50

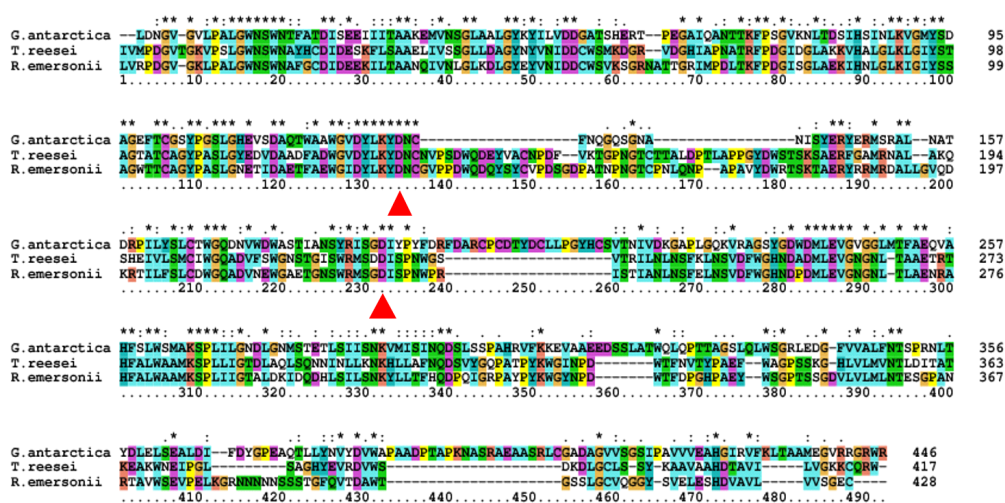


Figure 2. Multiple sequence alignment of  $\alpha$ -galactosidase from *G. antarctica* (psychrophilic), *T. reesei* (mesophilic), and *R. emersonii* (thermophilic). Catalytic residue label as a triangle, amino acid in same group label as ‘:’ and conserved amino acid label as ‘\*’.



stability on the surface of the proteins (Metpally & Reddy, 2009). Therefore, charged amino acids conserved in *T. reesei* and *R. emersonii* were substituted with small amino acids (Alanine, Proline, Isoleucine, and Serine).

Substitutions of glycine were found in *G. antarctica* species  $\alpha$ -galactosidase at positions 53, 102, 224, 228, 234, 237, 248, and 401 (Figure 2). This substitution may increase structural flexibility since the main chain of glycine is lack a side chain and this will allow chain rotations and dihedral angles. A structure that can change shape will reduce the energy required in confirmation modification to interact with the substrate (Thorvaldsen et al., 2007).

Other than the amino acid composition analysis, secondary structure was predicted using STRIDE to search for buried and exposed residue on secondary structure (Figure 3). Thorvaldsen group reported that more charged amino acid exposure will help an enzyme interact with a solvent (Thorvaldsen et al., 2007). The majority of exposed amino acids namely Asp, Glu, and Lys were higher in *G. antarctica*'s  $\alpha$ -galactosidase compared to *T. reesei* and *R. emersonii*'s. The increment of charged residues at the surface of a protein, especially the negatively charged amino acid is important to increase solvent interaction and structure flexibility.

### Homology Modeling of $\alpha$ -galactosidase

The structure of the enzyme will visualize the features of the enzyme in a 3D form. The structure of

$\alpha$ -galactosidase *T. reesei* was already solved in the PDB database. Therefore, the structure construction of  $\alpha$ -galactosidase was done only for *G. antarctica* and *R. emersonii*. There are more than 50% of identities in the sequence alignment from BLAST search against the PDB database. The structure of  $\alpha$ -galactosidase *G. antarctica* and *R. emersonii* were created by MODELLER using 3A5V and 1SZN as a template (Golubev et al., 2004).

Structure validation of  $\alpha$ -galactosidase from *G. antarctica*, *T. reesei*, and *R. emersonii* was evaluated using PROCHEK, VERIFY3D, ERRAT, and ProSA-Web. PROCHECK analysis revealed the distribution of amino acid backbone conformation that was evaluated by inspecting the Psi/Phi Ramachandran plot (Laskowski et al., 1993). The Ramachandran plot with more than 80% residues located in the most favored region showed a good structure of  $\alpha$ -galactosidase *G. antarctica* and *R. emersonii* while others are in the allowed, generous and disallowed region (Table 2). The structure of  $\alpha$ -galactosidase *T. reesei* (1SZN) was also analyzed for structure validation to make a comparison of all three structures whereby 90% of residues are in the most favored region (Table 2). Verify3D analyzed the compatibility of the structure to its amino acid sequence using a 3D profile derived from the atomic coordinate structure. The 3D profiles of proper protein structures match their sequence with a score of more than 80% indicating good structure quality (Lüthy et al., 1992) whereas  $\alpha$ -galactosidase of *G. antarctica*, *T. reesei* and *R. emersonii* structures have more than 87% residues complement with 1D-3D profile (Table 2).

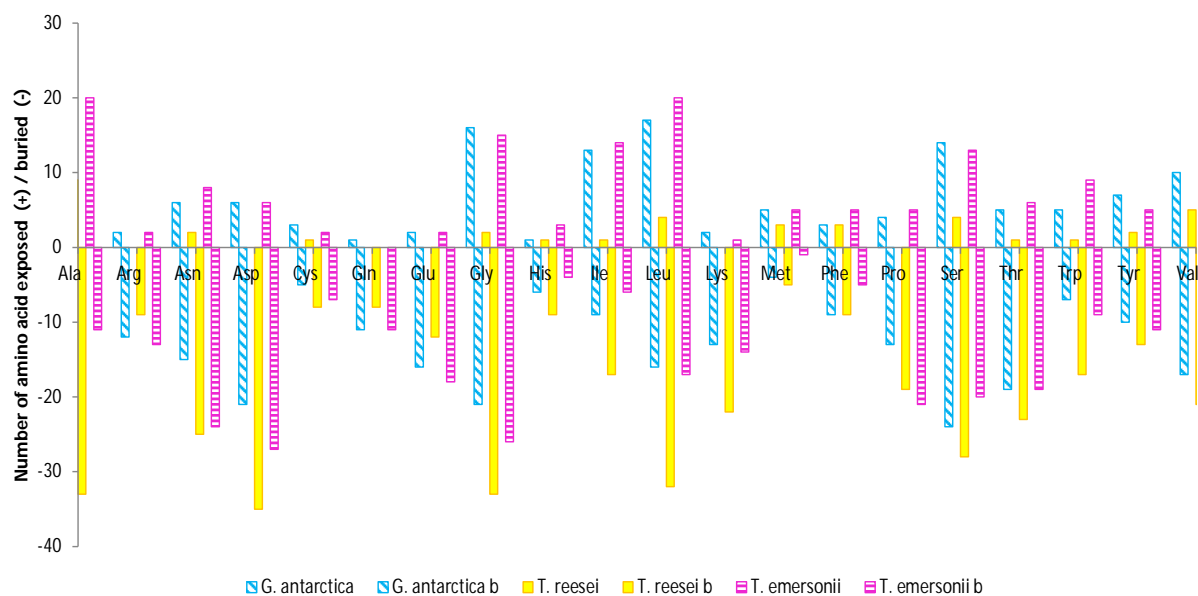


Figure 3. Analysis of amino acid composition of secondary structure for  $\alpha$ -galactosidase enzyme. Exposed (+) or Buried (b) (-) amino acid based on structure  $\alpha$ -galactosidase *G. antarctica*, *T. reesei*, and *R. emersonii*. The majority of exposed amino acids *G. antarctica* are Gly, Ser, Cys, Lys, Gln, and Tyr compared to  $\alpha$ -galactosidase *T. reesei* and *R. emersonii*. While, Ala, Ile, Ser, and Val are mostly buried amino acids in  $\alpha$ -galactosidase *G. antarctica* b compared to  $\alpha$ -galactosidase *T. reesei* b and *R. emersonii* b.

ERRAT which is useful for finding incorrectly-folded regions in preliminary protein models indicates an overall quality factor for non-bonded atomic interactions and higher scores mean better quality (Colovos & Yeates, 1993). This analysis showed that all  $\alpha$ -galactosidase structures have a good quality model with a score of more than 80%. ProSA-web is an interactive web service that recognizes errors in three-dimensional protein structures (Wiederstein & Sippl, 2007). ProSA-web analyzes all  $\alpha$ -galactosidase structures and denoted a value from -5 to -9 signifying the range within native conformations. The findings quoted above indicate that the protein model is reliable and of good quality.

### Structure Comparison Between $\alpha$ -galactosidase Psychrophilic Fungi, *G. antarctica* with $\alpha$ -galactosidase from Mesophilic and Thermophilic Fungi

Superimpose analysis of three structures using SuperPose (Maiti et al., 2004) and RMSD value show for  $\alpha$  carbon, backbone, and whole structure were 3.057 Å, 3.037 Å, and 3.130 Å, respectively. All the structures showed to have a very similar structure. Furthermore, two catalytic residues of  $\alpha$ -galactosidase that are Asparagine were conserved in *G. antarctica*, *T. reesei*, and *R. emersonii* at Asp129-Asp189, Asp132-Asp226, and Asp133-Asp229, respectively (Figure 4). The difference between  $\alpha$ -galactosidase *G. antarctica*

compared to  $\alpha$ -galactosidase *T. reesei* and *R. emersonii* is the presence of long and many loops (Figure 4). This suggests that  $\alpha$ -galactosidase *G. antarctica* is more flexible than  $\alpha$ -galactosidase *T. reesei* and *R. emersonii* (Georlette et al., 2003).

Hydrogen bond and disulfide bridge have been reported to play a significant role in protein structure stabilization (Tronelli et al., 2007). The disulfide bridge is an important feature in the folding process and analysis of structural properties and functions of a certain protein (Tronelli et al., 2007). *G. antarctica* is found to have less disulfide bridge and hydrogen bonds compared to *T. reesei* and *R. emersonii* (Table 2). This result suggested that *G. antarctica* has more flexible structure compared to *T. reesei* and *R. emersonii*.

The structure of *G. antarctica*  $\alpha$ -galactosidase contains three disulfide bridges namely Cys101-Cys131, Cys201-Cys215, and Cys208-Cys203 that are close around the catalytic site (Asp129 and Asp189) as shown in Table 2. Therefore, it is suggested that *G. antarctica* maintains its structural stability surrounding the catalytic site to ensure the functionality of the enzyme. Besides that, both  $\alpha$ -galactosidase *T. reesei* and *R. emersonii* have four disulfide bridges in which *T. reesei* at Cys24-Cys56, Cys104-Cys134, Cys160-Cys147, and Cys392-Cys414 position. While, disulfide bridges in *R. emersonii* were at Cys23-Cys55, Cys105-135, Cys148-Cys163, and Cys405-Cys428 position (Table 2). Three of these four disulfide bridges were located around the catalytic site (Asp132-Asp226 and Asp133-Asp229). Meanwhile, only one disulfide bridge is in  $\alpha$ -galactosidase *T. reesei* and *R. emersonii* near the N-terminal at Cys405-Cys428 and Cys392-Cys414 position, respectively. Thus,  $\alpha$ -galactosidase *T. reesei* and *R. emersonii* have a more rigid structure compared to  $\alpha$ -galactosidase *G. antarctica* structure.

Other than that, the hydrogen bond also helps stabilize the protein structure. The low number of hydrogen bonds will make the structure less rigid (Tronelli et al., 2007). These analyses show that *G. antarctica* has 510 hydrogen bonds while *T. reesei* and *R. emersonii* have 536 and 517, respectively. Therefore, the structure of  $\alpha$ -galactosidase *G. antarctica* is more flexible than  $\alpha$ -galactosidase *T. reesei* and *R. emersonii* because of the lower number of hydrogen bonds (Table 2).

To conclude, the *G. antarctica* enzyme has a high number of small amino acids and a low number of charged and aromatics amino acid group compared to *T. reesei* and *R. emersonii*. Meanwhile, a non-polar amino acid is higher in *G. antarctica* than in *T. reesei* and *R. emersonii*. Furthermore, *G. antarctica* also has a low number of disulfide bridges and hydrogen bonds.

Table 2 Structure validation  $\alpha$ -galactosidase from *G. antarctica*, *T. reesei* and *R. emersonii* analysis

	<i>G. antarctica</i> (Psychrophilic)	<i>T. reesei</i> (Mesophilic)	<i>R. emersonii</i> (Thermophilic)
PROCHECK	88.8%	94.2%	89.4%
VERIFY 3D	87.47%	96.41%	96.04%
Errat	80.05%	91.12%	79.76%
ProSA-Web	-5.67	-8.46	-9.1
RMSD	0.6 Å	0.5 Å	0.5 Å
z-score	7.8	-	7.9
Template/ Model	3a5v 2Å (Mesophilic fungi)	-	1szn 1.54Å (Mesophilic fungi)
SIM	52.6%	-	58.7%
Intramolecule hydrogen bond	510	536	517
Disulfide bond	3	4	4

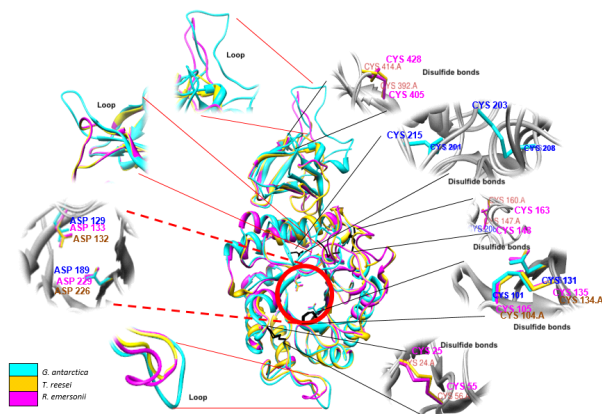


Figure 4. Structure comparison of  $\alpha$ -galactosidase *G. antarctica* (cyan), *T. reesei* (yellow), and *R. emersonii* (magenta) with conserved catalytic residue position (ASP-ASP) (zoom in red circle) while *G. antarctica* (cyan) has long and more loop and a smaller number of disulfide bonds (CYS-CYS) compared to *T. reesei* (yellow) and *R. emersonii* (magenta).

In conclusion, *G. antarctica* structure is more flexible with long and many loops and it also consists of amino acids with small side chains compared to *T. reesei* and *R. emersonii* (Figure 4).

### Molecular Dynamic (MD) Simulation of $\alpha$ -galactosidases

To confirm the structural flexibility of *G. antarctica*, MD analyses were carried out using GROMACS with specific parameters. The molecular dynamics were done for all structures  $\alpha$ -galactosidase *G. antarctica*, *T. reesei*, and *R. emersonii*. MD simulation was performed at room temperature (298 K) since the crystal structure from *T. reesei* is a mesophilic microorganism that grows at temperate temperature. Enzyme stability obtained from RMSD analysis showed a consistent plot. The  $\alpha$ -galactosidase *G. antarctica* has a very high RMSD plot showing that it is not very stable at room temperature because it is not able to grow above 20 °C and also this is a predicted model so there is a probability that the molecule in the structure is not stable. Furthermore,  $\alpha$ -galactosidase *T. reesei* is stable at 298 K since it is in crystal structure while  $\alpha$ -galactosidase *R. emersonii* is slightly high at starting point of simulation but becomes more stable along 20 ns. This could be because it is a predicted model which might not be stable at the start of a simulation (Figure 5a). Differences that stabilize the temperature of all  $\alpha$ -galactosidase structures may be because of the living environment whereby *G. antarctica*, *T. reesei*, and *R. emersonii* lives at low, temperate, and high temperature, respectively. This observation suggests that there is a correlation between protein stabilizing and the

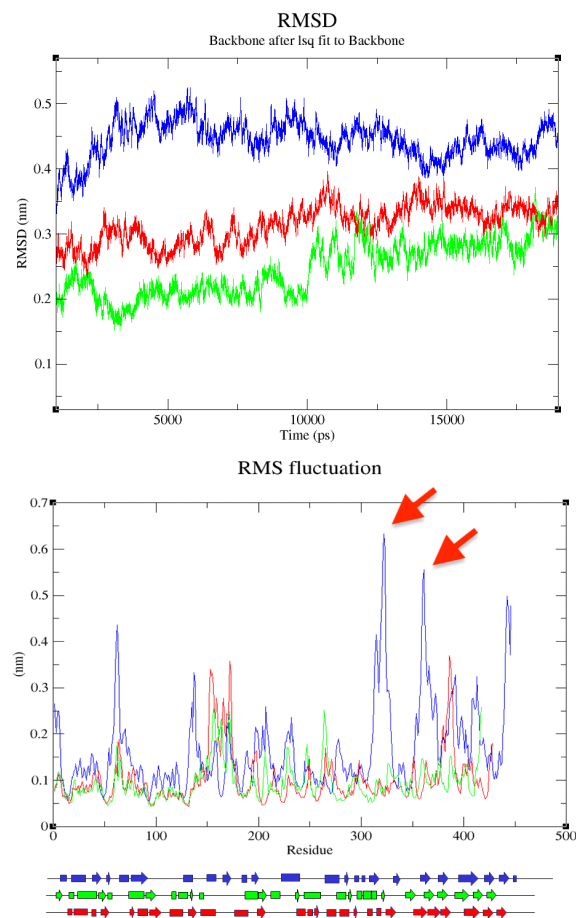


Figure 5. Molecular dynamic analyses on  $\alpha$ -galactosidase of *G. antarctica*, *T. reesei*, and *R. emersonii* using GROMACS: a) stability analysis based on RMSD analysis at a certain temperature and b) flexibility analysis based on RMSF analysis for  $\alpha$ -galactosidase *G. antarctica* (blue) more flexible with high peak (red arrow) compared to *T. reesei* (green) and *R. emersonii* (red) according to secondary structure (arrow shape as helices and rectangular as beta sheet).

temperature of the study microorganism. To evaluate and compare flexibility in the structure  $\alpha$ -galactosidase *G. antarctica*, *T. reesei* and *R. emersonii*, the RMSF analysis of C $\alpha$  atom of  $\alpha$ -galactosidase was plotted to study the average fluctuation of each residue during the simulation. This was analyzed at the final 5 ns because the system becomes stable after 15 ns simulation. RMSF plot shows that a higher peak at a certain region signifies a higher flexible region. The result from the RMSF analysis showed that *G. antarctica* is more flexible compared to *T. reesei* and *R. emersonii*. This is because most of the residue in *G. antarctica* has a higher rate of fluctuation compared to *T. reesei* and *R. emersonii* across different temperatures as shown in multiple peaks in Figure 5b. The B-factor in *G. antarctica* simulation is 69% higher compared to *T. reesei* and *R. emersonii* at only 17% and 15%,

respectively. A high number of B-factors also indicates higher structural flexibility. All of the results would suggest that  $\alpha$ -galactosidase *G. antarctica* has higher flexibility compared to *T. reesei* and *R. emersonii*. The result may prove that *G. antarctica* has a flexible enzyme structure in adaptation to extremely cold environments.

## Conclusion

Comparative analysis of the structure of psychrophilic, *G. antarctica*, against thermophilic, *R. emersonii*, and mesophilic, *T. reesei* enzyme showed that the conformational structure of psychrophilic enzymes is more flexible compared to thermophilic and mesophilic enzymes. There are factors involved in enzyme flexibility to increase small amino acids mainly Glycine and decrease collection of charged and aromatic amino acids. Besides that, the enzyme structure of *G. antarctica*  $\alpha$ -galactosidase was also found to have weak intramolecular interactions including the reduction of hydrogen bonding force and disulfide bond that contributes to the flexibility of its structure. Besides that, the result from molecular dynamics also showed that the enzyme of psychrophilic of *G. antarctica* shows high flexibility with many peaks compared to the thermophilic enzyme of *R. emersonii* and the mesophilic enzyme of *T. reesei*.

## Competing Interests

The authors declare that they have no competing interests.

## Author's Contributions

SMS, AMAM, and FDAB: Conceptualization; Data curation; Formal analysis; Investigation; Software; SMS and RM: Methodology; Validation; Visualization; SMS and NMM: Funding acquisition; NMM: Project administration; Resources; SMS and RM: Writing - Original Draft; AMAM, NMM, and FDAB: Supervision; Writing - Review & Editing.

## Acknowledgement

We would like to thank Dr. Muhammad Alif bin Mohammad Latif from Universiti Putra Malaysia for his guidance and Dr. Nooraisyah binti Mohammad Nor for improving the quality of the manuscript.

## Funding

This study was supported by grant No. 330830003 from the National Science Fellowship, Ministry Science, Technology, and Innovation. Furthermore, we would like to thank the Malaysia Genome Institute for access to facilities to complete this research.

## Supplementary Material

Supplementary material is not available for this article.

## References

- Adapa, V., Ramya, L. N., & Pulicherla, K. K. (2022). Cold-active enzymes: Enabling nonthermal processing in food industry. In *Microbial Extremozymes* (pp. 39–53). Elsevier.
- Álvarez-Cao, M.-E., Cerdán, M.-E., González-Siso, M.-I., & Becerra, M. (2019). Optimization of *Saccharomyces cerevisiae*  $\alpha$ -galactosidase production and application in the degradation of raffinose family oligosaccharides. *Microbial Cell Factories*, 18(1), 1–17.
- Bajaj, S., & Singh, D. K. (2015). Biodegradation of persistent organic pollutants in soil, water and pristine sites by cold-adapted microorganisms: Mini review. *International Biodeterioration and Biodegradation*, 100, 98–105. <https://doi.org/10.1016/j.ibiod.2015.02.023>
- Belcourt, L. A., & Labuza, T. P. (2007). Effect of raffinose on sucrose recrystallization and textural changes in soft cookies. *Journal of Food Science*, 72(1), C065–C071. <https://doi.org/10.1111/j.1750-3841.2006.00218.x>
- Bentahir, M., Feller, G., Aittaleb, M., Lamotte-Brasseur, J., Himri, T., Chessa, J. P., & Gerday, C. (2000). Structural, kinetic, and calorimetric characterization of the cold-active phosphoglycerate kinase from the antarctic *Pseudomonas* sp. TACII18. *Journal of Biological Chemistry*, 275(15), 11147–11153. <https://doi.org/10.1074/jbc.275.15.11147>
- Bhatia, S., Singh, A., Batra, N., & Singh, J. (2020). Microbial production and biotechnological applications of  $\alpha$ -galactosidase. *International Journal of Biological Macromolecules*, 150, 1294–1313. <https://doi.org/10.1016/j.ijbiomac.2019.10.140>
- Clarke, J. H., Davidson, K., Rixon, J. E., Halstead, J. R., Fransen, M. P., Gilbert, H. J., & Hazlewood, G. P. (2000). A comparison of enzyme-aided bleaching of softwood paper pulp using combinations of xylanase, mannanase and  $\alpha$ -galactosidase. *Applied Microbiology and Biotechnology*, 53(6), 661–667.
- Collins, T., Meuwis, M. A., Gerday, C., & Feller, G. (2003). Activity, stability and flexibility in glycosidases adapted to extreme thermal environments. *Journal of Molecular Biology*, 328(2), 419–428. [https://doi.org/10.1016/S0022-2836\(03\)00287-0](https://doi.org/10.1016/S0022-2836(03)00287-0)
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, 2(9), 1511–1519. <https://doi.org/10.1002/pro.5560020916>
- Davail, S., Feller, G., Narinx, E., & Gerday, C. (1994). Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile *Bacillus* TA41. *Journal of Biological Chemistry*, 269(26), 17448–17453.
- Dwyer, B., Hu, J., Madduri, A., Lundberg, D., Miller, B., Gill, J., Meiyappan, M., Pan, C., Miller, T., & Zhang, B. (2021). Affinity purification of human alpha galactosidase utilizing a novel small molecule biomimetic of alpha-D-galactose.



- Protein Expression and Purification*, 177, 105752. <https://doi.org/10.1016/j.pep.2020.105752>
- Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). [20] VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, 277, 396–404.
- Fabris, E., Bulfoni, M., Nencioni, A., & Nencioni, E. (2021). Intra-laboratory validation of alpha-galactosidase activity measurement in dietary supplements. *Molecules*, 26(6), 1–14. <https://doi.org/10.3390/molecules26061566>
- Fei, Y., Jiao, W., Wang, Y., Liang, J., Liu, G., & Li, L. (2020). Cloning and expression of a novel  $\alpha$ -galactosidase from *Lactobacillus amylolyticus* L6 with hydrolytic and transgalactosyl properties. *Plos One*, 15(7), e0235687.
- Firdaus-Raih, M., Hashim, N. H. F., Bharudin, I., Abu Bakar, M. F., Huang, K. K., Alias, H., Lee, B. K. B., Mat Isa, M. N., Mat-Sharani, S., Sulaiman, S., Tay, L. J., Zolkefli, R., Muhammad Noor, Y., Law, D. S. N., Abdul Rahman, S. H., Md-Illias, R., Abu Bakar, F. D., Najimudin, N., Abdul Murad, A. M., & Mahadi, N. M. (2018). The *Glaciozyma antarctica* genome reveals an array of systems that provide sustained responses towards temperature variations in a persistently cold habitat. *PLoS ONE*, 13(1). <https://doi.org/10.1371/journal.pone.0189947>
- Fujimoto, Z., Kaneko, S., Momma, M., Kobayashi, H., & Mizuno, H. (2003). Crystal structure of rice  $\alpha$ -galactosidase complexed with D-galactose. *Journal of Biological Chemistry*, 278(22), 20313–20318. <https://doi.org/10.1074/jbc.M302292200>
- 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.
- Georgette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V. N., Gerday, C., & Feller, G. (2003). Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *Journal of Biological Chemistry*, 278(39), 37015–37023. <https://doi.org/10.1074/jbc.M305142200>
- Germain, D. P., Hughes, D. A., Nicholls, K., Bichet, D. G., Giugliani, R., Wilcox, W. R., Feliciani, C., Shankar, S. P., Ezgu, F., & Amartino, H. (2016). Treatment of Fabry's disease with the pharmacologic chaperone migalastat. *New England Journal of Medicine*, 375(6), 545–555.
- Ghazi, S., Rooke, J. A., & Galbraith, H. (2003). Improvement of the nutritive value of soybean meal by protease and  $\alpha$ -galactosidase treatment in broiler cockerels and broiler chicks. *British Poultry Science*, 44(3), 410–418. <https://doi.org/10.1080/00071660310001598283>
- Golubev, A. M., Nagem, R. A. P., Brandão Neto, J. R., Neustroev, K. N., Eneyskaya, E. V., Kulminskaya, A. A., Shabalin, K. A., Savel'ev, A. N., & Polikarpov, I. (2004). Crystal structure of  $\alpha$ -galactosidase from *Trichoderma reesei* and its complex with galactose: Implications for catalytic mechanism. *Journal of Molecular Biology*, 339(2), 413–422. <https://doi.org/10.1016/j.jmb.2004.03.062>
- Heinig, M., & Frishman, D. (2004). STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Research*, 32(WEB SERVER ISS.), 500–502. <https://doi.org/10.1093/nar/gkh429>
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuik, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36(Web Server issue), 5–9. <https://doi.org/10.1093/nar/gkn201>
- 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. <https://doi.org/10.1107/s0021889892009944>
- Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, 8(4), 477–486.
- Liljestrom-Suominen, P. L., Joutsjoki, V., & Korhola, M. (1988). Construction of a Stable  $\alpha$ -Galactosidase-Producing Baker's Yeast Strain. *Applied and Environmental Microbiology*, 54(1), 245–249. <https://doi.org/10.1128/aem.54.1.245-249.1988>
- Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, 356(6364), 83–85.
- Maiti, R., Van Domselaar, G. H., Zhang, H., & Wishart, D. S. (2004). SuperPose: A simple server for sophisticated structural superposition. *Nucleic Acids Research*, 32(WEB SERVER ISS.), 590–594. <https://doi.org/10.1093/nar/gkh477>
- Metpally, R. P. R., & Reddy, B. V. B. (2009). Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. *BMC Genomics*, 10, 1–10. <https://doi.org/10.1186/1471-2164-10-11>
- Modrego, A., Amaranto, M., Godino, A., Mendoza, R., Barra, J. L., & Corchero, J. L. (2021). Human  $\alpha$ -galactosidase a mutants: Priceless tools to develop novel therapies for fabry disease. *International Journal of Molecular Sciences*, 22(12). <https://doi.org/10.3390/ijms22126518>
- Mohammadi, S., Parvizpour, S., Razmara, J., Abu Bakar, F. D., Illias, R. M., Mahadi, N. M., & Murad, A. M. A. (2018). Structure Prediction of a Novel Exo- $\beta$ -1,3-Glucanase: Insights into the Cold Adaptation of Psychrophilic Yeast *Glaciozyma antarctica* PII2. *Interdisciplinary Sciences: Computational Life Sciences*, 10(1), 157–168. <https://doi.org/10.1007/s12539-016-0180-9>
- Mulder, N., & Apweiler, R. (2007). Interpro and interproscan. In *Comparative genomics* (pp. 59–70). Springer.
- Parvizpour, S., Hussin, N., Shamsir, M. S., & Razmara, J. (2021). Psychrophilic enzymes: structural adaptation, pharmaceutical and industrial applications. *Applied Microbiology and Biotechnology*, 105(3), 899–907. <https://doi.org/10.1007/s00253-020-11074-0>
- Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785–786.
- Petersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Russell, R. J. M., Gerike, U., Danson, M. J., Hough, D. W., & Taylor, G. L. (1998). Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, 6(3), 351–361. [https://doi.org/10.1016/s0969-2126\(98\)00037-9](https://doi.org/10.1016/s0969-2126(98)00037-9)

- Sali, A. (2008). MODELLER A Program for Protein Structure Modeling Release 9v4, r6262. *Structure, September*.
- Sheridan, P. P., Panasik, N., Coombs, J. M., & Brenchley, J. E. (2000). Approaches for deciphering the structural basis of low temperature enzyme activity. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1543(2), 417–433. [https://doi.org/10.1016/S0167-4838\(00\)00237-5](https://doi.org/10.1016/S0167-4838(00)00237-5)
- Thorvaldsen, S., Hjerde, E., Fenton, C., & Willassen, N. P. (2007). Molecular characterization of cold adaptation based on ortholog protein sequences from *Vibrionaceae* species. *Extremophiles*, 11(5), 719–732.
- Tronelli, D., Maugini, E., Bossa, F., & Pascarella, S. (2007). Structural adaptation to low temperatures - Analysis of the subunit interface of oligomeric psychrophilic enzymes. *FEBS Journal*, 274(17), 4595–4608. <https://doi.org/10.1111/j.1742-4658.2007.05988.x>
- Tsaniklidis, G., Benovias, A., Delis, C., & Aivalakis, G. (2016). Acidic alpha galactosidase during the maturation and cold storage of cherry tomatoes. *Acta Physiologiae Plantarum*, 38(2), 57.
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(SUPPL.2), 407–410. <https://doi.org/10.1093/nar/gkm290>
- Xie, J., Wang, B., He, Z., & Pan, L. (2020). A thermophilic fungal GH36  $\alpha$ -galactosidase from *Lichtheimia ramosa* and its synergistic hydrolysis of locust bean gum. *Carbohydrate Research*, 491(January), 107911. <https://doi.org/10.1016/j.carres.2020.107911>
- Xu, H., Qin, Y., Huang, Z., & Liu, Z. (2014). Characterization and site-directed mutagenesis of an  $\alpha$ -galactosidase from the deep-sea bacterium *Bacillus megaterium*. *Enzyme and Microbial Technology*, 56, 46–52. <https://doi.org/10.1016/j.enzmictec.2014.01.004>
- Yusof, N. A., Hashim, N. H. F., & Bharudin, I. (2021). Cold adaptation strategies and the potential of psychrophilic enzymes from the antarctic yeast, *Glaciozyma antarctica* PI12. *Journal of Fungi*, 7(7). <https://doi.org/10.3390/jof7070528>