Analysis of the glycoside hydrolase family 8 catalytic core in cellulase-chitosanases from *Bacillus* species

Niveditha Prakash and Shubha Gopal*

*Department of studies in Microbiology, University of Mysore, Manasagangotri, Mysore, Karnataka, India

*Corresponding author: Shubha Gopal; e-mail: shubhagopal_mysore@yahoo.com

ABSTRACT

The glycoside hydrolase family 8 (GH-8) consists of bifunctional cellulase-chitosanases many of which are produced by species of *Bacillus*. Chitosanolytic enzymes can be useful in producing low molecular weight chitoooligosaccharides which have several applications. In addition, a bifunctional enzyme would be more beneficial than the use of two individual enzymes in the production of chitoooligosaccharides and in the degradation of the complex cellulose, chitin and chitosan rich biomass that occurs in nature. The crystal structure of a previously determined GH 8 chitosanase has revealed that the catalytic site is constructed on a scaffold of a double α6/α6 barrel which consists of six helix-loop-helix motifs. Bifunctional cellulase-chitosanases from *Bacillus* species isolated hitherto were analysed and were found to have the double α6/α6 barrel as previously predicted. The signature pattern of glycoside hydrolase family 8 (GH-8) was determined for the sequences. Some of the protein sequences among them were modelled and their closest structural analogs were determined. The structures were found to be related to endoglucanases, xylanases, epimerases and also terpenoid cyclases of which have the double α6/α6 barrel architecture. This study provides a detailed insight into the structure of cellulase-chitosanase catalytic core and identifies related enzymes with similar catalytic core thereby signifying a possible evolutionary relationship.

Keywords: Glycoside hydrolase family 8; *Bacillus*; six hairpin glycosidases; chitosanase-cellulases; homology modelling.

INTRODUCTION

Chitosan is hydrolysed by chitosanase [EC 3.2.1.132], an enzyme that catalyses the hydrolysis of glycosidic bonds of chitosan, the deacetylated derivative of chitin [1]. Chitosanases are classified under glycoside hydrolases, a diverse group of enzymes containing 93 families [2]. Bifunctional enzymes with cellulolytic and chitosanolytic activity mostly belong to glycoside hydrolase family 8 (GH-8) [3]. GH-8 is a chitosanase family which also has cellulase, xylanase, lichenase etc. Most of the bifunctional cellulase-chitosanases, particularly the ones produced by *Bacillus* species belong to glycoside hydrolase family 8. Their primary sequences show high homology with other glucanases of GH-8; xylanases, endoglucanases, lichenases and a few others [3]. Previously resolved crystal structure of GH-8 chitosanases catalytic domain from *Bacillus* sp. K17 by Adachi *et al.* [4] has revealed that the molecule consists of six α-helices forming a central barrel, surrounded by six other α-helices. Six repetitions of this helix-loop-helix motif form a typical α4/α6 double barrel structure. The structures have long loops with short β-strands and 3_10 helices protrude from the end of the outer helices and fold towards the inner helices forming a long cleft on the top of the double barrel suggestive of substrate binding pockets. On the bottom of the double barrel structure, short loops link the helices [4]. GH-8 chitosanases having the typical α4/α6 double barrel structure are called six hairpin glycosidases. The six-hairpin glycoside domain contains about seven alpha-hairpins arranged in closed circular assortment. In the present study, bifunctional cellulase-chitosanases that have been previously reported are structurally examined for presence of the six hairpin glycoside domain. The protein structures are generated using online homology modelling servers. GH-8 signature motif is determined for the sequences, their active site residue is predicted and their closest
structural analogs are detected. The presence of the a6/a6 double barrel architecture in proteins from different families although their active-site architecture may be different suggests the stability of the a6/a6 motif and the probable evolutionary relationship between different glycosyl hydrolases [5].

MATERIALS AND METHODS
Analysis of cellulase-chitosanase sequences from Bacillus species
Previously reported protein sequences of Bacillus species showing cellulase and chitosanase activities were retrieved from NCBI [www.ncbi.nlm.nih.gov/protein] with the following accession numbers: Chitosanase from Bacillus cereus H1 (Q84951) [6], Bacillus sp. strain KCTC 0377BP (QALZ1) [7] and cellulase-chitosanases from Bacillus thuringiensis subspecies (AB061887, AB061888, AB061889, AB061891, AB061892, AB061893, AB061894 and AB061895) [8].

Multiple sequence alignment and analysis of the glycosyl hydrolases family 8 signature motif
The protein sequences were aligned using Clustal Omega set at default parameters [www.ebi.ac.uk/Tools/msa/clustalo] [9-10]. The Multiple sequence alignment output from Clustal Omega was obtained using Jalview [http://www.jalview.org/] [11]. The protein sequences were analysed with Prosite [http://prosite.expasy.org/scanprosite/] [12] and the signature motifs of the enzymes were obtained. The position of the signature motif in the multiple sequence alignment was detected. Sequence logo was generated using Weblogo [http://weblogo.berkeley.edu/logo.cgi] [13].

Homology modelling and identification of structural homologs
Chitosanase-cellulase protein sequences of the above mentioned protein sequences were retrieved from NCBI [www.ncbi.nlm.nih.gov/protein]. Chitosanase from Bacillus cereus H1 (Q84951) [6], Bacillus sp. strain KCTC 0377BP (QALZ1) [7] and one from Bacillus thuringiensis subspecies- AB061887 (Bacillus thuringiensis serovar alesti) [8] were used for the homology modelling. Homology modelling was done using the Swiss model workspace in the automatic modelling mode [http://swissmodel.expasy.org/] [14-16]. The template used for homology modelling was 1V5CA. Protein data bank (PDB) homologs were determined for the sequences using the COFACTOR server [zhanglab.ccb.med.umich.edu/COFACTOR/] [17].

RESULTS AND DISCUSSION
Analysis of cellulase-chitosanase sequences from Bacillus species
Chitosanolytic enzymes are gaining importance as they can produce low molecular weight compounds including chitooligomers with high yield and low pollution. They have numerous applications in biomedical, pharmaceutical, agricultural and food fields [18]. The following previously isolated cellulase-chitosanases whose sequences were deposited in Genbank were selected for the in silico analysis. Chitosanase from Bacillus cereus H1, which showed CMC and glycol chitosan activity (Accession number Q84951) [6] and soil bacterium Bacillus sp. strain KCTC 0377BP isolated by Choi et al. [7] (Accession number QALZ1). Lee et al. In 2007 screened several Bacillus thuringiensis subspecies which produced bifunctional cellulase-chitosanases (Accession numbers: AB061887, AB061888, AB061889, AB061891, AB061892, AB061893, AB061894, AB061895) [8].

Multiple sequence alignment and study of the GH-8 signature motif
The retrieved protein sequences were aligned with Clustal Omega [www.ebi.ac.uk/Tools/msa/clustalo] [9-10] and the output of the multiple sequence alignment was obtained from Jalview [http://www.jalview.org/] [11]. The multiple sequence alignment output which contains the signature motif is shown in figure 1. The signature motif of the GH-8 family is a stretch of about 20 residues. The signature motif sequence has of two conserved aspartates. The first aspartate is believed to be a nucleophile in the catalytic mechanism [5]. The following consensus pattern was obtained:

A- [ST] -D-[AG]-D-x(2)-[IM]-A-x(3)-[LIVM]-[LIVMG]-x-A-x(3)-[FW], where the first D is the active site residue.

The signature motif was generated from Prosite [http://prosite.expasy.org/scanprosite/] [12] and the sequence logo was created. Sequence logo is a graphical representation of multiple sequence alignments. It has stacks of letters which are colour coded and represent amino acids at consecutive positions. Sequence logos are more precise than the consensus sequence. The total height of the logo is based on the degree of conservation of the sequence. The height of the amino acids in the stack pertains to the relative frequency of the amino acid at that position. The logos are read from the top of the stacks [20]. The colouring of the amino acids is based on their chemical properties. Polar amino acids have been coloured green (T- threonine, G- Glycine, Y- Tyrosine, S- Serine), Basic amino acids are coloured blue (K- Lysine, H- Histidine), Acidic amino acids are coloured red (D- Aspartic acid) and hydrophobic amino acids (A- Alanine, L- Leucine and I- Isoleucine) are coloured black [http://weblogo.berkeley.edu/info.html#color_sym].

Sequence logo was created using the online tool Weblogo [http://weblogo.berkeley.edu/logo.cgi] [13]. Sequence logo of the multiple sequence alignment is shown in figure 2.

Homology modelling of cellulase-chitosanase sequences
Homology modelling for the cellulase-chitosanase sequences was done in the automatic modelling mode
using the Swiss model server [http://swissmodel.expasy.org/][14-16]. Homology modelling revealed that all the selected cellulase-chitosanases sequences had the α6/α6 domain architecture which is typical of GH-8 members [5]. The homology models of the cellulase-chitosanase sequences with their signature motifs and catalytic domains are shown in figures 2, 3 and 4. Front view of the structures show the presence of the double barrel α6/α6 motif while the side views show the catalytic cleft and the N and C termini of the proteins. Signature motif of the proteins was seen in sequences 181-199. The first aspartate of the signature motif (Asp 183) was believed to be the active residue of the catalytic mechanism. The homology models were refined by using Modrefiner (http://zhanglab.ccb.med.umich.edu/ModRefiner/)[21]. The quality of the structures that were modeled was verified with PROCHECK obtained from PDBsum. [http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/][21].

Figure 1. A part of the multiple sequence alignment showing the GH-8 signature motif. The signature motif is coloured purple.

Figure 2. Sequence logo of the GH-8 signature motif. Sequence logo was generated from Weblogo [http://weblogo.berkeley.edu/logo.cgi]. Signature motif is found in the position 181-199.

Table 1. Plot statistics for Bacillus cereus H1 chitosanase

| Residues in most favoured regions [A,B,L] | 300 | 88.2% |
| Residues in additional allowed regions [a,b,l,p] | 39 | 11.5% |
| Residues in generously allowed regions [−a,−b,−l,−p] | 0 | 0.0% |
| Residues in disallowed regions | 1 | 0.3% |
| Number of non-glycine and non-proline residues | 340 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 1 |
| Number of glycine residues (shown as triangles) | 28 |
| Number of proline residues | 17 |
| Total number of residues | 386 |

Figure 3. Front view (a), side view (b), Ramachandran plot (c) and plot statistics for chitosanase from Bacillus cereus H1
Figure 4. Front view (d), side view (e), Ramachandran plot (f) and plot statistics for chitosanase Bacillus sp. strain KCTC 0377BP

Table 2. Plot statistics for Bacillus sp. strain KCTC 0377BP

| Residues in most favoured regions [A,R,L] | 305 | 89.7% |
| Residues in additional allowed regions [a,b,l,p] | 34 | 10% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 0 | 0.0% |
| Residues in disallowed regions | 1 | 0.3% |
| Number of non-glycine and non-proline residues | 340 | 100% |
| Number of end-residues (excl. Gly and Pro) | 1 |
| Number of glycine residues (shown as triangles) | 28 |
| Number of proline residues | 17 |
| Total number of residues | 386 |

Figure 5. Front view (g), side view (h), Ramachandran plot (i) and plot statistics for chitosanase Bacillus thuringiensis serovar alesti

Table 3. Plot statistics for chitosanase, Bacillus thuringiensis serovar alesti

| Residues in most favoured regions [A,R,L] | 304 | 89.4% |
| Residues in additional allowed regions [a,b,l,p] | 34 | 10.0% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 1 | 0.3% |
| Residues in disallowed regions | 1 | 0.3% |
| Number of non-glycine and non-proline residues | 340 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 1 |
| Number of glycine residues (shown as triangles) | 28 |
| Number of proline residues | 17 |
| Total number of residues | 386 |
Identification of structural homologs

In order to detect the occurrence of proteins with structural similarity, the COFACTOR server was used [zhanglab.ccmb.med.umich.edu/COFACTOR/] [16]. The top 10 identified structural analogs from PDB were retrieved and tabulated as shown in table 4. All the structures share the common α6/α6 domain. This clearly shows that the cellulase-chitosanases from the selected Bacillus species are structurally related to cellulases, xylanases, endoglucanases, epimerases and terpenoid cyclases from different organisms.

Table 4. Top 10 structural analogs of chitosanases from Bacillus cereus H1, Bacillus sp. strain KCTC 0377BP and Bacillus thuringiensis serovar alesti as identified from the COFACTOR server

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>structural analogs</th>
<th>name of the enzyme</th>
<th>structural analogs</th>
<th>name of the enzyme</th>
<th>structural analogs</th>
<th>Name of the enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1v5C8</td>
<td>Chitosanase</td>
<td>112aA</td>
<td>Processive endocellulase CelF</td>
<td>112aA</td>
<td>Processive endocellulase CelF</td>
</tr>
<tr>
<td>2</td>
<td>1cemA</td>
<td>Cellulase CelA</td>
<td>2afaB</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
<td>2afaB</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
</tr>
<tr>
<td>3</td>
<td>2b41A</td>
<td>Endo-1,4-β-xylanase</td>
<td>3gt5A</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
<td>3gt5A</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
</tr>
<tr>
<td>4</td>
<td>2drtaA</td>
<td>Xylanase Y</td>
<td>3pxqA</td>
<td>Endoglucanase</td>
<td>3pxqA</td>
<td>Endoglucanase</td>
</tr>
<tr>
<td>5</td>
<td>4fusA</td>
<td>Endoglucanase F</td>
<td>1wzzA</td>
<td>Probable endoglucanase (cmcAX)</td>
<td>1wzzA</td>
<td>Probable endoglucanase (cmcAX)</td>
</tr>
<tr>
<td>6</td>
<td>4e88A</td>
<td>Endoglucanase F</td>
<td>1fp3B</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
<td>1fp3B</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
</tr>
<tr>
<td>7</td>
<td>2bkbB</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
<td>3vw5A</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
<td>3vw5A</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
</tr>
<tr>
<td>8</td>
<td>2rgkC</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
<td>2zzrA</td>
<td>Glycosyl hydrolase</td>
<td>2zzrA</td>
<td>Glycosyl hydrolase</td>
</tr>
<tr>
<td>9</td>
<td>1l2aA</td>
<td>Processive endocellulase CelF</td>
<td>2fuaA</td>
<td>Glycosyl hydrolase</td>
<td>2fuaA</td>
<td>Glycosyl hydrolase</td>
</tr>
<tr>
<td>10</td>
<td>3gt5A</td>
<td>N-acetylglucosamine 2-epimerase (AGE)</td>
<td>1w6Ia</td>
<td>Terpenoid cyclase</td>
<td>1w6Ia</td>
<td>Terpenoid cyclase</td>
</tr>
</tbody>
</table>

CONCLUSION

GH-8 is a diverse group of multifunctional enzymes having chitosanases, cellulases, xylanases, lichenases etc. Cellulase-chitosanases belonging to this family are gaining importance as they can produce low molecular weight chitoooligosaccharides which are commercially important for the pharmaceutical, agricultural and food industries. Bacillus species which are common soil inhabitants are known to produce cellulase-chitosanases [3]. Several cellulase-chitosanases have been isolated so far, but their structural details are not clearly understood. The protein sequences of cellulase-chitosanases were shown to have a signature motif of about 20 amino acids were the first aspartate is probably indicates the evolutionary relationship among different glycoside hydrolases having the α6/α6 double barrel catalytic core and the evolutionary stability of the α6/α6 motif among the glycoside hydrolases of several species [5]. An attempt has been made in this study to understand the structures of these beneficial enzymes and also establish their similarity with enzymes having related structures and function.

REFERENCES


© 2014; AIZEON Publishers; All Rights Reserved
This is an Open Access article distributed under the terms of the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.