Appl Microbiol Biotechnol
DOI 10.1007/s00253-011-3346-8

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

Fusion of a family 9 cellulose-binding module improves catalytic potential of *Clostridium thermocellum* cellodextrin phosphorylase on insoluble cellulose

Xinhao Ye · Zhiguang Zhu · Chenming Zhang · Y.-H. Percival Zhang

Received: 31 March 2011 / Revised: 2 May 2011 / Accepted: 3 May 2011

Abstract *Clostridium thermocellum* cellodextrin phosphorylase (*Ct*CDP), a single-module protein without an apparent carbohydrate-binding module, has reported activities on soluble cellodextrin with a degree of polymerization (DP) from two to five. In this study, *Ct*CDP was first discovered to have weak activities on weakly water-soluble celloheptaose and insoluble regenerated amorphous cellulose (RAC). To enhance its activity on solid cellulosic materials, four cellulose binding modules, e.g., CBM3 (type A) from *C. thermocellum* CbhA, CBM4-2 (type B) from *Rhodothermus marinus* Xyn10A, CBM6 (type B) from *Cellvibrio mixtus* Cel5B, and CBM9-2 (type C) from *Thermotoga maritima* Xyn10A, were fused to the C terminus of *Ct*CDP. Fusion of any selected CBM with *Ct*CDP did not influence its kinetic parameters on cellobiose but affected the binding and catalytic properties on celloheptaose and RAC differently. Among them, addition of CBM9 to *Ct*CDP resulted in a 2.7-fold increase of catalytic efficiency for degrading celloheptaose. *Ct*CDP-CBM9 exhibited enhanced specific activities over 20% on the short-chain RAC (DP=14) and more than 50% on the long-chain RAC (DP=164). The chimeric protein *Ct*CDP-CBM9 would be the first step to construct a cellulose phosphorylase for in vitro hydrogen production from cellulose by synthetic pathway biotransformation (SyPaB).

Keywords Cellodextrin phosphorylase · Carbohydrate-binding module (CBM) · Cellulose · Protein engineering

X. Ye · Z. Zhu · C. Zhang · Y.-H. P. Zhang (✉)
Biological Systems Engineering Department, Virginia Tech,
Blacksburg, VA 24061, USA
e-mail: ypzhang@vt.edu

Y.-H. P. Zhang
Institute for Critical Technology and Applied Science (ICTAS),
Virginia Tech,
Blacksburg, VA 24061, USA

Y.-H. P. Zhang
DOE Bioenergy Science Center,
Oak Ridge, TN 37831, USA

Introduction

Cellodextrin phosphorylase (CDP, EC 2.4.1.49) is responsible for catalyzing phosphorolysis of soluble cellodextrins with degree of polymerization (DP) of two to five (Sheth and Alexander 1967; Schomburg et al. 2009). It belongs to glycoside hydrolase (GH) family 94 (Hidaka et al. 2004) and plays an important role in the energy-efficient metabolism of long-chain cellodextrins (Lou et al. 1996; Zhang and Lynd 2004a). CDP activity has been reported in *Fibrobacter succinogenes*, *Vibrio splendidus*, *Spirochaeta thermophile*, and several species of *Cellulomonas* and *Clostridia* (Wells et al. 1995; Reichenbecher et al. 1997; Sheth and Alexander 1967), but the CDP enzymes have only been isolated from *Clostridium thermocellum* (Arai et al. 1994) and *Clostridium stercorarium* (Reichenbecher et al. 1997). Recently, a recombinant cellodextrin phosphorylase from *C. thermocellum* (*Ct*CDP) was expressed in *Escherichia coli*, purified, and characterized in detail (Krishnareddy et al. 2002). The recombinant *Ct*CDP has been used for synthesis of cellulase inhibitors (Kawaguchi et al. 1998), in vitro enzymatic hydrogen production (Ye et al. 2009), and the production of glucose 1-phosphate, cellodextrin, and highly ordered cellulose (Samain et al. 1995; Hiraishi et al. 2009).

Glycoside hydrolases usually consist of a catalytic module and one or more non-catalytic carbohydrate-binding modules (CBM; Bourne and Henrissat 2001; Boraston et al. 2004).

Published online: 01 June 2011

CBM can enrich enzymes on the surface of solid substrates through affinity adsorption, so to enhance their activity (Carrard et al. 2000; Shoseyov et al. 2006). Currently, CBMs have been classified into 61 sequence-based families (www.cazy.org) and further grouped into three classes based on their structure, function, and ligand specificities (Boraston et al. 2004). Families 1, 2a, 3, 5, and 10 fit in type A CBMs that have a planar hydrophobic ligand-binding surface and bind to crystalline cellulose. Type B CBMs, comprising families 4, 6, 17, 28, etc., contain clefts that accommodate single polysaccharide chains, while the ligand-binding sites in Type C CBMs, comprising families 9, 13, 14, etc., interact with mono- or disaccharides (Boraston et al. 2004; van Bueren et al. 2005).

Unlike other glycoside hydrolases, the *C. thermocellum* CDP is a single modular protein and does not have the apparent CBM. Recently, several studies have shown that addition of a CBM to single domain enzymes increased their activity toward insoluble substrates (Kittur et al. 2003; Carrard et al. 2000). Ravalason et al. (2009) presented that addition of a family 1 CBM (type A) of *Aspergillus niger* to *Pycnoporus cinnabarinus* laccase significantly improved the delignification capabilities of laccase to softwood kraft pulp. *Ruminococcus albus* endoglucanase fused with a family 6 CBM (type B) of *C. stercorarium* xylanase A exhibited higher molar activity toward insoluble acid-swollen cellulose and ball-milled cellulose (Karita et al. 1996). Adding the CBM6 to *Bacillus halodurans* xylanase also improved the enzyme's activity toward insoluble xylan (Mangala et al. 2003). Moreover, Maglione et al. (1992) reported a tenfold increase in the specific activity of *Prevotella ruminicola* endoglucanase on insoluble cellulose after being fused with a family 2 CBM (type A) from *Thermomonospora fusca* endoglucanase. It thus would be of great interest to investigate if fusion of a CBM to *Ct*CDP could potentiate its catalytic activity on insoluble cellulose.

Creation of a non-natural cellulose phosphosphorylase would be extremely important for high-yield hydrogen production because it would be possible to produce low-cost hydrogen from less costly pretreated solid cellulosic materials rather than soluble cellodextrins without costly ATP consumption via synthetic pathway biotransformation (SyPaB) (Ye et al. 2009; Zhang 2010). In this study, we found that *Ct*CDP had weak activities on weakly water-soluble celloheptaose and insoluble regenerated amorphous cellulose (RAC). Then, *Ct*CDP was linked with four representatives of the three types of CBM (Table 1) to test whether fusion of a CBM could improve the catalytic potential of *Ct*CDP on cellulose. Among the CBMs selected, CBM3, derived from *C. thermocellum* CbhA, belongs to type A CBM and binds to the surface of insoluble cellulose (Carrard et al. 2000; Hong et al. 2007); CBM4-2 from *Rhodothermus marinus* Xyn10A and CBM6 from *Cellvibrio mixtus* Cel5B are type B CBMs, which interact with amorphous cellulose and/or cellodextrin (Simpson et al. 2002; Henshaw et al. 2004); CBM9-2 of Xyn10A from *Thermotoga maritima* pertains to type C CBM, which has affinity with a broad range of cellulose, e.g., amorphous and crystalline cellulose, cello-/xylo-oligomers, and soluble mono- and disaccharides (Notenboom et al. 2001; Boraston et al. 2001). Here, we described the production, isolation, and characterization of the chimeric proteins *Ct*CDP-CBM3, *Ct*CDP-CBM4, *Ct*CDP-CBM6, and *Ct*CDP-CBM9. The results demonstrated that *Ct*CDP-CBM9 promoted the catalytic activity toward short-chain and/or long-chain cellulose. The active *Ct*CDP-CBM9 would be the first step to develop high-activity cellulose phosphorylase in the future.

Materials and methods

Materials

All chemicals were reagent-grade, purchased from Sigma (St. Louis, MO, USA), unless otherwise noted. *C. thermocellum* ATCC 27405 genomic DNA was a gift from Dr. Jonathan Mielenz at the Oak Ridge National Laboratory (Oak Ridge, TN, USA). *E. coli* JM109 was used for plasmid manipulation; *E. coli* Rosetta BL21 (DE3) containing the gene expression plasmid was employed for producing the recombinant protein. The Luria-Bertani (LB) medium was used with 100 μg/mL ampicillin (sodium salt). Microcrystalline cellulose—Avicel PH105—was purchased

Table 1 The origins and functions of the CBMs used in this study

CBM	Type	Origin	Organism	GenBank ID	Binding site	Reference
CBM3	A	CbhA	*C. thermocellum*	ABN51651.1	Crystalline and amorphous cellulose	Carrard et al. (2000)
CBM4-2	B	Xyn10A	*R. marinus*	CAA72323.2	Amorphous cellulose and cellodextrin	Hachem et al. (2000)
CBM6	B	Cel5B	*C. mixtus*	AAB61462.2	Cello-oligosaccharides	Henshaw et al. (2004)
CBM9-2	C	Xyn10A	*T. maritima*	AAD35155.1	Amorphous and crystalline cellulose, soluble polysaccharides, mono- and disaccharides	Boraston et al. (2001)

from FMC (Philadelphia, PA, USA). Cellodextrins with DP from two to seven was prepared by mixed acid hydrolysis of Avicel and separated by large size chromatographic column, as described before (Zhang and Lynd 2003). RAC was prepared from Avicel after water slurring, cellulose dissolution in H_3PO_4, and regeneration in water (Zhang et al. 2006a). Two types of RAC were generated in terms of different number-average degrees of polymerization (DP). Insoluble RAC DP 164 was prepared in concentrated phosphoric acid with partial hydrolysis in an ice bath for 1 h, while slightly soluble RAC DP 14 was prepared by partial hydrolysis in concentrated phosphoric acid at 50 °C for 20 h (Zhang and Lynd 2004b).

Plasmid construction

A pair of primers (p1, 5′-TTAAGA CATATG ATTACT AAAGTA ACAGCG AG-3′, and p2, 5′-AATTTT CTCGAG GAGCTC GGATCC TTTAAA CTTAAG AGTCAC TATATG TTC-3′, restriction enzyme cutting sites underlined) was employed to amplify full-length *cdp* from *C. thermocellum* genomic DNA. The PCR product was digested by *Nde*I and *Xho*I and ligated with the digested plasmid pET21c to give plasmid pET-*cdp*. The gene fragment of *cbm3* was amplified from the gDNA of *C. thermocellum* by primers p3 (5′-TTAAGA GAGCTC GTACAG TATTTG TGCGAA AATACG-3′, *Sac*I site underlined) and p4 (5′-ATTATC CTCGAG TTCCAG CTGCAG ATAATG CTC-3′, *Xho*I site underlined). The amino acid sequences of CBM4, CBM6, and CBM9 were taken from GenBank (their accessible numbers were listed in Table 1) and back-translated into *E. coli* optimized nucleotide sequences by OPTIMIZER (http://genomes.urv.es/OPTIMIZER/; Puigbò et al. 2007). The genes were then synthesized in Virginia Bioinformatics Institute (Blacksburg, VA, USA). In particular, the synthesized fragment of *cbm6* was joined with a natural N-terminal linker of family 6 CBM from *C. thermocellum* (GenBank ID: AAB61462.2, residue 310–331) via a *Sac*I site (GAGCTC). The fragment was digested by *Bam*HI and *Xho*I and then inserted into the digested pET21-*cdp* plasmid, yielding the plasmid pET-*cdp-cbm6*. Later, *cbm6* was replaced by *cbm3*, *cbm4*, and *cbm9* through *Sac*I and *Xho*I sites, resulting to pET-*cdp-cbm3*, pET-*cdp-cbm4*, and pET-*cdp-cbm9*, respectively. Finally, the nucleotide sequences of the chimeric genes were validated by DNA sequencing in MCLAB (San Francisco, CA, USA).

Protein production and purification

The *E. coli* BL21 strain harboring the expression plasmid was grown in the LB medium at 37 °C with a rotary shaking rate of 250 rpm. Once the absorbance at 600 nm (OD_{600}) reached ~0.8, IPTG was added to a final concentration of 0.25 mM. After 20 h of cultivation at 16 °C, the *E. coli* cells were harvested by centrifugation and re-suspended in a 50-mM HEPES buffer (pH 7.2). The cells were then lysed by sonication. After centrifugation, the soluble His-tagged *Ct*CDP or a chimeric protein was adsorbed to the Bio-Rad Profinity IMAC Ni-resin (Hercules, CA, USA) and was eluted by a HEPES buffer (50 mM, pH 7.2) with 250 mM imidazole and 0.3 M NaCl. The eluate containing the purified proteins was then exchanged into 50 mM HEPES buffer (pH 7.2) through GE Healthcare PD-10 desalting columns (Piscataway, NJ, USA) and concentrated by Centriprep centrifugal filter tubes with a 50,000 molecular weight cut-off membrane (Millipore, MA, USA).

Activity assays

All enzymatic reactions were conducted in 5-mL glass tubes (12×75 mm, Fisher Scientific). The enzyme activities were assayed at 30 °C in 50 mM HEPES buffer (pH 7.2) containing 30 mM cellobiose, 10 mM glucose 1-phosphate, 5 mM DTT, and 1 mM Mg^{2+} unless otherwise noted. Enzyme concentrations were set at 2.5 mg L^{-1} for most assays. To measure the specific activities on RAC DP 14 and RAC DP 164, sugar concentration was increased to 6.3 g [glucose equivalent] L^{-1} and 7.5 g [glucose equivalent] L^{-1}, respectively. Enzyme concentrations were 0.01 g L^{-1} on RAC DP 14 and 0.05 g L^{-1} on RAC DP 164. The reactions were stopped by placing the tubes in a boiling water bath for 5 min. The temperature and pH optimum was determined as described elsewhere (Ye et al. 2010).

Enzyme kinetics

For the synthesis direction, kinetic parameters were determined based on the initial rates by measuring the release of inorganic phosphate (P_i). The reactions were conducted at 30 °C in a 50-mM HEPES buffer (pH 7.2) containing 1 mM Mg^{2+}, 5 mM DTT, and various substrate concentrations between 0.2 and 5 times of their respective K_m values. The product P_i released from glucose 1-phosphate was measured by the mild pH phosphate assay as described elsewhere (Ye et al. 2010). One unit of phosphorylase in the synthesis direction was defined as the amount of enzyme that generates 1 μmole of phosphate per minute. For the degradation direction, the reactions were conducted at 30 °C in a 50-mM HEPES buffer (pH 7.2) containing 1 mM Mg^{2+}, 5 mM DTT, and P_i at different concentrations. The product was measured by using a glucose hexokinase/glucose 6-phosphate dehydrogenase assay kit (Zhang and Lynd 2004a) supplemented with a recombinant *C. thermocellum* phosphoglucomutase (Wang and Zhang 2010). One unit of phosphorylase in the degradation direction is defined

as the amount of enzyme generating 1 μmole of glucose 1-phosphate per minute.

Binding assays

Binding assays were carried out in 2-ml glass screw-cap vials. Each vial contained 0.2 g RAC/L, 0.05–0.6 g L^{-1} protein, and 50 mM HEPES buffer (pH 7.2). The vials were incubated with slow rotation for 1 h at 30 °C. The binding substrate was pelleted by centrifugation (5 min, 16,000×*g*), and the supernatants were tested for protein content. The maximum adsorption capacity A_{max} was determined by regressing binding isotherm data to a modified Langmuir-type binding model as described previously (Hong et al. 2007).

Other assays

Protein concentrations were determined by the Bradford methods with bovine serum albumin (BSA) as the standard. The average DP of RAC was determined as described elsewhere (Zhang and Lynd 2005). The purity of the enzymes was checked by SDS-PAGE.

Results

Construction, production, and purification of the fusion proteins

Expression plasmid pET-*cdp* was first constructed to produce wild-type CDP. Since c*bm3* from the *C. thermocellum Cbh*A had been successfully expressed in *E. coli*, it was directly amplified from the genomic DNA of *C. thermocellum* without further optimization. After codon optimization, however, *cbm*4-2 from *R. marinus Xyn*10A, *cbm*6 from *C. mixtus Cel*5B, and *cbm*9-2 from *T. maritima Xyn*10A were synthesized to overcome the potential codon-bias problem (Kurland 1991). The DNA fragment of each CBM was inserted into pET-*cdp* to generate the chimeric gene, and a nature linker from *C. thermocellum* xylosidase (Cthe 2196) was placed in the junction between *cdp* and *cbm* (Fig. 1).

Around 40 mg of *Ct*CDP and four different fusion proteins were produced per liter LB medium. They were purified by immobilized metal affinity chromatography (IMAC) with yields of ca. 20%. SDS-PAGE analysis indicated that the purified proteins were homogenous, and that the observed molecular weights were consistent with the estimated values, i.e., 115 kDa for *Ct*CDP, 126 kDa for *Ct*CDP-CBM3, 134 kDa for *Ct*CDP-CBM4, 130 kDa for *Ct*CDP-CBM6, and 138 kDa of *Ct*CDP-CBM9 (Fig. 2).

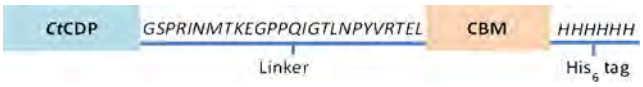

Fig. 1 A generalized structure of CDP-CBM fusion proteins. *CtCDP Clostridium thermocellum* cellodextrin phosphorylase, *CBM* carbohydrate-binding module

Basic characteristics of the fusion proteins

Since fusion of CBM may alter the activity and thermal stability of the catalytic domains (Kittur et al. 2003; Karita et al. 1996), the influences of pH and temperature on the activities of fusion proteins were examined. Every chimeric protein displayed the same pH and temperature optimum as the wild-type *Ct*CDP (data not shown). The maximum activities were detected at pH 7.0 and 60 °C, in agreement with previous report (Krishnareddy et al. 2002).

Kinetics of *Ct*CDP and fusion proteins were determined in both synthesis and degradation directions on cellobiose and celloheptaose. For cellobiose, similar values of K_m and k_{cat} were observed in the synthesis direction among the wild-type *Ct*CDP and the fusion proteins, whereas no activities were detected in the degradation direction (Table 2), consistent with the previous results (Sheth and Alexander 1967; Krishnareddy et al. 2002). The comparable kinetic parameters on cellobiose and optimal pH and temperature of *Ct*CDP and the fusion proteins suggested that fusion of CBM had little influence on the structure and function of the catalytic domain of *Ct*CDP on soluble substrate.

CBMs have various impacts on *Ct*CDP activities on celloheptaose (Table 3). With comparison to the wild-type *Ct*CDP, fusion of CBM3 decreased the turnover rates and the affinity between the enzyme and the substrate, resulting in reduced k_{cat} and increased K_m in both directions. In addition, the catalytic efficiencies (k_{cat}/K_m) were reduced from 2.41 to 0.56 mM^{-1} s^{-1} in the synthesis direction and from 2.27 to 1.00 mM^{-1} s^{-1} in the degradation direction. Similar effects were also observed in case of CBM6, even though family 6 CBM has different binding properties from CBM3. In contrast with *Ct*CDP, the catalytic efficiencies of *Ct*CDP-CBM6 were decreased to 0.41 and 0.90 mM^{-1} s^{-1} in the directions of glucan synthesis and degradation, respectively. Furthermore, addition of CBM4 and CBM9 attributed little to the turnover rate of *Ct*CDP, but promoted the Michaelis constant in different directions. Fusion of

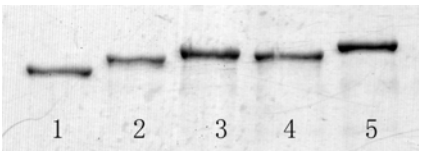

Fig. 2 Purity of the fusion proteins. *Lane 1*, CDP (MW, 115 kDa); *lane 2*, CDP-CBM3 (MW, 126 kDa); *lane 3*, CDP-CBM4 (MW, 134 kDa); *lane 4*, CDP-CBM6 (MW, 130 kDa); *lane 5*, CDP-CBM9 (MW, 138 kDa). *MW* molecular weight

Table 2 Kinetics of CDP-CBM fusion proteins on cellobiose

	Synthesis direction			Degradation direction		
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m
*Ct*CDP	0.78±0.02	2.67±0.04	3.42	ND	ND	–
*Ct*CDP-CBM3	0.59±0.03	1.83±0.08	3.10	ND	ND	–
*Ct*CDP-CBM4	0.53±0.01	2.07±0.08	3.90	ND	ND	–
*Ct*CDP-CBM6	0.76±0.02	2.08±0.05	2.73	ND	ND	–
*Ct*CDP-CBM9	1.08±0.01	2.55±0.02	2.36	ND	ND	–

ND Non Detectable

CBM4 caused a four-fold decrease of K_m in the synthesis direction, while addition of CBM9 lowered K_m by nearly three folds in the degradation direction. Consequently, the catalytic efficiencies were increased to 7.76 mM^{-1} s^{-1} for synthesis and to 6.08 mM^{-1} s^{-1} for degradation by the addition of family 4 and family 9 CBM, respectively.

Binding properties on insoluble cellulose

The binding properties of *Ct*CDP and the chimeric proteins on insoluble cellulose were studied by incubating the proteins with RAC. Figure 3 showed the binding isotherms of *Ct*CDP and the fusion proteins on RAC DP 14 and RAC DP 164. All of the fusion proteins had smaller dissociation constants (K_p) than *Ct*CDP. It suggested that CBMs enhanced the binding between enzyme and RAC. At 30 °C, the free energy of adsorption (ΔG) (Boraston et al. 2001) between *Ct*CDP and RAC DP 14 was −38.2 kJ mol^{-1}, very close to the energy change as *Ct*CDP bound to RAC DP 164 (ΔG=−38.1 kJ mol^{-1}; Table 4). However, the interaction between the fusion proteins with short-chain RAC had smaller energy change (ΔG) than that on long-chain RAC, except in the case of *Ct*CDP-CBM4 that preferentially bound to RAC DP 164.

Addition of CBM3, CBM6, or CBM9 increased the maximum adsorption capacities (A_{max}) of *Ct*CDP to both RACs. In particular, *Ct*CDP-CBM9 had A_{max} approximately six-fold and more than nine-fold higher than that of *Ct*CDP on RAC DP 14 and RAC DP 164, respectively (Table 4). By contrast, fusion of CBM4 lessened the binding capacities on the short-chain RAC (Fig. 3a), but promoted binding capacities to the long-chain RAC (Fig. 3b). The corresponding A_{max} were 0.32 μmol [*Ct*CDP-CBM4] g^{-1} [RAC DP 14] and 2.92 μmol [*Ct*CDP-CBM4] g^{-1} [RAC DP 164] (Table 4).

Activities on amorphous cellulose

Preliminary experiments had shown that wild-type *Ct*CDP was able to weakly phosphorolyse insoluble amorphous cellulose (Fig. 4). In order to obtain meaningful results for activity assay, the reactions on RAC DP 14 and RAC DP 164 were performed under different substrate and enzyme concentration. It was quite clear that *Ct*CDP, along with the chimeric proteins, had higher activities on RAC DP 14 than on RAC DP 164, especially in the degradation direction (Fig. 4). The difference may stem from the fact that reaction on RAC DP 14 started with a higher concentration of reducing end ([RE]=2.8 mM), where phosphorolysis happens (Samain et al. 1995), than that on RAC DP 164 in which [RE] was 0.28 mM.

On RAC DP 164, *Ct*CDP and its fusion proteins, like other phosphorylases (e.g., maltodextrin phosphorylase), had higher activities of extending the glucan chain than of degrading it (Fig. 4b) because the conversion of phosphate into glucose 1-phosphate is thermodynamically unfavorable (Nidetzky et al. 1996). Nevertheless, a reverse trend was found on RAC DP 14. Both *Ct*CDP and CBM-tagged *Ct*CDPs degraded the short-chain RAC more efficiently than extending it (Fig. 4a). As the slightly soluble RAC DP 14 was degraded, the cascade reaction generated shorter substrates which were more soluble and more accessible to CDP. Thus, the degradation reaction may be accelerated.

In addition, fusion of CBM3 and CBM6 reduced the activities of *Ct*CDP on amorphous cellulose. Fused CBM4

Table 3 Kinetics of CDP-CBM fusion proteins on celloheptaose

	Synthesis direction			Degradation direction		
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m
*Ct*CDP	1.29±0.02	3.11±0.05	2.41	1.82±0.03	4.14±0.10	2.27
*Ct*CDP-CBM3	1.93±0.04	1.08±0.02	0.56	2.71±0.02	2.70±0.01	1.00
*Ct*CDP-CBM4	0.29±0.05	2.25±0.05	7.76	1.56±0.01	4.53±0.03	2.90
*Ct*CDP-CBM6	5.24±0.05	2.16±0.02	0.41	1.53±0.01	1.37±0.02	0.90
*Ct*CDP-CBM9	1.46±0.02	2.98±0.04	2.04	0.62±0.02	3.77±0.06	6.08

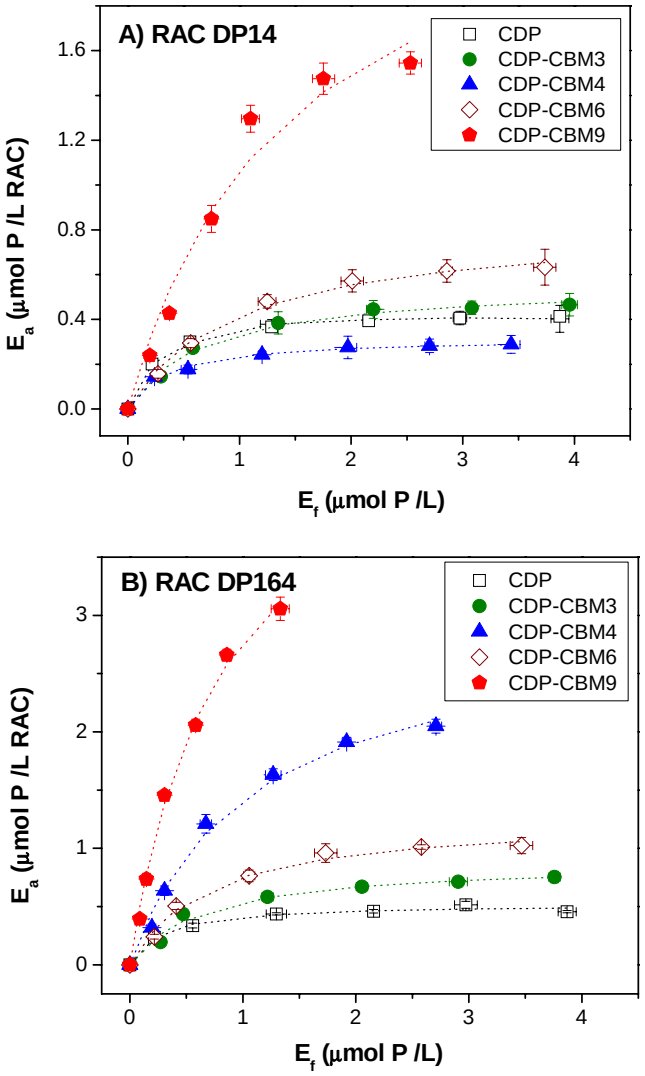

Fig. 3 Binding isotherms of CDP-CBM fusion proteins to RAC. **a** Adsorption on RAC DP 14. **b** Adsorption on RAC DP 164

had little impact on *Ct*CDP, although *Ct*CDP-CBM4 displayed nearly 10% increases of catalytic efficiencies on RAC DP 14. Comparatively, CBM9 improved the specific activities of *Ct*CDP in both synthesis and degradation directions over 20% on the short-chain RAC and more than 50% on the long-chain RAC. The specific activities of *Ct*CDP-CBM9 reached 18.9 and 5.9 U μmol^{-1} toward synthesis and degradation of RAC DP 14, respectively. One micromolar *Ct*CDP-CBM9 also possessed 4.1 U for extension and 1.2 U for degradation of RAC DP 164 (Fig. 4).

Discussion

Surface accessibility and end accessibility are recognized as two important factors in the enzymatic hydrolysis of cellulose (Rollin et al. 2010; Zhang and Lynd 2006). In nature, cellulose hydrolases developed a complex molecular architecture in which catalytic modules are appended to non-catalytic CBM that enrich enzyme onto its substrate, thereby increasing possibility of enzyme–substrate complex formation (McCartney et al. 2006). The *C. thermocellum* CDP, a single modular protein, does not have an apparent CBM. We discovered that it has a very weak activity toward amorphous cellulose. To mimic cellulase performance, we constructed four *Ct*CDP-CBM fusion proteins and characterized them. Among them, fusion of a family 9-2 CBM from *T. maritima* Xyn10A expands the catalytic potential of *Ct*CDP towards insoluble cellulose.

Being a type C CBM, the family 9 CBM from *T. maritima* has a broad binding specificity (Table 1). Addition of CBM9 to *Ct*CDP did not influence its kinetics on soluble cellobiose, but increased the binding capacities on insoluble cellulose. It likewise improved the catalytic activities, to different extents, on celloheptaose and RAC. The carbohydrate-binding site of CBM9 has been described as a "blind canyon," a groove on protein surface that is blocked at one end. Such a conformation accommodates only two sugar rings and makes CBM9 specifically bind to

Table 4 Modified Langmuir isotherm parameters for the adsorption of *Ct*CDP and its fusion proteins on regenerated amorphous cellulose (RAC)

	RAC DP 14			RAC DP 164		
	A_{max}[a] (μmol g^{-1})	K_p[b] (L μmol^{-1})	ΔG[c] (kJ mol^{-1})	A_{max}[a] (μmol g^{-1})	K_p[b] (L μmol^{-1})	ΔG[c] (kJ mol^{-1})
*Ct*CDP	0.44±0.04	3.83±0.35	−38.2	0.52±0.07	3.63±0.49	−38.1
*Ct*CDP-CBM3	0.56±0.03	1.51±0.08	−35.9	0.88±0.05	1.58±0.09	−36.0
*Ct*CDP-CBM4	0.32±0.02	2.90±0.18	−37.5	2.92±0.16	0.95±0.05	−34.7
*Ct*CDP-CBM6	0.82±0.06	1.03±0.07	−34.9	1.27±0.09	1.45±0.10	−35.8
*Ct*CDP-CBM9	2.55±0.10	0.71±0.03	−34.0	4.93±0.14	1.28±0.04	−35.4

[a] A_{max}, maximum adsorption capacity (μmol [enzyme] g^{-1} [RAC])

[b] K_p, dissociation constant (L μmol^{-1} [enzyme])

[c] ΔG, free energy of adsorption (kJ mol^{-1}). $\Delta G = -RT\cdot\ln(K_p)$ (Boraston et al. 2001), where R is a gas constant, and T represents temperature (K, T=303.15 K)

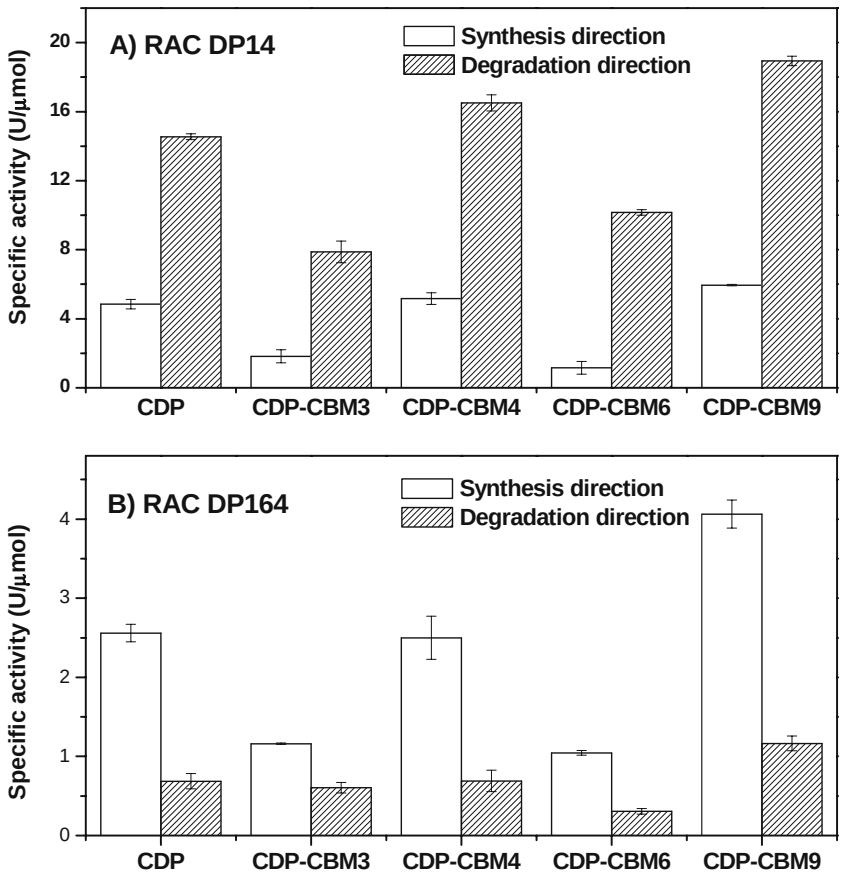

Fig. 4 Specific activities of CDP-CBM fusion proteins to regenerated amorphous cellulose. **a** Specific activities on RAC DP 14. **b** Specific activities on RAC DP 164

the reducing ends of sugars (Notenboom et al. 2001). Although both CBM9 and *Ct*CDP bind to the reducing end of cellulose, it was doubtful the sugar residues could be transferred from the binding site of CBM9 to the catalytic domain of *Ct*CDP. However, fusion of CBM9 may direct the enzyme toward certain reducing end-enriched area so that the fusion protein *Ct*CDP-CBM9 could show higher activities on insoluble cellulose than *Ct*CDP. Moreover, fusion of CBM9 considerably increased the specific activities on RAC DP 164 while only marginally improved the activities on RAC DP 14, suggesting that the enriching benefit of CBM9 fusion strategy was limited on RAC DP 14 because of much higher reducing end availability.

Our results also suggested that different CBMs, with different binding specificities, affected *Ct*CDP to different extents. For instance, fusion of family 3 CBM from *C. thermocellum Cbh*A improved enzyme binding to insoluble cellulose but significantly decreased the activities on both celloheptaose and RAC. CBM3, a typical type A CBM, targeted *Ct*CDP-CBM3 to the crystalline surface of cellulose where the trapped *Ct*CDP lost its freedom to capture the reducing ends of glucan chains, resulting in decreased activities. On the other hand, CBM3 cannot hold cellobiose. Therefore, *Ct*CDP-CBM3 and *Ct*CDP exhibited the similar kinetics toward cellobiose (Table 2).

Type B CBM6 from *C. mixtus* Cel5B has two potential ligand-binding sites (Henshaw et al. 2004). One (cleft A) is located within the loops connecting two β-sheets and recognizes two sugar residues from the non-reducing end, whereas the concave surface cleft B accommodates three to four saccharide units in both orientations. When *Ct*CDP-CBM6 was mixed with cellobiose, cleft A and *Ct*CDP may each adopt a cellobiose unit. Given that *Ct*CDP-CBM6 retained the activity of *Ct*CDP on cellobiose, it could be concluded that two binding processes happened in non-interacting sites and were independent of each other. On the contrary, fusion of CBM6 lowered the activities of *Ct*CDP on insoluble celluloses, including celloheptaose, RAC DP 14, and RAC DP 164. Since binding of CBM6 to insoluble cellulose involves synergistic interactions between cleft A and cleft B, cleft A may draw *Ct*CDP-CBM6 to the non-reducing termini once the fusion protein interacts with the cellulose. Similar to the case of *Ct*CDP-CBM3, it lessened the possibility of *Ct*CDP to reach the reducing end of cellulose, resulting in declined activities.

Through the comparison of the unequal impacts of fused CBMs on *Ct*CDP, it was concluded that fusion of CBM that anchors the enzyme at some specific positions may not promote the catalytic potential on the insoluble substrate. A desired CBM should direct the enzyme to the sites where

the enzyme would form active enzyme–substrate complex. The success of chimeric *Ct*CDP-CBM9 over the wild-type *Ct*CDP provided a good example here. Nevertheless, it was worth noting that substrate properties also exerted different effects on the fusion protein. For example, CBM 4-2 from *R. marinus Xyn*10A is a type B CBM interacting with a range of single-stranded polysaccharides, including amorphous (but not crystalline) cellulose, β-1,3-glucan, and xylan, with approximately equal affinity in both orientations (Hachem et al. 2000). Different from other fusion proteins, *Ct*CDP-CBM4 exhibited distinct binding and catalytic properties towards RAC DP 14 and RAC DP 164. On RAC DP 14, *Ct*CDP-CBM4 displayed a smaller A_{max} but higher activities than *Ct*CDP, whereas *Ct*CDP-CBM4 had a bigger A_{max} on RAC DP 164, but similar activities as *Ct*CDP in both synthesis and degradation directions. For the latter observation, an increase of binding capacities may not increase enzyme concentration around its target substrate sites (e.g., reducing ends of β-glucan chain), similar to the above discussion on *Ct*CDP-CBM3 and *Ct*CDP-CBM6. However, *Ct*CDP-CBM4 with low binding capacities but high catalytic activities on RAC DP 14 may be explained in respect of substrate properties, including solubility, chain length distribution, and accessibility of the reducing ends. Substrate solubility may play a critical role. Since RAC DP 14 is partially soluble (Zhang and Lynd 2004b) and the binding cleft of CBM4 prefers soluble substrate (Simpson et al. 2002; Johnson et al. 1996), the soluble fraction of RAC DP 14 attracted a large amount of *Ct*CDP-CBM4 in supernatant, leading to the decrease of A_{max} and the increase of K_p and absolute ΔG. Meanwhile, the interaction between soluble RAC with *Ct*CDP-CBM4 helped the enzyme access to the reducing ends and thus increased its activities.

Moreover, kinetic analysis demonstrated fusion of CBM4 and CBM9 promoted the enzyme affinities on celloheptaose in different directions. Note that CBM4 and CBM9 bind to different sites of cellulose (McLean et al. 2002; Blake et al. 2006). Above result implies CBM4 might direct the enzyme to the sites of celloheptaose where glucan synthesis was preferred, while CBM9 concentrates the fusion protein to the sites that favors glucan degradation. It was difficult to evaluate the validity of this hypothesis. However, it clearly suggested complex interactions between binding modules and catalytic modules in enzyme, between binding sites and catalytic sites in substrate, and between enzymes and substrates (Liu et al. 2010).

The linker joining *Ct*CDP with CBM is another factor that could potentially affect both binding and catalytic properties of the fusion protein (Liu et al. 2010). In this work, a natural linker was cloned from *C. thermocellum* Cthe 2196. It is a short and flexible peptide consisting of repeated units of proline, threonine, and glycine (Fig. 1). The sequence shows no putative consensus motif in the GenBank database. For construction of CBM-fused protein, the synthetic CBM domain routinely includes a fragment of its natural linker, if not all, to alleviate the possible steric hindrance of CBM on the activity of chimeric enzymes (Ravalason et al. 2009; Ahn et al. 2004; Kittur et al. 2003). However, it was still unclear the real effect of linker on the catalytic potential of chimeric enzymes. Concerted efforts would be focused on optimizing the linker length and composition for maximizing the catalytic potential of *Ct*CDP-CBM9.

Increasing enzyme activities on insoluble substrates is challenging (Zhang et al. 2006b; Wen et al. 2009). It is difficult to develop an active enzyme against insoluble substrates by directed evolution owing to the lack of reliable and efficient screening/selection methods (Liu et al. 2009). Since the mechanism how enzymes work with insoluble substrates is still unknown, engineering an enzyme toward insoluble substrates by rational design has been scarcely reported so far (Li et al. 2010). In this study, we discovered that *Ct*CDP had weak activities on celloheptaose and RAC. To test whether addition of a CBM could improve the catalytic potential of *Ct*CDP on insoluble cellulose, *Ct*CDP was linked with four representatives of the three types of CBM (Table 1). The results demonstrated that addition of a CBM 9-2 from *T. maritima* Xyn10A significantly improved the catalytic activity of *Ct*CDP on both short-chain and long-chain RAC. The chimeric protein will be the first step to construct a highly active cellulose phosphorylase for the in vitro hydrogen production by SyPaB. The enhanced activity on RAC also makes it possible to develop an efficient screening/selection method for directed evolution of *Ct*CDP-CBM9 (Liu et al. 2010; Zhang and Zhang 2011). The activity of *Ct*CDP-CBM9 on solid cellulose might be further enhanced by rational design of linker length and composition in the future.

Acknowledgments This work was not possible without support from the Biological Systems Engineering Department of Virginia Tech, the Air Force Office of Scientific Research (FA9550-08-1-0145), the USDA Biodesign and Bioprocess Center, and DOE BESC to YPZ.

References

Ahn JO, Choi ES, Lee HW, Hwang SH, Kim CS, Jang HW, Haam SJ, Jung JK (2004) Enhanced secretion of *Bacillus stearothermophilus* L1 lipase in *Saccharomyces cerevisiae* by translational fusion to cellulose-binding domain. Appl Microbiol Biotechnol 64:833–839

Arai M, Tanaka K, Kawaguchi T (1994) Purification and properties of cellodextrin phosphorylase from *Clostridium thermocellum*. J Ferment Bioeng 77:239–242

Blake AW, McCartney L, Flint JE, Bolam DN, Boraston AB, Gilbert HJ, Knox JP (2006) Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. J Biol Chem 281:29321–29329

Boraston AB, Creagh AL, Alam MM, Kormos JM, Tomme P, Haynes CA, Warren RAJ, Kilburn DG (2001) Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. Biochemistry 40:6240–6247

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781

Bourne Y, Henrissat B (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr Opin Struct Biol 11:593–600

Carrard G, Koivula A, Söderlund H, Béguin P (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. Proc Nat Acad Sci USA 97:10342–10347

Hachem MA, Nordberg Karlsson E, Bartonek-Roxâ E, Raghothama S, Simpson PJ, Gilbert HJ, Williamson MP, Holst O (2000) Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. Biochem J 345:53–60

Henshaw JL, Bolam DN, Pires VMR, Czjzek M, Henrissat B, Ferreira LMA, Fontes CMGA, Gilbert HJ (2004) The family 6 carbohydrate binding module *Cm*CBM6-2 contains two ligand-binding sites with distinct specificities. J Biol Chem 279:21552–21559

Hidaka M, Honda Y, Kitaoka M, Nirasawa S, Hayashi K, Wakagi T, Shoun H, Fushinobu S (2004) Chitobiose phosphorylase from *Vibrio proteolyticus*, a member of glycosyl transferase family 36, has a clan GH-L-like (α/α)6 barrel fold. Structure 12:937–947

Hiraishi M, Igarashi K, Kimura S, Wada M, Kitaoka M, Samejima M (2009) Synthesis of highly ordered cellulose II in vitro using cellodextrin phosphorylase. Carbohydr Res 344:2468–2473

Hong J, Ye X, Zhang YHP (2007) Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. Langmuir 23:12535–12540

Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP (1996) Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. Biochemistry 35:14381–14394

Karita S, Sakka K, Ohmiya K (1996) Cellulose-binding domains confer an enhanced activity against insoluble cellulose to *Ruminococcus albus* endoglucanase IV. J Ferment Bioeng 81:553–556

Kawaguchi T, Ikeuchi Y, Tsutsumi N, Kan A, Sumitani J-I, Arai M (1998) Cloning, nucleotide sequence, and expression of the *Clostridium thermocellum* cellodextrin phosphorylase gene and its application to synthesis of cellulase inhibitors. J Ferment Bioeng 85:144–149

Kittur FS, Mangala SL, Rus'd AA, Kitaoka M, Tsujibo H, Hayashi K (2003) Fusion of family 2b carbohydrate-binding module increases the catalytic activity of a xylanase from *Thermotoga maritima* to soluble xylan. FEBS Lett 549:147–151

Krishnareddy M, Kim YK, Kitaoka M, Mori Y, Hayashi K (2002) Cellodextrin phosphorylase from *Clostridium thermocellum* YM4 strain expressed in *Escherichia coli*. J Appl Glycosci 49:1–8

Kurland CG (1991) Codon bias and gene expression. FEBS Lett 285:165–169

Li Y, Irwin DC, Wilson DB (2010) Increased crystalline cellulose activity via combinations of amino acid changes in the family 9 catalytic domain and family 3c cellulose binding module of *Thermobifida fusca* Cel9A. Appl Environ Microbiol 76:2582–2588

Liu W, Hong J, Bevan DR, Zhang YHP (2009) Fast identification of thermostable β-glucosidase mutants on cellobiose by a novel combinatorial selection/screening approach. Biotechnol Bioeng 103:1087–1094

Liu W, Zhang X, Zhang Z, Zhang YHP (2010) Engineering of *Clostridium phytofermentans* endoglucanase Cel5A for improved thermostability. Appl Environ Microbiol 76:4914–4917

Lou J, Dawson K, Strobel H (1996) Role of phosphorolytic cleavage in cellobiose and cellodextrin metabolism by the ruminal bacterium *Prevotella ruminicola*. Appl Environ Microbiol 62:1770–1773

Maglione G, Matsushita O, Russell JB, Wilson DB (1992) Properties of a genetically reconstructed *Prevotella ruminicola* endoglucanase. Appl Environ Microbiol 58:3593–3597

Mangala SL, Kittur FS, Nishimoto M, Sakka K, Ohmiya K, Kitaoka M, Hayashi K (2003) Fusion of family VI cellulose binding domains to *Bacillus halodurans* xylanase increases its catalytic activity and substrate-binding capacity to insoluble xylan. J Mol Catal B Enzym 21:221–230

McCartney L, Blake AW, Flint J, Bolam DN, Boraston AB, Gilbert HJ, Knox JP (2006) Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. Proc Natl Acad Sci USA 103:4765–4770

McLean BW, Boraston AB, Brouwer D, Sanaie N, Fyfe CA, Warren RAJ, Kilburn DG, Haynes CA (2002) Carbohydrate-binding modules recognize fine substructures of cellulose. J Biol Chem 277:50245–50254

Nidetzky B, Weinhäusel A, Haltrich D, Kulbe KD, Schinzel R (1996) Maltodextrin phosphorylase from *Escherichia coli*: production and application for the synthesis of α-glucose-1-phosphate. Ann NY Acad Sci 782:208–218

Notenboom V, Boraston AB, Kilburn DG, Rose DR (2001) Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. Biochemistry 40:6248–6256

Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 35:W126–W131

Ravalason H, Herpoël-Gimbert I, Record E, Bertaud F, Grisel S, de Weert S, van den Hondel CAMJJ, Asther M, Petit-Conil M, Sigoillot J-C (2009) Fusion of a family 1 carbohydrate binding module of *Aspergillus niger* to the *Pycnoporus cinnabarinus* laccase for efficient softwood kraft pulp biobleaching. J Biotechnol 142:220–226

Reichenbecher M, Lottspeich F, Bronnenmeier K (1997) Purification and properties of a cellobiose phosphorylase (CepA) and a cellodextrin phosphorylase (CepB) from the cellulolytic thermophile *Clostridium stercorarium*. Eur J Biochem 247:262–267

Rollin JA, Zhu Z, Sathitsuksanoh N, Zhang YHP (2010) Increasing cellulose accessibility is more important than removing lignin: a comparison of cellulose solvent-based lignocellulose fractionation and soaking in aqueous ammonia. Biotechnol Bioeng 108:22–30

Samain E, Lancelon-Pin C, Férigo F, Moreau V, Chanzy H, Heyraud A, Driguez H (1995) Phosphorolytic synthesis of cellodextrins. Carbohydr Res 271:217–226

Schomburg D, Schomburg I, Chang A (2009) Class 2: transferases IV. In: Schomburg D, Schomburg I (eds) Springer handbook of enzymes, vol 31. Springer, New York, pp 434–438

Sheth K, Alexander JK (1967) Cellodextrin phosphorylase from *Clostridium thermocellum*. Biochim Biophys Acta, Gen Subj 148:808–810

Shoseyov O, Shani Z, Levy I (2006) Carbohydrate binding modules: biochemical properties and novel applications. Microbiol Mol Biol Rev 70:283–295

Simpson PJ, Jamieson SJ, Abou-Hachem M, Karlsson EN, Gilbert HJ, Holst O, Williamson MP (2002) The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. Biochemistry 41:5712–5719

van Bueren AL, Morland C, Gilbert HJ, Boraston AB (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of β-1,3-linked glucans by presenting a unique ligand binding surface. J Biol Chem 280:530–537

Wang Y, Zhang Y-HP (2010) A highly active phosphoglucomutase from *Clostridium thermocellum*: cloning, purification, characterization, and enhanced thermostability. J Appl Microbiol 108:39–46

Wells J, Russell J, Shi Y, Weimer P (1995) Cellodextrin efflux by the cellulolytic ruminal bacterium *Fibrobacter succinogenes* and its potential role in the growth of nonadherent bacteria. Appl Environ Microbiol 61:1757–1762

Wen F, Nair NU, Zhao HM (2009) Protein engineering in designing tailored enzymes and microorganisms for biofuels production. Curr Opin Biotechnol 20:412–419

Ye X, Wang Y, Hopkins RC, Adams MWW, Evans BR, Mielenz JR, Zhang YHP (2009) Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails. ChemSusChem 2:149–152

Ye X, Rollin J, Zhang Y-HP (2010) Thermophilic α-glucan phosphorylase from *Clostridium thermocellum*: cloning, characterization and enhanced thermostability. J Mol Catal B Enzym 65:110–116

Zhang YHP (2010) Renewable carbohydrates are a potential high density hydrogen carrier. Int J Hydrogen Energy 35:10334–10342

Zhang YHP, Lynd LR (2003) Cellodextrin preparation by mixed-acid hydrolysis and chromatographic separation. Anal Biochem 322:225–232

Zhang YHP, Lynd LR (2004a) Kinetics and relative importance of phosphorolytic and hydrolytic cleavage of cellodextrins and cellobiose in cell extracts of *Clostridium thermocellum*. Appl Environ Microbiol 70:1563–1569

Zhang YHP, Lynd LR (2004b) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng 88:797–824

Zhang YHP, Lynd LR (2005) Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis. Biomacromolecules 6:1510–1515

Zhang YHP, Lynd LR (2006) A functionally based model for hydrolysis of cellulose by fungal cellulase. Biotechnol Bioeng 94:888–898

Zhang XZ, Zhang YHP (2011) Simple, fast and high-efficiency transformation system for directed evolution of cellulase in *Bacillus subtilis*. Microb Biotechnol 4:98–105

Zhang YHP, Cui JB, Lynd LR, Kuang LR (2006a) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidences from enzymatic hydrolysis and supramolecular structure. Biomacromolecules 7:644–648

Zhang YHP, Himmel ME, Mielenz JR (2006b) Outlook for cellulase improvement: screening and selection strategies. Biotechnol Adv 24:452–481