Carbohydrate Research 418 (2015) 50–56

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Substrate-binding specificity of chitinase and chitosanase as revealed by active-site architecture analysis

Shijia Liu [a], Shangjin Shao [a], Linlin Li [a], Zhi Cheng [a], Li Tian [b], Peiji Gao [b], Lushan Wang [b,*]

[a] Taishan College, Shandong University, Jinan 250100, China
[b] The State Key Laboratory of Microbial Technology, Shandong University, Jinan 250100, China

ARTICLE INFO

Article history:
Received 10 July 2015
Received in revised form 3 October 2015
Accepted 6 October 2015
Available online 21 October 2015

Keywords:
Chitinase
Chitosanase
Active-site architecture
Substrate-binding specificity

ABSTRACT

Chitinases and chitosanases, referred to as chitinolytic enzymes, are two important categories of glycoside hydrolases (GH) that play a key role in degrading chitin and chitosan, two naturally abundant polysaccharides. Here, we investigate the active site architecture of the major chitosanase (GH8, GH46) and chitinase families (GH18, GH19). Both charged (Glu, His, Arg, Asp) and aromatic amino acids (Tyr, Trp, Phe) are observed with higher frequency within chitinolytic active sites as compared to elsewhere in the enzyme structure, indicating significant roles related to enzyme function. Hydrogen bonds between chitinolytic enzymes and the substrate C2 functional groups, i.e. amino groups and N-acetyl groups, drive substrate recognition, while non-specific CH–π interactions between aromatic residues and substrate mainly contribute to tighter binding and enhanced processivity evident in GH8 and GH18 enzymes. For different families of chitinolytic enzymes, the number, type, and position of substrate atoms bound in the active site vary, resulting in different substrate-binding specificities. The data presented here explain the synergistic action of multiple enzyme families at a molecular level and provide a more reasonable method for functional annotation, which can be further applied toward the practical engineering of chitinases and chitosanases.

1. Introduction

Advancements in high-throughput sequencing have exponentially increased the number of available sequences; however, few sequences have been functionally annotated. Thus, efficient functional annotation becomes an urgent task. In order to classify and systematize sequence information, protein databases, such as the CAZy database (http://www.cazy.org/),[1,2] classify enzymes into different families according to sequence similarity.

Chitin is a polymer of N-acetylglucosamine (GlcNAc) linked by β-1,4 glycosidic bonds. Both chitin and its partially deacetylated, water-soluble form, chitosan[3] are highly abundant and widely distributed throughout nature.[4–6] Hydrolyzed chitin and chitosan products and related materials are used in multiple fields, including agriculture,[7] medicine,[8] and industry.[9,10] Chitinolytic enzymes (chitinase and chitosanase) efficiently hydrolyze chitin and chitosan, with chitinase mainly cleaving the GlcNAc–GlcNAc and GlcNAc–GlcN bond[11] (EC 3.2.1.14) and chitosanase breaking the GlcNAc–GlcN, GlcN–GlcNAc, and GlcN–GlcN bonds[12] (EC 3.2.1.132). These catalytic- and glycosidic-bond hydrolysis mechanisms have been intensively studied; however, requirements to improve enzymatic efficiency call for further research into substrate-binding specificity.

For glycoside hydrolases, active site architecture is a part of enzyme which directly interacts with glycoside substrate, adopting the functions of substrate recognition and glycoside bond cleavage.[13] It is shared by all members of a protein family, which adopt over 30% of sequence identity.[1,2] Active site architecture constitutes approximately 2–3% of the total enzyme volume and is influenced by the length of the substrate considerably: an enzyme with a longer substrate has more interacting residues, thus larger active sites.[14] Using structural bioinformatics and statistical analysis, we are able to reveal the roles of key active-site residues involved in substrate binding. Similar research has been applied to other glycoside hydrolases (GH), including identification of spatial-position conservation of key active-site amino acids in GH13 enzymes[15–17] and exploration of relationships between structure and function in GH11 xylanases.[18]

Here, we analyze specific interactions between chitinolytic enzymes and substrates using structure-guided bioinformatics analysis to reveal distinct interactions among different families and propose practical instructions for protein engineering. Through similar interactions, such as hydrogen bonds and CH–π interactions, different families recognize different areas of the substrate, particularly the C2 functional group. The findings presented provide

* Corresponding author. The State Key Laboratory of Microbial Technology, Shandong University, Jinan 250100, China. Tel.: +86 131 5615 5829; fax: +86 531 8836 6202.
E-mail address: lswang@sdu.edu.cn (L. Wang).

http://dx.doi.org/10.1016/j.carres.2015.10.002
0008-6215/

insight into how different enzyme families are able to perform similar functions and offer guidance for increasing functional-annotation accuracy.

2. Experimental

2.1. Data selection

Chitinase and chitosanase family information was obtained from the CAZy database.[1,2] When selecting the enzyme families, the following criteria were adopted to make the alignments more accurate and convincing: the selected family for sequence alignment must contain over 20 available sequences; the target family for structure alignment must contain over 10 available PDB structures, and RMSD between members must not exceed 3 Å; if any structure or sequence contains known mutations, they must be corrected before the alignments. Apart from the two major families of chitosanase—GH8 and GH46, GH3, GH5, GH7, GH75 and GH80 also include members with chitosanase activity. However, among these five enzyme families, only GH7 contains one solved structure, and there were only 1, 2, 3, 16 and 4 characterized sequences by October 2015, so we omitted these families for further analysis. GH8 and 46 had 1 and 5 known structures, respectively, which is not adequate for a conceivable structural alignment, while 23 sequences were available for each family as of October 2015, thus sequence alignments alone were undertaken for GH8 and GH46 families. Similarly, for chitinase enzymes, two major enzyme families, GH18 and GH19, contained 411 sequences and 42 structures, as well as 173 sequences and 13 structures as of October 2015 (Table S1), meeting the above criteria, therefore both were chosen as the research object and undertaken sequence and structure alignments. Yet GH23, another enzyme family containing chitinase, was not chosen because it had only one characterized member.

All sequences downloaded from NCBI[19] and structures obtained from the Protein Data Bank (PDB)[20] with the selected EC number of the specific family were selected as the research sample. One PDB structure with whole-length ligand bound was chosen as the alignment template. If there were no structures with ligand bound available within a certain family, ligands from other enzyme structures within this family were obtained and inserted into the reference structure by molecular docking. Templates used for GH8, GH46, GH18, and GH19 were ChoK[21] (PDB: 1V5D) from *Bacillus* sp. strain K17 (docked with the ligand from PDB: 1KWF[22]), OU01[23] (PDB: 4OLT) from *Microbacterium* sp., ChiA[24] (PDB: 1EHN) from *Serratia marcescens*, and BcChi-A[25] (PDB: 3WH1) from *Bryum coronatum*, respectively.

2.2. Statistical analysis

InterPro[26] was used to detect and delete any non-catalytic domains. The selected model PDB structure was opened by PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC), then waters were removed and ligands were presented. The substrate was oriented with the non-reducing end on the left and reducing end on the right. The numbers denoting individual subsites increased from the non-reducing end to the reducing end, with the cleavage site in the middle. Residues located within 5 Å of the substrate were determined and further classified into the subsite with which they displayed the highest likelihood of binding, then were documented in accordance with the order of corresponding subsites. These residues made up of the active sites defined in the Introduction section. Multiple-sequence alignment and structure alignment were performed using ClustalW (gap open = 10.0, gap extend = 0.5)[27,28] and "Stamp Structural Alignment" tool of VMD,[29] respectively. The columns containing the active site amino acids were extracted from the alignment results and listed under the corresponding substrate subsites.

Both sequence and structure alignment were undertaken by GH18 and GH19. The results from structure alignment contained more information and better reflected the characteristic of each family, although there were fewer available structures than sequences. Therefore, only the results of structure alignment were further analyzed in detail in the text, while those of sequence alignment were shown in Fig. S2.

Alignment results were used to create an active-site architecture sequence profile using WebLogo,[30,31] showing residues characterized within different subsites. To assess sequence-profile accuracy, the score of each column was computed using WebLogo and compared to the conservation score obtained from the ConSurf Server[32–34] and Jalview.[35] Finally, conserved residues were chosen for further analysis. The processes involving data selection, active-site acquisition, multiple-sequence alignment, and WebLogo creation are outlined within a software at the website https://github.com/Stephen8554/MyUsefulTool.

3. Results and discussion

3.1. Amino acid preference in the active site architecture

The observed frequency with which each of the 20 standard amino acids are found within a general protein (calculated from the NCBI protein database[19]), selected chitinolytic enzymes, and chitinolytic enzyme active sites was calculated using BioEdit (Fig. 1A). The relative fold change of residue frequency within chitinolytic enzyme active sites as compared with the holoenzyme was further calculated (Fig. 1B). The residue frequency distribution observed within a general protein was similar to that of the chitinolytic enzymes ($r > 0.75$), while residue frequencies observed between chitinolytic enzyme active sites were less similar ($r < 0.35$). These data indicate that residue frequency between chitinolytic enzymes and other proteins shows little variation, while chitinolytic enzyme active-site composition differs significantly. This implies that amino acids found most frequently within the active site are likely involved in catalysis.

The residues Tyr, Glu, Trp, His, Arg, Asp, and Phe exhibited increased frequencies within the chitinolytic enzyme active site, whereas nonpolar amino acids, such as Val, Leu, Pro, and Ile, occurred with lower frequency as compared to elsewhere within the enzyme. Protein–ligand interactions primarily involve residues located on the protein surface, possibly explaining the lower occurrence of nonpolar residues within the active site.[36] However, hydrophobic aromatic residues, such as Trp, Phe, and Tyr, occurred with high frequency within the active site, possibly signifying important roles in enzyme function.

3.2. The number of residues bound to substrate subsites

Members of a certain protein family exhibit sequence identity of over 30%, therefore they share similar 3D structure backbones. The overall backbone of a protein determines the length of its catalytic cleft, which subsequently decides the maximum number of sugar rings that it can bind to, since the length of a pyranose ring is more or less 5.2 Å.[37–39] During evolution, different subsites display different conservation degrees, and subsites far away from the cleavage site tend to be less conservative than those near the cleavage site (Fig. 2). Although members within a certain enzyme family have various set of subsites, it displays a potential maximum set of subsites. And the maximum number of subsites can be inferred from the models which contain the longest substrate. From what is discussed above, the selected example is capable to represent the whole enzyme family to a large extent, for they are with the most numbers of substrates.