IJBPAS, April, 2024, 13(4): 1923-1948

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

IN SILICO ON STUDY CHITINASE AND CHITIN COMPLEX

KAUSHIK R[1], PANDEY J[2], BHARDWAJ N[3], KUMAR S AND RANA S[5]

1: Department of Biotechnology, Sharda University, Greater Noida (Uttar Pradesh), India

2: Department of Bioinformatics, Singhania University (Rajasthan), India

3: Department of Zoology, M. S. College, Saharanpur (Uttar Pradesh), India

4: Department of Chemistry, M. S. College, Saharanpur (Uttar Pradesh), India

5: Department of Microbiology, C. C. S. University, Meerut (Uttar Pradesh), India

***Corresponding Author: Dr. Nikunaj Bhardwaj: E Mail:**

Received 14th May 2023; Revised 15th July 2023; Accepted 15th Aug. 2023; Available online 1st April 2024

https://doi.org/10.31032/IJBPAS/2024/13.4.7964

ABSTRACT

Chitin, the second most abundant natural biopolymer, is composed of repeating units of N-acetylβ-D-glucosamine and primarily forms the structural component of protective biological matrices such as fungal cell walls and exoskeletons of insects. Chitinases are a ubiquitous class of extracellular enzymes that have gained attention in the past few years due to their wide range of biotechnological applications, especially in the field of agriculture for bio-control of fungal phytopathogens. They play an important role in the defence of organisms against chitin-containing parasites by hydrolyzing the β-1,4-linkages in chitin and hence act as anti-fungal as well as antibiofouling agents. Moreover, the effectiveness of conventional insecticides is increasingly compromised by the occurrence of resistance and thus, chitinases offer a potential alternative to the use of chemical fungicides. In recent years, thermostable enzymes isolated from thermophilic microorganisms have gained widespread attention in industrial, medical, environmental and biotechnological applications due to their inherent stability at high temperatures and a wide range of pH optima. Determination of the three-dimensional structure of a protein can provide important details about its biological functions and its mode of action. However, despite their significance, the precise three-dimensional structures of

most of the chitinases, including those isolated from *Thermomyces lanuginosusis* not fully characterized so far. Hence, the main focus of the present study was to gain a better understanding of the structural of chitinases computational techniques, and their relationship with their activity profiles. In silico protein modeling was helpful in predicting the 3D models of the novel chitinase class IV Brassica Juncea, followed by the prediction of its active sites. The presence of different amino acid was found to be essential for the activity of chitinase IV. A Molecular docking were performed between chitinase IV and Chitin.

Keywords: Chitin, chitinases, *Thermomyces lanuginosusis*

INTRODUCTION

Chitinases are a class of enzymes that hydrolyse the naturally occurring polysaccharide chitins, a linear polymer of β (1,4) N-acetyl β-D-glucosamine (Henrissat, 1999). Chitinases are ubiquitous and are present in all forms of life including bacteria, fungi, plants, animals and also in viruses. They form a class of highly conserved enzymes and exhibit diverse functions in these organisms. Chitinases possess an extraordinary ability to hydrolyse the highly insoluble chitin polymer directly to the lower molecular weight chitooligomers, which are widely used in agricultural, biotechnological, industrial and medical fields. In recent years chitinase has been the key focus of research owing to its important biophysiological functions and applications. They play a crucial role in the defence of plants against chitin-containing pathogens and also serves as an effective biocontrol agent against phytopathogens. Fungal chitinases primarily belong to the family of 18 glycosyl hydrolases (Henrissat B., 1999) and displays a high amino acid homology with class III plant chitinases (Hayes *et al*., 1994). The family of 18 fungal chitinases comprises of 5 domains viz., catalytic domain, N-terminal signal peptide region, chitin binding domain, serine/threonine rich region and C terminal extension region. The signal peptide, predicted at the N-terminus is an indicator of the protein secretion that targets the protein outside the cells through secretory pathways. The catalytic domain is the most important domain of this enzyme which is responsible for the chitin substrate hydrolysis. A comparison of the amino acid sequences of chitinases revealed two highly conserved motifs in the family of GH18 chitinases: DXXDXDXE and SXGG, corresponding to the catalytic domains and substrate-binding sites respectively (Henrissat and Bairoch, 1996). The residues Glu (E) and Asp (D) are

highly conserved within the catalytic domains of chitinases, indicating their direct involvement in the hydrolysis of the glycosidic bond.

In this study, the chitinase IV gene structure predictions of were performed using various algorithms of "Homology modelling" and "*ab initio*" methods. The models generated were further using the molecular docking. Active sites in the protein were predicted by bioinformatic tools as well as by comparing the amino acids sequence with other similar proteins. Analysis of the active sites was further confirmed by molecular docking.

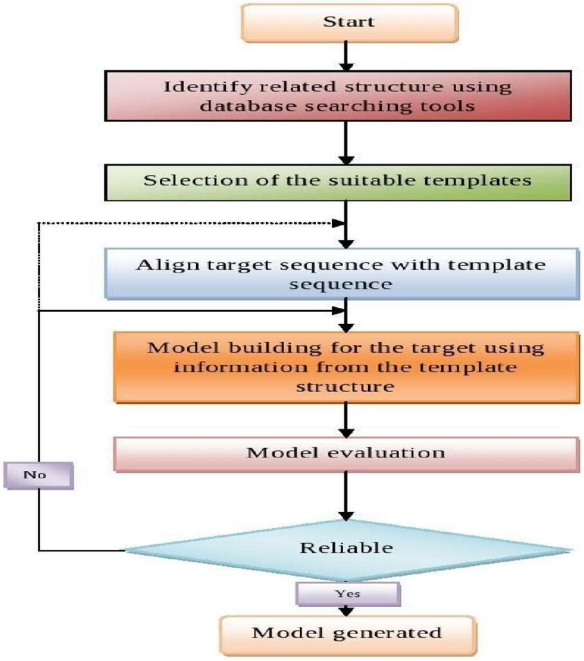

Figure 1: Computational workflow adopted for the reliable model generation using homology modelling

MATERIALS AND METHODS

Model generation of Chitinase IV through homology modelling by modeller

The initial step in homology modelling is to recognize the related protein structures to the target sequence in the structural databases like Protein Data Bank (PDB) (Berman *et al*., 2000), followed by the selection of template among those proteins. The numerous available protein structures and sequence databases facilitates this phase. The sequence of the target proteins is used for searching the template in the PDB (Berman *et al*., 2000), DALI (Holm and Rosenstrom, 2010), SCOP (Murzin *et al*., 1995) and CATH (Orengo *et al*., 1997). Depending upon the complexity of a genome, there is a 20% to 50% likelihood of calculating an associated protein structure for a respective sequence arbitrarily selected from the given genome (Fischer and Eisenberg,

1997). The protein comparison methods are divided into three major subdivisions which are helpful in similar fold identifications. The primary class are based on pairwise comparisons of sequences that includes the independent comparison of the query sequence with every sequence of the database (Apostolico and Giancarlo, 1998). Frequently used programs for pairwise comparison are FASTA (Pearson, 1998) and BLAST (Altschul *et al*., 1990). The secondary class of approaches are based on comparisons of multiple protein sequences to increase the accuracy of the template exploration (Altschul *et al*., 1997). PSI-BLAST (Altschul *et al*., 1997) are widely utilized for this purpose, which iteratively searches an array of homologs of the query sequence. Regarding a particular protein sequence, a preliminary array of protein homologs from a structural database is gathered, then a weighted multiple sequence alignments is constructed, thereafter a PSSM (position-specific scoring matrix) is created using this alignment, and the created matrix is utilized to explore the database for the homologs. The database searching steps are reiterated until no novel homologs are discovered. PSI-BLAST locate proteins of similar structures for about twofold as much sequences in comparison to BLAST (Park *et al*., 1998). An interrelated methodology (Rychlewski *et al*., 1998) also starts by discovering all sequences evidently linked to the query sequence to find the profile of the query sequence. Furthermore, comparable profiles are created for all identified homologs. Then templates are identified by evaluating the profile of the query sequences with every constructed profile for the established template structures. An additional alternative utilizes the multiple alignments of the protein sequence united with information of the structural elements calculated from the given protein sequences (Fischer and Eisenberg, 1997), which are helpful in the identification of significant relationships between the structures when the sequence identity falls under 25 % i.e., for identifying remote sequence–structure associations. The HMMER (Finn *et al*., 2011) is an HMM based software used for the profile-HMM comparison of the query protein with various protein databases. The “threading” or fold recognition methods are the listed third-class algorithms for structure prediction (Bowie *et al*., 1991) that utilize the pairwise likeness of a protein of known structure and a query sequence. The given methods are particularly helpful when there are no related sequences for the modelling query. The HHpred server (Soding *et al*.,2005) used for implementing fold recognition utilizes the pairwise

comparison of profile hidden Markov models (HMMs) for remote protein homology detection and structure prediction. Various other domain analysis tools were also used for the reliable template prediction. This analysis was performed using SMART (Schultz *et al*., 2000), InterProScan 5 (Jones *et al*., 2014), ProtoNet (Rappoport *et al*., 2012) and SYSTERS (Schultz *et al*., 2000) in order to identify the template more precisely. InterProScan 5 searches the similar signature in the interpro consortium databases. Similarly, Simple Modular Architecture Research Tool (SMART) annotates the genetically distant domains in the sequence of the given proteins and SYSTERS which uses the clustering-based algorithms for the function prediction of proteins. The Classify Your Protein module of ProtoNet was used to analyse the protein family of the respective proteins. The templates suitable for modelling the given proteins were selected on the basis of the following rules.

- The template showing the identity (>30 %) is considered as the reliable homolog. The template belonging to a similar subfamily to that of target sequence was selected on the basis of multiple sequence alignments and a phylogenetic analysis.
- The template descriptor factors like ligands, pH, solvent and quaternary interactions etc. should be evaluated to the required parameters for the model.
- The excellence of the experimental accuracy of the template structure in terms of NMR structure's number of restraints per residue and the R-factor and resolution of a crystallographic structure is considered as important factors for template selection.

The templates fulfilling the respective criteria were selected for the construction of the comparative model. Multiple templates can be used for the homology modelling as the utilization of various templates central from the sequence of the target usually amplifies the predicted model precision (Fiser and Sali, 2003).

Template-Target alignment

The predictions of homology modelling methods are based on folding assignments which makes an alignment among the template structures and sequence of the target. However, it is not considered as the reliable target–template alignment for homology modelling. The database probing methods are frequently adjusted to find the remote identity. Consequently, when particular templates are selected, a particular means should be utilized to perform the alignment of template structures

and the target sequence (Baxevanis, 1998). The alignment is approximately correct with identity over 40 % for closely related protein sequences, but if it is lower than 40 %, then the regions of minimal local sequence similarity turns out to be frequent (Saqi *et al*., 1998) (Rost, 1999). The alignments contain progressively more alignment errors and a number of gaps as the sequence identity decreases. In order to obtain the most accurate and highly reliable alignment the methods like CLUSTAL W (Thompson *et al*., 1994) and PRALINE (PRofile ALIgNEment) (Simossis and Heringa, 2005) were used. The latter method utilizes the secondary structure elements for the comparative-extended multiple sequence alignment (Simossis and Heringa, 2005).

Three-Dimensional (3D) Model Building

After a preliminary alignment of the target–template, an array of techniques can be utilized to predict a 3D model of the query protein. There are three methods widely used for the model building i.e. modelling using rigid body assembly (Blundell *et al*., 1987), segment matching (Claessens *et al*., 1989) and satisfaction of spatial restraints (Aszodi and Taylor, 1996). Modelling through assembly of rigid bodies form the 3-D model of the protein from a few of the rigid bodies generated from the alignment of the protein structures (Blundell *et al*., 1987). This method is centred on the inherent dissection into variable loops, conserved core regions and side chains of the various protein folds (Blundell *et al*., 1987). 3D model constructions by Coordinate Reconstruction or Segment Matching forms the basis on the assumption that the majority of the hexapeptide segments in the structure of proteins can be grouped into roughly 100 structural classes (Unger *et al*., 1989). Therefore, the atomic positions from the template structures can be utilized to build homology models by recognizing and collecting brief segments which mount these positions (Unger *et al*., 1989). The Cα atoms are assumed to be the guiding positions of the segments conserved in the aligned sequence of target and the structure of the template (Unger *et al*., 1989). The prediction of the protein structure using satisfaction of the spatial restraints produces many restraints or constraints obtained from the alignment with the template structure on the framework of the query sequence. The assumption that the equivalent angles and distances among aligned residues in the query and the template structures are alike, leads to the generation of the restraints. The stereochemical restraints on dihedral angles, bond angles, bond lengths and non-bonded atoms acquired from an empirical force field, supplemented these comparatively

derived restraints. The 3D model was then constructed by reducing the infringements of all the restraints. In this study, the MODELLER (Fiser and Sali, 2003) module present in DS was used to construct the three-dimensional structure of the query protein by satisfying the spatial restraint. The homology model was generated for the Chitinase Class IV from Brassica Juncea. The amino acid sequence was retrieved from UniPort database [1]. To retrieve amino acid sequence from UniProt we search by name "Chitinase Class 4 Brassica Juncea" we find one hit which is having UniPort id is "UniProtKB - A5JVZ1 (A5JVZ1_BRAJU)". Download the protein sequence run the BLAST search and select top 4 sequence similarity and maximum identity. Model were generated through using Modeller9.12 [6] Amino acid sequence was put into PIR format that is readable by modeller. Subsequently, a search for potentially was aligned with the template, and the model was constructed and evaluated on the basis of DOPE score.

Analysis of model constructed by modeler

An estimation of the precision of the predicted 3D models of the proteins is necessary for the understanding of derived structural information. The model evaluation was performed using whole structure as well as in distinct regions. The fold analysis was the first step using the energy-based Z-score, assuming that for a score less than 2.5, then the template and the query protein belong to the same fold (Sanchez and Sali, 1998). Various available methods were used for the evaluation of the model such as TM score and DOPE Profile. Moreover, the 3D models were further verified using the modules of the SAVES server (http://nihserver.mbi.ucla.edu/SAVES/) such as PROCHECK, VERIFY3D, WHAT_CHECK, PROVE and ERRAT modules. The PROCHECK assesses the stereo-chemical quality of the structure residue by residue in comparison with a well refined structure of the proteins with similar resolutions. Similarly, WHAT_CHECK algorithm performs the extensive calculation of the stereo-chemical parameters. ERRAT calculates the statistical parameters of the non-bonded interactions between different atom types, while PROVE computes the Z-score deviation by utilizing the comparisons with highly refined PDB structures and VERIFY_3D performs the compatibility of three-dimensional protein models with its own 1D structure i.e., amino acid sequences. If the evaluation score was not satisfactory, then the selected models were refined using GROMOS energy minimization algorithm implemented in DeepView (Kaplan and Littlejohn, 2001), CHARMm (Vanommeslaeghe *et al*., 2010)

energy minimization using ChiRotor algorithm of DS and SCWRL 4.0 (Wang *et al*., 2008). After model constructed through modeller, we select the best model on the basis of DOPE score, selected model will be analyzed by ERRAT and Ramachandran plot. ERRAT was calculated by SAVES v5.0 online program. This ERRAT sore is a program for verifying protein structure determined by crystallography and Ramachandran plot through RAMPAGE Ramachandran plot is used to show in theory which values, or conformations, of φ and ψ angles are possible for an amino-acid residue in a protein and show the empirical of data points observed in a single structure in usage for structure validation or else in a database of many structures and usually shown against for the theoretically favoured regions.

Finding Binding Pocket

Active or binding sites on the protein surfaces play a central role in the protein functions. The identification of those binding sites on the protein molecule is often the first step to study protein functions and structure-based drug design. In this study, the active site was predicted using various bioinformatics tools as well as by comparing the chitinase IV amino acids sequence with the template and other similar proteins. The active site pockets were predicted by metaPocket 2.0 (Zhang *et al*., 2011), COFACTOR (Roy *et al*., 2012) and COACH (Yang *et al*., 2013). The metaPocket 2.0 algorithm uses LIGSITEcs, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity and POCASA predictors to identify the pocket sites. COFACTOR identifies template proteins of similar folds and functional sites by threading the target structure through three representative template libraries that have known as protein–ligand binding interactions, Enzyme Commission number or Gene Ontology terms. COACH was based on two methods, (i) based on binding-specific substructure comparison (TM-SITE) and (ii) based on sequence profile alignment (S-SITE), for complementary binding site predictions. Finding the Binding pockets with the help of MetaPocket 2.0 meta server. Binding sites on the protein surfaces play important role in protein function. MetaPocket is an open source identify ligand binding sites on protein surface, which is based on a consensus method, in which the predicted binding sites from eight methods: LIGSITEcs, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity and POCASA are combined together to improve the prediction. For finding Binding pocket we used selected model which is created by modeller is used as input query and for output we find 05 pockets

by using default parameter. The docking studies were further performed with chitin.

Finding Conserved Domain and Motif

Finding the Domain in Chitinase Class IV Amino acid sequence we use the PROSITE tool [5] of ExPasY , which provides a resource for the identification and annotation of conserved regions in protein sequences, covering protein families, domains and motifs. PROSITE shows the protein domains, families and functional sites. To finding motif in chitinase 4 sequence we run the online motif tool. (https://www.genome.jp/tools/motif/)

Docking of Chitin into binding pocket

The characteristics of an enzyme derived from its amino acid sequence determine the shape of the enzymes. A closer fit between an active site of an enzyme and its substrate molecule increases the efficiency of a reaction (Sullivan and Holyoak, 2008). The active site is typically found in a 3D groove or pocket of the enzyme molecule aligned with amino acid residues. The molecular utility of a protein is generally limited to its active site residues, which may comprise of an interaction with small size ligands, nucleic acids or other proteins (Powers *et al*., 2006). Integrity of the essential structural constituent of the active site is vital for preserving the functional activity of the protein. The interaction studies between biomolecules are central to all biological processes. The interaction between biologically relevant molecules such as nucleic acids, proteins, carbohydrates and fatty acids play an important role in the process of signal transduction. Complex regulatory and metabolic interaction networks within the living systems are governed by these interactions. Experimental observations and computer based theoretical analysis are the main scientific tools for understanding this phenomenon. Molecular docking is a computational method which tries to predict the preferred orientation of one molecule to a second when bound to each other to form a stable complex (Kitchen *et al*., 2004). The docking of small molecules i.e. ligands against larger receptor molecules is a complex task. Docking between these respective molecules is often considered to be a *Lock and Key* mechanism, where the receptor and ligand do not change the conformation during binding. The ligands are believed to be more flexible and assume multiple conformations in solution space. Moreover, even though the receptors show well defined conformations, they can also show the alteration if the binding of ligand through *Induced Fit* mechanism. Knowledge of the preferred alignment in turn may be used to predict the strength of association or binding affinity between two molecules using scoring functions. Docking simulations are utilised for

rational drug design and virtual screening of the library of small compounds (Kitchen *et al*., 2004). Most of the docking algorithms generate a large number of likely conformations, some of which can be rejected immediately due to high energy clashes with the protein. We require a means to score each pose to identify those of most biological relevance. The free binding energy can be written as an additive equation of various intermolecular energies to reflect their relative contributions in binding.

ΔGbind = ΔGsolvent + ΔGconformation + ΔGinteractions + ΔGrotations + ΔGvibrations

Generally, the docking practices between ligands and the receptors can be executed in a rigid mode, where a single conformation is used. This methodology is of little importance in contrast to docking procedures. The universal assumption for docking studies is to grasp the protein in a strict rigid conformation and to dock a chain of ligand conformations against the active site of the receptor protein. Most of the docking algorithms are based on these assumptions. On the other hand flexible docking allows the optimization of a particular arrangement of side-chains during the process of molecular docking. Flexible docking also facilitates the specification of pregenerated conformations of the receptor that comprises backbone and side-chain flexibility. Flexible docking usually takes additional CPU time in comparison with the rigid docking practices. In this work, we predicted the active site of chitinase using the information present in the literature of similar proteins, various bioinformatics tools and docking studies. Chitinase Class IV protein 3-dimentaional structure was predicted by homology modeling and predicted protein structure was used for energy minimisation, Energy minimisation of modeled structure was performed by using UCSF Chimera [3] and prepared for docking. Autodock is a molecular simulation software, which carries out molecule docking studies through AutoDock Tools (ADT) 1.5.4. [3]. Docking enables us understanding the molecular interactions those take place between a ligand and corresponding receptor. In 3D structure of the chitin, hydrogen was added by employing (command line) ChemAxon (http://www.chemaxon.com) software-molconvert. We dock the chitin in 2 different pocket which is find in a MetaPocket 2.0. AutoDock Tools (ADT) 1.5.4. was used preparation of all input files. Polar Hydrogen's were added and partial atomic charges were assigned by Kollman charges method. The build structure was then saved in PDBQT format to be delivered to AutoDock tools as input file. The number of grid point in xyz

94×92×92 (x, y, and z) and grid box center is 12.41×-3.208×19.258 (x, y, and z) was assigned on the macromolecule binding pocket with the spacing of 0.375Å. All docking calculation parameter were kept as a default value. Ligands were docked using Lamarckian Genetic Algorithm, with initial population of 150 randomly placed individuals, a maximum number of 2500000 energy evaluation, a mutation rate of 0.02 and a crossover rate of 0.8. Total 10 docking conformation were run. The grid maps were calculated by Autogrid4 and docking procedure was performed by Autodock4. After docking extract the complex depends on binding score. Top BE score complex was used for find the hydrophobic, hydrogen bond interaction of chitin with active pocket by using Protein-Ligand Interaction Profiles online web server.

RESULT AND DISCUSSION

Retrieve sequence from Uniport

The amino acid sequence was retrieved from UniProt. UniProt ID: A5JVZ1 with the sequence as shown following **Figure 2**.

> tr|A5JVZ1|A5JVZ1_BRAJU Class IV chitinase OS=Brassica juncea OX=3707 PE=2 SV=1

MKYAKTTSRNDQFAVLLTALFFLILTVSKPVASQNCGCPPGLCCSTNGYCGTTDDYCGVG
CKEGPCKNSGPGDPTVSLEETVTPEFFNSILSQATGSDCKGRGFYTRETFIAAANSYSKFGA

SISKREIAAFFAHVTQETGFLCHIEEVDGPAKAAEYCNTTNTESPCAQGKGYYGRGAIQLS

WNYNYGPCGRDLNEDLLATPEKVAQDQVLAFKTAFWYWTTYVSSSFKSGFGATIKAVN
SRECTGGDSTEKAANRVRCFQDYCTKLGVQPGENLTC

Figure 2: Retrieved uniport sequence

Construction of Model through Modeller

MODELLER is a computer program for comparative protein structure modelling. Search the template using Chitinase Class IV Amino acid sequence through default modeller settings. The input is an alignment of a sequence to be modeled with the template structures, MODELLER then automatically calculates a model. Searching structures related to Chitinase Class IV. It is first necessary to convert the target Chitinase Class IV sequence into the 'PIR' format that is readable by MODELLER (.ali). MODELLER uses the PIR format to read and write sequences and alignments. The first line of the PIR-formatted sequence consists of >P1; followed by the identifier of the sequence as shown in following **Figure 3**.

>P1;BJ_ChiIV
Sequence : BJ_ChiIV : : : : : : : 0.00 : 0.00

MKYAKTTSRNDQFAVLLTALFFLILTVSKPVASQNCGCPPGLCCSTNGYCGTTDDY
CGVGCKEGPCKNSGPGDPTVSLEETVTPEFFNSILSQATGSDCKGRGFYTRETFIAAA
NSYSKFGASISKREIAAFFAHVTQETGFLCHIEEVDGPAKAAEYCNTTNTESPCAQG
KGYYGRGAIQLSWNYNYGPCGRDLNEDLLATPEKVAQDQVLAFKTAFWYWTTYV
SSSFKSGFGATIKAVNSRECTGGDSTEKAANRVRCFQDYCTKLGVQPGENLTC*

Figure 3: modeller pir format file

The sequence is identified by the code BJ ChiIV. The second line, consisting of ten fields separated by colons, usually contains details about the structure, if any. In the case of sequences with no structural information, only two of these fields are used: the first field should be sequence and the second should contain the model file name. The rest of the file contains the sequence with an asterisk (*) marking its end. After construction of basic modeller files need to search suitable template structure. A search for potentially related sequences of known structure can be performed using the command of MODELLER using build_profile.py script. And give a set of statistically significant alignments. Then select the templates from resulted templates. We select top 4 Template on the basis of maximum identity similarity, a sequence identity value above ~25% indicates a potential template.as shown in following **Table 1**.

Table 1: modeller selected template score

Template	Query Cover	Per. Ident
2dkv	86%	35.92%
4mck	71%	50.99%
5h7t	70%	46.83%
3hbd	72%	47.14%

Aligning TPRS sequence with template

To align the sequence of Chitinase Class IV with selected 4 template structures by use the aligned command in MODELLER. Sequence alignment with template as shown following **Figure 4**.

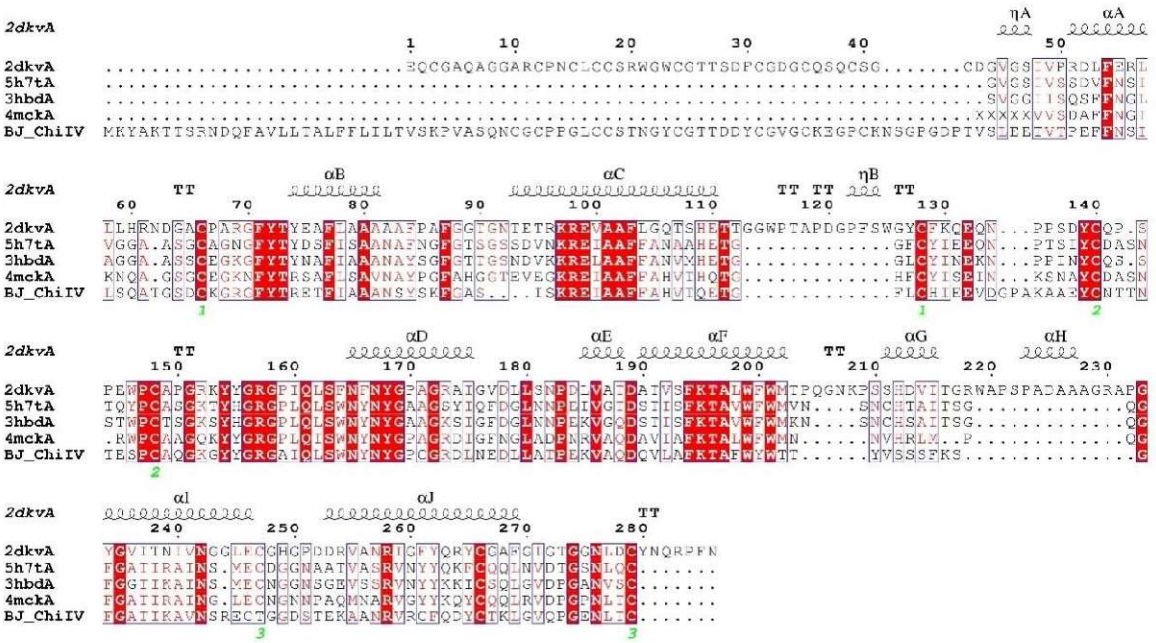

Figure 4: Aligning TPRS sequence with template

Model Building

Once a target-template alignment is constructed, MODELLER calculates a 3-D model of the target completely automatically, using model-single.py script we generate ten different models of Chitinase Class IV on basis of selected template structure.

Evaluating Model

From generated Ten model the best model can be selected by picking the model with the lowest value of the DOPE, top 5 DOPE score are shown as following **Table 2**.

Table 2: top 5 modeller score

Model	DOPE Score	GA341 score
BJ_ChiIV.B99990005.pdb	**-26845.06250**	**1.00000**
BJ_ChiIV.B99990004.pdb	**-26683.33008**	**1.00000**
BJ_ChiIV.B99990006.pdb	**-26655.59180**	**1.00000**
BJ_ChiIV.B99990008.pdb	**-26543.42578**	**1.00000**
BJ_ChiIV.B99990009.pdb	**-26486.24609**	**1.00000**

Among these 10 models we select the fifth model which DOPE score is -26845.06250Kcal/mol.

Once a final model is selected, we analysed the model, final model was shown in following **Figure 5**.

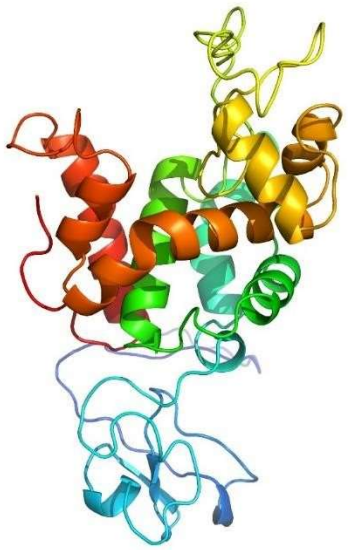

Figure 5: Generated Model Structure

Analysis of constructed model

The modeled structure was analyzed by Errate score and Ramachandran plot generated by SAVES v5.0. Regions of the structure that can be rejected at the 95% confidence level are yellow; 5% of a good protein structure is expected to have an error value above this level. Regions that can be rejected at the 99% level are shown in red. The model ERRAT quality factor scores of 72.201 revealed that the non-bonded interactions of model lie within a reasonable normal range **(Figure 6)**.

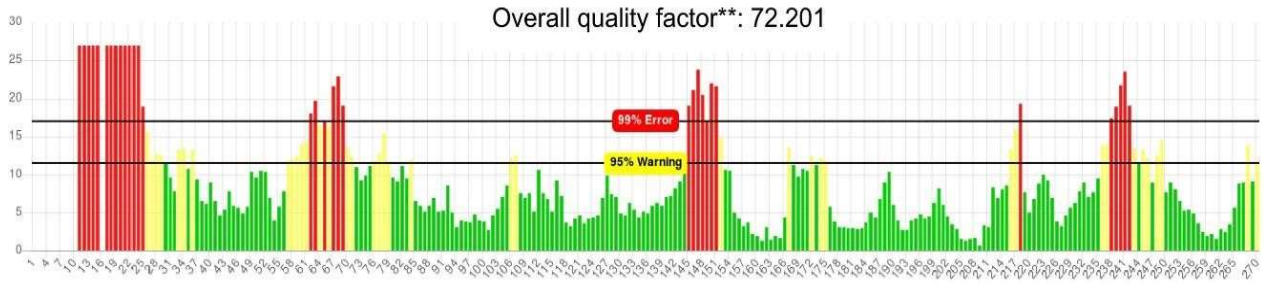

Figure 6: ERRAT score for selected model

After Errat we calculate the Ramachandran plot **(Figure 7)** which suggests that 90.7% residues are in the favoured region, and 1.3% of the modeled amino acids have disallowed geometry, as shown in following **Figure 7**. The obtained results suggest that the model of receptor is reliable and suitable for further studies.

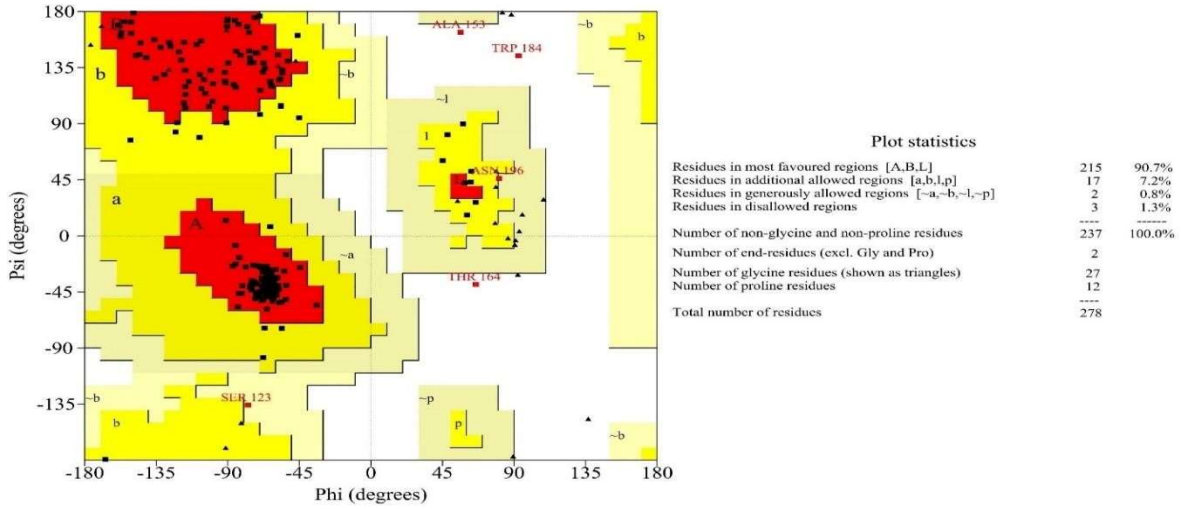

Figure 7: Ramachandran Plot for Selected Model

Finding active binding site of chitinase protein

MetaPocket 2.0 is a meta server shows the binding pocket on protein structure. we run the MetaPocket server by providing generated structure, we find total 5 Binding pocket in provided input structure as shown in following **Figure 8**.

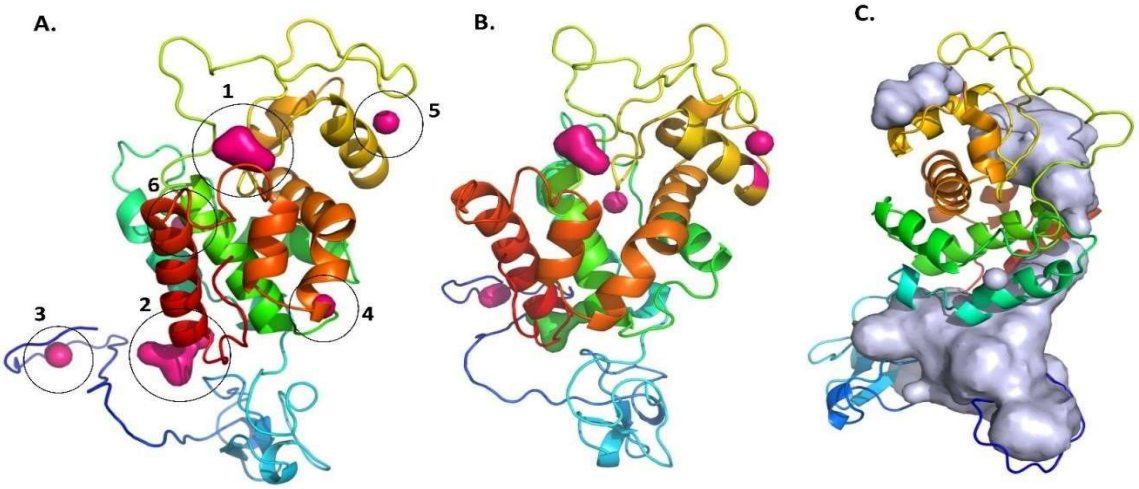

Figure 8: Top Four Binding Pocket Find By F-Pocket

By using MetaPocket server our protein successful run the CON, FPK, GHE, LCS, PAS, SFN method but QSF and PCS method were failed among total method. Above figure A and B shows the top 5 binding pocket in pink colour and **Figure 8 C** shows the cluster of all pocket found in MetaPocket which is having total Z- Score is 19.98 of pocket sites. As well as in COACH online server shows the amino acid present in binding pocket-1 which take

part in interaction these amino acids are as HIS135, GLU139, GLU148, TYR158, GLY176, ILE180, GLN181, LEU182, SER183, ASN187, TYR220, TRP221, ALA239, VAL240, ASN241, ARG257. To validate the above predicted result the docking analysis was performed.

Analysis of Conserved Domain

We run the PROSITE tool of ExPasY to finding the domain in Chitinase Class IV Amino acid sequence. In a given sequence of Chitinase Class IV we find the 3 hits, shown as in following figure. Which contains Chitin-binding type-1 domain, many plants respond to pathogenic attack by producing defense protein which is having capacity to bind to chitin, an N-acetylglucosamine polysaccharide present in cell wall of fungi and exoskeleton of insects. Fungi cell wall contains the polysaccharide and insect have exoskeleton. There are 30-43 amino acid residue of chitin binding protein having common structural motif, contain 4-disulfide core, called as chitin binding domain type-1. Following given figure shows the chitin binding domain type-1 at sequence 33 to 68 amino acid "SQNCG-------_CPPGLCCSTNGYCGTTDDYCGVGCKeGPCKN" which is having 9.454 score and di-sulfide bond form between amino acid shown in following **Figure 9**.

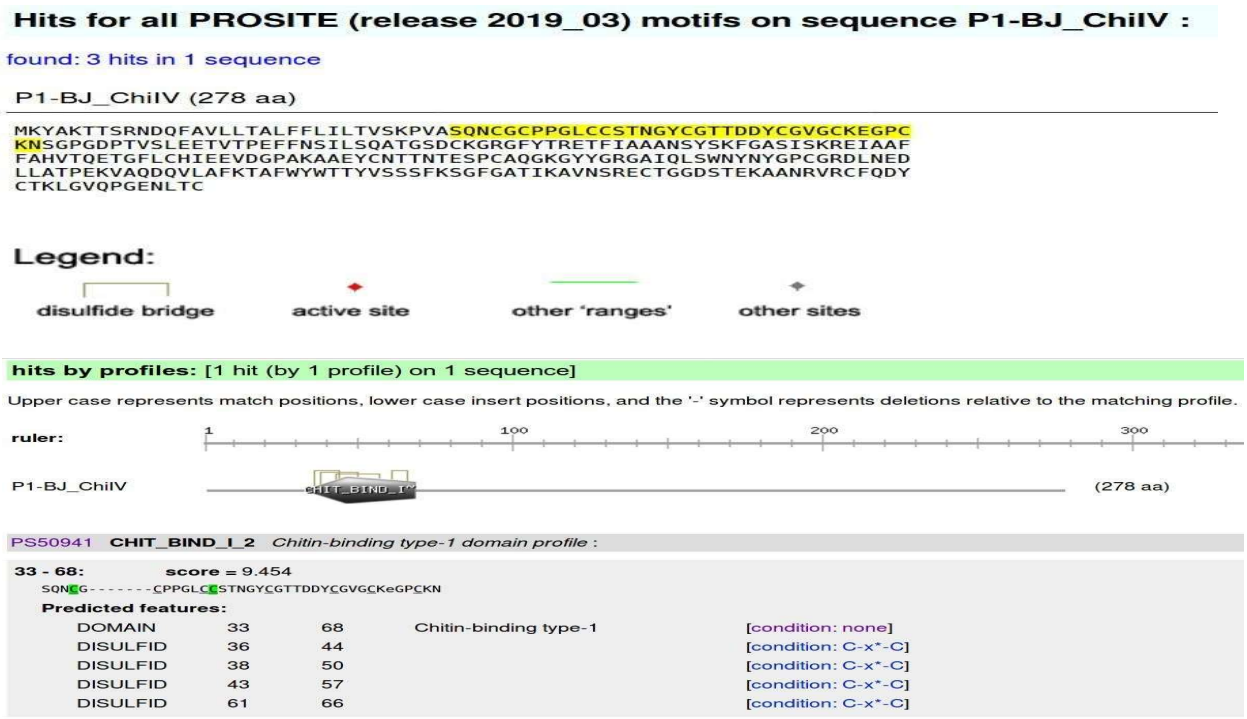

Figure 9

Also, given sequence find the chitin recognition domain or binding domain at 38-57 amino acid and another one chitinase family 19 at amino acid 99-127 on given sequence of Chitinase Class IV **(Figure 10)**.

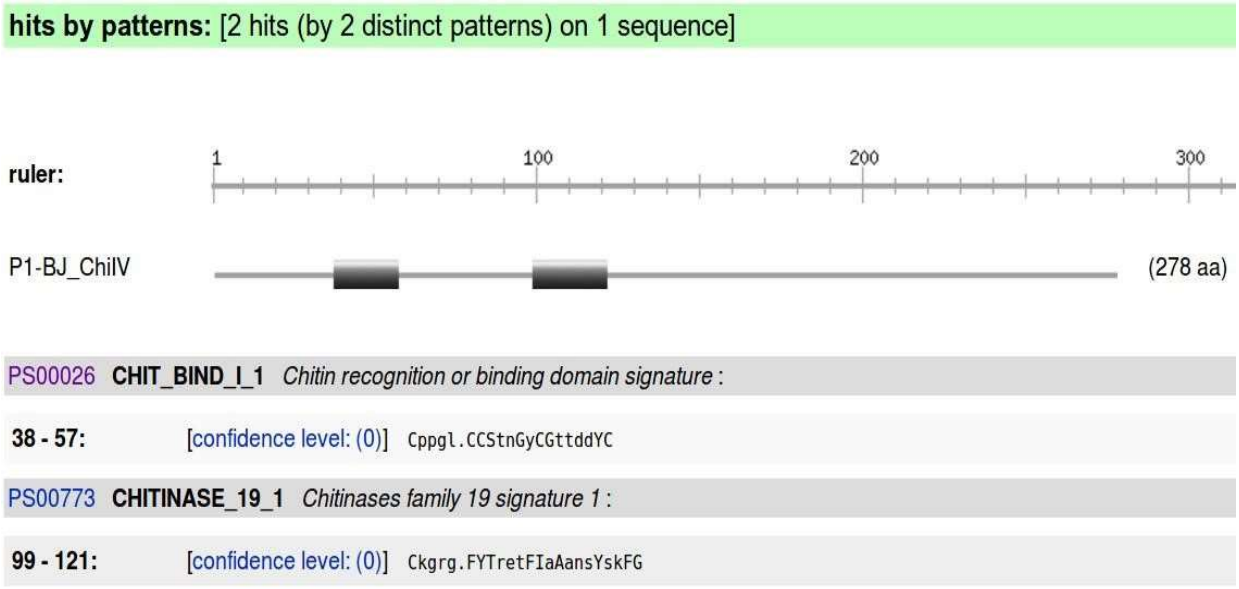

Figure 10

We run the online Motif tool to finding the motif in Chitinase Class IV Amino acid sequence. In a given sequence of Chitinase Class IV find motif as shown in following **Figure 11**.

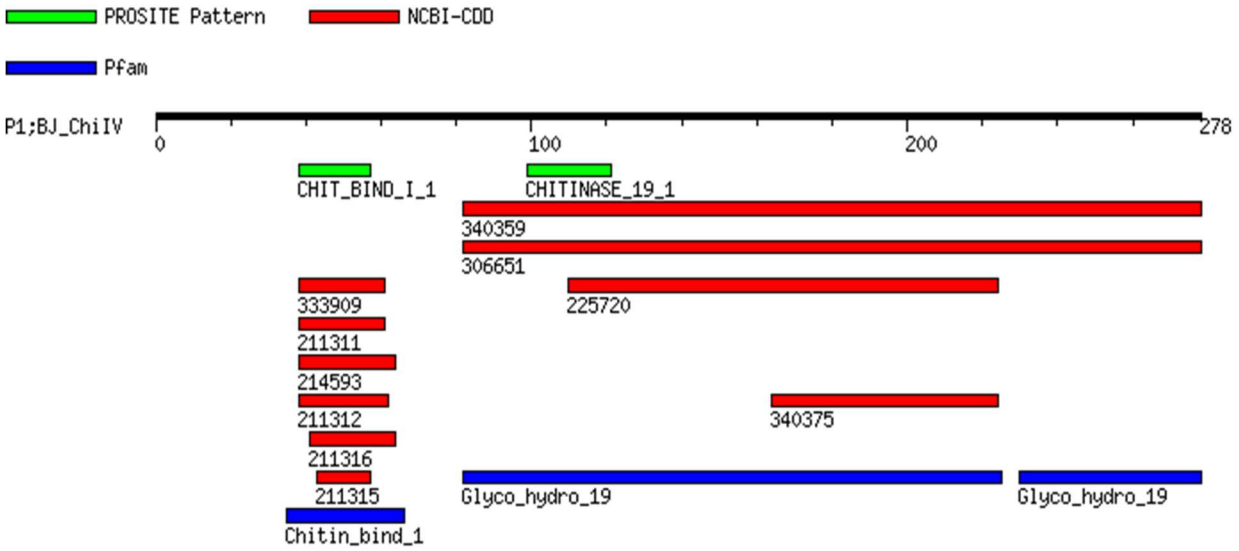

Figure 11

In a above **Figure 11** shows the PROSITE pattern, NCBI-CDD [2], and Pfam motifs. PROSITE pattern found 2 motifs shown in green colour, NCBI-CDD-10 motifs red in colour and Pfam-2 motif blue in colour and flowing **Table 3** give the ID and position of motifs.

Table 3: Conserved Domain Site Id and Position

S. No.	ID	Position
1.	Prosite: CHIT_BIND_I_1	38..57
2.	Prosite: CHITINASE_19_1	99..121
3.	NCBI-CDD ID:340359	82..278
4.	NCBI-CDD ID:306651	82..278
5.	NCBI-CDD ID:225720	110..224
6.	NCBI-CDD ID:333909	38..61
7.	NCBI-CDD ID:211311	38..61
8.	NCBI-CDD ID:214593	38..64
9.	NCBI-CDD ID:211312	38..62
10.	NCBI-CDD ID:340375	164..224
11.	NCBI-CDD ID:211316	41..64
12.	NCBI-CDD ID:211315	43..57
13.	Pfam: Glyco_hydro_19	82..225 and 230.278
14.	Pfam: Chitin_bind_1	35..66

Molecular Docking of Chitin

Oligosaccharide of chitin was docked in predicted Chitinase Class IV on predicted top 2 active site of protein. The chitin structure was draw in Marvin Sketch software of InstaJChem **(Figure 12)** and saved in ". mol2" format with 3D format by adding hydrogen bond.

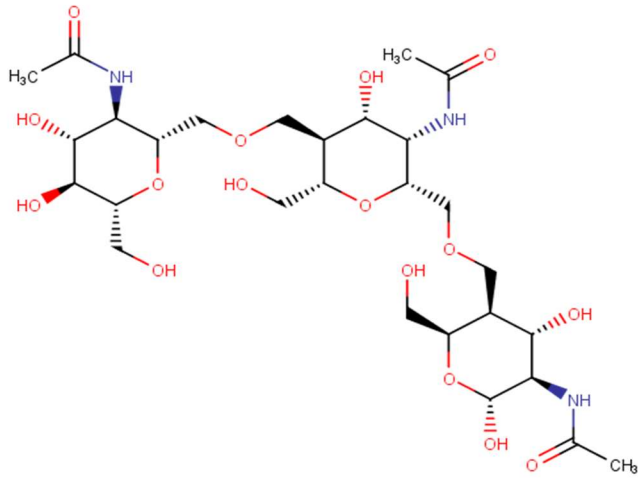

Figure 12: Chitin 2D Structure

Firstly, we perform docking for first predicted active site of protein shown in **Figure 13A**. The docking of chitin resulted binding energy is -2.09kcal/mol. The docked oligosaccharide Chitin find different interaction which hydrogen bond interaction, hydrophobic interaction and non-bonded interactions. Interactions are shown in following **Table 4** and **Figure 13A-D**.

Table 4: First Predicted Active Site Bonding Interaction with Amino Acid.

Bond	Amino Acid and number
Hydrophobic Interactions	186TYR, 224TYR
Hydrogen Bonds	139GLU, 181GLN, 183SER, 187ASN, 241ASN, 243ARG

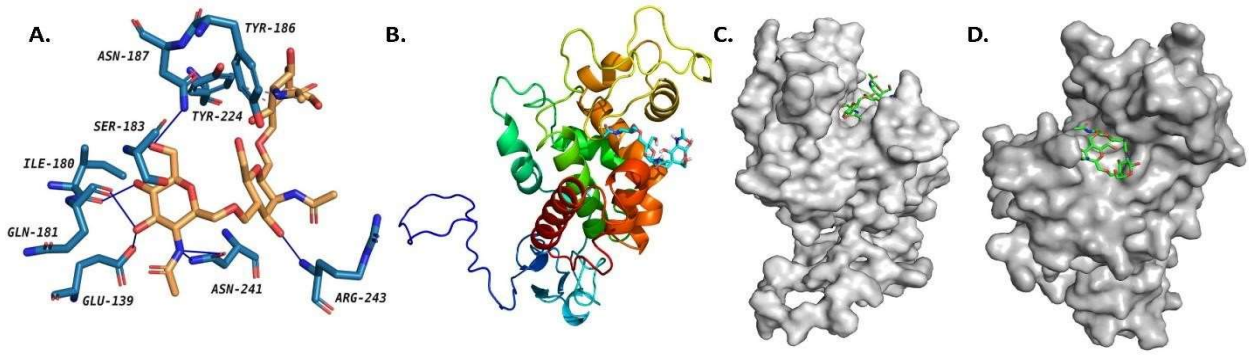

Figure 13: Chitin Interaction with Chitinase Class Iv First Pocket

Secondly, we perform docking for second top predicted active site of protein shown in **Fi gure 14 A**. The docking of chitin resulted binding energy is -3.95kcal/mol. different interaction are shown in following **Table 5 and Figure 14**.

Table 5: Second Predicted Active Site Bonding Interaction with Amino Acid.

Bond	Amino Acid and number
Hydrophobic Interactions	20LEU, 25LEU, 40PRO, 83THR, 85GLU
Hydrogen Bonds	18THR, 20LEU, 21PHE, 23LEU, 81THR, 1266LYS, 27LYS

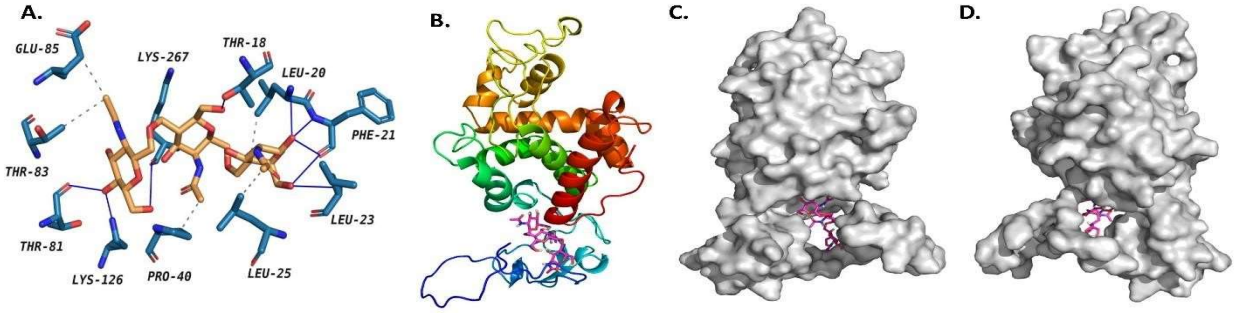

Figure 14: Chitin Interaction with Chitinase Class Iv Second Pocket

CONCLUSIONS

The sequence is identified by the code BJ_ChiIV. The templates from resulted templates were select, top 4 Template on the basis of maximum identity similarity. Align the sequence of Chitinase Class IV with selected 4 template structures by use the aligned command in MODELLER. The models were generated with high precision refinement using various available prediction algorithms. These models showed high structure similarities to the chitinase-like proteins with the characteristic (β/α)8 TIM barrel found in both proteins. From generated Ten model the best model can be selected based on the lowest value of the DOPE. Among these 10 models we select the fifth model which DOPE score is -26845.06250Kcal/mol. The modeled structure was analysed by Errate score and Ramachandran plot generated by SAVES v5.0. Regions that can be rejected at the 99% level. The model ERRAT quality factor scores of 72.20. After Errat we calculate the Ramachandran plot which suggests that 90.7% residues are in the favoured region, and 1.3% of the modeled amino acids have disallowed geometry. MetaPocket 2.0 shows the binding pocket on protein structure, server we find total 5 Binding pocket which is having having total Z- Score is 19.98 of pocket sites. The active site analysis is predicted by further structure refinement using the molecular. A closer inspection of the predicted 3D structures of chitinase IV showed that it contains binding pocket-1 which take part in interaction these amino acids are as HIS135, GLU139, GLU148, TYR158, GLY176, ILE180, GLN181, LEU182, SER183, ASN187, TYR220, TRP221, ALA239, VAL240, ASN241, ARG257 in the active site pocket. PROSITE pattern found 2 motifs, NCBI-CDD-10 motifs and Pfam-2 motif. These analyses of the active site were further confirmed by molecular docking studies using AutoDock. These analyses were helpful in understanding the active sites as this enzyme. The docking of chitin in pocket 1 shows the binding energy is -2.09kcal/mol. And in second pocket chitin resulted binding energy is -3.95kcal/mol.

REFERENCES

The UniProt Consortium, 'Uniprot: A Hub for Protein Information', Nucleic Acids Research, 43 (2014), D204-D12.

A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, L. Y. Geer, and S. H.

Bryant, 'Cdd/Sparcle: Functional Classification of Proteins Via Subfamily Domain Architectures', Nucleic Acids Res, 45 (2017), D200-D03.

E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T.

E. Ferrin, 'Ucsf Chimera--a Visualization System for Exploratory Research and Analysis', J Comput Chem, 25 (2004), 1605-12.

J. Pontius, J. Richelle, and S. J. Wodak, 'Deviations from Standard Atomic Volumes as a Quality Measure for Protein Crystal Structures', J Mol Biol, 264 (1996), 121-36.

Christian J. A. Sigrist, Edouard de Castro, Lorenzo Cerutti, Béatrice A. Cuche, Nicolas Hulo, Alan Bridge, Lydie Bougueleret, and Ioannis Xenarios, 'New and Continuing Developments at Prosite', Nucleic Acids Research, 41 (2012), D344-D47.

B. Webb, and A. Sali, 'Comparative Protein Structure Modeling Using Modeller', Curr Protoc Bioinformatics, 54 (2016), 5 6 1-5 6 37.

Zengming Zhang, Yu Li, Biaoyang Lin, Michael Schroeder, and Bingding Huang, 'Identification of Cavities on Protein Surface Using Multiple Computational Approaches for Drug Binding Site Prediction', Bioinformatics, 27 (2011), 2083-88.

C. J. & GEZELTER, J. D. 2006. Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics. J Chem Phys, 124, 234104.

FUZO, C. A. & DEGREVE, L. 2012. Effect of the thermostat in the molecular dynamics simulation on the folding of the model protein chignolin. J Mol Model, 18, 2785-94.

HOLDEN, Z. C., RICHARD, R. M. & HERBERT, J. M. 2013. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. J Chem Phys, 139, 244108.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: visual molecular dynamics. J Mol Graph, 14, 33-8, 27-8.

JORGENSEN, W. L. & TIRADO-RIVES, J. 2005. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proc Natl Acad Sci U S A, 102, 6665-70.

KARPLUS, M. & MCCAMMON, J. A. 2002. Molecular dynamics simulations of biomolecules. Nat Struct Biol, 9, 646-52.

KINI, R. M. & EVANS, H. J. 1991. Molecular modeling of proteins: a strategy for energy minimization by molecular mechanics in the AMBER force field. J Biomol Struct Dyn, 9, 475-88.

NGUYEN, T. T., VIET, M. H. & LI, M. S. 2014. Effects of water models on binding affinity:

evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase. ScientificWorldJournal, 2014, 536084.

NORBERTO DE SOUZA, O. & ORNSTEIN, R. L. 1999. Molecular dynamics simulations of a proteinprotein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. J Biomol Struct Dyn, 16, 1205-18.

VAN DER SPOEL, D., LINDAHL, E., HESS, B., GROENHOF, G., MARK, A. E. &

BERENDSEN, H. J. 2005. GROMACS: fast, flexible, and free. J Comput Chem, 26, 170118.

KITCHEN, D. B., DECORNEZ, H., FURR, J. R. & BAJORATH, J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov, 3, 935-49.

POWERS, R., COPELAND, J. C., GERMER, K., MERCIER, K. A., RAMANATHAN, V. & REVESZ, P. 2006. Comparison of protein active site structures for functional annotation of proteins and drug design. Proteins, 65, 124-35.

ROY, A., YANG, J. & ZHANG, Y. 2012. COFACTOR: an accurate comparative algorithm for structurebased protein function annotation. Nucleic Acids Res, 40, W471-7.

SULLIVAN, S. M. & HOLYOAK, T. 2008. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. Proc Natl Acad Sci U S A, 105, 13829-34.

SUZUKI, S., NAKANISHI, E., OHIRA, T., KAWACHI, R., NAGASAWA, H. & SAKUDA, S.

2006. Chitinase inhibitor allosamidin is a signal molecule for chitinase production in its producing Streptomyces I. Analysis of the chitinase whose production is promoted by allosamidin and growth

accelerating activity of allosamidin. J Antibiot (Tokyo), 59, 4029.

WU, G., ROBERTSON, D. H., BROOKS, C. L., 3RD & VIETH, M. 2003. Detailed analysis of gridbased molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. J Comput Chem, 24, 1549-62.

YANG, J., ROY, A. & ZHANG, Y. 2013. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics, 29, 2588-95.

ZHANG, Z., LI, Y., LIN, B., SCHROEDER, M. & HUANG, B. 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics, 27, 2083

HACKMAN, R. H. 1954. Studies on chitin. I. Enzymic degradation of chitin and chitin esters. Aust J Biol Sci, 7, 168-78.

HACKMAN, R. H. & GOLDBERG, M. 1965. Studies on chitin. VI. The nature of alpha- and betachitins. Aust J Biol Sci, 18, 935-46.

HAKI, G. D. & RAKSHIT, S. K. 2003. Developments in industrially important thermostable enzymes: a review. Bioresour Technol, 89, 17-34.

HARTL, L., ZACH, S. & SEIDL-SEIBOTH, V. 2012. Fungal chitinases: diversity, mechanistic properties and biotechnological potential. Appl Microbiol Biotechnol, 93, 533-43.

HAYES, C. K., KLEMSDAL, S., LORITO, M., DI PIETRO, A., PETERBAUER, C., NAKAS, J. P., TRONSMO, A. & HARMAN, G. E. 1994. Isolation and sequence of an endochitinaseencoding gene from a cDNA library of Trichoderma harzianum. Gene, 138, 143-8.

HENRISSAT, B. 1999. Classification of chitinases modules. EXS, 87, 137-56.

HENRISSAT, B. & BAIROCH, A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J, 293 (Pt 3), 781-8.

HENRISSAT, B. & BAIROCH, A. 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem J, 316 (Pt 2), 695-6.

HOLDEN, Z. C., RICHARD, R. M. & HERBERT, J. M. 2013. Periodic boundary conditions for QM/MM calculations: Ewald summation for

extended Gaussian basis sets. J Chem Phys, 139, 244108.

HOLM, L. & ROSENSTROM, P. 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res, 38, W545-9.

LOBANOV, M., BOGATYREVA, N. S. & GALZITSKAIA, O. V. 2008. [Radius of gyration is indicator of compactness of protein structure]. Mol Biol (Mosk), 42, 701-6.

LORITO, M., HAYES, C. K., DI PIETRO, A. & HARMAN, G. E. 1993. Biolistic transformation of Trichoderma harzianum and Gliocladium virens using plasmid and genomic DNA. Curr Genet, 24, 349-56.

LORITO, M., WOO, S. L., GARCIA, I., COLUCCI, G., HARMAN, G. E., PINTOR-TORO, J. A., FILIPPONE, E., MUCCIFORA, S., LAWRENCE, C. B., ZOINA, A., TUZUN, S. & SCALA, F. 1998. Genes from mycoparasitic fungi as a source for improving plant resistance to fungal pathogens. Proc Natl Acad Sci U S A, 95, 7860-5

MACHUQUEIRO, M. & BAPTISTA, A. M. 2006. Constant-pH molecular dynamics with ionic strength effects: protonation-conformation coupling in decalysine. J Phys Chem B, 110, 2927-33. MACHUQUEIRO, M. & BAPTISTA, A. M. 2007. The pH-dependent conformational states of kyotorphin: a constant-pH molecular dynamics study. Biophys J, 92, 1836-45.

MATTHEW, J. B., GURD, F. R., GARCIA-MORENO, B., FLANAGAN, M. A., MARCH, K. L. & SHIRE, S. J. 1985. pH-dependent processes in proteins. CRC Crit Rev Biochem, 18, 91197.

MAYER, R. T., MCCOLLUM, T. G., NIEDZ, R. P., HEARN, C. J., MCDONALD, R. E., BERDIS, E. & DOOSTDAR, H. 1996. Characterization of seven basic endochitinases isolated from cell cultures of Citrus sinensis (L.). Planta, 200, 289-95.

MENG ZHANG, A. K. P., ALGASAN GOVENDER, ZHENGXIANG WANG,SUREN SINGH, KUGENTHIREN PERMAUL 2014. The multi-chitinolytic enzyme system of the compostdwellingthermophilic fungus Thermomyces lanuginosus. Process Biochemistry.

MI, F. L., TAN, Y. C., LIANG, H. C., HUANG, R. N. & SUNG, H. W. 2001. In vitro evaluation of achitosan membrane cross-linked with genipin. J Biomater Sci Polym Ed, 12, 835-50.

MI, F. L., WONG, T. B. & SHYU, S. S. 1997. Sustained-release of oxytetracycline from chitosan microspheres prepared by interfacial acylation and spray hardening methods. *J Microencapsul,* 14**,** 577-91.

MINKE, R. & BLACKWELL, J. 1978. The structure of alpha-chitin. *J Mol Biol,* 120**,** 167-81.

MISURA, K. M., CHIVIAN, D., ROHL, C. A., KIM, D. E. & BAKER, D. 2006. Physically realistic homology models built with ROSETTA can be more accurate than their templates. Proc Natl Acad Sci U S A, 103, 5361-6.

MOHAMED, A. J., YU, L., BACKESJO, C. M., VARGAS, L., FARYAL, R., AINTS, A., CHRISTENSSON, B., BERGLOF, A., VIHINEN, M., NORE, B. F. & SMITH, C. I. 2009. Bruton's tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain. Immunol Rev, 228, 58-73.

MULLER, C. W., SCHLAUDERER, G. J., REINSTEIN, J. & SCHULZ, G. E. 1996. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure, 4, 147-56.

MURZIN, A. G., BRENNER, S. E., HUBBARD, T. & CHOTHIA, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol, 247, 536-40.

MUZZARELLI, R. A. 1999. Native, industrial and fossil chitins. EXS, 87, 1-6.

NGUYEN, T. T., VIET, M. H. & LI, M. S. 2014. Effects of water models on binding affinity:

evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase.

ScientificWorldJournal, 2014, 536084.

NORBERTO DE SOUZA, O. & ORNSTEIN, R. L. 1999. Molecular dynamics simulations of a proteinprotein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. J Biomol Struct Dyn, 16, 1205-18.

NOVAES-LEDIEU, M., MARTINEZ COBO, J. A. & GARCIA MENDOZA, C. 1987. The structure of the mycelial wall of Agaricus bisporus. Microbiologia, 3, 13-23.

ODIBO, F. J. C. & ULBRICH-HOFMANN, R. 2001. Thermostable α-Amylase and Glucoamylase from Thermomyces

lanuginosus F1. Acta Biotechnologica, 21, 141-153.

ORENGO, C. A., MICHIE, A. D., JONES, S., JONES, D. T., SWINDELLS, M. B. & THORNTON, J. M. 1997. CATH--a hierarchic classification of protein domain structures. Structure, 5, 1093-108.

PACE, C. N., SHIRLEY, B. A., MCNUTT, M. & GAJIWALA, K. 1996. Forces contributing to the conformational stability of proteins. FASEB J, 10, 75-83.

PARK, J., KARPLUS, K., BARRETT, C., HUGHEY, R., HAUSSLER, D., HUBBARD, T. & CHOTHIA, C. 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. J Mol Biol, 284, 1201-10.

PATIL, R. S., GHORMADE, V. V. & DESHPANDE, M. V. 2000. Chitinolytic enzymes: an exploration. Enzyme Microb Technol, 26, 473-483.

PEARSON, W. R. 1998. Empirical statistical estimates for sequence similarity searches. J Mol Biol, 276, 71-84.

POWERS, R., COPELAND, J. C., GERMER, K., MERCIER, K. A., RAMANATHAN, V. & REVESZ, P. 2006. Comparison of protein active site structures for functional annotation of proteins and drug design. Proteins, 65, 124-35.

RAPPOPORT, N., KARSENTY, S., STERN, A., LINIAL, N. & LINIAL, M. 2012. ProtoNet 6.0:

organizing 10 million protein sequences in a compact hierarchical family tree. Nucleic Acids Res, 40, D313-20.