Bioresource Technology 101 (2010) 2405–2411

Contents lists available at ScienceDirect

Bioresource Technology

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

Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases

E. Vlasenko, M. Schülein, J. Cherry, F. Xu *

Novozymes Inc., 1445 Drew Avenue, Davis, CA 95618, USA

ARTICLE INFO

Article history:
Received 28 April 2008
Accepted 12 October 2009
Available online 14 December 2009

Keywords:
Endoglucanase
Family
Specificity
Substrate

ABSTRACT

Endoglucanases are important enzymes for biomass conversion and other industrial processes. Determining the specificity of endoglucanases from various glycoside hydrolase families is of interest for bioinformatic functional prediction and substrate-tailored enzyme development. To do so, we characterized ~30 endoglucanases from six glycoside hydrolase families. For *p*-nitrophenyl cellobioside and lactoside, only family 7 enzymes showed significant activity. For xyloglucan, both family 7 and 12 enzymes showed significant activity. For xylan and arabinoxylan, only family 7 enzymes showed significant activity. For mannan and galactomannan, both family 5 and 9 enzymes showed significant activity. The difference in specificity was preliminarily attributed mainly to the structural difference of the enzymes' active sites. For family 7 endoglucanases, difference in thermal stability might affect their performance in hydrolyzing various (hemi)cellulose substrates. Phylogenetic analysis on the subfamily distribution of family 5 endoglucanases (in relation with mannanases) suggested that their mannanase side-activity might be the remnant of an ancestral multi-function enzyme. Similar analysis was also made with the xyloglucanase or arabionxylans side-activity of family 12 and 7 endoglucanases. The apparent dependence of the specificity on family (primary/tertiary structure) might assist us in better understanding the structure–function relationship of the enzymes, and developing more versatile biocatalysts for the utilization of biomass.

1. Introduction

In nature, the breakdown of highly heterogeneous (and often recalcitrant) lignocellulose is accomplished with high efficiency by a vast group of enzymes, O-glycoside hydrolases. Degradation and conversion of lignocellulosic biomass is attracting intensive attention because of its potential for the development of sustainable and environment-friendly energy and materials industries (for recent reviews, see Kumar et al., 2008; Jorgensen et al., 2007; Xu, 2004; Hilden and Johansson, 2004; Lynd et al., 2002).

For lignocellulolytic microbes, many carbohydrate-active enzymes are employed, in either "freely" dissociable forms (produced mostly by aerobic fungi) or as parts of a supramolecular assembly called cellulosome (produced mostly by anaerobic bacteria), to synergistically and codependently depolymerize lignocellulose. Major cellulose (β-1,4-glucan)-degrading enzymes include cellobiohydrolase (EC 3.2.1.91, targeting mainly crystalline cellulose), endo-1,4-β-endoglucanase (EG) (EC 3.2.1.4, targeting mainly amorphous cellulose), and β-glucosidase (EC 3.2.1.21, targeting cellobiose that is inhibitory for cellobiohydrolase and EG).

Based on sequence and 3-dimensional structure, EGs are grouped, along with other enzymes, into ~11 glycoside hydrolase (GH) families, including GH5, 6, 7, 8, 9, 12, 44, 45, 48, 51, and 74 (Cantarel et al., 2009; for updates, see http://www.cazy.org). An individual family often possesses unique tertiary structure, particularly at the active site. Tertiary structures of some families are highly similar (although their sequences may not be homologous), making them belong to a GH clan. In general, a globular EG molecule has a substrate-binding cleft. To be effectively bound, a cellodextrin molecule or cellulose chain segment interacts with multiple subsites (approximately 4–7) in the cleft, via regio-specific H-bonding and/or aromatic-carbohydrate ring π-stacking (based on London's dispersion or charge transfer) with the amino acid residues of EG. To reach the active transition state, EGs either stretched the target 1,4-glucosidic bond, while keeping the chair conformation of the glucosyl units at the catalytic "−1" and "+1" subsites (Davies et al., 1995), or bend the glucosidic bond to a quasi-axial position, while distorting the "−1" glucosyl from its chair form to a skew/boat form and displacing the "+1" glucosyl (Davies et al., 2003), via cooperative H-bonding and/or π-stacking networks. The glucosidic bond cleavage can be achieved by a mechanism that either inverts or retains the anomeric configuration, with the involvement of a glucosyl oxocarbenium intermediate as well as two key acidic amino acid residues serving as catalytic general acid and nucleophile. Because of the intricate nature of these interactions, any modifications on cellulose, such as steric substitution or OH-removal, may affect the catalysis of EG.

* Corresponding author. Tel.: +1 530 757 8100.
E-mail address: fxu@novozymes.com (F. Xu).

0960-8524/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biortech.2009.11.057

Based on their sequences and *in vivo* functions, different EGs may possess different specificities. For example, the GH5 family contains not only cellulases (Cel5) but also mannanases and xylanases (Cantarel et al., 2009). If all GH5 enzymes are derived from the divergence of an ancient gene (rather than the convergence of different ancient genes), than Cel5 EG may have significant activity on mannan or xylan. Having a broad or differential specificity may be beneficial for a biomass-hydrolyzing enzymatic system, because of the heterogeneity of the substrate. Indeed, many organisms employ more than one EG (Hilden and Johansson, 2004; Lynd et al., 2002).

In contrast to the vast, ever-growing information systematically gathered on the sequence and structure of cellulases, the progress on the comparative study that correlates cellulase's structure and specificity, which is of great importance for function annotation/prediction of genomes and other sequences, remains relatively slow (Mertz et al., 2005; Dias et al., 2004; Lawoko et al., 2000; Claeyssens and Henrissat, 1992). In this study, we investigated the activity of over 30 EGs (from six GH families) towards 12 (hemi)cellulosic substrates with different crystallinity, backbone alternation, and side chain substitution. Under our conditions, family-dependent non-cellulose specificities were observed for the EGs.

2. Methods

2.1. Reagents

Commercial chemicals used as buffers or substrates were of at least reagent grade.

Birchwood xylan (X), *p*-nitrophenyl-β-D-cellobioside (pNPC), and *p*-nitrophenyl-β-D-lactoside (pNPL) were obtained from Sigma–Aldrich (St. Louis, Missouri). Avicel (PH101 grade) was obtained from FMC. Carboxymethylcellulose (CMC) was obtained from Hercules (now part of Ashland, Wilmington, Delaware) (type 7L2, 0.7 average degree of substitution). Phosphoric acid-swollen cellulose (PASC) was prepared from Avicel, and bacterial cellulose (BC, Nata de Coco) was prepared as previously reported (Gilkes et al., 1992). Tamarind seed xyloglucan (XG, amyloid, Glc:Xyl:Gal = 45:35:16, approximately 78% 6-O-substitution of backbone β-1,4-glucan), wheat arabinoxylan (AX, medium viscosity, Xyl:Ara = 61:37, approximately 60% 2-/3-O-substitution of backbone β-1,4-xylan), β-1,4-D-mannan (M, approximately 15 degree of polymerization, borohydride-reduced, Man:Gal = 97:3), and carob galactomannan (GM, low viscosity, borohydride-reduced, Man:Gal = 78:22, approximately 20% 6-O-substitution of backbone β-1,4-mannan) were obtained from Megazyme. Acid pretreated corn stover (PCS, approximately 57% cellulose, 5% hemicellulose, and 28% lignin) was kindly provided by the US National Renewable Energy Laboratory (Golden, Colorado) and washed extensively with water before use.

2.2. Enzymes

Table 1 lists the EGs studied in this report. The recombinant *Acidothermus cellulolyticus* E1 catalytic core was kindly provided by the US National Renewable Energy Laboratory. The cloning, expression, and preparation of other EGs were previously described (Lassen et al., 2007; Harris et al., 2007; Schülein et al., 2007; Sharyo et al., 2002; Schülein and Bjornvad, 2000).

2.3. Assays

Protein concentration was determined by the BCA kit from Pierce (now part of Thermo Fisher Scientific, Rockford, Illinois). Reducing sugars were determined using *p*-hydroxybenzoic acid hydrazide (Harris et al., 2007), with glucose (0.005–0.1 g/L) as standard. Spectrophotometric measurement was made on a SpectraMAX microplate reader from Molecular Devices (Sunnyvale, California). Activity on pNPC and pNPL was determined with 2.5 mM pNPC or pNPL, 50 mM Na-acetate, pH 5, 40 °C, enzyme doses capable to hydrolyze ⩽8% substrate in 30 min, and absorption monitoring at 405 nm (absorptivity: 18 mM^{-1} cm^{-1}) after raising pH to 10 with Na_2CO_3 (Deshpande et al., 1984). Initial activity on PASC and CMC was determined, by reducing sugar release, with 2 g/L PASC or 5 g/L CMC, 0.05–5 mg/L enzyme, 0.5 g/L bovine serum albumin (BSA), 50 mM Na-acetate, pH 5, 50 °C, 30–90 min (to reach ⩽3% substrate hydrolysis). Prolonged polysaccharide hydrolysis was carried out with 5 g/L polysaccharide (0.9 g/L for BC), 25 mg/L enzyme (12.5 mg/L for PASC and CMC), 50 mM Na-acetate, pH 5, 50 °C, 1 mL total volume, in capped 1.2 mL 96 Deepwell plates from Eppendorf (Hauppauge, New York) with intermittent inversion. At selected times, 20 μL aliquots were removed, added to 180 μL of 102 mM Na_2CO_3 plus 58 mM $NaHCO_3$ in a MultiScreen HV 96-well filtration plate from Millipore (Billerica, Massachussetts), vacuum-filtered, and assayed for reducing sugar. Polymeric substrate conversion to reducing sugars was quantified by the following equation: conversion (in %) = [reducing sugars] (in g/L) × 162 ÷ ([starting substrate] (in g/L) × 180) × 100. Experimental data were presented as averages of duplicates (produced in the same microplates), with individual relative errors ⩽5%.

Thermal stability was determined by measuring the residual CMC-hydrolyzing activity of an enzyme after being incubated at 0.37–49 mg/L level in 50 mM Na-acetate of pH 5 (in the presence of 3 g/L BSA) at 40–80 °C for 3 h. Thermal activity was measured by the PASC-hydrolyzing activity of an enzyme with 2 g/L PASC

Table 1
Endoglucanases used in this study.

Abbreviation	Enzyme	CBM[a]	Thermophilic[d]
A Cel5A	*Acremonium* sp. CBS265.95 Cel5A	+	+
Aa Cel5B	*Aspergillus aculeatus* Cel5B		
Ac E1cd	*Acidothermus cellulolyticus* E1cd		+
B1 Cel5A	Basidiomycete CBS494.95 Cel5A[b]	+	
B2 Cel5B	Basidiomycete CBS495.95 Cel5B[c]	+	
Mt Cel5A	*Myceliophthora thermophila* Cel5A	+	+
Tt Cel5A	*Thielavia terrestris* Cel5A	+	+
Tf Cel6A	*Thermomonospora fusca* Cel6A	+	+
A Cel7A	*Acremonium* sp. CBS265.95 Cel7A		+
A Cel7B	*Acremonium* sp. CBS265.95 Cel7B		+
Cf Cel7	*Cladorrhinum foecundissimum* Cel7		
Fo Cel7B	*Fusarium oxysporium* Cel7B		
Hi Cel7B	*Humicola insolens* Cel7B		+
Tt Cel7C	*Thielavia terrestris* Cel7C	+	+
Tr Cel7B	*Trichoderma reesei* Cel7B	+	
Bl Cel9	*Bacillus licheniformis* Cel9	+	
Aa Cel12A	*Aspergillus aculeatus* Cel12A		
Aa Cel12B	*Aspergillus aculeatus* Cel12B		
Af Cel12A	*Aspergillus fumigatus* Cel12A		
Mt Cel12A	*Myceliophthora thermophila* Cel12A		+
A Cel45A	*Acremonium* sp. CBS265.95 Cel45A	+	+
Cb Cel45	*Chaetomium brasiliensis* Cel45	+	+
Cf Cel45A	*Cladorrhinum foecundissimum* Cel45A	+	
Cs Cel45A	*Crinipellis scabella* Cel45A		
Hg Cel45A	*Humicola grisea* Cel45A	+	+
Hi Cel45A	*Humicola insolens* Cel45A	+	+
Hi Cel45Ac	*Humicola insolens* Cel45A core		+
Mp Cel45A	*Macrophomina phaseolina* Cel 45A		
Sf Cel45A	*Sordaria fimicola* Cel45A	+	
Vc Cel45A	*Volutella colletotrichoides* Cel45A	+	
Tt Cel45	*Thielavia terrestris* Cel45	+	+

[a] Carbohydrate binding module (CBM): family 1, except Bl Cel9 (family 3b).
[b,c] Taxonomical identification of the two basidiomycetes remains inconclusive.
[d] The optimal and maximal growth temperatures are ~40–50 and 52–58 °C, respectively, for the thermophilic fungi, or ~55 and 65 °C, respectively, for the thermophilic actinobacteria, whose EGs were evaluated in this study.