Published in final edited form as:

Methods Mol Biol. 2014 ; 1118: 71–95. doi:10.1007/978-1-62703-782-2_5.

Cell-Free Translation of Biofuel Enzymes

Taichi E. Takasuka, Johnnie A. Walker, Lai F. Bergeman, Kirk A. Vander Meulen, Shin-ichi Makino, Nathaniel L. Elsen, and Brian G. Fox

Abstract

In nature, bacteria and fungi are able to utilize recalcitrant plant materials by secreting a diverse set of enzymes. While genomic sequencing efforts offer exhaustive lists of genes annotated as potential polysaccharide-degrading enzymes, biochemical and functional characterizations of the encoded proteins are still needed to realize the full potential of this natural genomic diversity. This chapter outlines an application of wheat germ cell-free translation to the study of biofuel enzymes using genes from *Clostridium thermocellum*, a model cellulolytic organism. Since wheat germ extract lacks enzymatic activities that can hydrolyze insoluble polysaccharide substrates and is likewise devoid of enzymes that consume the soluble sugar products, the cell-free translation reactions provide a clean background for production and study of the reactions of biofuel enzymes. Examples of assays performed with individual enzymes or with small sets of enzymes obtained directly from cell-free translation are provided.

Keywords

Wheat germ extract; Transcription; Translation; Glycoside hydrolase; Cellulase; Xylanase; Mannanase; Functional screen

1 Introduction

The availability of cheap energy has been a major factor supporting economic development worldwide during the past century [1, 2]. Among the many available energy sources, liquid fuels, which are primarily used for transportation, are the most valuable but are also increasingly problematic, with major perturbations in their supply and price occurring in the mid-1980s and again after 2000. Given this recurring state of affairs, efforts to secure new sources of liquid fuels have intensified worldwide, with ongoing research in academic, industrial, and government settings.

The Great Lakes Bioenergy Research Center is a multi-institution collaborative effort funded by the US Department of Energy, with a mission to perform basic research needed to generate and improve technology to convert cellulosic biomass to ethanol and advanced biofuels. Methods described in this chapter have been developed and carried out in the Great Lakes Bioenergy Research Center in order to evaluate the genomic contents of new isolates from highly cellulolytic environmental niches and to better understand the capabilities of

model cellulolytic organisms. This approach is also applicable to engineering of multi-protein mixtures for improved deconstruction of biomass.

1.1 Overview of Biomass Deconstruction Needed for Biofuel Production

The complexities of biomass, particularly the diversity of cellulosic and hemicellulosic substructures that make up a large portion of plants, make the economic utilization of cellulosic biomass a difficult challenge. Currently, biomass is chemically pretreated to achieve a partial decomposition or extraction of lignin, which exposes the entrained polysaccharide substructures and also alters its physical state [3]. After chemical pretreatment, the biomass is subjected to enzymatic hydrolysis to produce soluble sugar solutions suitable for fermentation by microbes into biofuels. Because of its complex structure, many different enzymes are necessary for efficient, high yield deconstruction of biomass [4].

1.2 Enzymes Needed to Deconstruct Cellulose

Natural cellulolytic organisms are known to deconstruct biomass materials into simple sugars [5]. To better understand this natural capability, the DOE Joint Genome Institute and other researchers have sequenced over 5,000 natural organisms and metagenomes, including an increasing number of proven cellulolytic microbes and fungi. Bioinformatics approaches have sorted the protein-coding sequences of these genomes into many putative cellulolytic functions. As one leading example, the Carbohydrate-Active Enzymes database (CAZy, <http://www.cazy.org> [6]) provides an intensively curated list that is separated into glycoside hydrolase (GH), glycoside transferase (GT), carbohydrate esterase (CE), polysaccharide lyase (PL), and carbohydrate-binding module (CBM) families.

Efficient hydrolysis of pure cellulose requires concerted activities provided by several different members of the GH family [7]. Endocellulases or endoglucanases attack crystalline regions in the interior of cellulose molecules, while exoglucanases or cellobiohydrolases degrade cellulose from either the reducing or nonreducing ends of a glucan strand to produce cellobiose. Cellobiosidases then hydrolyze cellobiose to glucose. Both endo- and exocellulases have associated CBM domains, which increase the efficiency of binding the catalytic GH domain to the insoluble polysaccharide. Related cascades of enzymes react with polysaccharides such as xylan, mannan, and pectin, which collectively represent the hemicellulose fractions. Deconstruction of biomass is potentiated when enzymes capable of reaction with the cellulose and hemicellulose fractions are simultaneously present [8].

Among cellulolytic microorganisms, *Clostridium thermocellum*, an anaerobic, thermophilic, Gram-positive bacterium, has demonstrated one of the fastest growth rates on crystalline cellulose [9]. Since it can also ferment sugars to ethanol, *C. thermocellum* has become an important model organism for biofuel research. *C. thermocellum* secretes dozens of cellulases, xylanases, mannanases, and other polysaccharide-degrading enzymes that allow the efficient conversion of polysaccharides into soluble oligosaccharides. Biochemical [10, 11], transcriptomic [12], and proteomic [13] studies have helped to establish the identities of these enzymes, and many have been purified and studied for their catalytic function and 3-dimensional structures. Additionally, *C. thermocellum* produces a unique extracellular

assembly termed the cellulosome, a multienzyme complex in which many different enzymes are recruited to a protein called scaffoldin through tight-binding interactions of dockerin and cohesin domains [14]. The formation of this large complex is believed to provide high efficiency in polysaccharide degradation by increasing local enzyme concentration. However, because the scaffoldin has a repetitive multi-domain structure of ~200 kDa, obtaining pure preparations of it through heterologous expression has been a significant challenge. Consequently, protein engineering has been used to develop more experimentally tractable substitutes for the scaffoldin and various assemblies of enzymes from *C. thermocellum* [15, 16].

1.3 Overview of Experimental Strategy and Method

In this chapter, we describe an approach for producing engineered fusions of GH families with CBM domains. Figures 1 and 2 show an overview of the plasmid constructs and cloning methods that were applied to genes from *C. thermocellum* to produce forms of the enzymes that efficiently react with insoluble substrates without a need for the scaffoldin. These general principles are applicable to the study of the catalytic and polysaccharide binding domains from other organisms.

Identification of candidate genes includes bioinformatics evaluation and, if available, correlation with experimental results such as transcriptomics or proteomics. Since Flexi Vector cloning uses SgfI and PmeI restriction sites [17], candidate genes containing these restriction sites will require mutagenesis (or gene synthesis) to eliminate them. Fortunately, most genomes contain a low frequency of these restriction sites [17]. Furthermore, since most biofuel enzymes are secreted from the natural host, they contain hydrophobic signal sequences at the N-terminus that should be removed in the design for cell-free translation. This simple excision can dramatically improve the solubility of translated proteins. Bioinformatics tools such as SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) are useful to identify possible positions for truncation of the sequence encoding a signal peptide from the native gene. In this work, we also removed nucleotide sequences encoding the C-terminal dockerin domains in order to facilitate our studies of catalytic potential outside of the cellulosomal format.

Endo, Sawasaki, and colleagues at Ehime University, Matsuyama, Japan, created an optimized wheat germ cell-free translation plasmid, pEU [18]. For the protocols presented here, the parental plasmid was modified by the UW Center for Eukaryotic Structural Genomics to create pEU-HSBC (Fig. 1). This Flexi Vector cloning-compatible plasmid contains an SP6 promoter at the 5' end of the critical transcription/translation region, followed by the tobacco mosaic virus translational enhancer (TMV-omega) sequence [19], and then nucleotide sequences encoding a start codon and Gly residue, a His6 purification tag, and an AIA sequence corresponding to the Flexi Vector SgfI restriction site [17]. The empty vector contains a toxic cassette consisting of the SacB gene [20] and chloramphenicol acetyl transferase genes. At the 3' end, pEU-HSBC provides a stop codon by use of the Flexi Vector PmeI site, which thus yields a candidate protein with the N-terminal tag MGHHHHHHAIA-provided by the vector and the remaining coding sequence produced by the PCR primer design and targeted gene amplification.

Another vector created by the Great Lakes Bioenergy Research Center, pEU-HSBC-CBM3a, creates a C-terminal fusion to a 40 amino acid linker and CBM3a domain from the *C. thermocellum* scaffoldin protein CipA (Cthe_3077). Upon translation, candidate genes cloned into pEU-HSBC-CBM3a yield an unnatural fusion protein with an N-terminal tag MGHHHHHHAIA-, the candidate domains, and then a fusion to CBM3a, which is known to bind to crystalline cellulose. Genes cloned according to the Flexi Vector method are easily transferred between different Flexi Vector plasmids [17], and the presence of a 3' pF1K homology sequence increases the transfer efficiency by inhibiting self-ligation of the two plasmid vector backbones during serial transfer of genes.

Primer design has an important role in this effort. By consideration of the domain structure of a gene of interest, domains of interest can be excised from the complete gene to assemble better performing constructs, to test specific aspects of the catalytic capabilities of multi-domain enzymes, or to obtain engineered assemblies. Figure 2 schematizes this primer design process as applied to Cthe_0032, a GH26 protein. In this work, the catalytic domain of the protein was cloned, expressed, and assayed. By use of the 2-stage PCR method, it is possible to build additional functionality at the 5' and 3' regions of the cloned gene. For example, inclusion of the coding sequence for a protease recognition site immediately after the SgfI site but before the coding sequence of the gene of interest is useful for removal of N-terminal tags from expressed proteins for structure determination studies [21]. Alternatively, omission of a stop codon from the coding sequence of the gene of interest can be used to create C-terminal fusion proteins such as some of the examples provided here (Table 3, CBM3a fusions).

Plasmid DNA purified according to the protocols given is the essential substrate for a transcription reaction, which provides an mRNA that becomes the substrate for cell-free translation. Both the transcription and translation reactions are carried out by robots [22], and the individual translation reactions are then suitable for assay with a wide variety of biofuel relevant substrates, can be combined into higher-order assemblies of enzymes, or can be subjected to additional purification if desired.

1.4 Protein Functional Assays

During our initial work on cell-free translation of biofuel enzymes, we discovered that wheat germ extract had no endogenous capability for hydrolysis of the insoluble substrates cellulose, xylan, or mannan and likewise no competing enzymatic reactions that would consume soluble sugars or contain other products such as NADH that would be produced from coupled enzyme assays. Moreover, a low-level soluble glucosidase activity present in the wheat germ extract was easily inactivated at temperatures higher than 45°C. Thus many proteins translated in wheat germ extract can be assayed for their various cellulolytic activities without a need to obtain purified preparations of target polypeptides.

Table 1 summarizes the substrates used in this work. Assays with small molecule fluorogenic analogs, purified polysaccharides, and biomass materials have all been carried out with enzymes expressed in cell-free translation. In many cases, assays of single enzymes are inadequate for some discovery purposes, particularly when synergistic enzyme activities are required to disassemble the complicated ultrastructure of biomass. With robotic cell-free

translation, new samples of enzymes can be rapidly obtained for combinatorial tests of synergistic activity or as replacements for individual enzymes in mixtures with established reactivity. We briefly introduce examples of this capability in the chapter.

The methylumbelliferyl glycosides are small molecule fluorogenic analogs used in enzyme screening to deduce the potential reactivity of biofuel enzymes with oligomeric substrates. Hydrolysis of the glycosidic bond releases the fluorophore 4-methylumbelliferone [23, 24]. The high sensitivity of the fluorescence detection, the short reaction time, and the small amounts of enzyme needed contribute to a broad glycosidase diagnostic utility of the small molecule analogs.

Insoluble substrates are most realistic for biofuel enzymology but also provide increasing experimental challenges [25]. While individual enzymes can often be reliably assayed using purified polysaccharides, studies of biomass most often require the combination of chemical pretreatment and mixtures of synergistic enzymes to provide sufficient reactivity. The determination of enzyme function with insoluble polysaccharides can be accomplished by a variety of methods [26]. While samples from cell-free translation are acceptable for most of these, we focus here on use of the dinitrosalicylic acid (DNS) assay [27] because of its broad applicability to a variety of products, relatively simple methodology, sufficient sensitivity, and low cost. When an enzyme hydrolyzes a polysaccharide, the reducing-end sugar product can react with the DNS reagent, producing a colored adduct with an absorbance maximum of ~540 nm. The color formation permits the use of optical spectroscopy for detection, and the method is well suited for high throughput because the sample handling process can be reduced to relatively few liquid handling steps. In addition, we have adapted a simple protein pull-down assay to assess the function of CBM domains [28] and show examples of this qualitative evaluation using pure cellulose as a binding substrate.

2 Materials

2.1 Equipment

1. The Protomist DT II benchtop cell-free translation robot (CellFree Sciences Co., Ltd., Yokohama, Japan) is used for preparative protein production. This robot uses a bilayer cell-free protein synthesis method [29],
2. The Protomist DT II software (CellFree Sciences Co., Ltd., Yokohama, Japan) is used to program and execute the robotic translation experiment.
3. A spectrometer is used to measure DNA concentrations in plasmid preparations.
4. The Criterion stain-free gel imaging system (Bio-Rad Laboratories, Inc., USA) is used for qualitative and quantitative measurement of protein production (*see Note*¹). Criterion Precast 4–20 % gels (#345-0426, Bio-Rad Laboratories, Inc., USA) are used for protein electrophoresis.

¹The Criterion stain-free gel imaging system is used to obtain a gel image, and the presence of translated proteins is determined by visual inspection of lanes containing control extract and translation reactions. In the stain-free analysis, tryptophan residues in the separated protein react with a trihalo compound such as trichloroethanol included in the gel matrix [30], and a UV-induced reaction produces a derivative that can be detected by the fluorescence imaging system [31, 32]. The concentration of identified translated proteins was determined by the Criterion software based on evaluation of volume and area of pixels contained in the identified

5. Image Lab software version 3.0 (Bio-Rad Laboratories, Inc., USA) is used to visualize and quantitate proteins separated by the Criterion imaging system.
6. A shaking device (suggested Microtube Rotisserie (LABQUAKE[®], Barnstead Thermolyne, NH, USA)) is used for enzyme assays.
7. A microplate reader is used for making fluorescence and optical density measurements of samples in either 96-well or 384-well microtiter plate formats (suggested instrument Infinite[®] M1000 Pro, TECAN Group Ltd., Switzerland).

2.2 Cloning Reagents

1. 18 MΩ water is used to prepare all buffers and other reagents.
2. The pEU plasmid is the parental plasmid ([18], CellFree Sciences Co., Ltd., Yokohama, Japan). The wheat germ cell-free translation plasmid pEU-HSBC, shown in Fig. 1, was created from this parent.
3. *C. thermocellum* ATCC 27405 genomic DNA was provided by Dr. Paul Weimer (US Department of Agriculture Dairy Forage Research Center, Madison, WI).
4. Herculan[®] II Fusion Enzyme with dNTP Combo (Agilent Technologies, Santa Clara, CA, USA) is used for PCR reactions.
5. 10× Flexi Enzyme Blend (Promega, Madison, WI, USA) containing optimized amounts of SgfI and PmeI endonucleases is used for restriction enzyme digestions.
6. T4 DNA Ligase (New England Biolabs Ltd., UK) is used for ligation reactions.
7. Chemically competent cells (suggested strain *Escherichia coli* 10G, Lucigen, Middleton, WI, USA) are used for transformation.
8. BigDye[®] v2.0 (Applied Biosystems, CA, USA) is used for DNA sequencing reactions.

2.3 Transcription and Translation Reagents

1. A Qiagen Miniprep kit (Qiagen, Germany) is used for plasmid DNA preparations.
2. Proteinase K (Sigma-Aldrich, St. Louis, MO, USA) is used to eliminate residual RNase from plasmid DNA preparations. The 10× Proteinase K buffer consists of 100 mM Tris-HCl, pH 8.0, containing 50 mM EDTA and 1 % (w/v) SDS.
3. Transcription buffer (TB) consists of 400 mM HEPES-KOH, pH 7.8, 100 mM magnesium acetate, 10 mM spermidine hydrochloride, and 50 mM DTT. This buffer is used in cell-free translation reactions carried out in the Protomist DT II benchtop robot.

translated protein band relative to the same signals for molecular weight size markers of the same approximate size and number of tryptophan residues (Fig. 3).

4. Nucleotide solution (NTPs) consists of 25 mM each of ATP, GTP, CTP, and UTP (Sigma-Aldrich, St. Louis, MO, USA) dissolved in Milli-Q water. This solution is used in the transcription stage of the cell-free translation reaction.
5. SP6 RNA polymerase and RNase inhibitor (Promega, Madison, WI, USA) are used in transcription reactions.
6. Translation buffer (40× solutions 1–4; CellFree Sciences Co., Ltd., Yokohama, Japan) is used with the Protomist DT II robot.
7. The amino acid mixture consists of 2 mM of each of the 20 L-amino acids.
8. Wheat germ cell-free extract, either WEPRO2240 or WEPRO2240H (CellFree Sciences Co., Ltd., Yokohama, Japan), is used in the Protomist DT II robot (*see Note*²).
9. A 50 mg/mL solution of creatine kinase (Roche Applied Sciences, Indianapolis, IN, USA) is prepared in Milli-Q water and used in translation reactions.

2.4 Protein Purification Reagents

1. Ni-Sepharose high-performance chromatography resin (suggested resin is GE Healthcare, Piscataway, NJ, USA) is used for purification of His-tagged proteins from the wheat germ cell-free translation reactions.
2. Purification loading buffer (100 mM MOPS, pH 7.4, containing 300 mM NaCl, 2 mM CaCl₂, and 25 mM imidazole) is used to load sample to the resin.
3. Purification washing buffer (100 mM MOPS, pH 7.4, containing 300 mM NaCl, 2 mM CaCl₂, and 50 mM imidazole) is used to reduce nonspecific binding of proteins to the resin.
4. Purification elution buffer (100 mM MOPS, pH 7.4, containing 300 mM NaCl, 2 mM CaCl₂, and 250 mM imidazole) is used to elute protein from the resin.
5. Laemmli sample buffer (#161-0737, Bio-Rad Laboratories, Inc., USA) is the 2× protein sample buffer.
6. Protein molecular weight size markers (suggested markers are Precision Plus Protein Unstained Standards, Bio-Rad Laboratories, Inc., USA) are used for internal standards in electrophoresis experiments.
7. VIVASPIN 500, 10,000 MWCO PES (Sartorius Stedim Biotech, Goettingen, Germany) is used for concentrating protein samples after purification (*see Note*³).
8. U-bottom 96-well plates are from Greiner Bio-One (Monroe, NC).

²Wheat germ extract 2240 (WEPRO2240) is used when protein purification is not planned. When His-tag purification of the translated protein is anticipated, wheat germ extract 2240H (WEPRO2240H) should be used. This extract has been specially prepared to reduce the level of endogenous wheat germ proteins that adventitiously bind to the metal-activated chromatography resin.

³If the translated protein possesses a cellulolytic activity or a cellulose binding function, then polyethersulfone (PES) membranes should be used in purification procedures instead of cellulose-derived materials. For example, proteins harboring CBM domains may bind to cellulose-derived materials and be lost during handling.

2.5 Enzyme Assay Substrates

1. The fluorogenic substrates 4-methylumbelliferyl- β -D-glucopyranoside (MUG), 4-methylumbelliferyl- β -D-cellobioside (MUC), 4-methylumbelliferyl- β -D-mannopyranoside (MUM), and 4-methylumbelliferyl- β -D-xylopyranoside (MUX) (Sigma-Aldrich, St. Louis, MO, USA) are used in small molecule analog assays. These substrates are prepared fresh for each experiment by dissolving an appropriate mass in a 50 mL solution of 0.1 M sodium phosphate buffer, pH 8.0, in order to achieve 0.2 mM substrate concentration.
2. Whatman filter paper 1001-185 (GE Healthcare, UK) is used as a representative substrate for cellulase assays. A standard paper hole punch tool is used to create ~0.7 cm diameter disks, corresponding to a weight of 3.0 mg.
3. Phosphoric acid-swollen cellulose (PASC) is used as a representative substrate for cellulase assays.
4. Birch xylan (Sigma-Aldrich, St. Louis, MO, USA) is used as a substrate for xylanase assays.
5. Mannan (Sigma-Aldrich, St. Louis, MO, USA) is used as a substrate for mannanase assay and mannan pull-down assay.
6. Ammonia fiber expansion (AFEX) switchgrass ground to 1 mm particle size and washed ionic liquid-treated switchgrass are used as biomass substrates (*see Notes* ⁴ and ⁵).
7. Sigmacell 20 (Sigma-Aldrich, St. Louis, MO, USA) is used as a substrate for the protein pull-down assay.

2.6 Enzyme Assay Reagents

1. A 50 mM sodium phosphate buffer, pH 8.0, is used for all reactions.
2. The colorimetric reagent 3,5-dinitrosalicylic acid (Sigma-Aldrich, St. Louis, MO, USA) is used in reducing sugar determinations. To prepare the stock reagent solution, 10.6 g of 3,5-dinitrosalicylic acid (Sigma Aldrich, St. Louis, MO) and 19.8 g of NaOH are dissolved in 1 L of Milli-Q water. Then 306 g of Rochelle salts (Sigma-Aldrich, St. Louis, MO, USA), 7.6 mL of phenol, and 8.3 g of sodium metabisulfite (Sigma-Aldrich, St. Louis, MO, USA) are added. The final volume of the reagent solution is adjusted to 1.5 L by addition of Milli-Q water. The reagent stock is stored at 4°C in foil-covered bottles until use.
3. For reducing sugar assays with reactions containing cellulose, β -glucosidase (Lucigen, Middleton, WI) is added to convert all soluble oligosaccharides to glucose. Although not essential, this addition allows the quantitation of total monomer sugar equivalents produced.

⁴AFEX-treated switchgrass was prepared as described elsewhere [33]. Samples of 1 mm ground material were used in assays. This material was the generous gift of Dr. Bruce Dale (Great Lakes Bioenergy Research Center, Michigan State University).

⁵IL-treated switchgrass was prepared as described elsewhere [34]. This material was the generous gift of Dr. Masood Hadi (US Department of Energy Joint BioEnergy Institute, Emeryville, CA).

2.7 Polyacrylamide Gel Silver Staining Reagents

1. A silver staining method was used to detect protein-polysaccharide interactions in polyacrylamide gels. Reagents used are sodium thiosulfate (Sigma-Aldrich, St. Louis, MO, USA), silver nitrate, AgNO₃ (Sigma-Aldrich, St. Louis, MO, USA), and formaldehyde (Fisher Scientific, Hampton, NH, USA).

3 Methods

3.1 Design of First-Step PCR Primers for Candidate Genes

Figure 2 summarizes how PCR primers are designed by using the *C. thermocellum* Cthe_0032 gene, Uniprot A3DBE4, as an example (see **Note 6**).

1. The boundaries of the translocation signal peptide (residues 1–32) are used to define the position for truncation at the 5′ end of the gene. This truncation removes the signal peptide, which is not necessary for cell-free translation. According to this criterion, the gene-specific portion of the forward primer is 5′-atgtattcccttctgtggac-3′ (Fig. 2b).
2. The first-step forward PCR primer is then completed by appending a nucleotide sequence encoding a portion of the tobacco etch virus protease recognition sequence to the 5′ end of the sequence identified in step 1. With this addition, the complete sequence of the first-step forward PCR primer is 5′-AACCTGTACTTCCAGTCCatgtattcccttctgtggac-3′, indicated as “1st PCR, partial TEV cleavage site + gene-specific sequence” in Fig. 2b.
3. The end of the GH26 domain (residue 506) is used to define the position for truncation at the 3′ end of the gene. This truncation removes the dockerin domain. According to this criterion, the first-step reverse complement primer for the gene-specific sequence is 5′-ttcatccaagtgattacata-3′.
4. For insertion of a stop codon at the end of the gene-specific sequence, the first-step reverse PCR primer is then completed by appending a nucleotide sequence encoding a stop codon and the PmeI restriction sequence to the 5′ end of the sequence identified in **step 3**. With this addition, the complete sequence of the first-step reverse PCR primer is 5′-GCTCGAATTCGTTTAACTAttcaccaagtgattacata-3′, indicated as “1st PCR, gene-specific sequence + PmeI site” in Fig. 2c. This primer is also shown in Table 2, Cthe_0032R.
5. For creation of a C-terminal fusion protein, the first-step reverse PCR primer is then completed by appending a nucleotide sequence encoding a PmeI restriction sequence to the 5′ end of the sequence identified in **step 1**. With this addition,

⁶Cthe_0032 has a structure consisting of a signal peptide, one GH26 and two dockerin 1 domains (Fig. 2a). Microarray studies showed the enzyme was highly expressed [12], and mass spectral studies showed it is highly abundant in the secreted proteome [13]. The translocation signal peptide was identified in Uniprot A3DBE4, as calculated using SignalP. The presence of a signal peptide is not needed for cell-free translation. This information was used to define a truncation at the 5′ end of the gene. Likewise, the end of the GH26 domain was also identified from the Uniprot record. The truncation at the 3′ end of the gene was designed to remove the dockerin domain in linker region between the GH26 and dockerin domains. This trimming of the natural gene encodes a variant Cthe_0032 of 327 residues.

the complete sequence of the first-step reverse PCR primer is 5'-GCTCGAATTCGTTTAAACCTcatccaaggtgattacata-3'. This primer is also shown in Table 2, Cthe_0032R_CBM3a.

3.2 PCR Cloning and Plasmid Preparation

1. A two-step PCR method is used to amplify genes for this cloning method (*see Note*⁷). The first-step PCR is performed using the forward and reverse primer pairs indicated in Table 2, with a reverse primer designed to create either a catalytic domain alone or a fusion protein. Each primer possesses ~20 nucleotides of gene-specific sequence and ~20 nucleotide overhang, respectively. The *C. thermocellum* genomic DNA is used as the amplification template. The PCR reaction contains 0.5 µL of 1 µg/µL *C. thermocellum* genomic DNA, 1.25 µL of 10 µM primer pair, 0.5 µL of 100 mM dNTP, 10 µL of 5× PCR buffer, and 1 µL of Herculase DNA polymerase in a total volume of 50 µL. The first-step PCR reaction cycle consists of 95°C for 5 min followed by 95°C for 30 s, a ramp from 50 to 60°C for 30 s, and 72°C for *X*min for 35 cycles, where *X* is increased by 1 min for each 1,000 base pairs in the template, and then the reaction cycle is ended by 72°C for 5 min.
2. The second-step PCR is performed using the first PCR product as a DNA template and the forward and reverse universal primers (Table 2). The forward universal primer is indicated in Fig. 2b as “2nd PCR with SgfI + TEV cleavage site,” while the reverse universal primer is indicated as “2nd PCR PmeI site.”
3. The PCR reaction contains 5 µL of 1st PCR reaction product, 2.5 µL of 10 µM primer pair, 2 µL of 100 mM dNTP, 20 µL of 5× PCR buffer, and 2 µL of Herculase DNA polymerase in a total volume of 100 µL. The second-step PCR reaction cycle consists of 95°C for 5 min followed by 95°C for 30 s, ramp from 50 to 60°C for 30 s, and 72°C for *X*min for 35 cycles, where *X* is increased by 1 min for each 1,000 base pairs in the template, and then the reaction cycle is ended by 72°C for 5 min.
4. Plasmid vector DNA (pEU-HSBC-CBM3a, *see Note*⁸) and PCR product are digested using the Flexi Enzyme Blend. A typical reaction contains 3.5 µL of the DNA product, 0.5 µL of the Flexi Enzyme Blend, and 1 µL of buffer in a total volume of 5 µL. The reaction is carried out at 37°C for 60 min and followed by 65°C for 20 min.

⁷This protocol is based on the two-step PCR Flexi cloning method developed by the University of Wisconsin Center for Eukaryotic Structural Genomics to amplify genes [17]. Figure 2b shows how primer sequences are aligned with the Cthe_0032 gene. In this example, the original start codon of the Cthe_0032 gene is included. This is not necessary, however, as the vector provides the start codon needed for cell-free translation prior to the His6 tag (Fig. 1). The TEV protease recognition sequence is ENLYFQ/S, with proteolysis occurring between amino acids Q and S [35].

⁸The plasmid pEU-HSBC is available by request from the NIH Protein Structure Initiative Materials Repository (<http://psi-mr.asu.edu>) and is used for translation of proteins containing an N-terminal MGHHHHHHAIA tag. The wheat germ cell-free translation plasmid pEU-HSBC-CBM3a was created from pEU-HSBC and is used in the translation of proteins containing the abovementioned N-terminal tag and a C-terminal tag consisting of a 40-amino acid linker and a CBM3a domain.

5. Ligation of the digested PCR product and the plasmid vector DNA (pEU-HSBC-CBM3a) is carried out using T4 DNA Ligase with 1 mM dATP in 1× ligation buffer for 12 h at 16°C.
6. The ligated sample is transformed into chemically competent *E. coli* 10G cells by a standard heat shock transformation protocol. Transformants are selected on Luria Bertani agar plates containing 100 µg/mL of ampicillin and 5% (w/v) of sucrose.
7. Single colonies are screened by colony PCR methods to identify transformants that contain an insert of the approximately correct size. Transformants with the correct insert size are then submitted for DNA sequencing, in this case at the University of Wisconsin Biotechnology Center. The nucleotide sequences are analyzed by software comparison of predicted and desired sequences (suggested DNASTAR software, v.10, DNASTAR, Madison, WI, USA).
8. A sequence-verified plasmid carrying the gene of interest is prepared for transcription by using a Qiagen Miniprep kit. The purified plasmid DNA is treated with proteinase K in 1 × Proteinase K buffer at 37°C for 1 h to eliminate residual RNase. After the digestion, the plasmid preparation is treated by phenol/chloroform extraction. The plasmid preparation is recovered from the aqueous phase of the phenol/chloroform extraction by ethanol precipitation with a 2.5× volume addition of absolute ethanol supplemented with a 1/10 volume of 3 M ammonium acetate. The precipitation reaction is incubated at –80°C for 1 h. The precipitated DNA is recovered by centrifugation at 16,100×*g* for 15 min. The amount of DNA is measured by spectrometry using A₂₆₀ measurement. Based on this measurement, the concentration of plasmid DNA is adjusted to 1 µg/µL.

3.3 Protein Cell-Free Translation

1. A 5 µL aliquot of a 1 µg/µL preparation of purified plasmid DNA is used for each transcription reaction carried out in the Protomist DT II.
2. The transcription mixture (45.25 µL total volume per reaction) consists of 29 µL of water, 10 µL of 5× of transcription buffer (CellFree Sciences Co., Ltd., Japan), 5 µL of 25 mM NTPs, 0.625 µL of SP6 RNA polymerase, and 0.625 µL of RNase inhibitor enzymes (80,000 units/mL).
3. The translation mixture (59.2 µL total volume per reaction) consists of 56 µL of wheat germ extract (CellFree Sciences Co., Ltd., Japan), 3 µL of 2 mM amino acid mix, and 0.2 µL of a 20 µg/µL solution of creatine kinase (Roche, USA).
4. The 1× translation buffer (1.1 mL per reaction) containing ATP, GTP, and 20 amino acids is prepared from a SUB-AMIX (CFS-SUB, CellFree Sciences Co., Ltd., Japan).
5. The plasmid DNA, transcription mixture, translation mixture, and translation buffer are placed onto the appropriate positions of the deck of the Protomist™ DT II robot as described in the instrumental manual. The Protomist™ DT II

carries out fully automated transcription and bilayer mode translation reactions in 28 h duty cycle [29].

3.4 Verification of Translated Protein in Wheat Germ Extract

1. Translated proteins are analyzed by using denaturing gel electrophoresis (suggested Bio-Rad Criterion system, *see* **Notes** ¹, ¹², ¹³). A 5 μ L sample of the translation reaction and an equal volume of 2 $\times$ protein sample buffer are mixed and heated at 95°C for 10 min. Electrophoresis is performed in constant voltage mode at 220 V for 50 min in 1 $\times$ Criterion buffer system.

3.5 Purification of His-Tagged Proteins

1. Ni-Sepharose resin is equilibrated with water to make an ~50% (v/v) slurry.
2. For each translation reaction from Protomist DT II robot, place 50 μ L of the Ni-Sepharose slurry into each well of the filter plate, retained in the U-bottom 96-well plate. Add 100 μ L of the Ni binding/washing buffer to each well. This combination of a filter plate and the U-bottom plates is the purification plate.
3. Add 100 μ L of the supernatant from the soluble translation product plate to each well of the purification plate.
4. Mix the solution in a plate shaker for 10 min at room temperature, taking caution to not spill or cross-contaminate the wells.
5. Centrifuge the plate (suggested JS 5.9 rotor and Avanti J-30I centrifuge for 1 min at 3,640 rpm, 2,500 $\times g$, and ambient temperature).
6. Add 150 μ L of Ni binding/washing buffer to wash out non-specifically bound proteins. Gently mix the solution using a plate shaker for 10 s at room temperature. Centrifuge the plate. Repeat this step three times. The filtrate should not contain the translated protein, but can be retained for analysis of binding efficiency and product yield if needed.
7. Place the filter plate onto a new U-bottom 96-well plate. Add 50 μ L of the Ni elution buffer into each well of the purification plate. Use a plate shaker to gently mix the solution for 1 min at room temperature to elute the bound proteins.
8. Centrifuge the plate and save the filtrate for analysis by denaturing gel electrophoresis (Fig. 3) and enzymatic assays.

¹²Our cumulative experience with wheat germ cell-free translation is that over 80% of protein targets from a wide range of prokaryotic and eukaryotic sources are expressed at levels that can be detected using Coomassie-stained denaturing gel electrophoresis. This overall percentage of successful expression has been observed in other research labs [37]. Enzymes and proteins in the size range of $M_r \approx 34$ –140 kDa have been produced using these methods.

¹³Some posttranslational modifications are possible in wheat germ extract, such as the removal of N-terminal methionine, Ser/Thr/Tyr phosphorylation, and slow formation of disulfide bonds. In general, glycosylation is not observed because neither the glycosyltransferases nor the nucleotide sugar substrates are present.

3.6 Preparation of Phosphoric Acid-Swollen Cellulose

1. Sigmacell 50 microcrystalline cellulose (20 g, Sigma, St. Louis, MO) is swollen in 800 g of cold (0°C) 80 % phosphoric acid, with rapid stirring with a plastic rod for ~1 h in an ice bath.
2. The cellulose is diluted with 2 L of cold water, thoroughly mixed, and allowed to settle, and the overlying liquid is removed by siphoning. This washing and settling procedure is repeated several times to reduce the acid content.
3. The cellulose slurry is neutralized with solid NaHCO₃, rinsed, decanted as above, and then secured inside bags formed from nylon-reinforced paper towelling. These bags are filled with ~1 L of distilled water, and the excess liquid is squeezed off; this process is repeated 20 times. The bags are then sealed, suspended in buckets containing 5 L of cold deionized water, and dialyzed for 10 days, with frequent changes of water. Prior to each change of water, the bags are tightly hand squeezed to facilitate removal of the equilibrated solutions.
4. After completion of dialysis (when the phosphate content reaches < 1 µg/L) the cellulose is lyophilized (*see Note* ⁹).

3.7 Fluorogenic Substrate Hydrolysis Assay

- 5 A 10 µL aliquot of cell-free translation reaction containing a translated enzyme is mixed with the fluorogenic substrate and assay buffer to give a total reaction volume of 100 µL. These reactions are conveniently prepared in a 96-well plate and include technical replicates.
- 6 The fluorogenic reaction is carried out at 37°C for 15 min in plate reader. The fluorescence measurement is performed using excitation at 360 nm and emission detection at 460 nm. A negative control is made using a blank wheat germ translation reaction.
- 7 Results from assays with a panel of fluorogenic substrates are summarized in Table 3. Enzymes that give a positive fluorescence increase after subtraction of the background fluorescence in the wheat germ extract control with the fluorogenic substrate are assigned to have positive reactivity with the indicated substrate (*see Note* ¹⁰).

3.8 Reducing Sugar Assays

1. A 50 µL aliquot of a cell-free translation reaction containing a translated enzyme is mixed with 50 µL of a 20 mg/mL suspension of pure polysaccharide substrate in 50 mM phosphate buffer, pH 8.0 (*see Note* ¹⁰). Pure polysaccharides suitable for use with this method include PASC, xylan, and mannan. These samples are

⁹X-ray diffraction analysis should be used to confirm that the lyophilized material obtained from this procedure is predominantly amorphous cellulose. A discussion of the method and analysis of X-ray powder diffraction of cellulose is found in [36].

¹⁰Enzyme assays for *C. thermocellum* enzymes have been performed in the pH range of 6.0–8.0; all assays described here were carried out at pH 8.0.

conveniently prepared in a 96-well microtiter plate and sealed with TempPlate sealing film (USA Scientific, Ocala, FL).

2. The reaction is carried out at 60°C for 20 h. A 30 μ L aliquot of the reaction and a 60 μ L aliquot of DNS reagent are combined in a 96-well PCR plate. After mixing, the sample is heated at 95°C for 10 min.
3. The absorbance at 540 nm is used to determine the amount of reducing sugar present (Fig. 4a). A water blank containing DNS reagent is used for the background subtraction. The amount of reducing sugar produced is determined by comparison to standard curves generated using the corresponding pure monomer sugar, i.e., glucose, xylose, or mannose.
4. Results of the DNS assay with pure polysaccharides are shown in Fig. 4a and summarized in Table 3.

3.9 Biomass Assays

1. A 50 μ L aliquot of cell-free translation reaction containing a translated enzyme is mixed with 50 μ L of a 20 mg/mL suspension of pretreated biomass in 50 mM phosphate buffer, pH 8.0 (*see Note*¹⁰). Biomass substrates used with this method include AFEX-SG and IL-SG.
2. The reaction is carried out at 60°C for 20 h. A 30 μ L aliquot of the reaction and a 60 μ L aliquot of DNS reagent are combined in a 96-well PCR plate. After mixing, the sample is heated at 95°C for 10 min.
3. Results of the DNS assay with pretreated biomass are shown in Fig. 4b and summarized in Table 3.
4. For combinatorial assays, 25 μ L of two separate cell-free translation reactions are mixed with 50 μ L of a 20 mg/mL suspension of pretreated biomass in 50 mM phosphate buffer, pH 8.0. The reaction is carried out as described in **step 2** above (*see Note*¹¹).

3.10 Protein-Polysaccharide Pull-Down Assay

1. Enzymes purified as described in Fig. 3 are used to test for binding to insoluble substrates Sigmacell 20 and PASC (Fig. 5). A 5 μ L aliquot of purified enzyme is mixed with 1 mg of each substrate in 5 μ L of total reaction in 50 mM phosphate buffer, pH 7.0, in a 96-well PCR plate. Reactions of protein without substrate are also prepared as a control for protein precipitation caused by aggregation or denaturation. Samples are incubated for 1 h at room temperature and then spun at 5,000 for 15 min at 4°C. The supernatant and pellet are carefully separated, and 10 μ L of 2 $\times$ sample buffer is added to the supernatant, while the pellet is suspended in 5 μ L of 2 $\times$ sample buffer and an equal volume of water. The samples are heated for 5 min at 95°C.

¹¹The synergy factor for reactions containing two enzymes is the increase in reactivity observed from the combination relative to the result expected by arithmetic combination of the reactivity observed as separate enzymes. Figure 4c provides an example where an endocellulase and a xylanase were combined. This gives rise to increased reaction, i.e., synergy.

2. Denatured samples are separated using polyacrylamide gel electrophoresis and analyzed as described above.
3. The gels are placed into fixing solution (50% methanol, 12% acetic acid, 0.5 mL/L 37% formaldehyde) for 1 h.
4. The gels are washed in 50% ethanol three times for 5 min and pretreated with a 0.2 g/L solution of $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$ for 1 min.
5. The gels are rinsed with Milli-Q water 3× for 20 s and stained in a solution containing 2 g/L AgNO_3 and 0.75 mL/L of 37 % formaldehyde for 30 min.
6. The gels are rinsed again with Milli-Q water 3× for 20 s; developed in a solution containing 60 g/L Na_2CO_3 , 0.5 mL/L of 37 % formaldehyde, and 4 mg/L of $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$ for 5–10 min; and rinsed with Milli-Q water 2× for 5 s.
7. The development is stopped with 50 % methanol (v/v) and 12% (w/v) acetic acid for 10 min. The gel is washed in 50 % (v/v) methanol for 20 min and imaged.

Acknowledgments

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494). The authors thank researchers in the University of Wisconsin Center for Eukaryotic Structural Genomics for advice and timely access to protocols and equipment (NIGMS U54 GM074901, U54 GM094584, U01 GM094622; J.L. Markley, G.N Phillips, B.G. Fox). The authors thank Dr. Paul Weimer (University of Wisconsin-Madison, Dept. of Bacteriology) for his scientific insights and the generous gifts of phosphoric acid-swollen cellulose and *C. thermocellum* genomic DNA, Dr. Bruce Dale for the generous gift of AFEX-pretreated switchgrass, Dr. Masood Hadi for the generous gift of ionic liquid-pretreated switchgrass, and Dr. George Phillips and Dr. Chris Bianchetti (Great Lakes Bioenergy Research Center) for the X-ray diffraction analysis of some cellulose substrates.

References

1. Dale BE. Cellulosic biofuels and the road to energy security. *Environ Sci Technol.* 2011; 45:9823. [PubMed: 22060184]
2. Hall, CAS., Klitgaard, KA. Energy and the wealth of nations: understanding the biophysical economy. Springer; Seacaucus, NJ: 2012. p. 07096-2485.
3. Li C, Cheng G, Balan V, et al. Influence of physico-chemical changes on enzymatic digestibility of ionic liquid and AFEX pretreated com stover. *Bioresour Technol.* 2011; 102:6928–6936. [PubMed: 21531133]
4. Merino ST, Cherry J. Progress and challenges in enzyme development for biomass utilization. *Adv Biochem Eng Biotechnol.* 2007; 108:95–120. [PubMed: 17594064]
5. Lynd LR, Weimer PJ, van Zyl WH, et al. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 2002; 66:506–577. [PubMed: 12209002]
6. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 2009; 37:D233–D238. [PubMed: 18838391]
7. Vuong TV, Wilson DB. Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnol Bioeng.* 2010; 107:195–205. [PubMed: 20552664]
8. Gao D, Uppugundla N, Chundawat SP, et al. Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol Biofuels.* 2011; 4:5. [PubMed: 21342516]
9. Demain AL, Newcomb M, Wu JH. Cellulase, clostridia, and ethanol. *Microbiol Mol Biol Rev.* 2005; 69:124–154. [PubMed: 15755956]

10. Gilbert HJ. Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Mol Microbiol.* 2007; 63:1568–1576. [PubMed: 17367380]
11. Bayer EA, Lamed R, White BA, et al. From cellulosomes to cellulosomics. *Chem Rev.* 2008; 8:364–377.
12. Riederer A, Takasuka TE, Makino S, et al. Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Appl Environ Microbiol.* 2011; 77:1243–1253. [PubMed: 21169455]
13. Raman B, Pan C, Hurst GB, et al. Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS One.* 2009; 4:e5271. [PubMed: 19384422]
14. Schwarz WH. The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol.* 2001; 56:634–649. [PubMed: 11601609]
15. Mingardon F, Chanal A, Lopez-Contreras AM, et al. Incorporation of fungal cellulases in bacterial mimcellulosomes yields viable, synergistically acting cellulolytic complexes. *Appl Environ Microbiol.* 2007; 73:3822–3832. [PubMed: 17468286]
16. Morais S, Barak Y, Caspi J, et al. Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. *MBio.* 2010; 1
17. Blommel PG, Martin PA, Wrobel RL, et al. High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system. *Protein Expr Purif.* 2006; 47:562–570. [PubMed: 16377204]
18. Sawasaki T, Ogasawara T, Morishita R, et al. A cell-free protein synthesis system for high-throughput proteomics. *Proc Natl Acad Sci USA.* 2002; 99:14652–14657. [PubMed: 12409616]
19. Gallie DR, Sleat DE, Watts JW, et al. The 5'-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts in vitro and in vivo. *Nucleic Acids Res.* 1987; 15:3257–3273. [PubMed: 3575095]
20. Gay P, Le Coq D, Steinmetz M, et al. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol.* 1985; 164:918–921. [PubMed: 2997137]
21. Blommel PG, Fox BG. A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr Purif.* 2007; 55:53–68. [PubMed: 17543538]
22. Beebe ET, Makino S, Nozawa A, et al. Robotic large-scale application of wheat cell-free translation to structural studies including membrane proteins. *N Biotechnol.* 2011; 28:239–249. [PubMed: 20637905]
23. Vantilbeurgh H, Claeysens M, Debruyne CK. The use of 4-methylumbelliferyl and other chromophoric glycosides in the study of cellulolytic enzymes. *FEBS Lett.* 1982; 149:152–156.
24. Potier M, Mameli L, Belisle M, et al. Fluorometric assay of neuraminidase with a sodium (4-methylumbelliferyl- α -D-N-acetyl-neuraminate) substrate. *Anal Biochem.* 1979; 94:287–296. [PubMed: 464297]
25. Chundawat SP, Balan V, Dale BE. High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass. *Biotechnol Bioeng.* 2008; 99:1281–1294. [PubMed: 18306256]
26. Sharrock KR. Cellulase assay methods: a review. *J Biochem Biophys Methods.* 1988; 17:81–105. [PubMed: 3063738]
27. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem.* 1959; 31:426–428.
28. Shoseyov O, Shani Z, Levy I. Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol Mol Biol Rev.* 2006; 70:283–295. [PubMed: 16760304]
29. Sawasaki T, Hasegawa Y, Tsuchimochi M, et al. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. *FEBS Lett.* 2002; 514:102–105. [PubMed: 11904190]
30. Edwards RA, Jickling G, Turner RJ. The light-induced reactions of tryptophan with halocompounds. *Photochem Photobiol.* 2002; 75:362–368. [PubMed: 12003125]
31. Kazmin D, Edwards RA, Turner RJ, et al. Visualization of proteins in acrylamide gels using ultraviolet illumination. *Anal Biochem.* 2002; 301:91–96. [PubMed: 11811971]
32. Ladner CL, Yang J, Turner RJ, et al. Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal Biochem.* 2004; 326:13–20. [PubMed: 14769330]

33. Balan V, Bals B, Chundawat SP, et al. Lignocellulosic biomass pretreatment using AFEX. *Methods Mol Biol.* 2009; 581:61–77. [PubMed: 19768616]
34. Li C, Knicrim B, Manisseri C, Arora R, et al. Comparison of dilute acid and ionic liquid pretreatment of switchgrass: biomass recalcitrance, delignification and enzymatic saccharification. *Bioresour Technol.* 2010; 101:4900–4906. [PubMed: 19945861]
35. Kapust RB, Tozser J, Copeland TD, et al. The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun.* 2002; 294:949–955. [PubMed: 12074568]
36. Park S, Baker JO, Himmel ME, et al. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnol Biofuels.* 2010; 3:10. [PubMed: 20497524]
37. Goshima N, Kawamura Y, Fukumoto A, et al. Human protein factory for converting the transcriptome into an in vitro-expressed proteome. *Nat Methods.* 2008; 5:1011–1017. [PubMed: 19054851]
38. Weimer PJ, Lopez-Guisa JM, French AD. Effect of cellulose fine structure on kinetics of its digestion by mixed ruminal microorganisms in vitro. *Appl Environ Microbiol.* 1990; 56:2421–2429. [PubMed: 2403252]
39. Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res.* 2010; 38:D211–D222. [PubMed: 19920124]

Fig. 1.

Maps of two cloning vectors used in this study. The *top portion* shows the nucleotide sequence of the SP6 promoter, TMV-omega sequence, the start codon (ATG) and sequence encoding a His6 tag, and the SgfI restriction site. Circular maps of pEU_HSBC and pEU_HSBC_CBM3a are shown below. In both plasmids, ampicillin resistance is used as the transformation selection marker. During cloning, the sacB-CAT cassette is replaced by the gene of interest, which can permit positive selection by growth on agar medium containing sucrose if desired. In pEU_HSBC_CBM3a, the 40-amino acid linker and CBM3a coding

sequence is placed in frame at the 3' end of sacB-CAT cassette so that the target protein is expressed as a fusion with the additional linker and CBM3a domain

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 2.

Design of primer pairs and a schematic of the cloning method, (a) The Cthe_0032 (ManA) gene is used as an example. The PFAM image of the domain structure [39] shows the predicted amino acid residue positions for the signal peptide (residues 1–32), the GH26 domain (residues 180–506, *green*), and the two dockerin domains (*red*). The forward primer (b) and reverse primer design (c) for PCR needed for two-step Flexi Vector transfer are shown with nucleotide sequences and corresponding amino acid sequences

Fig. 3.

SDS-PAGE analysis (a) of the wheat germ cell-free translation of *C. thermocellum* proteins and (b) purification of proteins containing an N-terminal His-tag. Lane 1, Cthe_0032; lane 2, Cthe_0212; lane 3, Cthe_0269; lane 4, Cthe_0270; lane 5, Cthe_0536; lane 6, Cthe_0543; lane 7, Cthe_0625; lane 8, Cthe_0797; lane 9, Cthe_0912; lane 10, Cthe_1256; lane 11, Cthe_1838; lane 12, Cthe_2872. The domains included in the translated proteins and results of catalytic assays on these samples are provided in Table 3

Fig. 4.

Representative endpoint assays of reducing sugar formation from *C. thermocellum* enzymes prepared by robotic cell-free translation and assayed without further purification. The reactions were carried out in a 384-well microtiter plate. The substrates used are indicated, and detection was by use of the DNS reagent, which reacts with reducing-end sugars, yielding a change in color from *pale yellow* to *red* (increase in absorbance at 540 nm). (a) Reactions of a set of individual cellulase and xylanase enzymes with pure polysaccharides, where each row contains the results for the panel of enzymes indicated with a different substrate, (b) The same set of single enzymes was tested for reaction with AFEX-SG and IL-SG. Table 3 gives a summary of the assay results, (c) Synergistic improvement in biomass deconstruction obtained from binary combinations of a cellulase with a xylanase. The influence of β -glucosidase in alleviation of product inhibition is also demonstrated

Fig. 5.

Pull-down assays of enzymes using Sigmacell 20 as the binding substrate. Five *C. thermocellum* proteins that have reactivity with insoluble cellulose were produced in cell-free translation reactions, purified as described, and tested for the ability to bind to cellulose. After introduction of the cellulose to the translation mixture and incubation, the pull-down sample was separated into pellet (P) and supernatant (S) fractions by centrifugation. Proteins with cellulose binding capacity were enriched in the pellet fraction after centrifugation (*marked by stars*). Soluble and pellet fractions were treated with denaturing buffer, separated by gel electrophoresis, and visualized by silver staining. For each protein, the control lane provides information on translated protein that pelleted in the absence of cellulose. Cthe_0536 lacks a CBM domain and did not bind to cellulose, while Cthe_0543 contained a natural CBM3c domain and was bound. Cthe_0625 also contained a natural CBM3c domain, but did not bind to cellulose. This second CBM3c domain has a different primary sequence than the similarly named domain in Cthe_0543. Surprisingly, Cthe_0797 formed a soluble complex with an endogenous wheat germ protein (*marked by triangle*); this interaction was broken when Cthe_0797 bound to cellulose. The multi-domain protein Cthe_1256 also exhibited binding to cellulose. In this figure, Cthe_0269 and Cthe_0412 were obtained by recombinant *E. coli* protein expression system. The Cthe_0269 is engineered such that additional CBM3a domain derived from Cthe_3077 scaffoldin is present at C-terminus. A native form of Cthe_0412 possesses a CBM4_9 domain at N-terminus. Pull-down assays are carried out for testing their binding capacity to two different forms of cellulose, Sigmacell 20 and PASC. Results indicate that both enzymes bind to two forms of cellulose. Apparently, Cthe_0269 binds more tightly to substrates than Cthe_0412, indicating that CBM3a binds to cellulose material more effectively than CBM4_9 domain

Table 1

Summary of the substrates used in enzyme assays

Substrate (abbreviation)	Assay type	Activity detected	Source
4-Methylumbelliferyl- β -D-glucopyranoside (MUG)	Fluorescence	β -Glucosidase	Sigma-Aldrich (M3633)
4-Methylumbelliferyl- β -D-cellobioside (MUC)	Fluorescence	Cellobiohydrolase	Sigma-Aldrich (M6018)
4-Methumbelliferyl- β -D-mannopyranoside (MUM)	Fluorescence	β -Mannosidase	Sigma-Aldrich (M0905)
4-Methylumbelliferyl- β -D-xylopyranoside (MUX)	Fluorescence	β -Xylosidase	Sigma-Aldrich (M7008)
Filter paper (FP)	DNS	Cellulase	GE Healthcare (Whatman 10001-185)
Phosphoric acid-swollen cellulose (PASC)	DNS	Cellulase	Prepared as described in Subheading 3.6; also <i>see</i> reference [38]
Birch xylan (xylan)	DNS	Xylanase	Sigma-Aldrich (X0502)
1, 4- β -D-Mannan (mannan)	DNS	Mannanase	Megazyme (90302b)
AFEX-treated switchgrass (AFEX-SG)	DNS	Biomass deconstruction	Great Lakes Bioenergy Research Center
Ionic liquid-treated switchgrass (IL-SG)	DNS	Biomass deconstruction	Joint Bioenergy Institute

Table 2

Oligonucleotides used for cloning

Primer name	Sequence ^a	Direction
Cthe_0032F	5'-AACCTGTACTTCCAGTCCatgtattccctctctgtggac-3'	Forward
Cthe_0032R	5'-GCTCGAATTCGTTTAAACTAttcaccaaggtgattacata-3'	Reverse
Cthe_0032R_CBM3a	5'-GCTCGAATTCGTTTAAACCttccaaggtgattacata-3'	Reverse
Cthe_0212F	5'-AACCTGTACTTCCAGTCCatgtcaagataactttccc-3'	Forward
Cthe_0212R	5'-GCTCGAATTCGTTTAAACTAttaaaacccgtgtttttgatta-3'	Reverse
Cthe_0269F	5'-AACCTGTACTTCCAGTCCatggcgcttttaacacaaa-3'	Forward
Cthe_0269R_CBM3a	5'-GCTCGAATTCGTTTAAACtctcgtttatgtacaacaagtg-3'	Reverse
Cthe_0270F	5'-AACCTGTACTTCCAGTCCatgaaaaaataccgttactta-3'	Forward
Cthe_0270R	5'-GCTCGAATTCGTTTAAACTAtcaatcatcaacaggtatattgt-3'	Reverse
Cthe_0405F	5'-AACCTGTACTTCCAGTCCatggatccgaacaatgacgactg-3'	Forward
Cthe_0405R_CBM3a	5'-GCTCGAATTCGTTTAAACtccatttgaaccaagaggtatc-3'	Reverse
Cthe_0412F	5'-AACCTGTACTTCCAGTCCatggacatcatcatcatca-3'	Forward
Cthe_0412R_CBM3a	5'-GCTCGAATTCGTTTAAACagctgtaaccatgcaaacg-3'	Reverse
Cthe_0536F	5'-AACCTGTACTTCCAGTCCatgggacatcatcatcatca-3'	Forward
Cthe_0536R	5'-GCTCGAATTCGTTTAAACTAtttatagggcaactactta-3'	Reverse
Cthe_0543F	5'-AACCTGTACTTCCAGTCCatgggacatcatcatcatca-3'	Forward
Cthe_0543R	5'-GCTCGAATTCGTTTAAACTAttactgttcagccggaattttt-3'	Reverse
Cthe_0578F	5'-AACCTGTACTTCCAGTCCatggactataactatggagaag-3'	Forward
Cthe_0578R_CBM3a	5'-GCTCGAATTCGTTTAAACggtaccattgggtctacacc-3'	Reverse
Cthe_0625F	5'-AACCTGTACTTCCAGTCCatgggacatcatcatcatca-3'	Forward
Cthe_0625R	5'-GCTCGAATTCGTTTAAACTActattctaccggaaatttatcta-3'	Reverse
Cthe_0797F	5'-AACCTGTACTTCCAGTCCatgggacatcatcatcatca-3'	Forward
Cthe_0797R	5'-GCTCGAATTCGTTTAAACTAtcttcaacgccgcactctc-3'	Reverse
Cthe_0912F	5'-AACCTGTACTTCCAGTCCatgggacatcatcatcatca-3'	Forward
Cthe_0912R	5'-GCTCGAATTCGTTTAAACTActcttcgattacagttcc-3'	Reverse
Cthe_1256F	5'-AACCTGTACTTCCAGTCCatggcggtagatatcaagaaaat-3'	Forward
Cthe_1256R	5'-GCTCGAATTCGTTTAAACTAtttccacgtgtttttttgt-3'	Reverse
Cthe_1838F	5'-AACCTGTACTTCCAGTCCatgggacatcatcatcatca-3'	Forward
Cthe_1838R_CBM3a	5'-GCTCGAATTCGTTTAAACgttaactatagcataaaatgc-3'	Reverse
Cthe_2872F	5'-AACCTGTACTTCCAGTCCatgggacatcatcatcatca-3'	Forward
Cthe_2872R	5'-GCTCGAATTCGTTTAAACTAtccgtagtactgcccgaggaaa-3'	Reverse
2nd universal primerF	5'-GGTTGCGATCGCCGAAAACCTGTACTTCCAG-3'	Forward
2nd universal primerR	5'-GTGTGAGCTCGAATTCGTTTAAACC-3'	Reverse

^aLowercase letters indicate the gene-specific sequence

Table 3

Summary of assay results obtained for individual *C. thermocellum* enzymes

Gene locus	Gene name	Domains produced in translation	His6 added ^a	CBM3a added ^b	Uniprot ID	FP	PASC	Xylan	Mannan	MUG	MUC	MUX	MUM	Measured function (s)
Cthe_0032		GH26			A3DBE4	-	-	-	+	-	-	-	-	Mannanase
Cthe_0040	celI	GH9 CBM3			Q02934	+	+	-	-	-	-	-	-	Endocellulase
Cthe_0212	bgIA	GH1			P26208	-	-	-	-	+	-	-	-	β-Glucosidase
Cthe_0269	celA	GH8		Yes	A3DC29	+	+	-	-	-	-	-	-	Endocellulase
Cthe_0270	chiA	GH18			A3DC30	-	-	-	-	-	-	-	-	None detected
Cthe_0405	celL	GH5		Yes	A3DCG4	+	+	-	-	-	+	-	-	Exocellulase
Cthe_0412	celK	CBM_4_9 CelDN GH9		Yes	A3DCHI	-	+	+	-	+	+	-	-	Exocellulase
Cthe_0536	celB	GH5	Yes		P04956	+	+	-	-	-	+	-	-	Exocellulase
Cthe_0543	celF	GH9 CBM3	Yes		P26224	+	+	-	-	-	-	-	-	Endocellulase
Cthe_0578	celR	GH9 CBM3		Yes	A3DCY5	+	-	-	-	+	-	-	-	Endocellulase
Cthe_0625	celQ	GH9 CBM3	Yes		A3DD31	+	+	-	-	-	-	-	-	Endocellulase
Cthe_0797	celE	GH5 Lipase_GDSL	Yes		A3DDK3	+	+	+	+	+	+	-	-	Exocellulase, mannanase, xylanase
Cthe_0912	xynY	CBM_4_9 GH10 CBM_4_9			A3DDW7	-	-	+	-	-	-	-	-	Xylanase
Cthe_1256	bgIB	GH3	Yes		P14002	-	-	-	-	+	+	+	-	β-Glucosidase, β-xylosidase
Cthe_1838	xynC	CBM_4_9 GH10		Yes	A3DGI0	-	-	+	-	-	-	-	-	Xylanase
Cthe_2872	celG	GH5			Q05332	+	+	-	-	-	+	-	-	Exocellulase

^aProteins produced using pEU-His-FV, which adds an N-terminal MGHHHHHHA1A tag

^bProtein produced using pEU-His-FV-CBM3a, which gives a native N-terminus and adds fusion of a 40-amino acid linker followed by the CBM3a domain from Cthe_3077 to the C-terminus