Bioinformation by Biomedical Informatics Publishing Group *open access*
www.bioinformation.net **Hypothesis**

Comparative characterization of commercially important xylanase enzymes

Neelima Arora, Amit Kumar Banerjee, Srilaxmi Mutyala, Upadhyayula Suryanarayana Murty*

Bioinformatics Group, Biology Division, Indian Institute of Chemical Technology, Hyderabad-500007, A.P., India.
U.S.N Murty*-Email: murty_usn@yahoo.com, *Corresponding author

Received April 30, 2009; accepted June 07, 2009; published August 05, 2009

Abstract:
Xylanase is an industrially important enzyme having wide range of applications especially in paper industry. It is crucial to gain an understanding about the structure and functional aspects of various xylanases produced from diverse sources. In this study, a bioinformatics and molecular modeling approach was adopted to explore properties and structure of xylanases. Physico-chemical properties were predicted and prediction of motifs, disulfide bridges and secondary structure was performed for functional characterization. Apart from these analyses, three dimensional structures were constructed and stereo-chemical quality was evaluated by different structure validation tools. Comparative catalytic site analysis and assessment was performed to extract information about the important residues. Asn72 was found to be the common residue in the active sites of the proteins P35809 and Q12603.

Keywords: Xylanase, Bioinformatics, sequence analysis, structure prediction; function

Background:
Burgeoning rise in demand for industrial enzymes is anticipated to touch the mark of $5.1 billion soon. With the advent of technological advancements, the remarkable growth in biotechnological use of Xylanases has spurred new interest in these enzymes [**1, 2**]. Market trends reveal that xylanase and cellulose takes the major chunk of share amounting to 20% of the world enzyme market, together with pectinases [**3**]. Xylanases are glycosidases (O-glycoside hydrolases, EC 3.2.1.x) which degrade the linear polysaccharide beta-1, 4-xylan into xylose and produced by a myriad group of microrganisms belonging to varied genera and species of bacteria, actinomycetes and fungi [**2, 4, 5, 6**].

This group of enzymes has been attracting a lot of attention in the recent past due to its probable applicability in a spectrum of industrial processes [**1**]. Xylanases are mainly exploited in the Kraft process for the removal of the lignin-carbohydrate complexes [**7, 8, 9**]. Other important processes where xylanases are used frequently in extraction and preparation of beverages [**4**]; clarification of juices [**10**]; detergents [**11**]; generation of protoplast in plant cells [**12**]; production of pharmacologically active polysaccharides for use as antimicrobial agents [**13**] or antioxidants [**14**]; production of surfactants [**15**] and bioconversion of lignocellulosic materials to fuels. Broadly xylanases are classified under two classes: Family 10 (F) and Family 11 (G), based on hydrophobic cluster analysis and sequence homology [**16**, **17**, **18**]. Xylanases differ in their physicochemical properties, structures, specific activities, thermo stability and yields, thus providing a great deal of choice in their potential usage. In this paper, we report the *in silico* analysis and characterization studies on 8 xylanases from various organisms.

Methodology:
Xylanase protein sequences were retrieved from the SWISS-PROT, a public domain protein database [**19**].During the sequence retrieval process, key word 'Xylanase' was used. The database search yielded 76 xylanase protein sequences. Sequences representing putative, partial, precursor and fragment of Xylanase protein were excluded from the study. Hence, 8 unique proteins were retrieved and considered for this study (**Table 1 in supplementary material**). The selected xylanase protein sequences were retrieved in FASTA format and used for further analysis.

Physico-chemical characterization:
Theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient [**20**], instability index [**21**], aliphatic index [**22**] and grand average hydropathy (GRAVY) [**23**] were computed using the Expasy's ProtParam server [**24**] (http://us.expasy.org/tools/protparam.html) (Table 3 in **supplementary material**). Amino acid composition of the protein sequences can reveal their nature; hence, amino acid composition was also computed (**Table 2 in supplementary material**). PROTSCALE (http://us.expasy.org/tools/protscale.html) was used to calculate the number of codons, bulkiness, polarity, refractivity, recognition factors, hydrophobicity, transmembrane tendency, percent buried residues, percent accessible residues, average area buried, average flexibility, relative mutability and number of amino acids.

Secondary structure prediction:
SOPMA [**25**] was employed for calculating the secondary structural features of the selected target protein sequences considered for this study (Table 4 in **supplementary material**).

Functional characterization:
Disulphide bonds are important in determining the functional linkages, so, SS bonds were analyzed using the primary protein sequence data with the help of CYS_REC (**Table 7 in supplementary data**). CYS_REC identifies the positions of cysteines, total number of cysteines present and computes the most probable SS bond pattern of pairs in the protein sequence. Motifs in the considered sequences were scanned using Motif Search (Table 5 in **supplementary material**) [**26**]. SOSUI server [**27**] was used to predict the transmembrane tendency of the proteins considered for this study (**Table 6 in supplementary material**). Hydrophobicity score and plot was obtained using Kyte and Doolittle method keeping a window size of 7 (**Figure 1**).

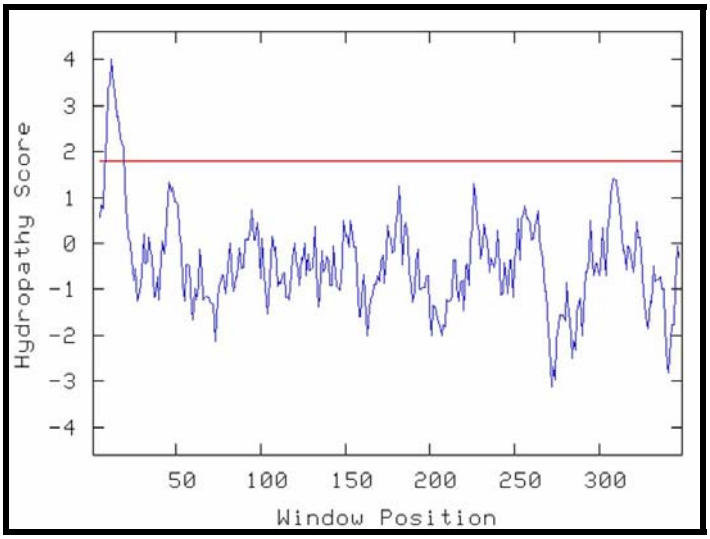

Figure 1: Kyte and Doolittle hydropathy plot for xylanase of *Dictyoglomus thermophilum*

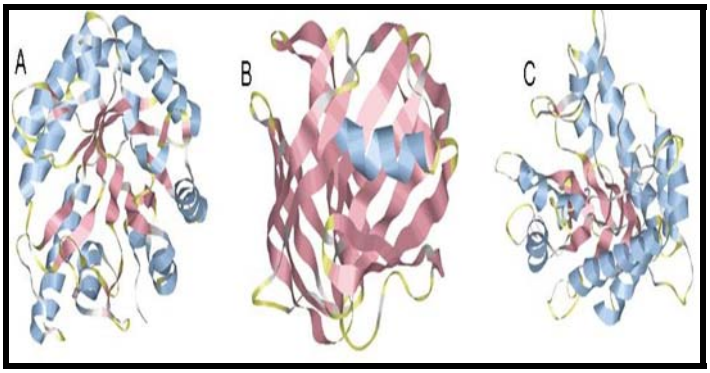

Figure 2: Modeled Structure of xylanase proteins (A) P35809 (B) Q12603 (C) P26223

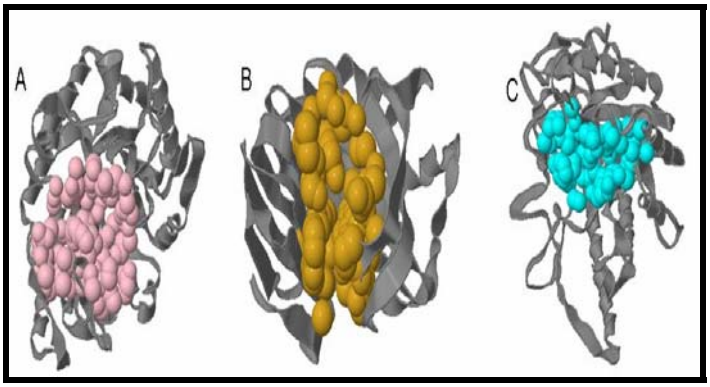

Figure 3: Active site of modeled proteins (A) P35809 (B) Q12603 (C) P26223 as predicted by CASTp server.

Tertiary structure prediction and structure Validation:
Since the crystal structures for P35809, Q12603, P26223 and P48791 are not available, SWISSMODEL [**28**] was used to model the 3D structure of these proteins based on the best template. The details of template and the criteria used for selection are listed in **Table 8 in supplementary material**. No model could be built for P48791 using the first approach mode of SWISSMODEL. Structure validation tools like ERRAT [**29**], PROVE [**30**], PROCHECK [**31**], WHATCHECK [**32**] and Verify 3D [**33**] were employed to evaluate the stereochemistry and quality of the models (**Table 9 in supplementary material**).

Active site analysis:
Possible catalytic sites were assessed and explored applying CASTp [**34**]. Out of many binding sites predicted, active site was selected on the basis of maximum surface area and volume. Important residues involved in active sites were identified and compared for the modeled proteins (**Table 10 in supplementary material**).

Discussion:
Amino acid composition determines the fundamental properties of the enzyme. The amino acid composition of xylanase sequences is represented in Table 2 (see **supplementary material)**. Isoelectric point (pI) is the pH at which net charge existing on the protein is zero. The pI values of all protein sequences are in the range of 4.78-8.71 indicating that all considered xylanase sequences are acidic except P56588 and P48793. The calculated isoelectric point (pI) will be useful because at pI, solubility is least and mobility in an electro focusing system is zero. The instability index which gives clue about the stability of a protein *in vitro* can be calculated using equation 1 (see supplementary material). All the considered sequences were classified as stable with value ranging from 13.57 to 37.23 as a value > 40 indicates an unstable protein.

The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains is regarded as a positive factor for the increase of thermal stability of globular proteins. It can be calculated by equation 2 **(see supplementary material).** Aliphatic index for the xylanase sequences ranged from 48.71- 87.76. The very high aliphatic index of all xylanase sequences indicates that these xylanases may be stable for a wide temperature range. From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine does not absorb appreciably at wavelengths >260 nm, while cystine does) at a given wavelength, the extinction coefficient of the native protein in water can be computed using equation 3 **(see supplementary material).** The computed protein concentration and extinction coefficients help in the quantitative study of protein–protein and protein–ligand interactions in solution. The Grand Average hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. GRAVY indices of xylanases are ranging from -0.608 to -0.173. This low range of value indicates the possibility of better interaction with water. The secondary structure indicates whether a given amino acid lies in a helix, strand or coil. Secondary structure features as predicted using SOPMA are represented in **Table 4 (see supplementary).** The results revealed that random coils dominated among secondary structure elements followed by alpha helix, extended strand and beta turns in P40942, P81536, P35809, P26223, P48791, and P48793 while alpha helix outnumbered random coils in Q12603, P56588. A set of conserved amino acid residues located in vicinity that provides clues to the functions is termed as motif. Motifs predicted using Motif Search is shown in **Table 5 (see supplementary material)**.

It was found that P40942 and P26223 contained Glycosyl hydrolases family 10 motif. Glycosyl hydrolases family 11 contains two signature motif viz signature 1 that spans upto 11 residues and signature 2 motif of l2 residues. The average length of the motif predicted was 11 in both Glycosyl hydrolases family 10. Motifs could not be predicted for Q12603 and P48791. SOSUI distinguishes between membrane and soluble proteins and predicts the transmembrane helices from amino acid sequences quickly with high precision. Xylanase from *Dictyoglomus thermophilum* was classified as membrane protein by SOSUI server while all other xylanases were predicted to be soluble proteins (**Table 7 in supplementary material**). The transmemebrane region predicted was found to be rich in hydrophobic amino acids and it is also evident in Kyte and Doolittle mean hydrophobicity profile generated using online tool (http://gcat.davidson.edu/rakarnik/kyte-doolittle.htm) (**Figure 1**) in which many points lie above the 0.0 line and a clear peak was observed in plot that indicates about the plausible transmembrane region. As disulphide bridges play an important role in determining the thermostability of these enzymes, CYS_REC was used to determine the Cysteine residues and disulphide bonds. CYS_REC predicted no Cystine residues in P40942 and P48793. Possible pairing and pattern with probability indicated by scores are presented in **Table 7 (see supplementary material**). Since there is lack of experimental structures for 4 Xylanases considered, SWISSMODEL was used to predict the 3D structures of proteins. 1YNA_A (*Thermomyces lanuginosus*), 1N82 (*Bacillus stearothermophilus*), 2F8Q (*Bacillus* sp.ng-27) were selected as templates from PDB database for P35809, Q12603 and P26223 respectively based on sequence identity (Table 8 in supplementary material). The final modeled structures are shown in **Figure 2.** The predicted structures were validated using various structure validation servers. 90.5%, 81.4%, 82.5% % of amino acids lie in the most favored regions of Ramachandran Plot as revealed by PROCHECK analysis for the structure modeled for P35809, Q12603 and P26223 respectively. The predicted structures conformed well to the stereochemistry indicating reasonably good quality and were used for further analysis. These structures will provide a good foundation for functional analysis in dearth of experimentally derived crystal structures. It is important to explore and characterize the active site of an enzyme for understanding its interactions with substrate. CASTp was used to investigate the possible

binding sites (**Table 10 in supplementary material** and **Figure 3**). The results indicate that Asn72 was present in active site of both P35809, Q12603 thus implying its essential role in enzyme–substrate interactions.

Conclusion:

For obtaining desirable results in industrial application, it is essential to manipulate the characteristic properties of enzyme which is a tedious task. Protein engineering techniques used to achieve this goal require a sound knowledge about the protein both at sequence and structure level. In this study, 8 xylanase sequences were selected to acquire an understanding about their physico-chemical properties and various protein structure levels by using *in silico* techniques. Primary structure analysis reveals that most of the xylanase under study are hydrophobic in nature and three of them contain disulphide linkages. Secondary structure analysis established that in most of the sequences, random coils dominated among secondary structure elements followed by alpha helix, extended strand and beta turns. Three dimensional structures were predicted for proteins where such data was unavailable and active sites were explored for determining important residues. This study will provide an insight about the physiochemical properties and function of xylanases and thus aid in formulating their uses in industries.

Acknowledgement:

The authors are grateful to Dr. J. S. Yadav, Director, Indian Institute of Chemical Technology for his constant support and encouragement during the study. AKB thanks Council of Scientific and Industrial Research (CSIR), Govt. of India, for the Senior Research Fellowship (SRF).

References:

[1] M.K. Bhat, *Biotech.Adv.* 18: 355–383(2000) [PMID: 14538100]
[2] Q.K. Beg *et al.*, *Appl. Microbiol. Biotechnol.* 56: 326–338 (2001) [PMID: 11548999]
[3] L.T.M. Polizeli *et al., Appl. Microbiol. Biotechnol.* 67:577–591(2005) [PMID: 15944805]
[4] K.K.Y. Wong *et al., Microbiol. Rev.*52: 305–317 (1988) [PMID: 3141761]
[5] R.C. Kuhad & A. Singh, *Crit. Rev. Biotechnol.* 13:151–172 (1993) [PMID:]
[6] R.C. Kuhad *et al., Adv. Biochem. Eng. Biotechnol.*57: 45–125 (1997) [PMID:]
[7] M.G. Paice *et al., Enzyme Microb. Technol.* 14: 272–276 (1992) [PMID:]
[8] J. Buchert *et al., Biores Technol.* 50: 65-72 (1994) [PMID:]
[9] M.L. Nikupaavola *et al., Biores Technol.* 50: 73-77 (1994) [PMID:]
[10] P. Biely, *Trends Biotechnol.* 3: 286-290(1985) [PMID:]
[11] B. K.Kumar *et al., J. Ind. Microbiol. Biotechnol.* 31(2): 83-87 (2004) [PMID: 14986151]
[12] N. Kulkarni *et al., FEMS Microbiol. Rev.*23: 411–456(1999) [PMID: 10422261]
[13] P. Christakopoulos *et al., Int. J. Biol. Macromol.* 31: 171–175(2003) [PMID: 12568925]
[14] P. Katapodis *et al., Eur. J. Nutr.* 42: 55–60(2003) [PMID: 12594542]
[15] S. Matsumura *et al., Biotechnol. Lett.* 21: 17–22 (1999) [PMID:]
[16] G. Davies & B. Henrissat, *Structure* 3: 853-859 (1995) [PMID: 8535779]
[17] B. Henrissat & A. Bairoch, *Biochem J.* 293,781–788(1993) [PMID: 8352747]
[18] Sapag *et al., J Biotechnol.* 95:109–131 (2002) [PMID: 11911922]
[19] Bairoch & R. Apweiler, *Nucleic Acids Res.* 28: 45-48(2000) [PMID: 10592178]
[20] S.C. Gill & P.H. Von Hippel, *Anal. Biochem.*182: 319-326(1989) [PMID: 2610349]
[21] K. Guruprasad *et al., Protein Engineering* 4(2): 155-161(1990) [PMID: 2075190]
[22] A.J. Ikai, *J.Biochem.* 88: 1895-1898 (1980) [PMID: 7462208]
[23] J. Kyte & R.F. Doolittle, *J. Mol. Biol.* 157: 105-132 (1982) [PMID: 7108955]
[24] E. Gasteiger et al., *Protein Identification and Analysis Tools on the ExPASy Server* (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press.571-607 (2005) [PMID:]
[25] Geourjon & G. Deleage, *Comput Appl Biosci.* 11(6): 681-684 (1995) [PMID: 8808585]
[26] L. Falquet *et al., Nucl. Acids Res.* 30(1): 235-238 (2002) [PMID: 11752303]
[27] T. Hirokawa *et al., Bioinformatics* 14:378-379 (1998) [PMID: 9632836]
[28] N. Guex & M.C. Peitsch, *Electrophoresis* 18: 2714-2723 (1997) [PMID:]
[29] Colovos & T.O. Yeates, *Protein Sci.* 2: 1511-1519 (1993) [PMID: 8401235]
[30] J. Pontius *et al., J. Mol. Biol.* 264: 121–136 (1996) [PMID: 8950272]
[31] R.A. Laskowski *et al., J. Appl. Cryst.* 26:283-291 (1993) [PMID:]
[32] R.W.W. Hooft *et al., Nature* 381: 272 (1996) [PMID: 8692262]
[33] Eisenberg *et al., Methods Enzymol.* 277:396-404 (1997) [PMID: 9379925]
[34] J. Dundas *et al., Nucl. Acids Res.* 34:W116-W118 (2006) [PMID: 16844972]

Edited by P. Kangueane
Citation: Arora *et al.,* Bioinformation 3(10): 446-453 (2009)

Bioinformation by Biomedical Informatics Publishing Group *open access*
www.bioinformation.net **Hypothesis**

Supplementary material

Equation 1

$$II = (10/L) * \sum_{i=1}^{i=L-1} DIWV(x(i)x(i+1))$$

Where, L denotes length of sequence, DIWV(x (i) x (i+1)) is the instability weight value for the dipeptide starting in position i.

Equation 2

Aliphatic index = X (Ala) + a*X (Val) + b*X (Leu) + b*X (Ile)
X (Ala), *X (Val)*, *X (Ile)* and *X (Leu)* are the amino acid compositional fractions.

Equation 3

E (Prot) = N (Tyr)*Ext (Tyr) + N (Trp)*Ext (Trp) + N (Cystine)*Ext (Cystine)
Where (for proteins in water measured at 280 nm): N= number, Ext (Tyr) = 1490, Ext (Trp) = 5500, Ext (Cystine) = 125; Extinction coefficients of considered xylanases at 280 nm range from 57995 to 121630 M^{-1} cm^{-1} assuming all cysteine residues appear as half cystines. The high extinction coefficient of some sequences indicates presence of high concentration of Cys, Trp and Tyr.

Table 1: Xylanase sequences considered for the study

S. No.	Accession No.	Length	Description	Function	Organism
1	P40942	387	Thermostable celloxylanase	Active toward xylan, carboxymethylcellulose, P-nitrophenyl-beta-D-xylopyranoside and P-nitrophenyl-beta-D-cellobioside	*Clostridium stercoarium*
2	Q12603	352	Beta-1,4-xylanase	-	*Dictyoglomus thermophilum*
3	P81536	194	Endo-1,4-beta-xylanase	-	*Paecilomyces variotii*
4	P56588	302	Endo-1,4-beta-xylanase	-	*Penicillium simplicissimum*
5	P35809	197	Endo-1,4-beta-xylanase A	Hydrolyzes xylans into xylobiose and xylose	*Schizophyllum commune*
6	P26223	635	Endo-1,4-beta xylanase B	B.fibrisolvens is located in the rumen of ruminant animals, where it contributes to the animal's digestion of plant material by hydrolyzing hemicellulose with its xylanases	*Butyrivibrio fibrisolvens*
7	P48791	319	Beta-xylosidase	Exoxylanase capable of acting on certain xylans and xylooligosaccharides.	*Prevotella ruminicola*
8	P48793	190	Endo-1,4-beta-xylanase	-	*Trichoderma harzianum*

Table 2: Amino acid composition of considered xylanase sequences

A.A. composition	P40942	Q12603	P81536	P56588	P35809	P26223	P48791	P48793
Ala	0.067	0.057	0.046	0.109	0.061	0.076	0.06	0.047
Arg	0.049	0.037	0.031	0.023	0.02	0.043	0.019	0.032
Asn	0.059	0.051	0.072	0.079	0.066	0.054	0.019	0.1
Asp	0.075	0.06	0.052	0.06	0.041	0.065	0.11	0.021
Cys	0	0.003	0.01	0.007	0.01	0.014	0.022	0
Gln	0.031	0.02	0.036	0.026	0.036	0.022	0.025	0.032
Glu	0.07	0.085	0.026	0.026	0.025	0.079	0.06	0.021
Gly	0.059	0.057	0.06	0.086	0.052	0.054	0.088	0.142
His	0.013	0.034	0.021	0.02	0.015	0.028	0.041	0.021
Ile	0.078	0.065	0.031	0.07	0.041	0.052	0.044	0.053
Leu	0.088	0.088	0.036	0.073	0.036	0.074	0.053	0.026
Lys	0.07	0.097	0.01	0.066	0.025	0.06	0.078	0.021
Met	0.026	0.017	0	0.01	0.005	0.031	0.022	0.005
Phe	0.047	0.054	0.026	0.03	0.015	0.06	0.047	0.037
Pro	0.054	0.051	0.031	0.03	0.036	0.049	0.066	0.032

Ser	0.047	0.023	0.013	0.086	0.042	0.058	0.041	0.126
Thr	0.049	0.034	0.008	0.063	0.002	0.054	0.034	0.089
Trp	0.023	0.034	0.041	0.023	0.041	0.019	0.025	0.032
Tyr	0.041	0.057	0.088	0.043	0.086	0.058	0.075	0.095
Val	0.054	0.077	0.062	0.07	0.046	0.052	0.072	0.068
Pyl	0	0	0	0	0	0	0	0
Sec	0	0	0	0	0	0	0	0

Table 3: Parameters computed using Expasy's ProtParam tool

Gene ID	**P40942**	**Q12603**	**P81536**	**P56588**	**P35809**	**P26223**	**P48791**	**P48793**
No. of codon	3.4445	3.6665	3.611	3.8885	3.5	3.278	3.3335	3.6665
Bulkiness	14.5485	17.1665	13.943	13.7185	13.7905	14.44	14.4965	13.9925
Polarity	17.518	16.8865	12.1815	14.0755	9.2735	17.2795	17.0295	12.4295
Refractivity	15.9255	18.432	15.3205	13.1035	14.8145	18.749	18.3295	16.5105
Recognition factors	89.2775	86.889	92	92.0555	90.111	87.3335	87.944	90.5555
Hydrophobicity	-0.639	0.439	-0.822	-0.1725	-0.4835	-0.6225	-0.0555	-0.639
Transmembrane tendency	-0.69	-0.1735	-0.618	-0.355	-0.545	-0.783	-0.492	-0.3955
% buried residues	5.4665	6.683	6.7275	7.189	6.5	5.572	6.556	6.2
% accessible residues	5.5555	6.05	6.2555	5.7445	6.0665	5.9835	5.2885	5.9555
Average Area buried	129.622	136.333	118.5775	116.7165	117.6555	127.0555	129.628	125.661
Average flexibility	0.4365	0.4385	0.444	0.447	0.447	0.4385	0.4305	0.438
Relative mutability	81.1665	73.778	79.778	78.3335	75.389	74	69.1665	81.722
No. of A.A.	387	352	194	302	197	635	319	190
M.W	44377.5	41485.6	20947.3	32551	20978.5	73184.6	36406.7	20703.4
pI	5.24	6.33	4.78	7.87	5.16	5.19	4.82	8.71
"-" charged residues	56	51	15	26	13	91	54	8
"+" charged residues	46	47	8	27	9	65	31	10
Extinction coefficients	73340	95800	69455	57995	69455	121630	80135	59820
Instability index	36.69	28.95	27.83	21.82	28.13	37.23	33.78	13.57
Aliphatic index	86.95	87.76	48.71	86.62	49.04	71.76	64.76	55.37
GRAVY	-0.403	-0.415	-0.608	-0.173	-0.58	-0.426	-0.546	-0.499

Table 4: Calculated secondary structure elements

Secondary structure	**P40942**	**Q12603**	**P81536**	**P56588**	**P35809**	**P26223**	**P48791**	**P48793**
Alpha helix	37.21%	41.19%	6.19%	40.40%	6.09%	33.70%	8.15%	5.79%
310 helix	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Extended strand	14.99%	16.19%	36.60%	18.21%	36.04%	19.84%	27.90%	37.37%
Beta turn	6.72%	6.25%	9.79%	7.28%	10.66%	10.55%	6.90%	12.11%
Bend region	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Random coil	41.09%	36.36%	47.42%	34.11%	47.21%	35.91%	57.05%	44.74%
Ambigous states	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 5: Details of Motifs predicted in considered sequences

Protein ID	**Motif Found**	**Motif ID**	**Description**	**Start**	**End**
P40942	GLYCOSYL_HYDROL_F10	PS00591	Glycosyl hydrolases family 10 active site.	286	296
Q12603	No motif was found	-	-	-	-

P81536	GLYCOSYL_HYDROL_F11_2	PS00777	Glycosyl hydrolases family 11 active site signature 2.	175	186
	GLYCOSYL_HYDROL_F11_1	PS00776	Glycosyl hydrolases family 11 active site signature 1.	83	93
P56588	GLYCOSYL_HYDROL_F10	PS00591	Glycosyl hydrolases family 10 active site.	231	241
P35809	GLYCOSYL_HYDROL_F11_2	PS00777	Glycosyl hydrolases family 11 active site signature 2.	181	192
	GLYCOSYL_HYDROL_F11_1	PS00776	Glycosyl hydrolases family 11 active site signature 1.	84	94
P26223	GLYCOSYL_HYDROL_F10	PS00591	Glycosyl hydrolases family 10 active site.	248	258
P48791	No motif was found	-	-	-	-
P48793	GLYCOSYL_HYDROL_F11_2	PS00777	Glycosyl hydrolases family 11 active site signature 2.	174	185
	GLYCOSYL_HYDROL_F11_1	PS00776	Glycosyl hydrolases family 11 active site signature 1.	83	93

Table 6: Prediction of hydrophobicity and transmembrane range for considered sequences

Accession No.	Average hydrophobicity	Type of Protein	Transmembrane range	Length	Type
P40942	-0.403359	Soluble	-	-	-
Q12603	-0.415057	Transmembrane	RFSILVLLLILL TFSLGFLKEEA	27	Primary
P81536	-0.608248	Soluble	-	-	-
P56588	-0.172848	Soluble	-	-	-
P35809	-0.579696	Soluble	-	-	-
P26223	-0.425669	Soluble	-	-	-
P48791	-0.546081	Soluble	-	-	-
P48793	-0.499474	Soluble	-	-	-

Table 7: Details of patterns of Cystine–Cystine binding

Accession No.	Cystine residues	Status	Score	Probable pattern
P40942	0	_	_	_
Q12603	1	CYS 174 is not SS-bounded	Score= -35.7	-
P81536	2	CYS 110 is SS-bounded	Score= 75.4	-
		CYS 154 is SS-bounded	Score= 60.3	
P56588	2	CYS 256 is SS-bounded	Score= 61.4	-
		CYS 262 is SS-bounded	Score= 48.3	
P35809	2	CYS 111 is SS-bounded	Score= 27.9	-
		CYS 160 is SS-bounded	Score= 19.8	
P26223	9	CYS 104 is probably not SS-bounded	Score= -11.3	The most probable pattern of pairs: 40-404, 104-417, 414-491, 416-565
		CYS 325 is not SS-bounded	Score= -25.3	
		CYS 404 is probably not SS-bounded	Score= -9.5	
		CYS 414 is probably not SS-bounded	Score= -13.2	
		CYS 416 is probably not SS-bounded	Score= -13.2	
		CYS 417 is probably not SS-bounded	Score= -8.7	
		CYS 491 is not SS-bounded	Score= -16.6	
		CYS 565 is not SS-bounded	Score= -18.3	
P48791	7	CYS 37 is probably not SS-bounded	Score= -10.8	The most probable pattern of pairs: 37-290, 67-246
		CYS 54 is not SS-bounded	Score= -31.5	
		CYS 67 is not SS-bounded	Score= -18.1	
		CYS 88 is probably not SS-bounded	Score= -14.2	
		CYS 138 is not SS-bounded	Score= -24.7	
		CYS 246 is probably SS-bounded	Score= 2.3	
		CYS 290 is probably not SS-bounded	Score= -9.8	
P48793	0	_	_	_

Table 8: Details of template selected for selected Xylanase sequences

Accession No.	Template (resolution)	Modeled residue range	Identity with target	E-value
P35809	1YNA_A (1.55 Å)	3 to 196	61.34	2.39e-51
Q12603	1N82_A(1.45 Å)	31 to 351	48.012	0.00e-1
P26223	2F8Q (2.20 Å)	2 to 335	33.333	0.00e-1

Table 9: Validation of modeled protein structure using different tools

Server	Property	Q12603	P35809	P26223
ERRAT	Overall Quality factor	72.524	84.409	49.080
PROVE	Z-Score Mean	1.682	3.096	-0.026

	Z-Score std deviation	37.731	52.180	1.768
	Z−score RMS	37.755	52.337	1.767
PROCHECK	Residues in most favored regions	90.5%	81.4%	82.5%
	Residues in additional allowed regions	9.2%	17.3%	14.9%
	Residues in generously allowed regions	0.4%	1.3%	2.3%
	Residues in disallowed regions	0.0%	0.0%	0.3%
WHATCHECK	RMS Z-score for bond lengths	0.788	0.723	0.848
	RMS Z-score for bond angles	1.138	1.010	1.469
	Omega angle restraints	1.003	1.203	1.014
	Side chain planarity	1.562	1.249	2.953
	Improper dihedral distribution	1.492	1.292	1.657
	Inside/Outside distribution	1.012	0.978	1.128
	2nd generation packing quality	-0.839	-1.346	-2.052
	Ramachandran plot appearance	-0.618	-1.488	-1.863
	chi-1/chi-2 rotamer normality	1.913	-0.593	0.834
	Backbone conformation	-3.683	-5.960	-8.812
Verify 3D	(3D-1D score > 0.2)	95.03%	100.0%	91.94%

Table 10: Total number of binding sites, selected active site and residues involved

Accession No.	Total number of Binding sites	Area of selected Binding site	Volume of selected Binding Site	Residues involved in active site
P35809	68	743.3	1268.8	E71, N72, K75, H108, , V112, H113, I114, Q115, N154, E155, D159, R166,Q231, H233, W234, W312, Y318, T319, ,W320, l321, y322, f323, W324, P325, V326, R327, R329, D331
Q12603	31	562.8	894.5	T7, S16, W18, N45, V47, G48,Y94, N72, S73, Y74, Y78, W80, E87, Y89, Y97, P99, R123, P127, S128, I129, F135, Q137, W139, ,Y178, I180, A182, E184, Y186,
P26223	60	670	937.3	K45, W94, H90, N149, E150, Q197, A196, T195, E194, Y193, M223, Q224, S225, H226, L227, L228, H231, P232, W299, A240, E255, D257, M258, W307, L308, F311, R312, S316,