Journal of Structural Biology 180 (2012) 447–457

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

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

Crystal structure of 1,3Gal43A, an exo-β-1,3-galactanase from *Clostridium thermocellum*

Daohua Jiang [a,1], Junping Fan [b,1], Xianping Wang [b], Yan Zhao [b], Bo Huang [b,*], Jianfeng Liu [a], Xuejun C. Zhang [b,*]

[a] *Sino-France Laboratory for Drug Screening, Key Laboratory of Molecular Biophysics of Ministry of Education, School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China*
[b] *National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China*

ARTICLE INFO

Article history:
Received 15 June 2012
Received in revised form 9 August 2012
Accepted 13 August 2012
Available online 29 August 2012

Keywords:
Ct1,3Gal43A
GH43
CBM13
Crystal structure
Enzyme activity

ABSTRACT

Glycoside hydrolase family 43 (GH43) consists of a variety of enzymes distributed widely in prokaryotes and eukaryotes. The mechanism by which GH43 enzymes hydrolyze oligosaccharides requires three essential acidic amino acid residues. However, one of them is thought to be missing in galactan β-1,3-galactosidases from the GH43 family. Ct1,3Gal43A, from *Clostridium thermocellum*, is comprised of a GH43 domain, a CBM13 domain, and a dockerin domain and exhibits an unusual ability to hydrolyze β-1,3-galactan in the presence of a β-1,6 linked branch. Here, we present its crystal structure at 2.7 Å resolution and complex structures of the enzyme with several substrates and analogs. Two modes of substrate binding were observed at the β site of the CtCBM13 domain, and one galactobiose molecule was found in an "L" shaped pocket of the CtGH43 domain, which appears large enough to accommodate two more galactose units. In addition, we found that mutating Glu112 to Gln or Ala eliminated the galactan hydrolysis activity of Ct1,3Gal43A while did not disrupt its ligand binding ability. Combining this results and the crystal structure we identified Glu112 in Ct1,3Gal43A as the 'missing' essential acidic residue in galactan β-1,3-galactosidases. Structural information presented here also suggests a mechanism by which Ct1,3Gal43A bypasses β-1,6 linked branches in the substrate and another mechanism by which the substrate is delivered '*in trans*' from the CBM13 domain to the catalytic GH43 domain.

1. Introduction

Celluloses form the most abundant biological carbon source on our planet and can be biochemically converted into bio-fuel. Some micro-organisms have been found to assemble a type of micro-machine called cellulosomes that degrades different types of cellulose (Bayer et al., 1994; Felix and Ljungdahl, 1993). Cellulosomes were first discovered by Bayer and colleagues in 1983 on the surface of the thermophilic anaerobic bacteria, *Clostridium thermocellum* (ATCC_27405) (Lamed et al., 1983). Cellulosomes are huge protein complexes with a total molecular weight ranging from 2×10^6 to 6×10^6 Da. They contain 14–50 subunits with molecular weights ranging from 40 to over 200 kDa. There are mainly two types of cellulosome component, namely a scaffold protein (scaffoldin) and a group of cellulose enzymes, called cellulases. One end of scaffoldin is anchored to the bacterial surface through a pair of interacting protein domains, *i.e.* dockerin (type II) and cohesin. There are up to nine tandem-linked type I docking receptors responsible for recruiting a variety of cellulases along the peptide chain of scaffoldin (Bayer et al., 2004; Xu et al., 2004). In general, each cellulase contains three functional domains, namely a catalytic domain responsible for the hydrolysis of glycosidic bonds, a carbohydrate binding module (CBM) responsible for targeting a type of cellulose, and a type I dockerin domain responsible for anchoring the enzyme to the scaffold protein (Carvalho et al., 2003). A cellulosome may contain many types of glycoside hydrolases (Doi and Kosugi, 2004; Shoham et al., 1999), the degradation specificity of which towards different types of celluloses is mainly determined by their CBM binding specificities and the substrate specificities of their catalytic domains (Schwarz, 2001).

All known glycosidases, including those forming cellulosomes, have been systematically classified in the Carbohydrate-Active EnZymes database (CAZy) (Cantarel et al., 2009). Based on the primary sequences of their catalytic subunits, glycosidases are classified into 130 families (GH1-130). Similarly, their carbohydrate binding modules are classified into 64 families (CBM1-64). Different

Abbreviations: Ct1,3Gal43A, *Clostridium thermocellum* exo-β-13-galactanase; r.m.s.d., root mean square deviation; PDB, Protein Databank; GH43, glycoside hydrolase family 43; CBM, carbohydrate binding module; CAZy, Carbohydrate-Active EnZymes database.

* Corresponding authors. Fax: +86 10 64889186.
E-mail addresses: huangb@moon.ibp.ac.cn (B. Huang), zhangc@ibp.ac.cn (X.C. Zhang).

[1] These authors contributed equally to this work

1047-8477/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2012.08.005

combinations of catalytic domains with CBM domains produce a large number of glycosidases targeting a variety of cellulose glycosidic bonds. Furthermore, assembly of different types of subunits at the cellulosome level makes it possible for a cellulosome to efficiently degrade celluloses. For example, a combination of endo-glucanase and exo-glucanase is necessary for efficient degradation of a complicated glucan complex. Cellulosomes thus provide an ideal platform for their component enzymes to perform synergetic functions.

Exo-β-1,3-galactanase (Cthe_0661) from *C. thermocellum* has a typical cellulosomal cellulase structure (Ichinose et al., 2006). Its CAZy name is Ct1,3Gal43A, reflecting its organism of origin (Ct), the enzyme type (1,3Gal), and the classification of its catalytic domain (family GH43). It consists of 571 amino acid residues, including a signal peptide of 30 amino acid residues for secretion, an N-terminal catalytic domain (residues 31–350), a CBM domain (351–493), and a C-terminal type I dockerin domain. Ct1,3Gal43A shows significant affinity to galactose-containing polysaccharides and catalyzes hydrolysis only of β-1,3-linked galactosyl oligosaccharides and polysaccharides with an optimal activity at pH 6.0 and 50 °C (Ichinose et al., 2006). Amino acid sequence analysis of Ct1,3Gal43A revealed that while the catalytic domain belongs to the glycosidase GH43 family, the carbohydrate binding domain belongs to the CBM13 family (Ichinose et al., 2006). Currently, there are about 2129 GH43-containing proteins and about 1565 CBM13-containing proteins in the CAZy database, of which over 90 proteins contain both a GH43 catalytic domain and a CBM13 carbohydrate binding domain. However, three dimensional details of this type of GH43–CBM13 joint protein are unavailable.

Ct1,3Gal43A specifically hydrolyses β-1,3 glycosidic bonds in galactose-based oligosaccharides or polysaccharides, and shows maximum activity towards β-1,3-galactotetraose (*i.e.* galactose × 4). Furthermore, Ct1,3Gal43A can perform β-1–3 linked main chain hydrolysis in the presence of a β-1,6 linked branch (Ichinose et al., 2006). The current hypothesis on the mechanism of GH43 enzyme hydrolysis of glycosidic chains asserts that three acidic amino acid residues respectively play the critical roles of general acid, general base, and pK_a modulator of the general acid. However, the position of the "general base" residue in the GH43 domain is mutated to Gly in galactan β-1,3-galactosidases, including Ct1,3Gal43A, leaving only two of the three essential acidic residues (Ichinose et al., 2006). Nevertheless, these enzymes are still functional.

To investigate the structural basis underlying the function of Ct1,3Gal43A, we determined its crystal structure, and complex structures of the enzyme with various substrates and analogs. We also performed functional studies including mutagenesis, thermostability based ligand binding analysis, and activity assays. Together with information from functional studies and crystal structures, we identified the 'missing' general base in Ct1,3Gal43A and perhaps other GH43 β-1,3-galactosidases, reconciling the substrate hydrolytic mechanism of Ct1,3Gal43A with the current hypothesis on the GH43 enzymatic mechanism. Our results also provide insights into the structural basis of substrate recognition, especially on how the enzyme accommodates branched substrates during hydrolysis. Moreover, we unexpectedly discovered an interaction between the CBM domain and the GH43 domain of Ct1,3Gal43A which suggests a possible '*trans*' substrate-delivery mechanism.

2. Methods

2.1. Protein expression and purification

The Ct1,3Gal43A gene (GenBank ID: 125712750, Cthe_0661) was cloned from the genome of *C. thermocellum* (ATCC catalog No, 27405), with the first 90 nucleic acids which correspond to a predicted signal peptide being deleted (Ichinose et al., 2006). Primers for the polymerase chain reaction (PCR) were the following: (Forward) 5′-CG GAA TTC GCG GCA GAA GGG GTT ATA GT-3′ (an *Eco*RI restriction site is underlined) and (Reverse) 5′-CCG CTC GAG TTA CAA ATC CAC TTC CAT AAG C-3′ (an *Xho*I restriction site is underlined). Both the PCR product and expression vector, pET28a (Novagen, Germany) were digested with *Eco*RI and *Xho*I (Takara, Japan) at 37 °C overnight. After extraction from the DNA gel, they were ligated at 16 °C overnight. The ligation reaction product was used to transform BL21 (DE3) competent cells. The entire coding region was confirmed by DNA sequencing.

Protein expression was carried out in Luria–Bertani (LB) medium. Cells were cultured at 37 °C until an $O.D._{600nm}$ of about 0.8 was reached, and the temperature was then reduced to 16 °C. Thirty minutes later, the cell culture was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.5 mM and then harvested 18 h later.

Five grams of cell pellet (from 1 L) were resuspended in 50 ml buffer containing 20 mM Tris–HCl (pH 8.0) and 500 mM NaCl and were lysed using a French Press high pressure instrument at 10,000 psi in three passes. Cell debris was removed by centrifugation at 30,000*g* for 40 min at 4 °C. The supernatant containing the N-terminal poly-His tagged protein sample was loaded on a 3-ml affinity column of Ni^{2+}–NTA agarose (Qiagen, Germany). After a 20× column volume wash, the Ct1,3Gal43A protein was eluted from the affinity resin using 20 mM Tris–HCl (pH 8.0), 500 mM NaCl, and 500 mM imidazole and concentrated to 10 mg/ml before further purification by gel filtration. Note that the optical extinction coefficient for 280 nm absorption was estimated to be 138,005 M^{-1} cm^{-1}, calculated with the online utility ProtParam (Wilkins et al., 1999). The buffer for gel filtration chromatography (Superdex-200, GE Healthcare) contained 20 mM Tris–HCl (pH 8.0) and 100 mM NaCl. Peak fractions were collected (monitored by 280 nm absorption).

Seleno-methionine substituted protein was produced using the expression strain BL21(DE3) described above by metabolic inhibition (Doublie, 1997). The protein sample was purified using a similar method as that for the native protein, except that 2 mM β-mercaptoethanol was added to all buffers to maintain reducing conditions for the seleno-methionine residues.

2.2. Crystallization, data collection, and structure determination

For crystallization, the His-tag containing protein sample from gel filtration was concentrated to 50 mg/ml in a buffer containing 20 mM Tris–HCl (pH 8.0) and 100 mM NaCl. An initial crystallization condition was obtained by screening commercial crystallization kits using the hanging-drop vapor-diffusion method. Optimized crystals were grown at 16 °C for 3 d after mixing 1 μl protein sample with 1 μl reservoir solution composed of 100 mM sodium acetate (pH 4.5) and 2.9–3.3 M sodium chloride. Se-Met-derived crystals were grown at 4 °C under a condition similar to that described above, except that 100 mM sodium acetate (final concentration) was added. Before freezing in liquid nitrogen for storage, shipping, and data collection, both native and Se-Met derivative crystals were soaked in the mother liquor supplemented with 20% glycerol (v/v/) for 20 min. Ligands were soaked into the crystals by transferring Ct1,3Gal43A crystals into a solution of the mother liquid supplemented with the sugar ligand to be tested at a concentration ranging from 3% to 5% (w/v). Crystals were soaked for 30 min before data collection.

Diffraction data were collected up to 2.7 Å resolution for the native crystal and to 3.2 Å for the Se-Met derivative crystal at the BL17U beam line of the SSRF synchrotron facility (Shanghai, China) and at the 17A beam line of the Photon Factory (Tsukuba, Japan), respectively, both at a temperature of 100 K. Data were indexed,