

ICSET-2014 [6th – 7th May 2014]
International Conference on Science, Engineering and Technology

Bioinformatics analyses of a Thermophilic Fungal Recombinant Chitinase

Muthu Prasad and Peramachi Palanivelu*

Department of Molecular Microbiology, School of Biotechnology,
Madurai Kamaraj University, Madurai – 625 021, India

*Corres.author: ppmkupp@gmail.com

Abstract : A thermophilic fungal chitinase was cloned, sequenced and overexpressed in *Saccharomyces cerevisiae*. Multiple sequence alignment of the recombinant chitinase revealed that the catalytic domain is 100% conserved with other family GH18 chitinases. Phylogenetic analysis revealed that the recombinant chitinase exhibited a very close proximity to the chitinases of other thermophilic fungi and higher fungi. Glycosylation of the recombinant chitinase was analyzed by *in silico* methods and confirmed by experimental data. Homology modeling of the chitinase showed the typical $(\alpha/\beta)_8$ TIM barrel structure. Molecular docking of the enzyme revealed the involvement of a glutamic acid of the catalytic domain in the catalytic process.

Keywords: Chitin, fungal recombinant chitinase, GH18 family, glycosylation, homology modeling, molecular docking.

Introduction

Chitin is the most abundant, natural, cationic biopolymer present in the world. It is a homopolymer of β , 1-4 linked *N*-acetyl D-glucosamine residues. The chitinases hydrolyze the chitin polymers which are broadly classified into exo- and endo-chitinases. The endochitinases are further classified into two different glycosyl hydrolase families, viz., GH18 and GH19, based on their catalytic domain, catalytic mechanism, three dimensional structures, and sensitivity to inhibitors.

Chitinases are ubiquitous enzymes and perform several functions in different species. They hydrolyze the chitin and contribute to the recycling and effective utilization of the most abundant marine waste for various applications¹. The GH18 family chitinases contains $(\alpha/\beta)_8$ TIM barrel structures composing of α -helices and β strands. The chitinases have two highly conserved motifs viz., *SxGG* and *DxxDxDxE*, corresponding to substrate binding and catalytic domains, respectively². The GH18 family chitinases are predominantly characterized for degradation of chitins to produce industrially and pharmaceutically important, chito-oligosaccharides and *N*-acetyl D-glucosamine (NAG).

Chitinases are produced by fungi during all phases of their growth and these enzymes also contribute to morphogenetic and pathogenic processes such as spore germination, hyphal growth and branching and mycoparasitism³. In this study, the chitinase gene was isolated from the thermophilic fungus, *Thermomyces lanuginosus* RM-B, cloned and overexpressed in *S. cerevisiae* and characterized further^{4,5}. This thermophilic fungal chitinase is identified as GH18 family type. Most reliable method on the prediction utilized bioinformatics analysis, based on the sequence and three dimensional structure of the chitinase. Experimental data were used to validate the predictions, using web servers. This communication describes, sequence

alignment, prediction of conserved motifs and glycosylation sites, evolutionary relationship and secondary structure analysis, homology modeling and docking of the recombinant chitinase of *T. lanuginosus* RM-B with NAG.

Materials and Methods

Sequence alignment and phylogenetic analysis of fungal chitinases

Fungal chitinase protein sequences were retrieved from NCBI and used for sequence alignments and phylogenetic analysis. Pairwise sequence alignment (PSA) of chitinase sequences of *T. lanuginosus* RM-B with *T. lanuginosus* SY2 was carried out using EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The recombinant chitinase of *T. lanuginosus* RM-B with other fungal chitinases were aligned using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), an advanced Multiple Sequence Alignment (MSA) program. It uses seeded guide trees and Hidden Markov Model (HMM) profile-profile techniques to generate alignments. The phylogenetic tree was constructed using Phylogeny (<http://www.phylogeny.fr/>) by maximum likelihood method implemented in the PhyML program (v3.0 aLRT) and the graphical representation and phylogenetic tree were created using TreeDyn (v198.3).

Prediction of glycosylation sites in chitinase

Purified recombinant chitinase of *T. lanuginosus* RM-B was found to be a glycoprotein using PAS staining method (data not shown). In order to validate this experimental data, the possible glycosylation sites in the chitinase sequence were analyzed using bioinformatics tools. The tools used for predictions were www.cbs.dtu.dk/services/NetNGlyc/ for *N*-linked glycosylation and www.cbs.dtu.dk/services/NetOGlyc/ for *O*-linked glycosylation. PSIPRED protein sequence analysis workbench (<http://bioinf.cs.ucl.ac.uk/psipred/>) was also used to predict glycosylation sites in the recombinant chitinase sequence.

Homology modeling and validation of chitinase structure

Phyre² (Protein Homology/analogy Recognition Engine) V 2.0 and FUGUE, Sequence-structure homology were used for recognizing distant homologues by sequence-structure comparison. FUGUE scans a database of structural profiles, calculates the sequence-structure compatibility scores and produces a list of potential homologues and alignments to the chitinase. Homology model was generated for the recombinant chitinase of *T. lanuginosus* RM-B using an automated protein structure homology-modeling server, SWISS-MODEL beta (<http://beta.swissmodel.expasy.org/>). Homology model of the chitinase three dimensional structure was validated using Protein Structure Validation Suite (PSVS) server. Ramachandran plot was used to validate the three dimensional structure, by calculating allowed and disallowed regions. *T. lanuginosus* RM-B recombinant chitinase was superimposed with the template using UCSF Chimera and RMS deviations between these protein structures were calculated.

Prediction of secondary structure of chitinase

Secondary structure of the recombinant chitinase was calculated by SOPMA (Self-Optimized Prediction Method with Alignment: http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). It is a protein secondary structure prediction tool in ExPASy, which helps to predict the secondary structure of a protein from the amino acid sequences. Sequence Annotated by Structure (SAS) was used to measure the α -helices and β -strands of the chitinase, based on the homology model. SAS (<https://www.ebi.ac.uk/thornton-srv/databases/sas>) is a tool for applying structural information to a given protein sequence against all the proteins of known 3D structure in protein data bank (PDB).

The Define Secondary Structure of Proteins (DSSP) database was also used for prediction of secondary structures of the chitinase. It provides secondary structure assignments for all of the entries in the PDB. DSSP analysis tool summarizes secondary structure information such as α -helix, 3- and 5-helices, β -strands, β -bridges, hydrogen bonded turns, bends and other non-structural information of the chitinase. This program works by calculating the most likely secondary structure assignment for the given 3D structure of the chitinase by reading the position of the atoms followed by calculation of the H-bond energy between all the atoms. The best two H-bonds for each atom are used to determine the most likely class of secondary structure for each residue in the protein.

The PDBsum (<https://www.ebi.ac.uk/pdbsum/>) is a pictorial database which provides an overview of the contents of the 3D structure of the chitinase. It shows the protein chains, ligand interactions and contact of metal ions and schematic diagrams of their interactions. Topology, secondary structure plot, folds matching, surface cleft analysis, functional 3D templates and possible homologues in the PDB of the chitinase were

obtained using ProFunc (<https://www.ebi.ac.uk/thornton-srv/databases/ProFunc>). ProFunc is a server of PDBsum page helps to analyze and predict protein structure and function from the 3D structure⁶.

Docking of chitinase with NAG

Docking of the recombinant chitinase of *T. lanuginosus* RMB with the ligand, NAG (monomer unit of the chitin polymer), was carried out using GOLD (Genetic Optimization for Ligand Docking, version 5.2.1) docking software⁷. Energy was minimized for both the protein and the ligand using In Sight II tool. The structure of chitinase and the ligand were checked for errors and the protonation states were assigned at biological pH. Hydrogens added protein system was relaxed by keeping the rest of the system static, using 2000 steps of conjugate gradient method.

The binding pocket was defined using Glu126 residue within 10Å distance. GOLD was run comprising 100000 genetic operations on an initial population of 100 members divided into five subpopulations (number of islands 5). The annealing parameters of fitness function were set at 4.0 for van der Waals' forces and 2.5 for hydrogen bonding. A niche size of 2 and a selection pressure of 1.1 were used. The early termination option was turned off and top ten best solutions were analyzed.

Results and Discussion

Sequence alignment and analysis of chitinase sequences

CAZy (Carbohydrate-Active enZYme) database contains all the information of enzymes related to carbohydrates, which involves in degradation, modification or generation of glycosidic bonds. It also comprises genomic, structural and biochemical features of carbohydrate active enzymes. CAZy database grouped all the chitinases under GH18 and GH19 family types⁸. Chitinases of GH18 family is an ancient gene family which occurs in a broad range of organisms such as archaea, prokaryotes and eukaryotes⁹. In this study, a GH18 family chitinase was studied from a thermophilic fungus and the characteristics of that chitinase is explained. Protein sequences of fungal chitinases were retrieved from NCBI protein database and these sequences were aligned by BioEdit multiple sequence alignment programs (Fig. 1). MSA output helped to find identity, similarity and conserved sequences of the fungal chitinases. Sequence similarity and identity based search, identified the conserved domains of the fungal chitinases. Catalytic domain of GH18 family chitinases, $DxxDxDxE^2$ was found to be highly conserved in all fungal chitinases used in this study. Substrate binding domain (SBD), $SxGG$ was ascertained to be conserved in ~70% of the fungal chitinases used in this analysis and in the remaining sequences, most of them were replaced with amino acids of same property i.e., serine was replaced by alanine (23%) or with the hydrophobic residue, methionine (7%) and isoleucine of the SBD was replaced with leucine in chitinases of *Ajellomyces dermatitidis* (gi|531982779), *Aspergillus niger* (gi|145250221), *Colletotrichum gramicola* (gi|310801258) and *Penicillium digitatum* (gi|425775892).

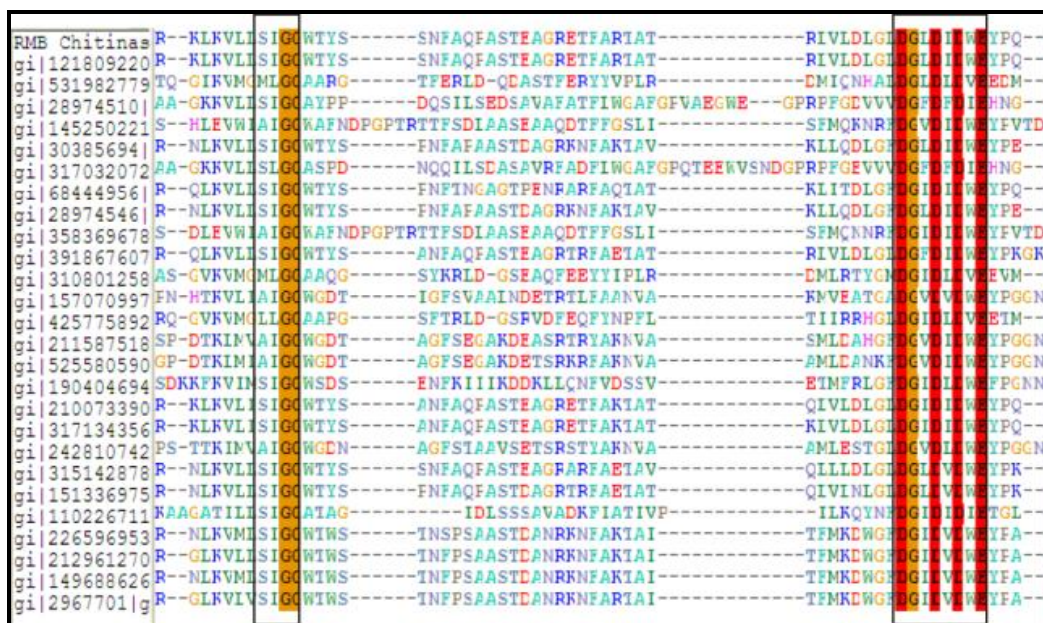


Fig. 1) MSA of the fungal chitinases

Fungal chitinase sequences were used to find out the evolutionary relationship of the *T. lanuginosus* RM-B chitinase sequence with other chitinases from fungal sources. Phylogenetic analysis of fungal chitinases

revealed that *T. lanuginosus* RM-B chitinase is closely related to the chitinases of other thermophilic fungi such as *T. lanuginosus* SY2, *Thermoascus aurantiacus* var. *aurantiacus* and *T. aurantiacus* var. *levisporus*. Among the different fungal chitinases analyzed, chitinases of higher fungi (especially Ascomycetes) found to have in close agreement with *T. lanuginosus* RM-B chitinase (Fig. 2).

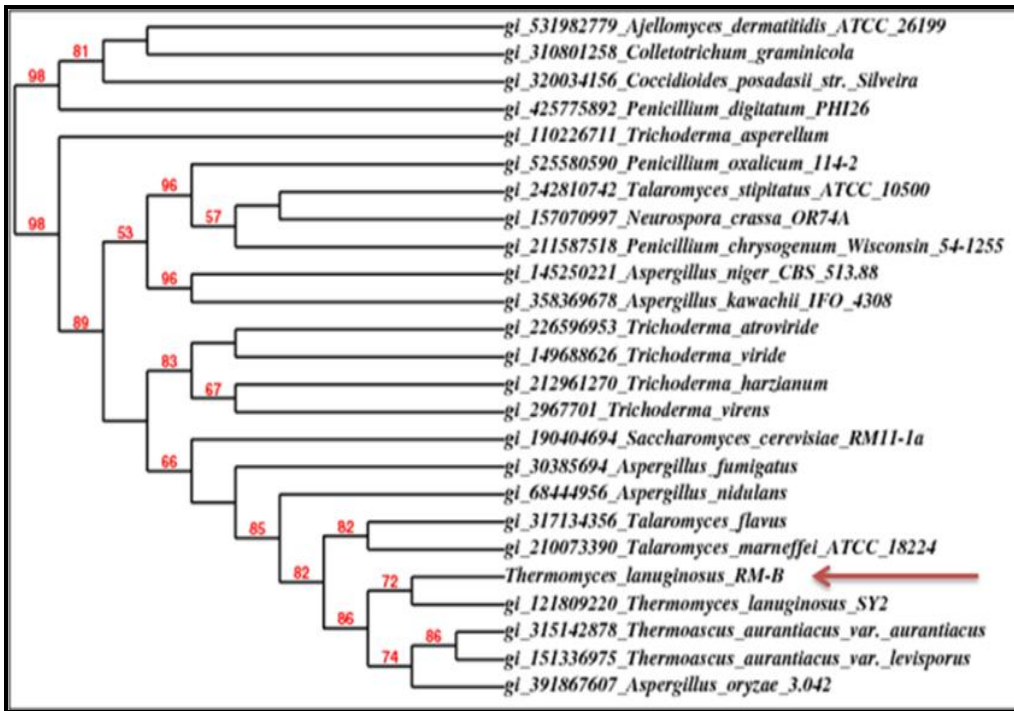


Fig. 2) Evolutionary relationship between *T. lanuginosus* RM-B and other fungal chitinases

Similarity regions of the recombinant chitinase of *T. lanuginosus* RM-B with chitinase of *T. lanuginosus* SY2 were evaluated by PSA using EMBOSS Needle program (Fig. 3). Conserved residues of catalytic domains of GH18 family chitinases found to be conserved in both *T. lanuginosus* RM-B and *T. lanuginosus* SY2 chitinases and have an identity of 99% with 87% query coverage. In fact, the overexpressed recombinant chitinase of *T. lanuginosus* RM-B is truncated at N-terminal region of the protein and show variations only in five residues (F/L, A/P, E/K, A/G and N/Y) when compared to the chitinase of *T. lanuginosus* SY2; but these variations did not affect the function of the enzyme. Replacement of such amino acid residues favors for hydrophobicity which may contribute to the thermostability of the enzyme (data not shown).

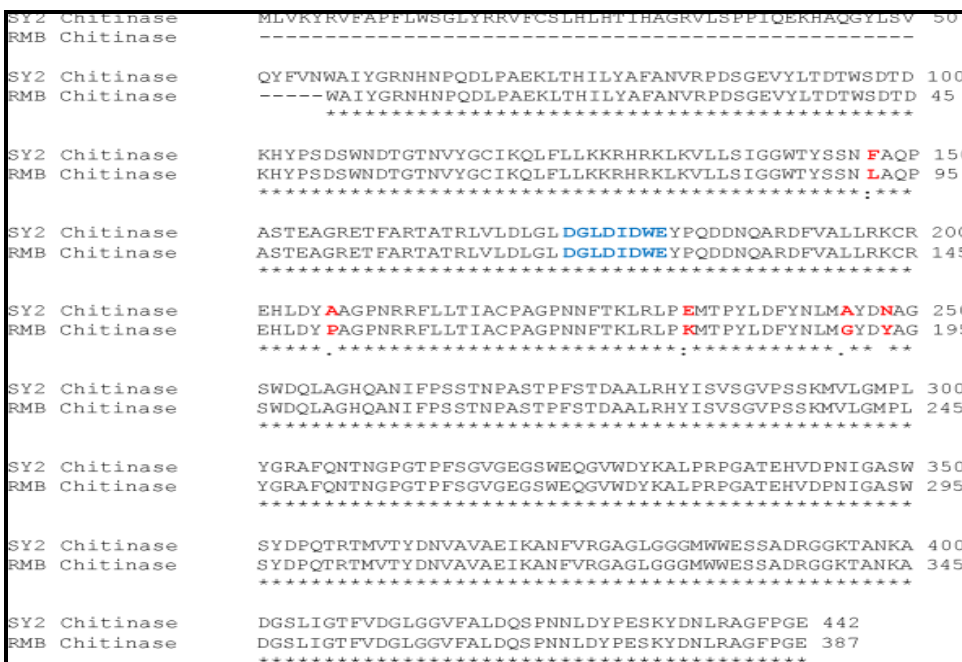


Fig. 3) PSA of the chitinases of *T. lanuginosus* RM-B with *T. lanuginosus* SY2

Prediction of glycosylation sites in the recombinant chitinase

Two most common glycosylation sites that occur in proteins are *N*-linked and *O*-linked glycosylations. *N*-linked glycosylation has a motif called as sequon, NxT/S and *O*-linked glycosylation has a motif of S/T¹⁰. The recombinant chitinase found to have two sequons at positions 54 (NDTG) and 170 (NFTK) for *N*-linked glycosylation and one *O*-linked glycosylation site at position 221 (STDA). PSIPRED protein sequence analysis workbench also predicted the same glycosylation sites in the chitinase sequence. Position dependent feature prediction mapped onto the sequence is shown in Fig. 4. The line height of the glycosylation features reflects the confidence of the residue prediction.

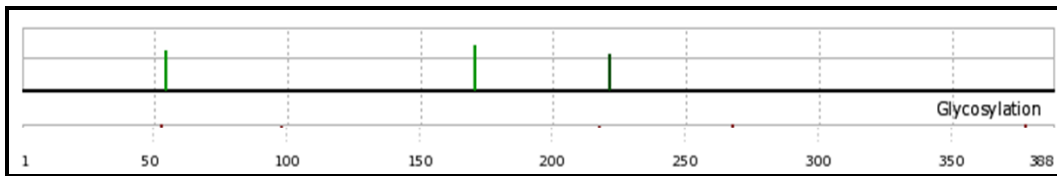


Fig. 4) Prediction of the glycosylation sites in the protein sequence of the recombinant chitinase

Homology modeling, validation and analysis of the 3D structure of chitinase

Homology modeling was initiated with FUGUE fold recognition using PSI-BLAST to collect sequence homologues and Phyre² was also used for further corroborating the homology prediction¹¹. Both the tools were suggested the same template, i.e., crystal structure of chitinase (2.1Å) of *Aspergillus fumigatus* (PDB ID: 1WNO) with 100% confidence. The *T. lanuginosus* RM-B chitinase showed 63.83% identity with the chitinase of *A. fumigatus*. Homology model for *T. lanuginosus* RM-B chitinase was generated using beta-SWISS-MODEL¹² based on the template suggested by FUGUE and Phyre². Chitinases of *Vibrio carchariae*¹³ and *Munduca sexta*¹⁴ were modeled using the same. Sequence alignment of the model with the template is illustrated in Fig. 5.

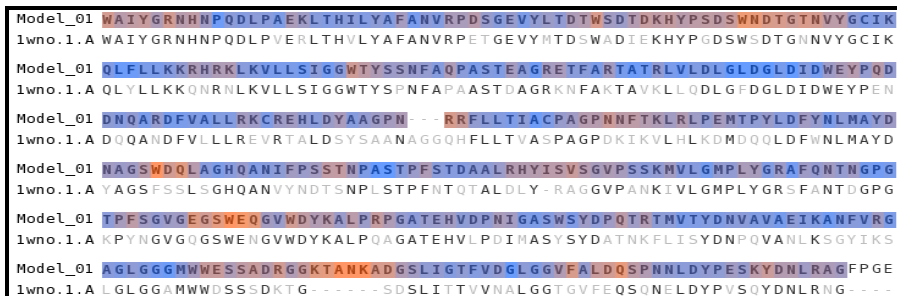


Fig. 5) Model-Template alignment (Matched sequences are indicated in bold)

Homology based model was viewed and analyzed using PyMOL molecular visualization tool and the chitinase 3D model is shown in Fig. 6. Active site and substrate binding residues were identified to be present in the catalytic cleft. The inner core of the chitinase structure was found to be surrounded with β -stands, which plays an important role in enzyme stability. The recombinant chitinase model exhibits a $(\beta/\alpha)_8$ TIM barrel structure which is conserved in all GH18 family chitinases¹⁵.

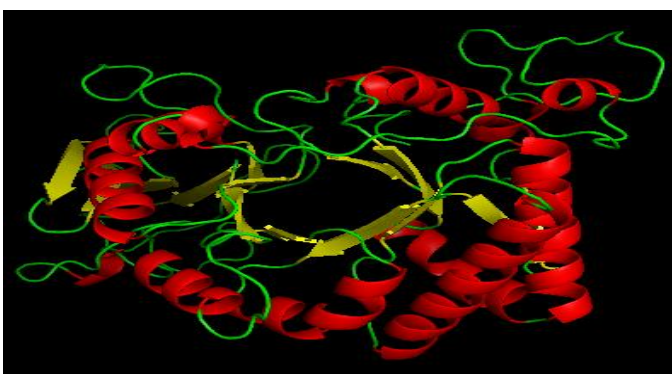


Fig. 6) Homology model of *T. lanuginosus* RM-B chitinase structure

The three dimensional model of the chitinase was validated using PSVS tool and the quality assessment score of the modeled protein and its template are summarized in *Table I*. Ramachandran plot depicted only 4 residues (1.3%) viz., Ser50, Tyr127, Arg156 and Glu332 in the disallowed regions (Fig. 7). As these residues are found in the loop regions, they do not affect the function of the protein.

Table I. Quality assessment scores of homology model and template

| Procheck scores | Template (1WNO) | Chitinase model |
|----------------------------------|-----------------|-----------------|
| Most favored regions (%) | 88.6 | 88.1 |
| Additionally allowed regions (%) | 10.1 | 9.7 |
| Generously allowed regions (%) | 1.0 | 0.9 |
| Disallowed regions (%) | 0.3 | 1.3 |

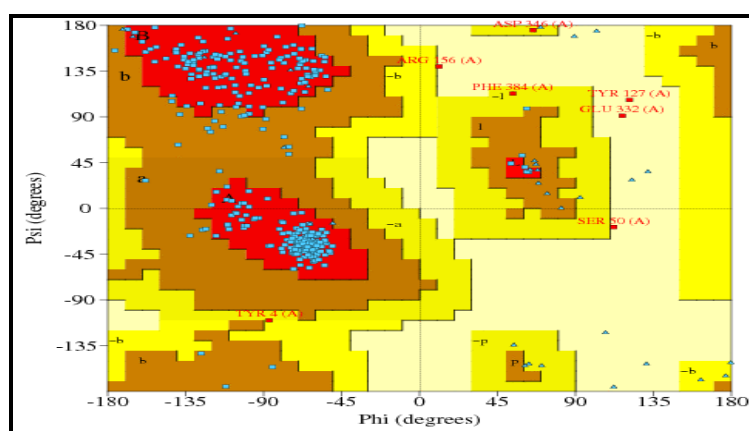


Fig. 7) Ramachandran plot of the *T. lanuginosus* RM-B chitinase

The homology model of the *T. lanuginosus* RM-B chitinase was superimposed with the template, crystal structure of the chitinase of *A. fumigatus* using UCSF Chimera 1.8.1 (Fig. 8). The superimposed structures showed RMSD value of 0.132Å for the complete protein structure of the chitinases.

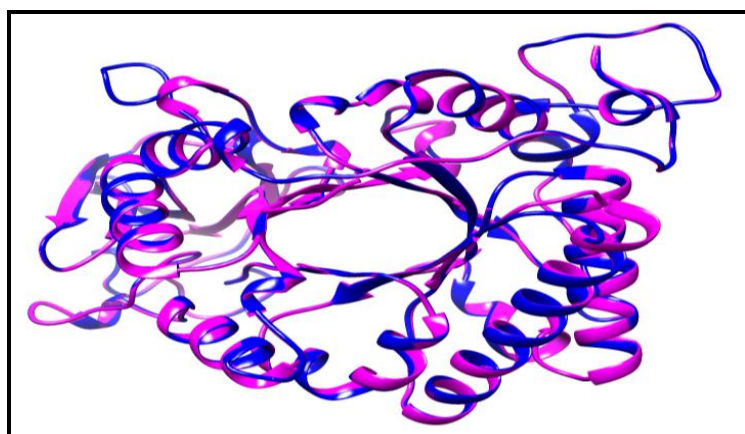


Fig. 8) Superimposition of *T. lanuginosus* RM-B chitinase (blue) with the template (magenta)

Homology model of the chitinase was used to retrieve conserved residues, analyze secondary structures, topology predictions and interaction of protein with ligands by docking. Secondary structure plot of the chitinase model was obtained from ProFunc of PDBsum and illustrated in Fig. 9a and Fig. 9b which show the topology of the chitinase structures. The chitinase model consists of eight α -helices and eight β -strands with four β -turns. Conserved catalytic residues, typical of the GH18 family chitinases were found in the β -strand regions of the recombinant chitinase.

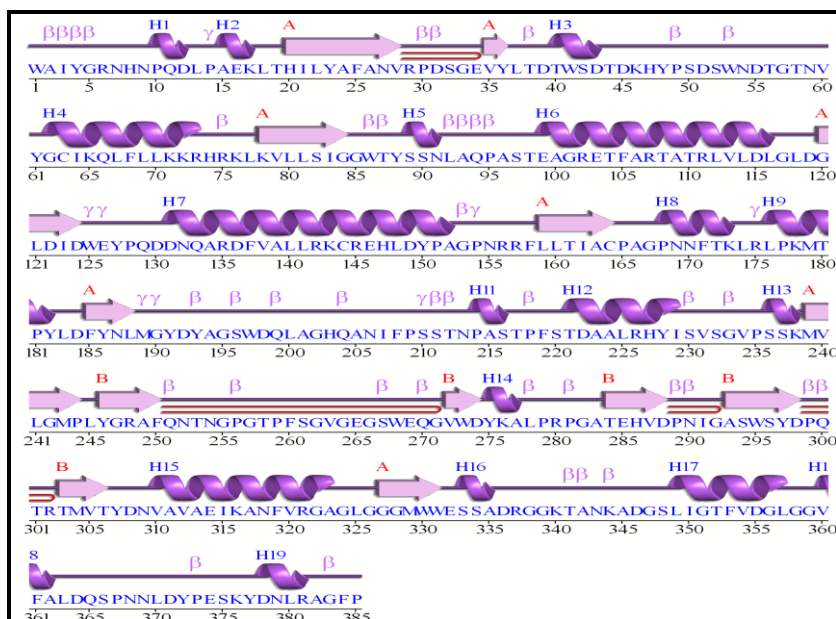


Fig. 9a Secondary structure plot of the 3D structure of *T. lanuginosus* RM-B chitinase

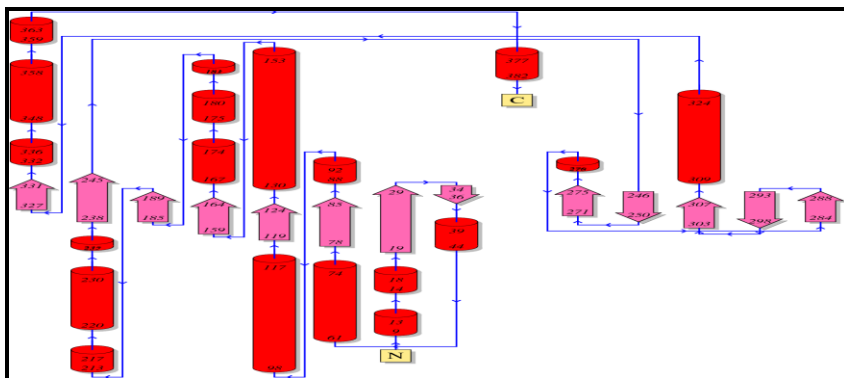


Fig. 9b) Topology of the *T. lanuginosus* RM-B chitinase structure

Analysis of secondary structure of the chitinase

Three different strategies are used to evaluate the secondary structure of the *T. lanuginosus* RM-B chitinase. They are sequence based, structure based and CD spectroscopy based analyses (data not shown) and the results are summarized in Table 2. Secondary structure of the chitinase was analyzed by a wet lab experiment and also predicted using the bioinformatics sequence based SOPMA and structure based DSSP and PDBsum tools. *In silico* prediction of the α -helices and β -strands of the chitinase were found to be closely similar to the experimental data (Table II). Homology model based prediction observed to have similar output from both DSSP and PDBsum.

Table II. Summary of secondary structure analysis of *T. lanuginosus* RM-B chitinase

| Secondary structure | Method of Analysis | | | |
|---------------------|--------------------|-------|------|--------|
| | CD Spectroscopy | SOPMA | DSSP | PDBsum |
| α -helix (%) | 28.04 | 33.07 | 33.5 | 32.7 |
| β -strand (%) | 21.48 | 21.71 | 17.4 | 17.4 |
| Others (%) | 50.48 | 45.22 | 49.1 | 49.9 |

Fig. 10a and 10b are the predicted secondary structures of the chitinase, based on the sequence and the three dimensional structure, respectively. Sequence Annotated by Structure (SAS) is also utilized to predict the secondary structure and functional properties of the chitinase using both the structure and sequence of the chitinase (Fig. 10c).

SAS predicts the catalytic residues, residues contact to ligand and metal, PROSITE patterns for conservation and active sites from PDB records. SAS showed the active site residues of chitinase as Leu118, Asp119, Gly120, Leu121, Asp122, Leu123, Asp124, Trp125 and Glu126. The divalent metal ions such as Mg^{2+} and Zn^{2+} found to have contact with Ser97, Asn132 and Lys46, Gln94, Glu101, Ala194, Asp198, respectively. According to the experimental data, binding of Zn^{2+} with the respective residues of chitinase helps to enhance the activity of enzyme by facilitating feasible interaction of substrate into the catalytic core of the enzyme but Mg^{2+} found to play a negative role.

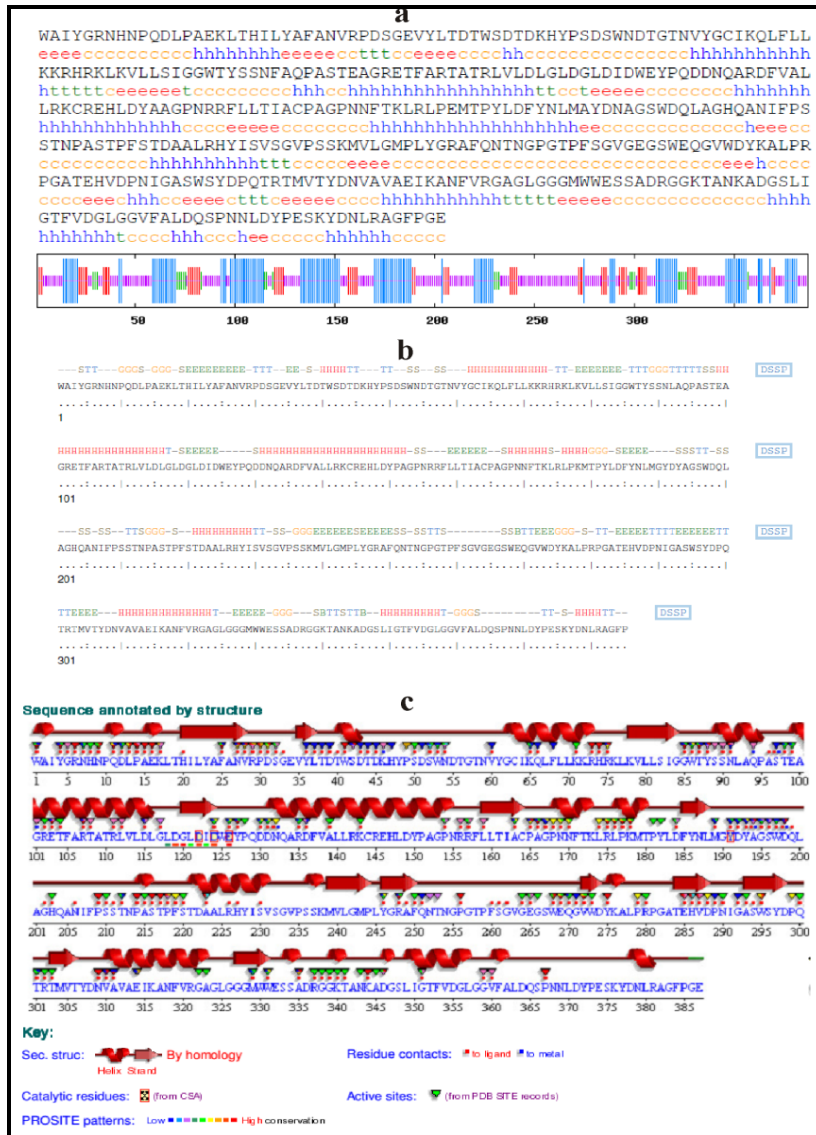


Fig. 10a, b and c) Secondary structure predictions of the chitinase

Docking of chitinase with NAG

GOLD docking software was used to dock the homology model of the chitinase of *T. lanuginosus* RM-B with NAG (substrate). The best dock pose showed CHEMPLP score as 56.40. UCSF Chimera was used to view the interaction of ligand with the residues and the interactions of chitinase with NAG and shown in Fig. 11. Recently homology model of a human chitinase like protein, CHI3L2 was reported¹⁶ and the modeled protein was docked with a chito-oligosaccharide. Catalytic domain residues of the chitinase viz., Asp124 and Glu126 were found to interact by hydrogen bonding with amino group and 1-4 linkages of NAG, respectively. Distance between the oxygen of Glu126 and hydrogen at C-4 of NAG was found to be 2.644 Å. Two dimensional interaction of NAG with chitinase was viewed using Discovery Studio visualizer 3.5 (Fig. 12). The chitinase model found to have hydrogen bonding with Asp124, Glu126, Tyr191 and Asp192. There are eight residues interacting with NAG by electrostatic interactions and five residues interact by van der Waals' forces.

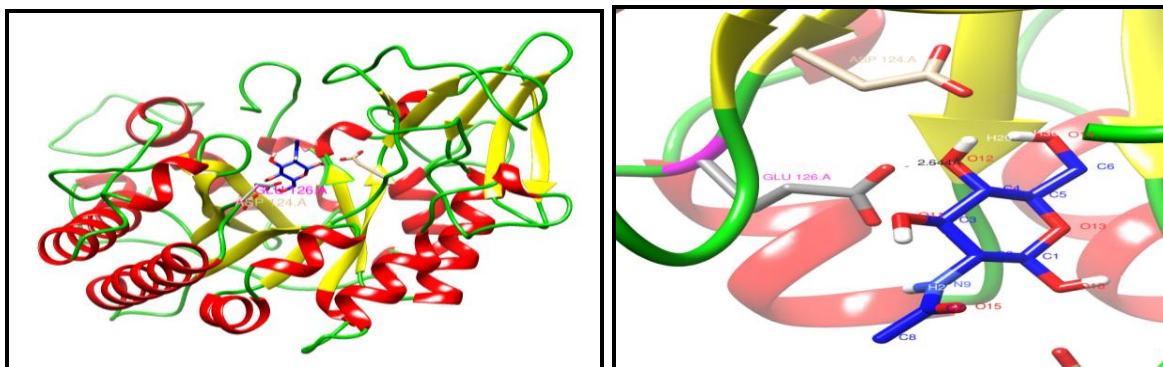


Fig. 11) Docking of chitinase with NAG

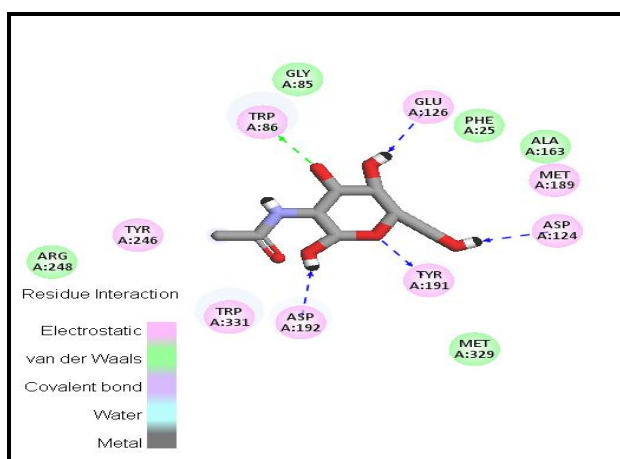


Fig. 12) Two dimensional interaction of chitinase with NAG

Summary

The recombinant chitinase of *T. lanuginosus* RM-B was found to belong to GH18 family. MSA showed that the substrate binding and catalytic domains are highly conserved. Phylogenetic analysis showed that this chitinase has close proximity with chitinases of higher fungi and other thermophilic fungi. The secondary structure of the chitinase was analyzed using secondary structure prediction tools as well as CD spectroscopy. Homology model of the chitinase revealed a typical $(\alpha/\beta)_8$ TIM barrel structure, topological and other properties of the chitinase. Molecular docking of the chitinase with NAG showed that the interactions of the conserved residues of the chitinase with NAG and found that a glutamic acid residue involves in catalytic reaction, which is the most common proton donor in GH18 family chitinases.

Acknowledgements

Authors would like to thank Dr. S. Krishnaswamy, Coordinator, Centre for Excellence in Bioinformatics, School of Biotechnology, Madurai Kamaraj University. Authors also wish to thank the University Grants Commission, New Delhi, for financial support.

References

1. Cohen-Kupiec, R. and Chet, I., The molecular biology of chitin digestion, *Curr. Opin. Biotechnol.*, 1998, 9, 270-277.
2. Henrissat, B., A classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochem. J.*, 1991, 280, 309-316.
3. Seidl, V., Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions, *Fung. Boil. Rev.*, 2008, 22, 36-42.
4. Palanivelu, P. and Vaishna, L., Molecular Cloning and Characterization of a Chitinase Gene from the Thermophilic Fungus, *Thermomyces lanuginosus*-RMB, *IOSR J. Pharm. Biol. Sci.*, 2013, 5(2), 99-107.

5. Prasad, M. and Palanivelu, P., Overexpression of a chitinase gene from the thermophilic fungus, *Thermomyces lanuginosus* in *Saccharomyces cerevisiae* and characterization of the recombinant chitinase, *J. Microb. Biochem. Technol.*, 2012, 4, 86-91.
6. Laskowski, R.A., Watson, J.D. and Thornton, J.M., ProFunc: a server for predicting protein functions from 3D structure, *Nucleic Acids Res.*, 2005, 33 Web Server issue: W89-W93.
7. Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. and Taylor R.D., Improved protein-ligand docking using GOLD, *Proteins*, 2003, 52, 609-623.
8. Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics, *Nucleic Acids Res.*, 2009, 37, Database issue D233-D238.
9. Funkhouser, J.D. and Aronson Jr, N.N., Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC Evol. Biol.*, 2007, 7, 96.
10. Roth, Z., Yehezkel, G. and Khalaila, I., Identification and Quantification of Protein Glycosylation, *Int. J. Carbohydr. Chem.*, 2012, ID 640923.
11. Kelley, L.A. and Strenberg, M.J.E., Protein structure prediction on the web: a case study using Phyre server, *Nature Protocols*, 2009, 4, 363-371.
12. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T., The SWISS-MODEL Workspace: a web-based environment for protein structure homology modeling, *Bioinformatics*, 2006, 22, 195-201.
13. Pantoom, S., Songsiriritthigul, C. and Suginta, W., The effects of the surface-exposed residues on the binding and hydrolytic activities of *Vibrio carchariae* chitinase A, *BMC Biochem.*, 2008, 9:2.
14. Huang, X., Zhang, H., Zen, K., Muthukrishnan, S. and Kramer, K.J., Homology modeling of the insect chitinase catalytic domain–oligosaccharide complex and the role of a putative active site tryptophan in catalysis, *Insect Biochem. Mol. Bio.*, 2000, 30, 107–117.
15. Robertus, J.D. and Monzingo, A.F., The structure and action of chitinases, *EXS*, 1999, 87, 125-135.
16. Iershov, A., Odynets, K., Kornelyuk, A. and Kavsan V., Homology modeling of 3D structure of human chitinase like protein, CHI3L2, *Cent. Eur. J. Biol.*, 2010, 5(4), 407–420.
