IN SILICO ANALYSIS AND 3D STRUCTURE PREDICTION OF A CHITINASE FROM PSYCHROPHILIC YEAST *Glaciozyma antarctica* PI12

YUSOF, N.Y.[1,2], FIRDAUS-RAIH, M.[1], MAHADI, N.M.[3], ILLIAS, R.M.[4], ABU BAKAR, F.D.[1] and MURAD, A.M.A.[1*]

[1]*School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
[2]*Institute for Research in Molecular Medicine, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia*
[3]*Malaysia Genome Institute, Jalan Bangi Lama, 43000 Kajang, Selangor, Malaysia*
[4]*Department of Bioprocess Engineering, Faculty of Chemical Engineering, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia*
[*]*E-mail: munir@ukm.edu.my*

ABSTRACT

Chitinases are a group of glycosyl hydrolases that are essential to recycle chitin presence in nature. The aim of this work is to characterise the sequence of a chitinase isolated from a psychrophilic yeast, *Glaciozyma antarctica* PI12 and to predict and analyse the 3-dimensional protein structure. The cDNA for the *G. antarctica* chitinase gene, *GaCHT43*, with a length of 1,176 bp was reverse transcribed from mRNA, cloned and sequenced. The gene encodes a mature protein of 391 amino acids with an expected molecular weight of 43 kDa. Sequence analysis showed that GaCh43 has high similarity to the endochitinase family 18 proteins of other fungi. A three-dimensional (3D) model of GaCht43 was built by homology modelling with *Aspergillus fumigatus* chitinase (1W9P) as the template. Validation analysis via PROCHECK, VERIFY3D and ERRAT showed that the GaCht43 model surpassed the quality requirements and was accepted for further analysis. GaCht43 contained chitinase conserved regions, SxGG and DxxDxDxE that are required in the catalytic mechanism. Analysis of the GaCht43 structure showed the presence of extra loop regions compared to mesophilic chitinases, which might contribute to the flexibility of the protein.

Key words: Chitinase, *Glaciozyma antarctica*, protein structure, molecular modelling

INTRODUCTION

Chitin, the second most abundant organic compound in nature after cellulose, is a linear homopolysaccharide of β-N-acetylglucosamine (GlcNac). It occurs mainly in the exoskeleton of crustaceans and arthropods as well as the major component of fungal cell wall (Gohel *et al*., 2005). The degradation of chitin is performed by chitinases, members of glycoside hydrolases that hydrolyse the β-1, 4-glycosidic linkages between the GlcNAc residues of chitin polymer to produce GlcNac monomers. Chitin hydrolysis is performed by three groups of chitinases (Li, 2006). Endochitinases cleave randomly within the chitin molecules to produce GlcNac chains with low molecular weight, such as chitotetraose and chitotriose. Exochitinases catalyse the progressive release of a dimer, chitobiose, starting at the non-reducing end of the chitin molecule. Chitobiases hydrolyse chitobiose to monomers of GlcNac. These enzymes are widely distributed in nature and occur in various organisms including fungi, bacteria, plants, insects and animals. Chitinases are grouped into family 18 or family 19 glycosyl hydrolases (GH) based on their amino acid sequence (Davis & Henrissat, 1995). Chitinases of family 18 are mainly found in fungi, bacteria and animals, as well as class III or IV plant chitinases. Family 19, on the other hand, includes class I, II, or IV chitinases from plants and chitinases present in some *Actinobacteria* and *Streptomyces* (Kawase *et al*., 2004).

---

* To whom correspondence should be addressed.

Glaciozyma antarctica is an obligate psychrophilic yeast that was isolated from sea ice, collected near Casey Research Station, Antarctica (Hashim *et al*., 2013). The genome of this yeast has been sequenced, and a genome mining effort to identify *G. antarctica* glycoside hydrolase has revealed the presence of seven genes encoding chitinases (Nooraisyah *et al*., 2015). These chitinases could work in tandem to degrade chitin waste that might be present in the sea ice as a source of carbon for *G. antarctica*. These enzymes are expected to survive and function at cold temperatures and are expected to show distinct structural characteristics from their mesophilic counterparts. Hence, understanding the structure of these cold-active chitinases would illuminate how these enzymes adapt to cold temperatures, which in turn could be used in several biotechnological applications. Chitin biodegradation by cold-active chitinases could be utilised for various applications including the biocontrol of plant pathogens in cold environments, bioconversion of chitin-rich waste at low temperatures and the biocontrol of microbial spoilage of refrigerated food.

In this work, we analysed the gene and amino acid sequences of GaCht43, one of the chitinases produced by *G. antarctica*. To understand the structural properties that allow the enzyme to survive at low temperature, a 3-dimensional (3D) structure of the protein was predicted via a homology modelling approach, analysed and compared to mesophilic chitinase.

MATERIALS AND METHODS

GaCHT43 cDNA amplification

Glaciozyma antarctica was a kind gift from Professor Dr. Nazalan Najimudin of Universiti Sains Malaysia, Penang, Malaysia (Hashim *et al*., 2013). The full-length sequence for the chitinase (reference number in *G. antarctica* database: LAN_10_172) was obtained from the *G. antarctica* genome database, available at http://27.126.156.144/glaciozyma_antarctica/index2.php. RNA extraction was carried out using TRIzol® Reagent (Invitrogen, USA) as described by Bharuddin *et al* (2014). The cDNA encoding *GaCHT43* was isolated and amplified using reverse transcription-PCR (RT-PCR). The forward and reverse primers used to amplify the cDNA were 5'-TCC CCA CCC AAG GTT GAG AT-3' and 5'- GAA ACT ACC CAA AAA CGT GTC CA-3', respectively. PCR was performed using the following cycles: 95ºC for 2 min; 30 cycles of denaturation at 95ºC for 1 min, annealing

at 56.7ºC for 1 min, extension at 72ºC for 1 min; final extension at 72ºC for 10 min. Subsequently, the PCR product was purified from the gel using a QIAquick PCR purification kit (Qiagen, Germany), cloned into the pGEM-T Easy vector (Promega, USA) and transformed into *Escherichia coli* DH5α. The nucleotide was sequenced using the Sanger Platform at the Malaysian Genome Institute for sequence validation. The plasmid extraction was carried out using the Wizard® plus SV MiniPreps DNA Purification System (Promega, USA).

GaCht43 sequence analysis

The GaCht43 amino acid sequence (391 residues) was analysed using BLAST (Altschul *et al*., 1990) to determine sequence similarity. Subsequently, Interproscan (Gough *et al*., 2001), PFam HMM (Finn *et al*., 2005) and NCBI CDART (Conserved Domain Architecture Retrieval Tool) (Geer *et al*., 2002) analyses were performed to determine the conserved domains in the protein. Amino acid residues were also analysed using the Protparam (Gasteiger *et al*., 2005) to elucidate various physico-chemical properties of the protein and SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP) (Nordahl *et al*., 2011) to determine the presence of the signal peptide. Multiple sequence alignment was carried out using MUSCLE (Edgar, 2004).

Modeling of GaCht43 tertiary structure

The search for an appropriate template and the prediction for the secondary structure of GaCht43 were performed using the programs BLAST-PDB and HHPRED (Soding *et al*., 2006). The 3D building of the GaCht43 model was performed using MODELLER9v11 (Eswar *et al*., 2007). One-hundred models were produced and the model with the lowest objective function values was chosen for subsequent analysis. The models energy minimisation was carried out using the steepest descent algorithm available in GROMOS from Deepview (Guex & Peitsch, 1997).

Evaluation of GaCht43 structure model

To evaluate the quality of the modelled structure, GaCht43 model was analysed based on the geometric quality of the backbone conformation, the residue interaction and contacts and the energy profile of the structure using three different methods: PROCHECK (Lovell *et al*., 2003), ERRAT (Tomii *et al*., 2005) and VERIFY 3D (Bowie *et al*., 1991). Superimposition between GaCht43 model and its template and comparative analysis were done using CHIMERA USCF (Pettersen *et al*., 2004).
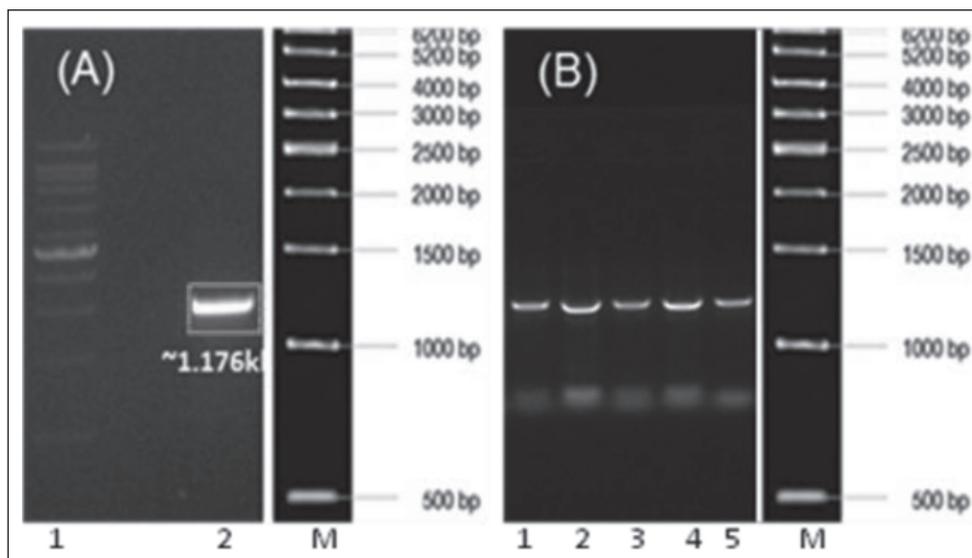
Fig. 1. Gel electrophoresis analysis of *GaCHT43* PCR products.
(A) *GaCHT43* cDNA after reverse transcription from total RNA. Lane 1: 1 kb DNA marker (Vivantis, USA); lane 2: amplified cDNA of *GaCHT43*. (B) *GaCHT43* PCR products amplified from cloning vector for sequence verification. Lane 1 – 5: amplification from cloning vectors isolated from different positive transformants; M: 1 kb DNA marker (Vivantis, USA).

RESULTS AND DISCUSSION

Cloning and GaCht43 sequence analysis

The *G. antarctica GaCHT43* nucleotide sequence was retrieved from the *G. antarctica* PI12 genome database. Amplification of the cDNA via reverse transcription PCR produced a 1,176 bp PCR amplicon (Figure 1A). The PCR products (Figure 1B) were cloned into the cloning vector pGEM-T®Easy, and all five clones were sequenced. Only clones 1, 3 and 5 were perfectly sequenced, whereas the others showed evidence of mutation.

GaCht43 consists of 391 amino acids and the calculated molecular weight of the protein 43 kDa. It theoretical pI is 5.8, as analysed via ProtParam. SignalP prediction showed that the GaCht43 protein lacks the N-terminal putative signal peptide, suggesting that it is probably localised intra-cellularly. Interproscan, Pfam HMM and NCBI CDART analysis revealed that GaCht43 is a member of the glycoside hydrolase family 18 (GH18) superfamily. BLAST search analysis revealed that the GaCht43 amino acid has 55% identity to endochitinase from *Neofusicoccum parvum* UCRNP2.

GaCht43 sequence alignment and domain analysis

The catalytic domain multiple sequence alignment was performed using the MUSCLE program (Figure 2). The catalytic domain of five eukaryotic chitinases that show high sequence identity to the chitinase of *G. antarctica* were selected and aligned with GaCht43 catalytic domain. This catalytic domain, which was highly conserved among the family 18 chitinases, contained two conserved amino acid regions, SxGG and DxxDxDxE. These regions may constitute the catalytic pocket of the enzyme (Terwisscha van Scheltinga *et al*., 1996). Importantly, residues such as Asp135, Asp138, Asp140 and Glu142 that are essential for chitinase activity were identified in the GaCht43 catalytic domain. These residues have been reported to be important for the catalytic activity and to support the structure of the enzyme (Liu *et al*., 2008).

3D model development

The HHPRED and BLAST-PDB results showed that GaCht43 has 51% and 50% structure similarity, respectively, to the solved structure of *Aspergillus fumigatus* chitinase (1W9P). Hence, the structure of GaCht43 was predicted by homology modelling based on 1W9P using MODELLERv9.11. MODELLER is a program that attempts to satisfy spatial restraints, by which a set of geometrical criteria is used to create a probability density function of the location of each atom in the protein. Models with high violations of the restraints are considered poor (Guex & Peitsch, 1997). The model with the lowest objective function was selected as the best 3D model and subjected to energy minimisation using GROMOS to prevent poor molecular contacts (Ramli *et al*., 2012).
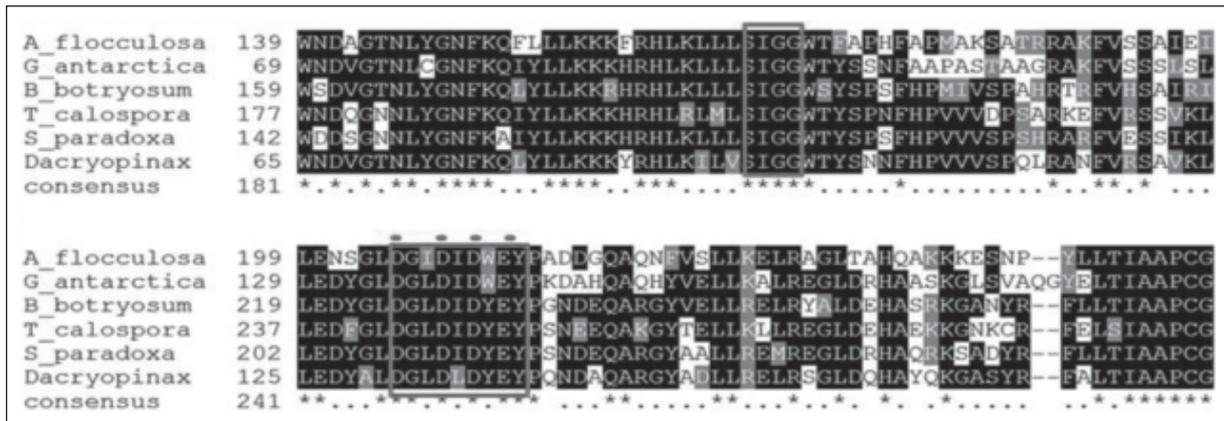
Fig. 2. Multiple sequence alignment of the chitinase catalytic domains of different fungi. Analysis was performed using MUSCLE. Highly conserved amino acid residues, SxGG and DxxDxDxE, are in red boxed. Amino acid residues important for catalytic activity are marked with blue dots. The amino acid sequences used in this analysis are the chitinase catalytic domain of GaCht43, *Anthracocystis flocculosa* PF-1, *Botryobasidium botryosum FD-172, Tulasnella calospora* MUT 4182, *Schizopora paradoxa* and *Dacryopinax* sp.
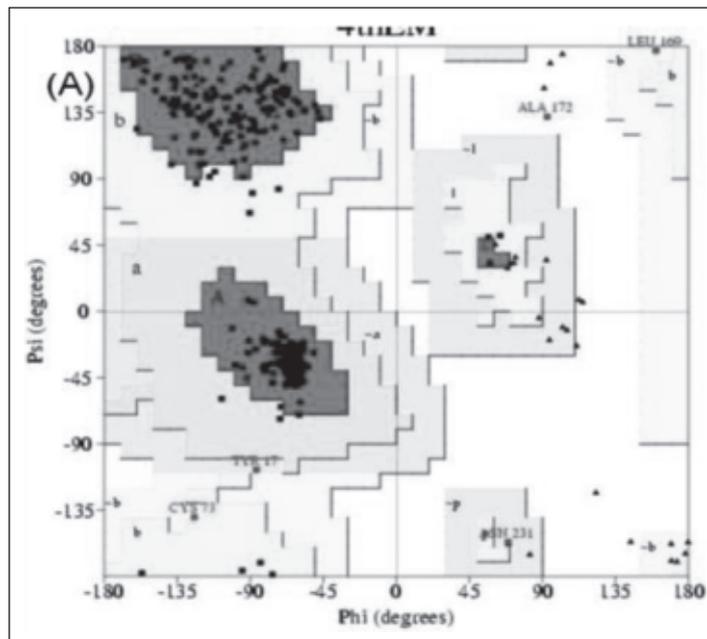
Fig. 3. Evaluation of the GaCht43 model using a Ramachandran plot. A total of 93.5% of the residues was located in the most favoured regions, and only Ala172 was located in the disallowed region.

Model validation and assessment

PROCHECK, VERIFY3D and ERRAT were used to evaluate the quality of the model. The PROCHECK program was utilised to calculate the residue-by-residue stereochemical quality of GaCht43 to construct the Ramachandran plot. Ramachandran plot is a way to visualize energetically allowed regions for backbone dihedral angles ψ (psi) against φ (phi) of amino acid residues in protein structure (Ramachandran *et al.*, 1963). The plot of the GaCht43 model illustrated that only Ala172 was located in a disallowed region; 93.5%

of the residues were positioned in the most favoured regions and the rest of the residues located in the additional and generously allowed regions (Figure 3). A plot score of more than 90% indicates that the predicted model is of satisfactory quality.

The VERIFY 3D analysis showed that 97.42% of the residues had an average 3D-1D score greater than 0.2. The analysis shows that only 2.58% of the residues did not complement to the 3D-1D profile. The VERIFY 3D program validates the structure by a statistical approach to establish the compatibility of an atomic model (3D) with its own amino acid

sequence (1D) by assigning a structural class based on location and environment and comparing the results to better structures (Luthy *et al.*, 1992). A VERIFY 3D score greater than 80% indicated a high quality model.

The verification of the predicted GaCht43 structure using ERRAT resulted in a score of 77.51%. ERRAT was designed to detect local errors within the geometry in a protein structure that could indicate the overall quality factor for non-bonded atomic interactions and is calculated by comparison with statistics from highly refined structures (Ramli *et al.*, 2012). The high ERRAT score indicated good quality; the basic requirement for a high quality model is more than 50% (Chaitanya *et al.*, 2010).

Overall, based on the evaluation program scores, it was concluded that the predicted GaCht43 structure was reasonably good.

GaCht43 structure analysis

The model protein, which belongs to the class 18 family of chitinases, has two signature sequences (XXGG and DXXDXDXE), corresponding to residues 94-97 and 131-138 (Figure 4A), which are shown as balls and sticks. In other reported chitinases, these residues lie along the barrel strands and help to form the active site cleft at the carboxyl end of the β-barrel (Hurtado-Guerrero & Van Aalten, 2007). The catalytic domain of GaCht43 has a conserved $(\beta/\alpha)_8$ fold. This fold is known as a TIM barrel and includes an alternating pattern of α-helices and β-strands within a single domain. The TIM barrel is commonly found among various glycosyl hydrolases of family 18 for which crystal structures have been solved (Hurtado-Guerrero & Van Aalten, 2007). The β-strands that form the core of the enzyme are labelled S1-S15, and the strands

are each followed by α-helices labelled H1-H15 (Figure 4A).

The superimposition of the GaCht43 model with its template, 1W9P, yielded an RMSD of 0.094 Å (Figure 4B) covering almost 100% of the backbone atoms. This observation indicates a good overall structure alignment. The proposed catalytic residues for the modelled GaCht43 structures (Asp131-Glu137) (represented as sticks) are found at the same locations as in 1W9P (Figure 4C), which indicates that the modelled chitinase structure is an acceptable representation of the actual protein's structure. GaCht43 was found to possess other surface loops that are absent in the mesophilic 1W9P structure. Figure 5B shows three locations in GaCht43 that contain more surface loops: section I (residues Ser148-Asp150), section II (residues Asn267-Met345) and section III (residues Asp385-Val409). These loops may be responsible for the flexibility of GaCht43, hence leading to the cold adaptation of this enzyme. A similar study by Alveraz *et al* (1998) revealed that triosephosphate isomerase of the psychrophilic bacterium *Vibrio marinus* has more and longer loops connecting the α-helices and β-sheets compared to its homologues from the mesophilic organisms. Longer surface loops were found to enhance protein flexibility and decrease their stability (Siddiqui & Cavicchioli, 2006).

CONCLUSION

The cDNA of *GaCht43* from *G. antarctica* PI12 was amplified and sequenced. Sequence and domain analysis of the protein strongly suggest that it is an endochitinase from the G18 family. Analysis of the
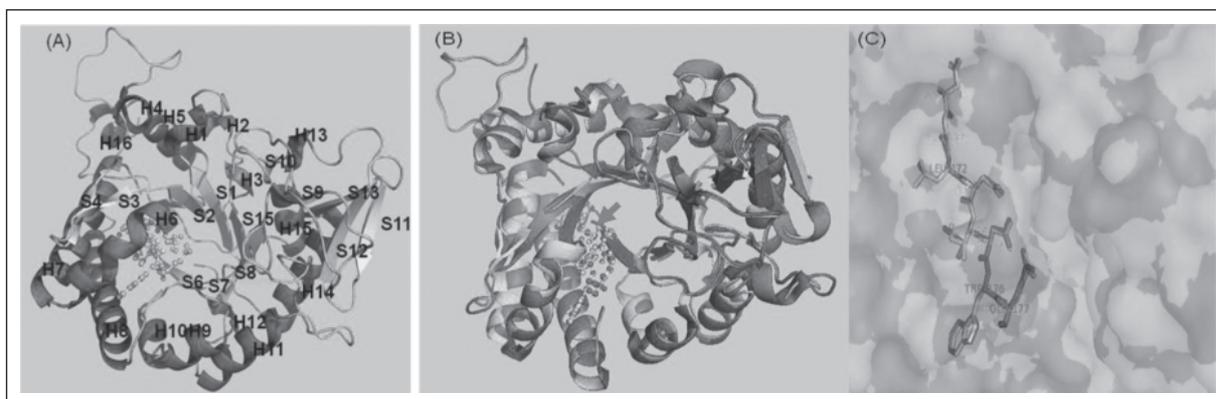
Fig. 4. (A) A 3D structure of GaCht43 with the β-strands S1 to S15 and the α-helices H1 to H16 labelled. Two signature sequences (XXGG and DXXDXDXE), corresponding to residues 94-97 and 131-138 of GaCht43, are represented as balls and sticks. (B) Superposition of the refined GaCht43 model (cyan ribbon) and its template, 1W9P (magenta ribbon). Both GaCht43 and 1W9P structures are presented in a cartoon representation. The catalytic domains of both protein models are depicted as balls and sticks (red arrow). (C) Closer view of GaCht43 catalytic residues (Asp131-Glu138) compared to 1W9P (Asp170-Glu177).
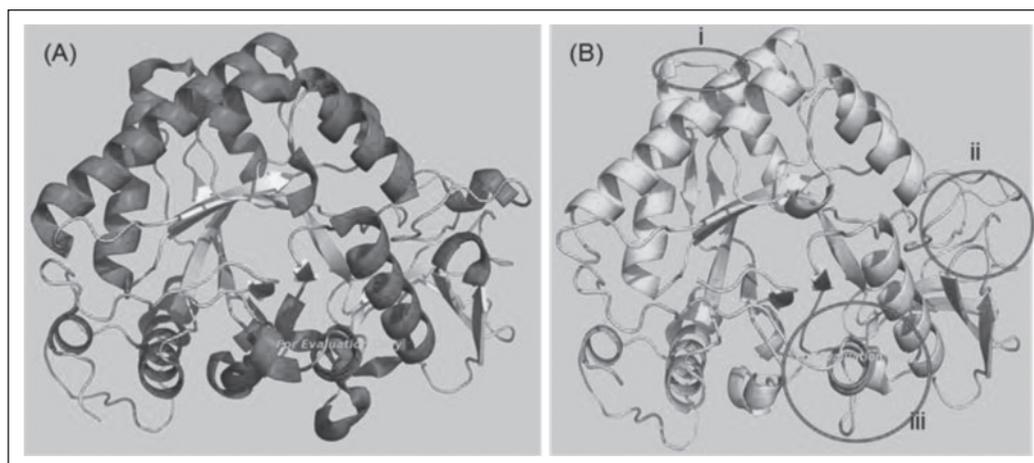
Fig. 5 (A) 3D structure model of 1W9P used as a template. GaCht43 was found to possess extra surface loops that are absent in the mesophilic 1W9P structure. (B) Three locations in the GaCht43 model that contain more surface loops: section I (residues Ser148-Asp150), section II (residues Asn267-Met345) and section III (residues Asp385-Val409).

gene sequence and 3D model of GaCht43 indicate two conserved signature regions that are important for the chitinase catalytic mechanism. The 3D model also revealed the presence of additional loops in three different regions of the GaCht43 model compared to its template, which might be responsible for the flexibility of the structure, contributing to its adaptation to low temperature.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic alignment search tool. *Journal of Molecular Biology,* **215**: 403-10.

Alveraz, M., Zeelen, J.P., Mainfroid, V., Rentier-Delrue, F., Martial, J.A., Wyns, L., Wierenga, R.K. & Maes, D. 1998. Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. *Journal of Biological Chemistry*, **273**: 2199-2206.

Bharudin, I., Zaki, N.Z., Bakar, F.D.A., Mahadi, N.M., Najimudin, N., Illias, R.M. & Murad, A.M.A. 2014. Comparison of RNA extraction methods for transcript analysis from psychrophilic yeast, *Glaciozyma antarctica*. *Malaysian Applied Biology*, **43**: 71-79.

Bowie, U., Luthy, R. & Eisenberg, D. 1991. A method to identify protein sequences that fold into a known three dimensional structure. *Science*, **253**: 164-170.

Chaitanya, M., Babajan, B., Anuradha, C.M., Naveen, M., Rajasekhar, C., Madhusudana, P. & Kumar, C.S. 2010. Exploring the molecular basis of selective binding of *Mycobacterium tuberculosis* Asp kinase toward its natural substrates and feedback inhibitors: A docking and molecular dynamics study. *Journal of Modeling*, **16**: 1357-1367.

Davis, G. & Henrissat, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**: 853-859.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32(5)**: 1792-1797.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U. & Sali, A. 2007. Comparative protein structure modelling using MODELLER. *Current Protocol Protein Science*, **2**: 1-30.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer E.L. & Bateman, A. 2005. Pfam clans, web tools and services. *Nucleic Acids Research*, **34**: D247-D251.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, A.R., Appel, R.D. & Bairoch, A. 2005. Protein identification and analysis tools on ExPASY server. In: Walker, J.M. (ed.) *The Proteomics Protocols Handbook*, Humana Press, Totowa, New Jersey, pp. 571-607.

Geer, L., Domrachev, M., Lipman, D.J. & Bryant, S.H. (2002). CDART: Protein homology by domain architecture. *Genome Research*, **12(10)**: 1619-1623.

Gohel, V., Vyas, P. & Chatpar, H.S. 2005. Activity staining method for chitinase on chitin agar plate through polyacrylamide gel electrophoresis. *African Journal of Biotechnology*, **4**: 87-90.

Gough, J., Karplus, K., Hughey, R. & Chothia, C. 2001. Assignment of homology to genome sequence using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology,* **313**: 903-919.

Guex, N. & Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis,* **18**: 2714-2723.

Hashim, N., Bharudin, I., Nguong, D., Higa, S., Bakar, F., Nathan, S., Rabu, A., Kawahara, H., Illias, R., Najimudin, N., Mahadi, N. & Murad, A.M.A. 2013. Characterization of Afp1, an antifreeze protein from the psychrophilic yeast *Glaciozyma antarctica* PI12. *Extremophiles,* **17**: 63-73.

Hurtado-Guerrero, R. & Van Aalten, D.M.F. 2007. Structure of *Saccharomyces cerevisiae* chitinase 1 and screening-based discovery of potent inhibitors. *Chemistry and Biology*, **14**: 589-599.

Kawase, T., Saito, A., Sato, T., Kanai, R., Fujii, T., Nikaidou, N., Miyashita, K. & Watanabe, T. 2004. Distribution and phylogenetic analysis of family 19 chitinases in *Actinobacteria*. *Applied and Environmental Microbiology,* **70**: 1135-1144.

Li, D.C. 2006. Review of fungal chitinases. *Mycopathologia,* **161**: 345-360.

Liu, Z.H., Yang, Q., Hu, S., Zhang, J.D. & Ma, J. 2008. Cloning and characterization of a novel chitinase gene (chi46) from *Chaetomium globosum* and identification of its biological activity. *Applied Microbiology and Biotechnology*, **80**: 241-252.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S. & Richardson, D.C. 2003. Structure validation by cα geometry: Φ, ψ and cβ. *Proteins: Structure, Function and Boinformatics,* **50**: 437-450.

Luthy, R., Bowie, J.U. & Eisenberg, D. 1992. Assessment of protein models with three-dimensional profiles. *Nature*, **356**: 83-85.

Nooraisyah, M.N., Siti Nur Hasanah, M.Y., Mahadi, N.M., Abu Bakar, F.D. & Murad, A.M.A. 2015. Genome mining for glycoside hydrolases from the psychrophilic yeast *Glaciozyma antarctica* PI12. *Malaysian Applied Biology,* **44**: 167-172.

Nordahl, T., Brunak, S., von Heijne, G. & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**: 785-786.

Pettersen, 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.

Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan V. 1963. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology,* **7**: 95-99.

Ramli, A.N.M., Mahadi, N.M., Shamsir, M.S., Rabu, A., Joyce-Tan, K.H., Murad, A.M.A. & Illias, R.M. 2012. Structural prediction of a novel chitinase from the psychrophilic *Galciozyma antarctica* PI12 and an analysis of its structural properties and function. *Journal of Computer-Aided Molecular Design*, **26**: 947-961.

Siddiqui, K.S. & Cavicchioli, R. 2006. Cold adapted enzymes. *Annual Review of Biochemistry*, **75**: 403-433.

Soding, J, Remmert, M. & Biegert, A. 2006. HHrep: de *novo* protein repeat detection and the origin of TIM barrels. *Nucleic Acids Research,* **34**: 137-142.

Terwisscha van Scheltinga, A.C., Hennig, M. & Dijkstra, B.W. 1996. The 1.8Å resolution structure of Hevamine, a plant chitinase/lysome, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. *Journal of Molecular Biology,* **262**: 243-257.

Tomii, K., Hirokawa, T. & Motono, C. 2005. Protein structure prediction using a variety of profile libraries and 3D verification. *Protein: Structure, Function, and Bioinformatics*, **61**: 114-121.