JOURNAL OF PURE AND APPLIED MICROBIOLOGY, Nov 2015. Vol. **9**(Spl. Edn. 2), p. 321-326

Characterization of Xylanase Protein Sequences of *Bacillus cereus*: An In-Silico Study

Asish Mandal

Post Graduate Department of Botany, Ramananda College, Bishnupur, Bankura - 722 122, India.

(Received: 21 March 2015; accepted: 15 May 2015)

The use of hemicellulose enzymes is recently gained considerable interest in different industries. The challenges in incorporating the enzyme to the industrial sectors are to have an active and stable form of it at high temperature and wide range of pH conditions. In this study xylanase protein sequences of 10 strains of *Bacillus cereus* were retrieved from the database of NCBI. These sequences were analysed by computational methods for different physical and chemical properties, multiple sequence alignment, phylogenetic tree construction to predict the final motif and the evolutionary relationship among them. The multiple sequence alignment of these xylanase protein sequences showed conserved regions at different stretches with maximum homology from amino acid residues 123-129, 147-154, 178-182 which could be used for designing probes specific for xylanase producing strains.

Key words: *Bacillus cereus,* xylanases, motif, protein designing.

Xylan, the major component of hemicellulose, is composed of a linear backbone of 1, 4-β-linked-D-xylopyranosyl units that often has side chains of arabinosyl and methylglucuronyl substituents[1]. Xylanases play a key role in xylan hydrolysis into xylo-oligosaccharides. Xylanases are extra-cellular enzymes produced by microorganisms such as bacteria, fungi and some yeasts[2]. The enzyme is also found in protozoa, insects, crustaceans, snails, seaweeds and also some seeds of plants during the germination phase in the soil[3]. Microbial xylanases basically consist of two enzymes i.e. endo 1-4-β-D xylan-xylanohydrolase (EC-3.2.1.8), which primarily cleave β-1, 4 linked xylan backbone and β-xylosidase (EC-3.2.1.37), which hydrolyses xylooligomers[4]. This microbial enzyme has tremendous industrial applications. The most important one is in the process of biobleaching and biopulping. In these processes xylanase hydrolyses xylan and facilitates release of lignin from paper pulp and thereby reduces the use of chlorine[2, 5]. The other uses are in the saccharification of xylan from agro-wastes and agro-foods that intensify the preparation of bio-fuels and will be the potential use of this enzyme in modern biotechnology[6, 7]. Xylanases are also widely applied in food, animal feed[8] and have a worldwide market of around $ 200 million each year[9]. Along with cellulase and pectinase it occupies about 20% of global enzyme market[5]. However, high cost and low yields of xylanase from microbes have been the main constrains for its industrial exploitation. In industries, bacterial xylanases are more fascinating for their more alkali tolerance and thermostability than fungal sources[2]. Higher levels of xylanase activity at alkaline pH and at high temperature are reported mainly from *Bacillus* species[7].

The production, purification and biochemical characterization of xylanase have been extensively studied. Few number of xylanase genes have been cloned and sequenced. The motif

* To whom all correspondence should be addressed.
Tel.: +91-9434242741;
E-mail: mandalasish71@gmail.com

specificity of a set of protein sequences indicates a fundamental relationship among them. This study reports in silico analysis of a set of xylanase protein sequences from the bacterial source organisms for physico chemical analysis, super family search, phylogenetic tree construction, multiple sequence alignment, homology search and motif analysis using various bioinformatics tools to find out the functional motif.

MATERIALS AND METHODS

Sequence retrieval

A total of 10 xylanase protein sequences (minimum 97% identity) of *Bacillus cereus* were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/).

Physiochemical parameter determination

Physiochemical properties of selected protein sequences were determined by ProtParam software of ExPASy server[10]. Amino acid number, molecular weight (kilodalton), pI value, instability index, aliphatic index and grand average hydropathicity (GRAVY) of these protein sequences were calculated by using the tool.

Sequence alignment and phylogenetic tree construction

Clustal W2[11] was used for multiple sequence alignment and viewed in CLC –Bio sequence viewer. Phylip-3.69[12]. was used for phylogram construction by Nighbor-Joining (NJ) method using 100 bootstrap values. The dendrogram was edited by Dendroscope[13].

Protein family and superfamily search

SUPERFAMILY (SCOP domain searching tool based on Hidden Markov Model library [10] was used to determine the superfamily and family of selected xylanase protein sequences.

Motif finding

Pfam[14] (http.//www.sanger.ac.uk/software/pfam/search.html) was used for conserve domain finding. BLOCK MAKER (htttp://block.fhcrc.org/blocks/blockmkr/make_blocks.html)was used to identify protein blocks and MEME Suite (http://meme-suite.org/tools/meme) was used for motif finding.

Transmembrane regions searching

TMHMM (http:www.cbs.dtu.dk/service/TMHMM-2.0) was used for the searching of transmembrane regions of the xylanse protein sequences.

Table 1. Physiochemical parameters of xylanase protein sequences from *Bacillus cereus*

Sl no.	Accession no.	No. of amino acids	Molecular weight (kilodalton)	Theoretical pI	Total number of negatively charged residues (Asp + Glu)	Total number of positively charged residues (Arg + Lys)	Instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
1	KGT44235.1	234	26635.6	9.25	24	30	31.86	93.33	-0.324
2	KGT44347.1	275	31687.8	6.32	29	27	32.50	71.96	-0.485
3	KGT45614.1	360	40687.2	6.82	52	51	31.21	70.97	-0.616
4	KIZ27252.1	275	31654.8	6.60	28	27	34.11	73.75	-0.457
5	EEK60769.1	299	34628.9	8.66	39	42	32.79	84.11	-0.485
6	EEK61058.1	244	28170.4	8.75	29	32	27.74	89.84	-0.352
7	EEK61587.1	275	30824.2	6.31	35	31	34.96	82.95	-0.360
8	EEK62498.1	273	30677.2	8.47	31	33	46.06	82.20	-0.362
9	EEL15731.1	299	34634.9	8.85	38	42	33.08	82.81	-0.479
10	EEL16060.1	241	27742.0	9.22	26	32	28.14	91.78	-0.327

RESULTS AND DISCUSSION

All of the 10 *Bacillus cereus* xylanase proteins have some different amino acid sequences. The accession numbers of protein sequences are listed in Table 1. As all of the protein sequences are taken from the same species they are belonging to one family and one super family (Table 2).

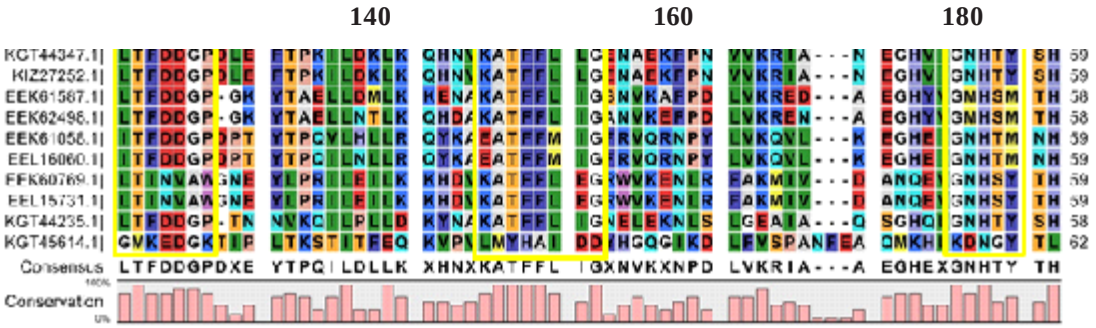

Fig. 1. Conserved regions (123-129, 147-154 and 178-182) of xylanase protein sequences from selected *Bacillus cereus*

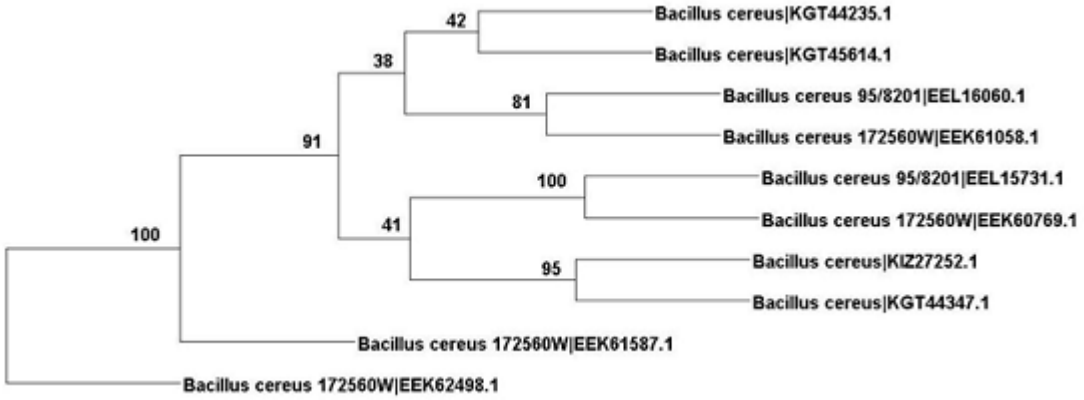

Fig. 2. Phylogenetic tree of xylanase protein sequences from *Bacillus cereus* strains

Table 2. Superfamily and Family of xylanase protein sequences from *Bacillus cereus*

Accession no.	Superfamily	Family	Domain region
KGT44235.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	6-229
KGT44347.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	41-268
KGT45614.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	162-332
KIZ27252.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	41-268
EEK60769.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	79-294
EEK61058.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	40-241
EEK61587.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	65-264
EEK62498.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	63-264
EEL15731.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	84-294
EEL16060.1	Glycoside hydrolase/deacetylase	NodB-like polysaccharide deacetylase	37-238

Table 3. Motif froms 7 xylanase protein sequences

Accession No.	Motifs A (width = 13)	Motifs B (width = 22)	Motifs C (width = 11)	Motifs D (width = 14)
EEK61058.1	NEKIIAITFDDGP	YTPQVLHLLRQYKAEATFFMIG	GHEIGNHTMNH	VKEPLLFRPPGGYI
EEK61587.1	VRKVAYLTFDDGP	YTAELLDMLKKENAKATFFLIG	GHYVGMHSMTH	GKSPVLTRPSYGSM
EEK62498.1	ERKVAYLTFDDGP	YTAELLNTLKQHDAKATFFLIG	GHYVGMHSMTH	GKSPKLTRPPYGSM
EEL16060.1	NEKIIAITFDDGP	YTPQILNLLRQYKAEATFFMIG	GHEIGNHTMNH	VKEPLLFRPPGGYI
KGT44235.1	NEKVIALTFDDGP	NVKQILPLLDKYNAKATFFLIG	GHQLGNHTYSH	FTGEIDFRPPNGKX
KGT44347.1	NKAEVALTFDDGP	FTPKILDKLKQHNVKATFFLLG	GHVIGNHTYSH	GYAPKFIRPPYGEI
KIZ27252.1	NKAEVALTFDDGP	FTPKILDKLKQHNVKATFFLLG	GHVIGNHTYSH	GYAPKFIRPPYGEI

Table 4. Functional similarity of 4 different motifs with the existing sequence data in the NCBI database

Motif No.	Description of matched sequence from BLAST	Maximum score	Total score	Query coverage	E- value	Identity	Accession
1	Polysaccharide deacetylase [*Bacillus thuringiensis*serovar sotto str. T04001]	45.2	45.2	100%	3e-04	100%	EEM40786.1
2	Chitooligosaccharidedeacetylase [*Bacillus cereus*]	76.6	76.6	100%	2e-14	100%	WP_048544063.1
3	Polysaccharide deacetylase [*Bacillus thuringiensis*serovar sotto str. T04001]	40.9	40.9	100%	0.006	100%	EEM40786.1
4	Peptidoglycan N-acetylglucosaminedeacetylase	46.9	46.9	100%	1e-04	100%	WP_042876839.1

Table 5. Transmembrane regions of selected xylanase protein from *Bacillus cereus*

Accession no.	N - terminal	C - terminal	Transmembrane region	Length
KGT44235.1	5	23	IIITIVTLFFIITALFGTY	19
KGT44347.1	-	-	-	-
KGT45614.1	-	-	-	-
KIZ27252.1	-	-	-	-
EEK60769.1	5	24	ILAYICIFSLYVSLGSYSVF	20
EEK61058.1	9	26	IFLFVFSLLCAVHIFQVE	18
EEK61587.1	-	-	-	-
EEK62498.1	7	29	IKQIVVVLIAIAAVAIGYYMFQS	23
EEL15731.1	5	24	ILAYICIFFLYVSVGSYSVF	20
EEL16060.1	-	-	-	-

Physiochemical parameter analysis

The physiochemical parameter for these bacterial protein sequences are listed in Table 1. The total number of amino acid residues ranged from 234 to 360 and the molecular weights ranged from 26635.6 to 40687.2, pI values range from 6.31 to 9.25. Negatively charged (aspartic acid and glutamic acid) and positively charged residues (arginine and glycine) and grand average hydropathicity (GRAVY) of these protein sequence showed considerable variability. The alipathic indices of these proteins have a simple range of 82±10. The measurement of the relative volume occupied by the alipathic amino acid residues (alanine, valine, leucine, and isoleucine) are directly proportional to index value. The index may be regarded as a positive factor for the increase of thermostabiity of globular proteins[15]. The Instability index[16] is a measure of proteins, used to determine whether it will be stable in a test tube. If the index is less than 40, then it is probably stable in the test tube[17]. If it is greater than 40, it is probably not stable. In the present study the instability indices are lower than 40 except EEK62489. So most of them are considered to have greater half-lives.

Multiple sequence alignment and homology search

The multiple sequence alignment of these proteins showed conserve regions with maximum homology from 123-129, 147-154 and 178-182 (Fig. 1). These regions could have role in conformation of the proteins and could be used for designing degenerate primers or probes for PCR based amplification of xylanse sequence from the source organism[18].

Superfamily and family search and phylogenetic tree construction

All of the organisms showed similarity with Glycoside hydrolase/deacetylase superfamily and NodB-like polysaccharide deacetylase family (Table 2). As two sequence sets such as EEK60769.1 & EEL15731.1 on the other hand KGT44235.1 & KGT45614.1 showed some sequence level dissimilarity in the multiple sequence alignment, they were found to create two different clusters during phylogenetic tree construction, having bootstrap value of 100 and 42 respectively (Fig. 2).

Motif sequence analysis

High level sequence similarity was observed for a total of seven sequences except KGT45614.1, EEK60769.1 andEEL15731.1. This result was also reflected in conserved motif finding. Four conserved motifs were found for all the seven sequences (Table 3). Functional similarities were deduced for all the four motif sequences using Protein BLAST (Table 4).

Among the 10 selected protein sequences 5 contain transmembrane regions (Table 5). These regions have less similarity. Generally xylanses are extracellular proteins (Beg et al.2001) and these portion probably have less importance on the mode of action of the enzymes. The consensus region is found in the regions 123-129, 147-154 and 178-182 and from these regions specific primmer can be designed for individual xylanase producing bacteria.

CONCLUSION

The present study revealed the motif based similarity at sequence level and can be utilized for designing the strategy for cloning of xylanase genes. The whole protein sequence based phylogenetic trees focuses the relation among the organisms. Physicochemical parameters can be used to predict the molecular nature of the enzyme for industrial exploitation.

ACKNOWLEDGEMENTS

The author is grateful to the BIF centre, Department of Microbiology, Vidyasagar University, West Bengal India, for providing the computation facilities.

REFERENCES

1. Mandal, A., Kar, S., Das Mohapatra, P. K., Maity, C., Mondal, K. C., Pati, B. R. Xylanase production under submerged fermentation by newly isolated *Bacillus cereus* BSA1: parametric optimization of cultural conditions. *J. Pure Appl. Microbiol.*, 2008; **2**(1):155-60.
2. Beg, Q. K., Kappor, M., Mahajan, L., Hoondal, G. S. Microbial xylanases and their industrial application: a review. *Appl. Microbiol. Biotechnol.*, 2001; **56**: 326-38.
3. Wong, K. K. Y., Tan, L. U. L., Saddler, J. N. Multiplicity of â-1,4- xylanase in microorganisms, functions and applications. *Microbiol.*, 1988; **52**: 305-17.
4. Liu, C., Sun, Z. T., Du, J. H., Wang, J. Response surface optimization of fermentation conditions for producing xylanase by *Aspergillus niger* SL-05. *J. Ind. Microbiol. Biotechnol.*, 2008; : 703-11.
5. Polizeli M. L. M., Rizzatti, A. C. S., Monti, R., Trenzi, H. F., Jorge, J. A., Amorin, D. S. Xylanases from fungi: properties and industrial applications. *Appl. Microbiol. Biotechnol.*, 2005; **67**: 577-91.
6. Bernier, R. J., Michel, D., Lubomir, J., Michael, G. P. Isolation and characterization of a xylanase from *Bacillus subtilis*. *Appl. Environ. Microbiol.*, 1983; **46**: 511-4.
7. Subramaniyan, S., Prema, P. Biotechnology of microbial xylanases: enzymology, molecular biology, and application. *Critical Rev. Biotechnol.*, 2002; **22**: 33-64.
8. Subramaniyan, S. P., Prema, P., Ramakrishna, S. V. Isolation and screening for alkaline thermostable xylanases. *J. Basic. Microbiol.*, 1997; **37**: 431-7.
9. Katapodis, P., Christakopoulou, V., Kekos, D., Christakopoulos, P. Optimization of xylanase production by *Chaetomium thermophilum* in wheat straw using response surface methodology. *Biochem. Eng. J*., 2007; **35**:136-41.
10. Gough J., Karplus K., Hughey R., Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Molecular Biol*., 2001; **313** (4): 903-19.
11. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., William, H. M., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. E., Gibson, T. J., Higgins, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics.*, 2007; **23**: 2947-48.
12. Tuimala, J 2006, A primer to phylogenic analysis using the PHYLIP package, 5th edn, CSC Scientific computing Ltd, at http://koti.mbnet.fi/tuimala/oppaat/phylip2.pdf
13. Huson, D. H., Richter, D. C., Rausch, C., Dejzulion, T., Franz, M., Rupp, R. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinform.*, 2007; **8**: 460-4.
14. Bateman, A., Coin L, Durbin, R, Finn, R. D., Hollich, V, Jones, S. G., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme DJ, Yeats C, Eddy S. R. The Pfam protein families database. *Nucleic acids Res.*, 2004; **32**:138-41.
15. Ikai, A. Thermostability and aliphatic index of globular proteins. *J. Biochem.*, 1980; **88**:1895-8.
16. Guruprasad, K., Reddy, B. V., Pandit, M. W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*., 1990; **4**:155-61.
17. Rogers, S., Well, R., Rechsteiner, M. Amino acid sequence common to rapidly degraded proteins: the PEST hypothesis. *Sci*. 1986; **234**: 364-8.
18. Banerjee, A., Jana, A., Pati, B. R., Mondal, K. C., Das Mohapatra, P. K. Characterization of tannase protein sequences of bacteria and fungi: an in silico study. *J. Protein.*, 2012; **31**: 306-27.