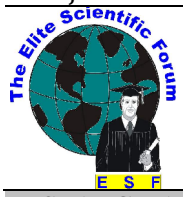
Pak. j. life soc. Sci. (2013), 11(3): 190-195 E-ISSN: 2221-7630;P-ISSN: 1727-4915



Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

RESEARCH ARTICLE

Functional Annotation of the *cda*1 Gene from *Bacillus thuringiensis* through Homology Modeling and Molecular Docking

Muhammad Ali[1,2], Muhammad Aamer Mehmood[1]*, Khadim Hussain[1], Ujala Sehar[1], Rana Rehan Khalid[1] and Zahid Iqbal[3]
[1]Department of Bioinformatics and Biotechnology, Faculty of Science and Technology, Government College University, Faisalabad, Pakistan
[2]Center for Bioinformatics, University of Saarland, Saarbrucken, Germany
[3]Department of Pharmacology, Al-Nafees Medical College, Isra University Islamabad Campus, Pakistan

ARTICLE INFO
Received: Jun 16, 2013
Accepted: Aug 28, 2013
Online: Sep 19, 2013

Keywords
B. thuringiensis
Chitin deacetylase
Homology modeling
Molecular docking

***Corresponding Author:**
draamer@gcuf.edu.pk

ABSTRACT

Insecticidal proteins from *Bacillus thuringiensis* have been exploited to develop insect-resistance in crop plants. We are using *B. thuringiensis* serovar konkukian S4 as a model organism to elucidate the substrate utilization pathways in bacteria. Functional annotation of the putative genes involved in recalcitrant substrate degradation is vital in this context. Genome sequence analyses of the serovar konkukian have shown the presence of a probable chitin deacetylase/xylanase gene named as *cda*1. The protein sequence of the said gene was retrieved from UniProt database to describe its function using *in silico* tools. Primary structure analyses show that the protein CDA1 is hydrophobic in nature due to the high content of non-polar amino acid residues. The instability (Ii) and aliphatic indices are computed to be 23.73 and 93.6, respectively by Expasy's ProtParam that classify the protein as stable for a wide range of temperatures. Secondary structure analyses predicted that CDA1 has predominantly α-helical structure. Domain and motif prediction by Pfam has shown the protein to contain one potential domain called polysaccharide deacetylase 1 and two potential motifs, respectively. The three dimensional structure of the CDA1 was predicted using online available servers like Phyre, SWISS-MODEL and PS^2 followed by Modeller 9.9 through homology modeling. To elucidate the function of the protein, molecular docking was performed by AutoDock Vina using chitin and xylan oligomers as substrate. Moreover, the protein was compared with several chitin deacetylases and xylanases through comparative phylogenic analyses. Our results have demonstrated that CDA1 is a chitin deacetylase. This is the very first report on this gene from *B. thuringiensis*.

INTRODUCTION

Online services and computational packages are heavily being used nowadays for the characterization of proteins. Various structural and physicochemical properties of proteins can be better exploited by using computational tools. For the purpose of protein structure prediction and identification plenty of tools are available on World Wide Web which can either be used online or as standalone service. The basic primary sequence analysis of protein can yield information about sequence length, contribution of individual residues along with physico-chemical properties of protein like atomic composition, molecular weight, extinction coefficient, theoretical isoelectric point (*p*I), estimated half life, instability index and many more parameters. A molecule's functional, chemical and physical properties can solely be determined and characterized using amino acid sequence information.

Although precise and accurate structure of proteins can be guaranteed by experimental methods yet they have the disadvantage of being expensive, time consuming and large amount of purified protein is required for this purpose. Computational methods are an excellent and cost effective alternative, in this context. Despite of the fact that they are not as much reliable as experimental ones, still they can provide us nearly exact structure of proteins besides the deep understanding of structure-

function relationship of protein at almost no cost. By far the most authentic and precise solution to the problem of protein structure prediction is template-based modeling (Wu et al., 2008). Extensive expertise in the structural biology and the use of extremely particularized computer programs are the prerequisite of modeling of protein structures for each of the individual steps of the modeling process (Tramontano et al., 2001). The modern era of automated modeling facility with integrated expert knowledge and easy to use interface was came into existence 15 years ago with the establishment of SWISS-MODEL server (Peitsch, 1996). Prediction of three-dimensional (3-D) protein structure of CDA1 by online available servers is verified and followed by the template-based structure prediction which is also called homology modeling as described previously (Sehar et al., 2013). Homology modeling is the prediction of three dimensional structure of a target protein from the amino acid sequence (primary structure) of a homologous (template) protein for which an X-ray or NMR structure is available. Furthermore, the important cellular process of binding of CDA1 with chitin is carried out by protein complexes, for this purpose of providing physical pictures of interacting proteins, many computational protein-protein prediction methods have been developed in the past but AutoDock Vina is used due to most reliable and efficient algorithm used in the backend which yields the near optimal results.

The enzyme chitin deacetylase (EC 3.5.1.41) reported in various fungi, marine bacteria and insects, is involved in the hydrolysis of the N-acetomide chitin and chitosan bonds, yielding glucosamine and acetic acid (Zhao et al., 2010). Considering the role of chitosan in prevention of infection in fishes and crustaceans, chitin deacetylase has vital importance in defense system of marine life. Beside this, it also has immense applications in food products, pharmaceuticles, cosmetics and biomedicine (Tsigos et al., 2000).

We are using *Bacillus thuringiensis* serovar konkukian to understand the biodegradation pathways of the naturally abundant and recalcitrant substrates in prokaryotes. Chitin is a polymer of N-acetyl glucosamine and is ubiquitous in nature. Previosuly, we have characterized two chitinases (Chi74, Chi39) and two chitin binding proteins (CBP50, CBP24) from the serovar konkukian (Mehmood et al., 2010; Mehmood et al., 2011; Mehmood et al., 2012; Sehar et al., 2013). The present study involves the *in silico* characterization of a *cda*1 gene which is putatively identified as either a xylanase or a chitin deacetylase gene. We have characterized *cda*1 gene on the basis of 3D structural, physico-chemical properties and phylogenetic relationships using latest bioinformatics tools and up to date online servers.

MATERIALS AND METHODS

Sequence retrieval

Amino acid sequence of the putative xylanase/chitin deacetylase (CDA1) was obtained from UniProt which is a central, authoritative resource for protein sequences and also gives functional information (Apweiler et al., 2004). UniProt gives complete, richly classified, precisely and completely annotated protein sequence knowledge, with widespread query interfaces and cross-references.

CDA1 Sequence analyses

A domain is an area in a protein sequence that has a particular 3D structure and has a definite function, as CDA1 contains only one functionally putative domain hence domain boundaries were directly taken from Pfam annotations (Finn et al., 2010). Motif detection was performed using Pfam service and two putative motifs were found at different locations. Signal peptide was removed by SignalP 4.0 server, the removal of signal peptide makes certain that the structure prediction is in agreement with the native protein structure (Jannick et al., 2004). The physico-chemical parameters of the protein sequence were suggested by Protparam analysis tool available at Expasy (Gasteiger et al., 2005). Information about protein's secondary structure was gained by PSIPRED, which predicts helices, coils and strands in target sequence with a confidence value (Buchan et al., 2010).

Sequence alignment and phylogenetic analyses

Comparative modeling usually begins with search for known protein structure (Template structure) in Protein Data Bank (PDB) based on the domain polysaccharide deacetylase-1, we have found in CDA1 (Target sequence). Domain wise multiple sequence alignment of chitin deacetylase was performed using ClustalW, multiple sequence alignments are essential for protein function, structure prediction and phylogeny inference. The template selected for target sequence was the one having the highest sequence similarity (32%) and also the similar domain (Ploysaccharide deacetylase1).

Homology Modeling and structure verification

To get a better idea about the 3-D structure of CDA1, The 3-D structure was built by means of the Modeller 9.9 along with subsequent online servers like Phyre, SWISS-MODEL and PS^2. By using template based approach the domain in target sequence was modeled using template composed of homologous polysaccharide deacetylase-1 in *Bacillus anthracis*. Homology model of domain particularly have high structure-sequence compatibility score. The resulting 3-D model evaluation, widespread checking of many stereo-chemical parameters of the residues and energy minimization in the target model were done by PROCHECK (Laskowski et al., 1996), WHAT_CHECK (Vriend, 1990) and VERIFY3D

(Bowie et al., 1991), respectively. The structures were visualized by Pymol which is a molecular visualization system on an open-source foundation. The stereo chemical quality was accessed by the analysis of Ramachandran plot (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) of the predicted target structure.

Substrate preference analyses through molecular docking

After the successful prediction of 3-D structure and an insight into residue by residue information of amino acids of CDA1 the docking was performed by using AutoDock Vina (Trott and Olson, 2010) to study the enzyme-substrate specific interactions. To evaluate the substrate preference by CDA1, hexamer of xylan and chitin were used.

RESULTS AND DISCUSSION

The FASTA sequence of CDA1 enzyme from *Bacillus thuringiensis* was obtained by UniProt, the primary sequence analysis of CDA1 which exploited basic physico-chemical properties like isoelectric point, destruction coefficient, amino acid and atomic compositions (Table 1: Table 2). Proteins demonstrate their function when the proteins are folded properly to form appropriate three dimensional structures. The instability (Ii) and aliphatic indices are computed to be 23.73 and 93.6, respectively revealed the protein as stable for a wide range of temperature. Thus, information of the three-dimensional structure is essential for studying metabolic mechanism of the proteins. The 3D structure of CDA1 was predicted through homology modeling, the models generated were checked and assessed by ERRAT, this evaluation is based on an algorithm for protein structure validation, and it works by analyzing the statistics of non bonded interactions between different atom types. Crystal structures of 2J13 were used as structural templates for the homology modeling by MODELLER 9.9 program (Oberbarnscheidt et al., 2007). The best quality model structure was selected from the obtained model ensemble (Fig. 2). Furthermore, the model was evaluated for possible combinations of ψ and φ torsion angles using the Ramchandran plots (Fig. 1), the evaluation of the results have shown that the predicted CDA1 model generated by Modeller program is high quality.

To perfectly predict the binding activity of CDA1 either with chitin or xylan, the binding domain of CDA1 was firstly identified by Pocket-Finder, and then that particular region was selected to perform docking by

Table 1: Parameters computed using Expasy's ProtParam tool.

Amino acid Composition	Frequency	Percentage
Ala (A)	15	5.9 %
Arg (R)	8	3.1%
Asn (N)	15	5.9%
Asp (D)	15	5.9%
Cys (C)	0	0.0%
Gln (Q)	8	3.1%
Glu (E)	9	3.5%
Gly (G)	13	5.1%
His (H)	6	2.4%
Ile (I)	19	7.5%
Leu (L)	27	10.6%
Lys (K)	24	9.4%
Met (M)	2	0.8%
Phe (F)	14	5.5%
Pro (P)	7	2.8%
Ser (S)	26	10.2%
Thr 9T)	19	7.5%
Trp (w)	5	2.0%
Tyr (Y)	7	2.8%
Val (V)	15	5.9%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

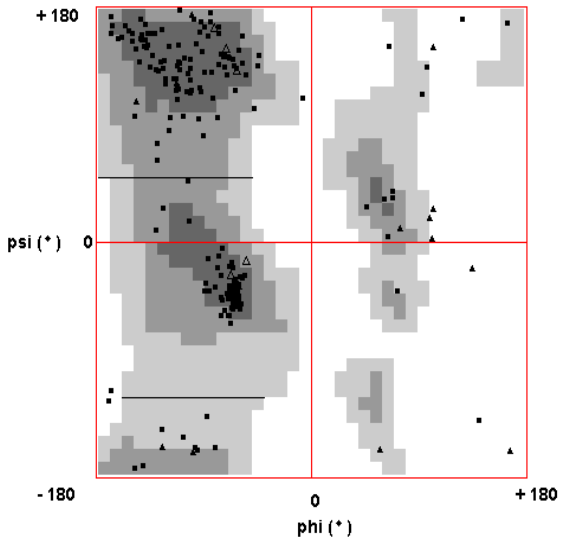


Fig. 1: Ramachandran Plot of CDA1

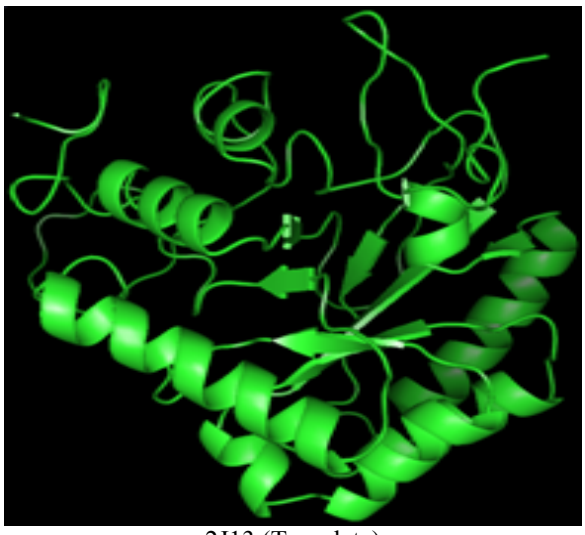
Presence of dark spots in the allowed region (Dark shaded area in each block) indicates the high quality prediction of the 3D structure.

Table 2: Parameters computed Expasy's ProtParam tool

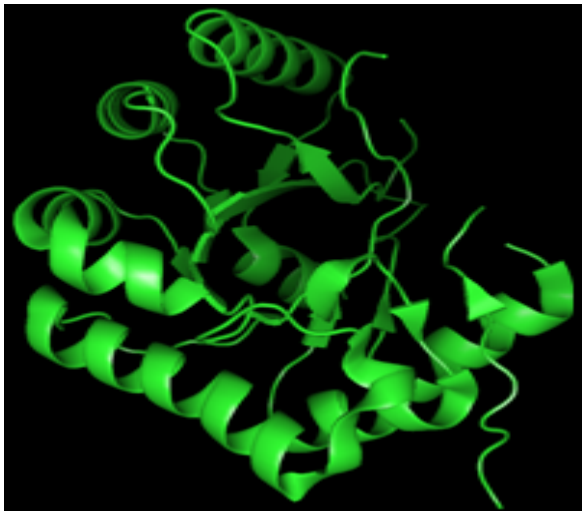
No. of Amino acids	254
Molecular weight	28522.6
Theoretical pI (pI)	9.46
- Charged residues	24
+ Charged residues	32
Formula	$C_{1298}H_{2044}N_{342}O_{378}S_2$
Extinction coefficient	37930 $m^{-1}cm^{-1}$
Instability index (Ii)	23.73
Aliphatic index (Ai)	93.66

Table 3: Binding energy parameters

With Chitin hexamer ligand			With Xylan hexamer ligand		
Affinity (Kcalmol^{-1})	Distance from RMSD l.b	Best mode RMSD u.b	Affinity Kcalmol^{-1}	Distance from RMSD l.b	Best mode RMSD u.b
-6.5	0.000	0.000	-6.3	0.000	0.000
-6.4	2.608	13.912	-5.6	7.527	12.750
-6.4	2.940	14.715	-5.4	21.553	24.822
-5.9	6.453	11.991	-5.4	2.923	11.540
-5.8	5.423	12.840	-5.2	23.949	27.376

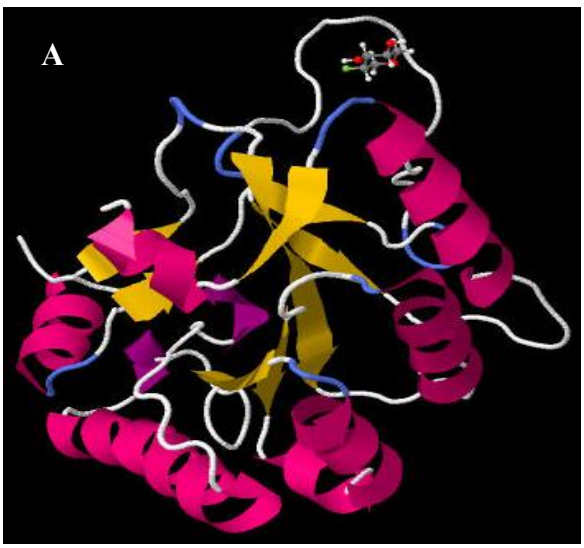


2J13 (Template)



CDA1 Model by Modeller (Target)

Fig. 2: Homology Modeling (Template vs. Target)
Figure shows the homology between the 3D structure of template and predicted structure of CDA1 enzyme.



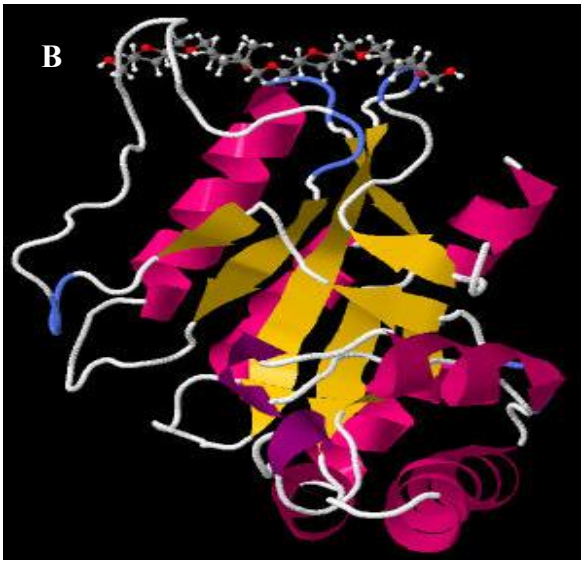


Fig. 3: (A); Docking of Xylan ligand with CDA1 (B); Docking of chitin hexamer ligand with CDA1
Figures represent docking of the ligand within active site of the enzyme. Pink represents the α-helices, yellow represents the β-sheets and gray represents the loops in the enzyme. Ligands are present at the top of the protein structure showing interaction.

AutoDock Vina. The best ligand binding pose with the least Glide/IFD score or energy was chosen taking into account the previous docking studies on enzymes of *E. coli* (Wojciechowski et al, 2005). Ligands selected for

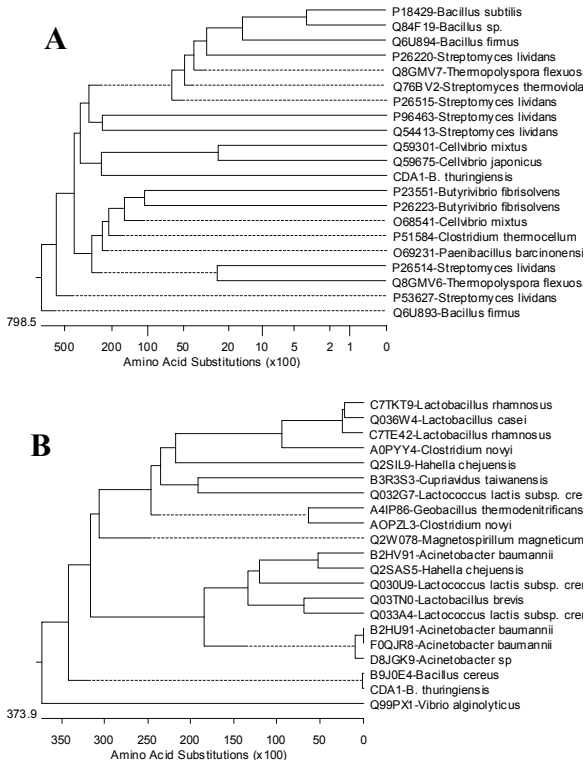


Fig. 4: Phylogetic analyses of the CDA1
The tree presents amino acid sequence based similarity of CDA1 with xylanase (A) and chitin deacetylase (B).

this purpose are a naturally occurring chitin product called N-Acetylchitohexaose ($C_{48}H_{80}N_6O_{31}$) and Xylan. The docking has shown strong binding of target protein with both chitin (Fig. 3A) and xylan (Fig. 3B) ligands, nevertheless the score was better in case of chitin ligand. AutoDock randomly dock the ligand with target protein structure at various orientations and gives a defined number of most favorable ligand-target complexes. In this case ten successful ligand-target complexes were generated with acceptable energy values out of which the best five are presented (Table 3), the most negative affinity value is consider as the best possible binding orientation between ligand and target. Infect, it represents the Gibbs free energy of the target sequence which is given by $\Delta G=-RTlnK$, as much as ΔG is low, the better will be the binding between ligand-target complex, for this experiment the best affinity got is -6.5 (K-cal/mol) and -6.3 (K-cal/mol) in case of chitin and xylan ligands respectively which proves the preferable binding of ligand against N-Acetylchitohexaose ligand. Hence this work proves that CDA1 is a chitin binding protein, the residues involved in the binding of CDA1 with chitin products are within the domain sequence, modeled by homology modeling and are well know.

To support this assumption we also performed multiple sequence alignment of target protein, specifically the catalytic domain with the catalytic domains of the chitin deacetylases and xylanases separately and constructed phylogenetic trees. We found more conserved residue and similarity of target protein (Q6HPL9) against chitin deacetylase genes (Q06702, O13842) as compared to xylanase genes **(**Q45518, P09850) of closely related species as shown in phylogenetic tree (Fig. 4A & 4B). The ClustalW2 phylogram also proves our assumption that although CDA1 is having similarity with xylanase gene but actually it is a functional homolog of chitin deacetylase.

We also compared our predicted 3D target protein structure with crystal structure of chitin deacetylase from the fungal pathogen *Colletotrichum lindemuthianum* (2IW0) and mesophilic xylanase A from *Bacillus subtilis* 1A1 (1XXN) available (Blair et al., 2006; Murakmi et al., 2005) in PDB (Fig. 5) and found our target protein to be structurally more similar to chitin deacetylase (2IW0) as compared to xylanase (1XXN). Like our predicted structure the 2IW0 also contain multiple alpha helix and beta sheets while 1XXN

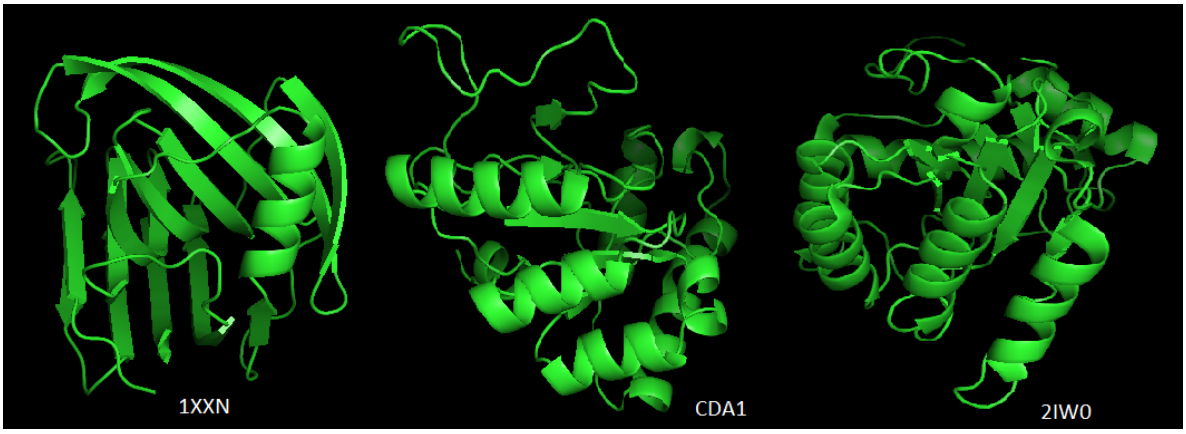


Fig. 5: Structural comparison of CDA1 (*B. thuringiensis*), 1XXN (*B. subtilis*) and 2IW0 (*C. lindemuthianum*)

contains only one alpha helix and more beta sheets (Murakmi et al., 2005), this comparison also supports our hypothesis that CDA1 is a chitin deacetylase but not a xylanase. Molecular characterization of the CDA1 enzyme through protein expression and enzymological studies may be conducted in the future to confirm the chitin deacetylase activity.

REFERENCES

Apweiler R, A Bairoch, CH Wu, WC Barker, B Boeckmann, S Ferro, E Gasteiger, H Huang, R Lopez, M Magrane, MJ Martin, DA Natale, C O'Donovan, N Redaschi and LS Yeh, 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Research, 32: 115-119.

Bendtsen JD, H Nielsen, G Heijne and S Brunak, 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology, 340: 783-795.

Blair DE, O Hekmat, AW Schüttelkopf, B Shrestha, K Tokuyasu, SG Withers and DMvan Aalten, 2006. Structure and mechanism of chitin deacetylase from the fungal pathogen *Colletotrichum lindemuthianum*. Biochemistry, 45: 9416-9426.

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

Buchan DW, SM Ward, AE Lobley, TC Nugent, K Bryson and DT Jones, 2010. Protein annotation and modelling servers at University College London. Nucleic Acids Research, 38: 563-568.

Finn RD, J Mistry, J Tate, P Coggill, A Heger, JE Pollington, OL Gavin, P Gunesekaran, G Ceric, K Forslund, L Holm, EL Sonnhammer, SR Eddy and A Bateman, 2010. The Pfam protein families database. Nucleic Acids Research, 38: 211-222.

Gasteiger E, C Hoogland, A Gattiker, S Duvaud, MR Wilkins, RD Appel and A Bairoch, 2005. *The Proteomics Protocols Handbook*, John M. Walker edition, Humana Press, pp: 571-607.

Laskowski RA, JA Rullmannn, MW Macarthur, R Kaptein and JM Thornton, 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of Biomoecular NMR, 8: 477-486.

Mehmood MA, X Xiao, FY Hafeez, Y Gai and F Wang, 2011. Molecular characterization of the modular chitin binding protein Cbp50 from *Bacillus thuringiensis* serovar *konkukian*. Antonie Van Leeuwenhoek Journal of Microbiology, 100: 445-453.

Mehmood MA, X Xiao, FY Hafeez, G Yingbao and F Wang, 2010. Molecular characterization of an endochitinase from *Bacillus thuringiensis* subsp. *konkukian*. World Journal of Microbiology and Biotechnology, 26: 2171-2178.

Mehmood MA, M Latif and FY Hafeez, 2012. Heterologous expression and haracterization of an antifungal chitinase Chi39 from *Bacillus thuringiensis* serovar *konkukian.* Pakistan Journal of Life and Social Sciences. 10: 116-122.

Murakami MT, Ruller R, Ward RJ and Arni RK, 2005. Crystallization and preliminary X-ray crystallographic studies of the mesophilic xylanase A from *Bacillus subtilis* 1A1. Acta Crystallogrphy Section F: Structural Biology and Crystallization Communication, 61: 219-220.

Oberbarnscheidt L, EJ Taylor, GJ Davies and TM Gloster, 2007. Structure of a carbohydrate esterase from *Bacillus anthracis*. Proteins, 66: 250-252.

Peitsch MC, 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. Biochemical Society Transactions, 24: 274-279.

Sehar U, MA Mehmood, S Nawaz, K Hussain, I Sohail, MR Tabassum, SS Gill and A Saqib, 2013. Three dimensional (3D) structure prediction and substrate-protein interaction study of the chitin binding protein CBP24 from *B. thuringiensis*. Bioinformation, 9: 725-729.

Tramontano A, R Leplae and V Morea, 2001. Analysis and assessment of comparative modeling predictions in CASP4. Proteins, 45: 22-38.

Trott O and AJ Olson, 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31: 455-461.

Tsigos I, A Martinou, D Kafetzopoulos and V Bouriotis, 2000. Chitin deacetylases: new, versatile tools in biotechnology. Trends in Biotechnology, 18: 305-312.

Vriend G, 1990. WHAT IF: a molecular modeling and drug design program. Journal of Molecular Graphics, 8: 52-56.

Wojciechowski M, S Milewski, J Mazerski and E Borowski, 2005. Glucosamine-6-phosphate synthase, a novel target for antifungal agents. Molecular modelling studies in drug design. Acta Biochimica Polonica, 52: 647-653.

Wu Y, G Preston and AE Bianco, 2008. Chitinase is stored and secreted from the inner body of microfilariae and has a role in exsheathment in the parasitic nematode *Brugia malayi*. Molecular and Biochemical Parasitology, 161: 55-62

Zhao Y, R Park, AA Riccardo and RAA Muzzarelli, 2010. Chitin Deacetylases: Properties and Applications. Marine Drugs, 8: 24-46.