Biochimica et Biophysica Acta 1814 (2011) 1832–1840

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Diverse substrate recognition mechanism revealed by *Thermotoga maritima* Cel5A structures in complex with cellotetraose, cellobiose and mannotriose☆,☆☆

Tzu-Hui Wu [a,1], Chun-Hsiang Huang [c,1], Tzu-Ping Ko [d], Hui-Lin Lai [g], Yanhe Ma [c], Chun-Chi Chen [e], Ya-Shan Cheng [a,g], Je-Ruei Liu [a,b,f,*], Rey-Ting Guo [c,**]

[a] *Institute of Biotechnology, National Taiwan University, Taipei 106, Taiwan*
[b] *Department of Animal Science and Technology, National Taiwan University, Taipei 106, Taiwan*
[c] *Industrial Enzymes National Engineering Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China*
[d] *Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan*
[e] *Institute of Biomedical Science, Academia Sinica, Taipei 115, Taiwan*
[f] *Agricultural Biotechnology Research Center, Academia Sinica, Taipei 115, Taiwan*
[g] *Genozyme biotechnology Inc., Taipei 106, Taiwan*

ARTICLE INFO

Article history:
Received 24 February 2011
Received in revised form 2 July 2011
Accepted 21 July 2011
Available online 4 August 2011

Keywords:
Cellulose
Cellulase
Biofuel
Crystal structure
Synchrotron radiation

ABSTRACT

The hyperthermophilic endoglucanase Cel5A from *Thermotoga maritima* can find applications in lignocellulosic biofuel production, because it catalyzes the hydrolysis of glucan- and mannan-based polysaccharides. Here, we report the crystal structures in apo-form and in complex with three ligands, cellotetraose, cellobiose and mannotriose, at 1.29 Å to 2.40 Å resolution. The open carbohydrate-binding cavity which can accommodate oligosaccharide substrates with extensively branched chains explained the dual specificity of the enzyme. Combining our structural information and the previous kinetic data, it is suggested that this enzyme prefers β-glucosyl and β-mannosyl moieties at the reducing end and uses two conserved catalytic residues, E253 (nucleophile) and E136 (general acid/base), to hydrolyze the glycosidic bonds. Moreover, our results also suggest that the wide spectrum of *Tm*_Cel5A substrates might be due to the lack of steric hindrance around the C2-hydroxyl group of the glucose or mannose unit from active-site residues.

1. Introduction

Due to increasing energy consumption worldwide and limited crude oil supply, the need for alternative energy source has drawn enormous attention. Renewable energy such as biofuel produced from lignocellulosic biomass has significant potential to meet the need [1]. Lignocellulosic biomass contains complex carbohydrate molecules with various chemical linkages, and is an abundant carbohydrate source of energy from plants. By breaking down the chemical linkages, monomeric sugars are released and then fermented into ethanol as biofuel. Therefore, the carbohydrate source from plants, such as lignocellulosic biomass, as biofuel not only provides a solution for energy shortage but also reduces greenhouse gas emissions [2,3]. Hydrolysis of a complex carbohydrate trapped inside the lignocellulose to obtain monomeric sugar compositions requires an array of enzymes acting synergistically to cleave the various chemical linkages [4]. At least three cellulolytic enzymes, including endoglucanase, exoglucanase and β-glucosidase, are required for complete hydrolysis of cellulose into monosaccharides [5].

Endoglucanase are widespread among GH families (now 125 GH families), such as families 5–9, 12, 44, 45, 48, 51, 61, 74 and 124. (http://www.cazy.org) [5,6]. Among these, the GH5 family possesses a variety of enzymatic activities, which can function as cellulase, mannanase, galactomannanase and xylanase [6,7]. In addition, the structural studies have revealed that the GH5 family shares a structural feature of a $(\beta/\alpha)_8$ topology with two glutamates on strands β-4 and β-7, acting as the acid/base and nucleophile, respectively, to cleave the glucosidic bonds [8]. The structure of the substrate and the amino acid residues in the active site of the GH5 enzymes were analyzed to show how different

Abbreviations: BGC, β-D-glucose; CBI, cellobiose; CTT, cellotetraose; MAT, mannotriose; PEG, polyethyleneglycol; PCR, polymerase chain reaction; RMSD, root mean square deviation; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; Tris, tris (hydroxymethyl) aminomethane; XG, xyloglucan

☆ This work was supported by grants from the National Science Council of Taiwan (NSC98-3114-B-002-003 and NSC98-2313-B-002-033-MY3 to JRL), National Basic Research Program of China (2011CB710800 to RTG) and Tianjin Municipal Science and Technology Commission (10ZCKFSY06000 to RTG).

☆☆ The atomic coordinates and structure factors for the wild-type *Thermotoga maritima* Cel5A (PDB IDs: 3AMC and 3AMD), E253A mutant in complex with cellotetraose (PDB ID: 3AZT), cellobiose (PDB ID: 3AZR) and mannotriose (PDB ID: 3AZS), and E136A mutant in complex with cellotetraose (PDB ID: 3AMG) and mannotriose (PDB ID: 3AOF) have been deposited in the RCSB Protein Data Bank.

* Correspondence to: J.-R. Liu, Institute of Biotechnology, National Taiwan University, Taipei 106, Taiwan. Tel.: +886 2 3366 6011; fax: +886 2 3366 6001.
** Corresponding author. Tel.: +86 022 84861999; fax: +86 022 24828701.
E-mail addresses: jrliu@ntu.edu.tw (J.-R. Liu), guo_rt@tib.cas.cn (R.-T. Guo).

[1] These two authors contributed equally.

1570-9639/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2011.07.020

properties of the active sites, even by one distinct amino acid, can lead to different substrate binding specificities [7,9,10].

Thermotoga maritima, which is an obligately anaerobic, heterotrophic and hyperthermophilic bacterium [11], has been demonstrated to contain a large number of glycoside hydrolase [12]. The *Cel5A* gene from *Thermotoga maritima* encoding endo-β-1,4-glucanase of GH5 has been demonstrated to exhibit both endoglucanase and exoglucanase activities [13,14] with a broad substrate spectrum including mannan, galactomannan, glucomannan, β-glucan, carboxymethyl cellulose, xyloglucan and lichenan [15], although the relative activity varies with the different compositions of the substrates [13,14]. The characteristics of thermostability and dual-function to hydrolyze both glucan- and mannan-based polysaccharides make *Tm*_Cel5A an excellent enzyme for industrial applications.

Recently, the crystal structure of the apo-form of *Tm*_Cel5A (PDB ID: 3MMU and 3MMW) was determined and compared to the mesophilic homologue Cel5A from *C. cellulolyticum* (*Cc*_Cel5A), providing several important structural features to explain why *Tm*_Cel5A has higher thermostability [16]. In their crystallization condition, cadmium chloride was necessary. One Cd^{2+} ion was observed in the active site and occupied the site for real substrate binding. The complex structure of *Tm*_Cel5A and substrate was not obtained in this crystallization condition [16]. In the mean time, we have solved the *Tm*_Cel5A structures in the apo-form by the multiple-wavelength anomalous diffraction (MAD) method. Here we present the crystal structures of not only the E253A mutant in complex with cellotetraose (CTT), cellobiose (CBI) and mannotriose (MAT), but also the E136A mutant in complex with cellobiose (CBI) and mannotriose (MAT). Previous results demonstrated that *Tm*_Cel5A can hydrolyze cellodextrins, pretreated switchgrass and Avicel to glucose, cellobiose, and cellotriose [17]. Actually, the bound CTT, CBI and MAT may be more appropriately called products. To avoid confusion, we will use ligand instead of substrate/product in this paper. These enzyme–ligand complex structures further elucidate how *Tm*_Cel5A can have diverse substrate specificity.

2. Materials and methods

2.1. Material

ExSel high fidelity DNA polymerase was obtained from Bertec Enterprise Co. The plasmid mini-prep kit, DNA gel extraction kit, and Ni–NTA resin were purchased from GeneMark, Viogene, and Qiagen, respectively. Factor Xa and the protein expression kit (including the pET 32 Xa/LIC vector and competent BL21 (DE3) cells) were obtained from Novagen. *Thermotoga maritima* genomic DNA was purchased from ATCC (ATCC accession number 43589). All commercial buffers and reagents were of the highest grade possible.

2.2. Protein expression and purification

The gene encoding Cel5A was amplified from *Thermotoga maritima* genomic DNA by polymerase chain reaction (PCR) with forward primer 5′-GGTATTGAGGGTCGCATGGGTGTTGATCCTTTTGAAAGG-3′ and reverse primer 5′-GAGGAGAAGCCCGGTTATTCAATGCTATCTCCTAT-3′ and cloned into the pET 32 Xa/LIC vector. Then the recombinant Cel5A plasmid was transformed into *E. coli* BL21 (DE3) for protein expression.

Table 1
Summary of data processing and refinement statistics.

Name	Apo-form1	Apo-form2	E253A-CBI	E253A-MAT	E253A-CTT	E136A-CTT	E136A-MAT
PDB ID	3AMC	3AMD	3AZR	3AZS	3AZT	3AMG	3AOF
Data collection							
Wavelength (Å)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Resolution (Å)	25–1.40 (1.45–1.40)	25–2.0 (2.07–2.00)	25–1.71 (1.77–1.71)	25–1.69 (1.75–1.69)	25–1.80 (1.86–1.80)	25–2.40 (2.49–2.40)	25–1.29 (1.34–1.29)
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell a/b/c (Å), β (°)	63.0/78.3/63.0 β = 97.2	82.3/75.3/93.6 β = 90.4	62.4/77.1/62.6 β = 97.5	62.7/78.0/92.9 β = 97.2	82.4/76.2/94.3 β = 90.5	61.0/73.4/62.2 β = 97.6	62.7/77.9/63.0 β = 97.2
No. of measured reflections	448,560 (44,863)	372,638 (37,018)	216,412 (21,298)	274,832 (26,432)	412,701 (36,939)	78,947 (6827)	975,862 (97,056)
No. of unique reflections	118,042 (11,806)	77,633 (7712)	62,921 (6264)	67,472 (6608)	106,844 (10,554)	21,337 (2008)	147,549 (14,486)
Completeness (%)	99.4 (100)	99.7 (99.9)	99.5 (99.8)	99.4 (97.4)	99.2 (98.0)	98.7 (93.4)	97.5 (95.7)
R_{merge} (%)[a]	4.8 (18.6)	8.5 (40.1)	5.3 (47.9)	4.3 (12.4)	5.6 (44.2)	4.2 (28.3)	3.9 (32.9)
Mean $I/\sigma(I)$	27.6 (6.7)	18.5 (3.8)	23.0 (2.6)	30.1 (10.8)	22.0 (2.4)	30.5 (3.6)	43.3 (6.3)
Multiplicity	3.8 (3.8)	4.8 (4.8)	3.5 (3.4)	4.1 (4.0)	3.9 (3.5)	3.7 (3.4)	6.6 (6.7)
Refinement							
No. reflection used	116,630 (11,023)	74,149 (6914)	60,662 (5184)	66,584 (6574)	100,259 (8428)	21,315 (2403)	147,539 (4567)
R_{factor} (%)	18.5 (24.2)	16.8 (20.2)	20.4 (28.9)	15.5 (15.6)	22.1 (35.5)	22.8 (32.7)	17.2 (23.2)
R_{free} (%)	21.0 (27.7)	22.0 (26.6)	24.9 (32.9)	18.9 (20.3)	26.3 (38.9)	28.4 (38.1)	18.8 (26.5)
No. protein atoms	5176	10,440	5203	5184	10,405	5040	5170
No. ligand atoms	–	–	70	68	90	35	34
No. water molecules	1161	761	581	674	913	123	904
R.M.S.D. bond angles (°)	2.089	1.844	1.654	1.672	1.777	1.355	1.006
R.M.S.D. bond lengths (Å)	0.023	0.019	0.015	0.015	0.015	0.018	0.004
Average B factor ($Å^2$)	15.7	20.2	22.6	15.3	30.5	70.4	15.9
Ramachandran plot							
Most favored (%)	90.1	89.9	88.9	89.7	89.3	85.9	89.0
Additionally allowed (%)	9.5	9.8	10.4	9.9	9.9	13.7	10.4
Generously allowed (%)	0.4	0.3	0.7	0.4	0.7	0.4	0.5
Disallowed (%)	0.0	0.0	0.0	0.0	0.1	0.0	0.0

Values in parentheses are for the highest resolution shell.
[a] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - <I(hkl)>| \sum_{hkl}\sum_i I_i(hkl)$.