APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 2007, p. 3165–3172 Vol. 73, No. 10
0099-2240/07/$08.00+0 doi:10.1128/AEM.02960-06

Processivity, Substrate Binding, and Mechanism of Cellulose Hydrolysis by *Thermobifida fusca* Cel9A▽

Yongchao Li,[1] Diana C. Irwin,[2] and David B. Wilson[2]*

Field of Microbiology, Cornell University, Ithaca, New York 14850,[1] and Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853[2]

Received 20 December 2006/Accepted 10 March 2007

Thermobifida fusca **Cel9A-90 is a processive endoglucanase consisting of a family 9 catalytic domain (CD), a family 3c cellulose binding module (CBM3c), a fibronectin III-like domain, and a family 2 CBM. This enzyme has the highest activity of any individual *T. fusca* enzyme on crystalline substrates, particularly bacterial cellulose (BC). Mutations were introduced into the CD or the CBM3c of Cel9A-68 using site-directed mutagenesis. The mutant enzymes were expressed in *Escherichia coli*; purified; and tested for activity on four substrates, ligand binding, and processivity. The results show that H125 and Y206 play an important role in activity by forming a hydrogen bonding network with the catalytic base, D58; another important supporting residue, D55; and Glc(−1) O1. R378, a residue interacting with Glc(+1), plays an important role in processivity. Several enzymes with mutations in the subsites Glc(−2) to Glc(−4) had less than 15% activity on BC and markedly reduced processivity. Mutant enzymes with severalfold-higher activity on carboxymethyl cellulose (CMC) were found in the subsites from Glc(−2) to Glc(−4). The CBM3c mutant enzymes, Y520A, R557A/E559A, and R563A, had decreased activity on BC but had wild-type or improved processivity. Mutation of D513, a conserved residue at the end of the CBM, increased activity on crystalline cellulose. Previous work showed that deletion of the CBM3c abolished crystalline activity and processivity. This study shows that it is residues in the catalytic cleft that control processivity while the CBM3c is important for loose binding of the enzyme to the crystalline cellulose substrate.**

Cellulose, the most abundant polymer in nature, is a linear homopolymer of β1-4-linked glucose residues. It is a renewable carbon source that can be hydrolyzed to sugars, which can then be fermented to ethanol or other chemicals to replace nonrenewable petroleum (12). However, cellulose is difficult to degrade because the molecules form tightly packed, extensively hydrogen-bonded microfibrils which are embedded in the plant cell wall matrix. The detailed molecular mechanism of crystalline cellulose and plant biomass degradation by glycoside hydrolases is still not well understood.

Thermobifida fusca is a filamentous soil bacterium that grows at 50°C in defined medium and can utilize cellulose as its sole carbon source. It is a major degrader of plant cell walls in heated organic materials such as compost piles and rotting hay. Six different cellulases have been purified and characterized in our lab using standard purification methods. Among these six cellulases, three of them are endocellulases, Cel9B, Cel6A, and Cel5A (7, 8); two of them are exocellulases, Cel6B and Cel48A (6, 20); and one is a processive endocellulase, Cel9A (5, 7).

Thermobifida fusca Cel9A-90 consists of a family 9 catalytic domain (CD), a family 3c cellulose binding module (CBM3c), a fibronectin III-like domain, and a family 2 CBM. It has the highest activity of any individual *T. fusca* enzyme on crystalline substrates. Cel9A has relatively high activity on carboxymethyl cellulose (CMC), which is a feature of an endocellulase. However, it produces 87% soluble products from filter paper (FP), a form of crystalline cellulose, and only 13% insoluble products (19). True endocellulases produce 30 to 40% insoluble products from FP, and exocellulases produce only 5 to 8% insoluble products; thus, Cel9A is quite processive (19). Another interesting feature of Cel9A is that it can synergize with both reducing and nonreducing end-directed exocellulases and with endocellulases, while true endocellulases do not synergize with each other. This evidence confirms that Cel9A possesses the properties of both endo- and exocellulases (19). Cel9A-68 consists of the family 9 CD rigidly attached to the CBM3c by a short linker. The crystal structure of this enzyme was determined by X-ray crystallography at 1.9-Å resolution (Protein Data Bank code 4TF4) (13). The CD has a shallow open active site cleft with six known glucose binding sites, numbered from −4 to +2, aligned with a flat binding surface of the CBM3c (Fig. 1). The linker region between the CD and this CBM (G445 to P461) lies along the opposite side of the CD from the active cleft with many contacts with the CD and the CBM. The CBM consists of a 10β-stranded sandwich motif (13). It binds weakly to crystalline cellulose and is thought to play a role in catalysis by feeding a single cellulose molecule into the catalytic cleft. The CBM is critical for hydrolysis and for processivity, as a cloned enzyme construction of the family 9 CD with the linker Cel9A-51 but without this CBM had very low activity on CMC, bacterial cellulose (BC), and swollen cellulose (SWC) and showed no processivity (4) (Table 1).

Family 9 cellulases perform catalysis with inversion of configuration at the anomeric carbon (15). In the work of Zhou et al. (21), E424 was shown to be the catalytic acid and the residues D55 and D58 were shown to act in concert to function as the catalytic base. In the structure (4TF4), which has cellotetraose in the Glc(−1) to Glc(−4) subsites and cellobiose in the Glc(+1) to Glc(+2) subsites, a hydrogen bonding network

* Corresponding author. Mailing address: 458 Biotechnology Building, Cornell University, Ithaca, NY 14850. Phone: (607) 255-5706. Fax: (607) 255-2428. E-mail: dbw3@cornell.edu.
▽ Published ahead of print on 16 March 2007.

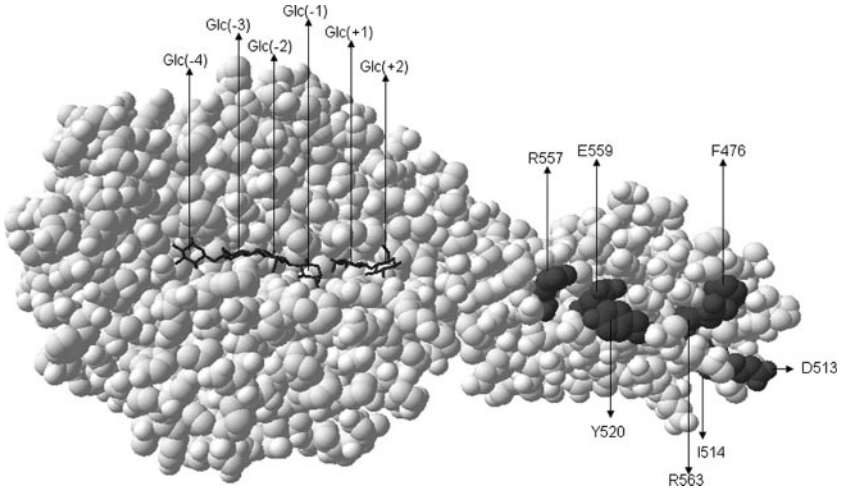

FIG. 1. Space-filled structure of Cel9A-68 (4TF4) with six glucose molecules in the catalytic cleft. This enzyme product structure was obtained after soaking the crystals in cellopentaose (G5). G5 was cleaved to G4 and G1 or G3 and G2, and density for six glucose residues can be seen in the structure because G5 can bind from the position of Glc(−3) to Glc(+2) or from Glc(−4) to Glc(+1). The CBM residues chosen for mutation are shown in dark gray.

can be seen between Y206, D55, Glc(−1) O1, D58, and H125. The Y206S mutant enzyme was shown to have almost no activity and zero processivity, and it was postulated that this charged network is essential for catalytic activity (21). Y318, a residue which hydrogen bonds to Glc(−3), was found to be essential for processivity (21). The work of Zhou et al. raised intriguing questions about the complex catalytic and processive mechanisms of this enzyme. Is H125 a crucial residue for activity, and does it also function in substrate binding? What other amino acid residues in the catalytic cleft are important for processivity? Does the CBM3c contribute to processivity as hypothesized, and which residues are important for this? In this work we investigated the role of the CBM3c in processivity by mutation of amino acids thought to be on the binding face of this module. H125 was mutated to determine its role in the proposed charged network at the catalytic center, and additional residues were chosen to further investigate the mechanism of processivity and the binding of the substrate into the catalytic cleft.

MATERIALS AND METHODS

Strains and plasmids. *Escherichia coli* XL10 (Stratagene) was used as the host strain for plasmid extraction, and *E. coli* BL21(DE3) (Stratagene) was used as the host strain for protein expression. The mature Cel9A-68 gene product (613 amino acids) with the *T. fusca* Cel6A signal sequence (MRMSPRPLRALLGA AAAALVSAAALAFPSQAA) in the pET-26b$^+$ vector (Novagen) was used as the template for mutagenesis.

Mutagenesis and DNA manipulations. Standard nucleic acid techniques were used as described by Sambrook et al. (14). All mutant genes were generated using the QuikChange site-directed mutagenesis method (Stratagene). Two oligonucleotide primers for each mutation were designed, complementary to each other on opposite strands at the site of the desired mutation. The length of both primers was 40 to 45 bases; G or C was on each terminus, and one restriction enzyme site was created or deleted for screening. QuikChange PCR was performed under the following conditions: (i) denaturation at 95°C for 2 min; (ii) addition of *Pfu* polymerase (Stratagene); (iii) 18 cycles of denaturation at 95°C for 60 s, annealing at 60°C for 50 s, and extension at 68°C for 8 min; and (iv) a final step at 68°C for 7 min. The sample (25 μl) was then incubated with 20 units of DpnI for 3 h at 37°C to remove the methylated template, and the sample was concentrated by ethanol precipitation at −20°C overnight. The PCR product was transformed into *E. coli* XL10, and plasmid minipreparations of individual colonies were screened by restriction enzyme digestion for the altered site. The DNA sequences of the candidate plasmids were determined by the Cornell Biotechnology Resource Center using a Perkin-Elmer/Applied Biosystems automated sequencer to confirm the presence of the desired mutation and to confirm that no other mutations were present. Plasmid DNA was transformed into *E. coli* BL21(DE3) for protein expression.

Enzyme production and purification. *E. coli* BL21(DE3) strains were grown at 37°C overnight in 30 ml of Luria broth with 60-μg/ml kanamycin. The overnight culture was transferred into 1 liter of M9 medium containing 0.8% glucose and 60-μg/ml kanamycin. After growth at 30°C to an optical density at 600 nm (OD_{600}) of about 0.8, isopropylthio-β-D-galactoside was added to the culture to 0.8 mM and growth was continued at 30°C overnight. The proteins were purified from the culture supernatants on a phenyl-Sepharose column, followed by a Q-Sepharose column, as described in previous work (4). Protein purity was assessed on sodium dodecyl sulfate gels, and the protein concentrations were measured by absorbance at OD_{280} using molar extinction coefficients of 165,480 for Cel9A-68 and 129,110 for Cel9A-51, which were determined from the predicted amino acid compositions using the DNA Star Protean program.

Activity assays. The activities of Cel9A-68 wild-type and mutant enzymes were determined concurrently on CMC (10 mg/ml), BC (2.5 mg/ml), phosphoric acid-SWC (2.5 mg/ml), and Whatman FP no. 1 (8 mg/ml). All assays were run in a reaction volume of 0.4 ml in triplicate at 50°C for 16 h in 0.05 M sodium acetate, 0.015 M $CaCl_2$ (pH 5.5), using from 0.5 to 600 pmol of protein as needed. Reducing sugars were measured using the dinitrosalicylic acid (DNS) reagent (3) and a glucose standard curve using Sigma 635-100 glucose standard solution. Thin-layer chromatography of the BC assay products under these conditions showed that the major product was cellobiose along with approximately 10% glucose and a small amount of cellotriose for some mutants (data not shown). To calculate the activity, the OD_{600} values were converted to micromoles of cellobiose by comparing a cellobiose standard curve and a glucose standard curve. A major advantage of the DNS reagent is that it does not have a blank with the amount of enzyme used and it fits the assay range well. The disadvantage of the DNS reagent, used with a cellobiose standard, is that glucose will be about 30% underestimated and cellotriose will be slightly overestimated. Mutant enzyme activities were determined concurrently with those of the wild type; the nanomoles of protein used (X) were plotted versus the OD_{600} (Y) and fitted to the equation $Y = m_1X/(m_2 + X)$ using the program KaleidaGraph from Synergy Software. This curve was used to determine the amount of enzyme required to digest the substrate to the target value of 5% for CMC and FP (0.585 μmol and 0.467 μmol cellobiose, respectively) and 10% for SWC and BC (0.292 μmol cellobiose). For those enzymes which could not reach the target digestion, the activity was calculated from the micromoles of cellobiose produced by 0.6 nmol of enzyme. Although the estimated overall error of CMC and FP activities is around 5% and that for SWC and BC is around 10% from previous work, the

TABLE 1. Activity and binding characteristics of mutant enzymes

Enzyme	Activity on substrate[a] (μmol cellobiose min^{-1} μmol $enzyme^{-1}$):				Processivity[b] (ratio of soluble/insoluble reducing ends produced)	Binding to:	
	CMC	SWC	BC	FP		BC[b] (% bound)	MUG3 (dissociation constant [μM])
Cel9A-68 amino acids 1–613							
Wild type	279	10.5	3.3	0.28	3.1	15	0.57
Key catalytic residue mutants							
E424A	0.28[d]	0.03[d]	0.01[d]			40	1.23[d]
D55A	0.6[d]	0.14[d]	0.03[d]				2.21[d]
D55C	9.3	0.3				32	
D58A	0.9[d]	0.21[d]	0.07[d]				0.50[d]
Y206S	1.0[d]	0.15[d]	0.040[d]	0.06[d]	0[d]	5	Poor[d]
H125N	7.9	0.4	0.11	0.07	0.3	25	
H125A	6.7	0.7	0.09	0.05	0	36	0.73
CD mutants interacting with Glc(−1) to Glc(−4)							
F205A	96	4.4	0.33	0.14	1.4	11	2.30
Y429A	58	3.0	0.37	0.15	1.3	26	0.26
W313G	10	0.6	0.10	0.08	1.0	13	Poor
W313D	8.4	0.5	0.08	0.06	0.9	16	Poor
W209S	629	10.5	0.46	0.18	2.6	21	
W256A	1,080	7.4	0.26	0.13	1.8	16	
R317K	443	9.2	0.44	0.18	2.2	19	
D261A	758	9.6	0.49	0.17	2.6	20	
D261N	1,510	10.0	0.37	0.16	1.6	18	
CD mutants interacting with Glc(+1), Glc(+2), and putative Glc(+3)							
R378K	270	9.6	3.4	0.26	4.3	20	
W128A	32	0.4	0.25	0.12	1.9	18	
CD double mutant							
D261A/R378K	503	3.0	0.18	0.10	1.3	18	
CBM mutants							
Y520A	302	8.2	1.1	0.22	2.7	14	
D513A	279	12.1	4.0	0.30	3.7	20	
I514H	290	9.6	3.7	0.29	3.4	23	
R557A/E559A	252	7.9	0.48	0.21	3.3	15	
F476A	270	11	2.6	0.28	4.5	21	
R563A	270	7.4	1.7	0.26	4.0	20	
Cel9A-51 amino acids 1–463							
Wild type	24	0.6	0.2[c]	0.13[c]	0.6[c]	8	
D55C	0.9	0.1				10	

[a] The average coefficients of variation were 3%, 4%, 3%, and 2% for CMC, SWC, BC, and FP, respectively.
[b] The average coefficients of variation for processivity and BC binding were each 3%.
[c] Data from reference 4.
[d] Data from reference 21.

calculated coefficients of variation for a given assay are below 5% as noted in Table 1.

Processivity was determined from the distribution of reducing ends between the FP (insoluble) and the supernatant (soluble) following a published procedure (5). Assays were performed on FP using 0.6 nmol of enzyme per reaction; the supernatant was separated from the FP circle, and the reducing sugar content of each fraction was determined with the DNS reagent (3).

Binding assays. Binding to BC was measured in 2.2-ml "low-adhesion" tubes (USA Scientific) by adding 2 nmol of enzyme to 2.5 mg/ml BC in a total volume of 0.5 ml of 50 mM sodium acetate buffer, pH 5.5, plus 10% glycerol. A rack of tubes was placed on a Nutator rocking table (Clay-Adams) at 4°C for 1 h. The tubes were then centrifuged for 5 min at 13,000 rpm, and the top 300 μl of liquid was transferred to 1.5-ml Eppendorf tubes with a gel loading pipette tip. The tubes were centrifuged again for 5 min at 13,000 rpm, and the top 150 μl of liquid was transferred to clean 0.5-ml Eppendorf tubes with a gel loading pipette tip. The amount of unbound protein was determined by the protein concentration calculated from the OD_{280}; control reactions without BC and also with BC alone were performed for every assay.

Ligand binding studies. Dissociation constants for the binding of 4-methylumbelliferyl β-cellotrioside (MUG3) to Cel9A-68 and to mutant enzymes were determined by direct fluorescence quenching titrations using an Aminco SLM 8000C spectrofluorimeter and were calculated as described previously (1). The initial concentration of MUG3 was 1.8 μM in 903 μl of 50 mM sodium acetate buffer, pH 5.5, and 100 μM protein was added at 3.5 μl/min. The temperature was maintained at 6.5°C, and excitation was at 316 nm, while emission was measured at 360 nm.

Azide rescue studies. The azide activity rescue studies were run in a reaction volume of 0.4 ml in triplicate at 50°C for 16 h in 0.05 M sodium acetate (pH 5.5), with 0.4 nmol of protein and addition of 0.075 M to 0.6 M sodium azide; 200 μl of 5.0-mg/ml SWC was used as substrate in the reactions. Reducing sugars were measured using the DNS method, and the activity was determined from the micromoles of cellobiose produced by 0.4 nmol of enzyme.

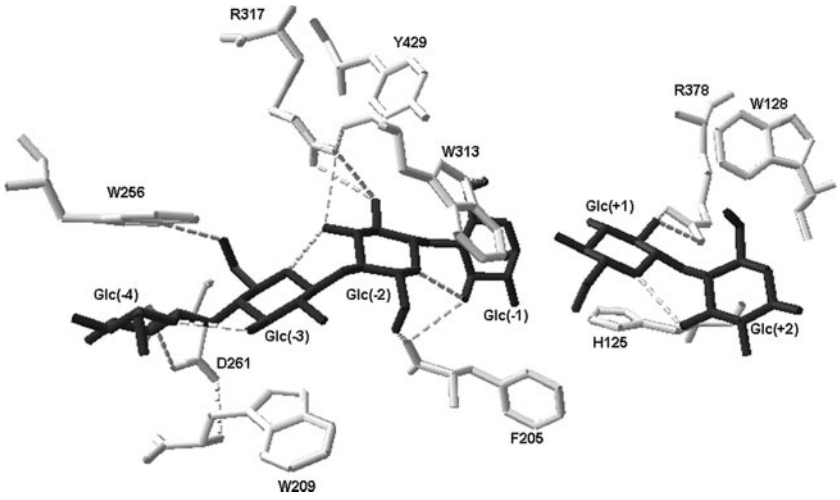

FIG. 2. Cel9A-68 product complex structure (4TF4) showing amino acids chosen for mutation around the catalytic center in light gray and the six sugar residues in dark gray. Hydrogen bonds are shown as dashed lines. Cleavage occurs between Glc(−1) and Glc(+1).

Accession numbers. The sequence files for Cel9A (formerly called E4) are L20093 and P26221. The structure files for Cel9A-68 are 1JS4, 1TF4, 3TF4, and 4TF4.

RESULTS

Tests used to compare wild-type and mutant enzymes. Mutant enzyme assays were performed using four different substrates. CMC is a soluble substrate that endocellulases readily degrade. SWC is insoluble, but its cellulose molecules are more hydrated than those of crystalline cellulose and have fewer hydrogen bonds to each other. Therefore, it is much easier for an enzyme to bind a single chain of SWC into its catalytic site. BC and FP are both crystalline substrates, where the cellulose chains are highly ordered into microfibrils by hydrogen bonding and hydrophobic stacking. Processivity was measured by comparing the ratio of soluble reducing ends in the supernatant to the insoluble reducing ends left in an FP, which reflects the ability of a cellulase to "process" along the cellulose chain: that is, to make a hydrolytic cleavage, release the product, and then move the same molecule further along the active site and into position for another hydrolytic cleavage without releasing the substrate. This test is useful for comparison purposes only, as there is no "standard curve" for the absolute amount of insoluble reducing ends. The binding of Cel9A-68 to BC is influenced by a combination of activity and affinity, which makes this test complicated to interpret. Wild-type Cel9A-68 does not bind measurably to BC at room temperature (4); thus, we performed the binding assays at 4°C to minimize (but not totally eliminate) the effect of activity. At 4°C, wild-type Cel9A-68 has about 15% binding to BC; the E424A catalytic acid mutant enzyme has 40% binding; while the Y206S enzyme, which has very low activity, also has very poor binding of 5%. These values represent the most likely maximum and minimum binding to BC in the current assay. MUG3 undergoes a measurable fluorescence change upon binding into the Glc(−2) to Glc(+1) sites of Cel9A-68 (21). Therefore, fluorescence titration binding assays reflect the active-site binding affinity for small, soluble ligands.

Activity of CD mutant enzymes. The selection of CD target residues focused on the catalytic site, especially those residues with hydrogen bonding or hydrophobic stacking to one of the six sugar residues in the active cleft. The residues chosen are listed in Table 1 and illustrated in Fig. 2. Data for some mutated residues previously published but discussed in this paper are included in Table 1 and referenced.

The catalytic base was known from the work of Zhou et al. (21) to be either D55 or D58. An azide rescue test was performed to determine which one was the acceptor of the nucleophilic water hydrogen and thus the actual catalytic base (16). The activity of the D58A enzyme was improved in a dose-dependent manner in the presence of sodium azide while the activity of the D55A enzyme remained unchanged, indicating that D58 is the catalytic base (Fig. 3). However, D55 is an essential supporting residue based on previous data (21). The fact that the activity of the E424A enzyme was not improved in the presence of sodium azide supports the idea that the addition of exogenous nucleophile can reactivate a catalytic base mutant but not a catalytic acid mutant in inverting glycoside hydrolases (16).

H125 makes one hydrogen bond to D58 OD2 and is part of a hydrogen bond network that could stabilize the interaction of D55 and D58 with the nucleophilic water molecule (Fig. 4). It can be predicted from the crystal structure that both mutations (H125A and H125N) disrupt the hydrogen bond between D58 and H125. In addition there is another hydrogen bond predicted to be formed between N125 OD1 and Glc(+2) O4. Half of the family 9 members have an N at position 125, and only one in five has an H (10). Both of the mutant enzymes, the H125A and H125N mutants, lost more than 93% of their activity on SWC, CMC, and BC and bound more tightly to BC (Table 1). The H125A mutant enzyme had almost normal MUG3 binding, showing that the loss in activity was not due to disruption of the substrate binding site. In addition, the H125A mutant has zero processivity, as does the Y206S mutant (21). These results show conclusively that the charged network

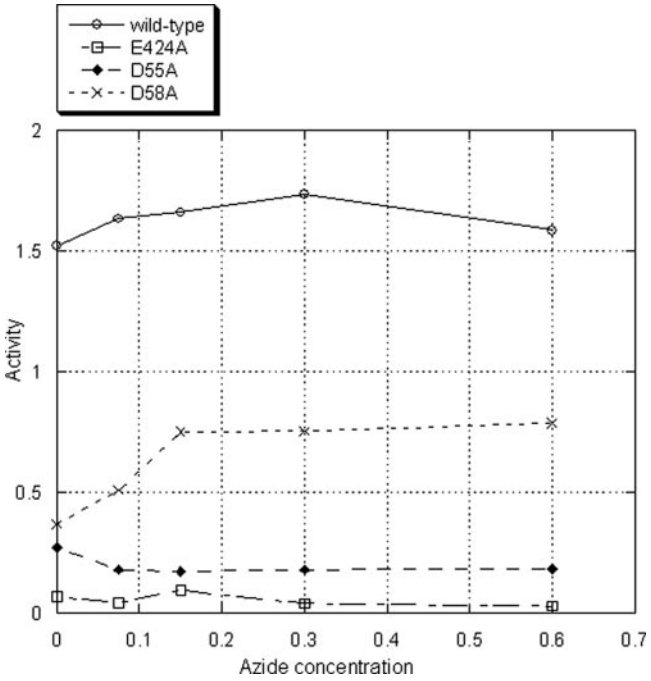

FIG. 3. Sodium azide rescue test. The samples were treated with various concentrations of azide (0 M to 0.6 M) for 16 h, and then SWC activity was assayed and calculated [μmol of cellobiose/(min · μmol of enzyme)]. The average coefficients of variation were 4%, 5%, 3%, and 4% for wild-type and E424, D55, and D58 mutant enzymes, respectively.

formed between Y206, D55, Glc(−1), D58, and H125 is essential for catalytic function (Fig. 4).

Another important residue is W313 (Fig. 2), which is hydrogen bonded to Glc(−1) O6. The activities of the W313G and W313D mutant enzymes are below 10% of that of the wild type, their processivity is reduced dramatically, and MUG3 binding is very poor for both mutant enzymes. This is good evidence that both the hydrogen bond and the hydrophobic interaction of this tryptophan side chain are very important to produce the strained configuration of Glc(−1) which aids in catalysis and processivity.

F205 and Y429 (Fig. 2) do not have direct interactions with any sugar residues, even though they are adjacent to the catalytic site. The F205A mutant enzyme has lower activity on all substrates, lower processivity, and significantly lower oligosaccharide binding. F205 has one hydrogen bond from its backbone carbonyl group shared between Glc(−2) O6 and Glc(−1) O3, which is consistent with lower binding. The Y429A mutant has normal binding but low crystalline cellulose activity and impaired processivity.

W209 appears to stack against Glc(−3), and W256 stacks against Glc(−4) and is hydrogen bonded to the Glc(−3) O6 (Fig. 2). D261 is at the bottom of the substrate cleft and is hydrogen bonded to Glc(−4) O2 and to S210 OG. D261 and S210 work together to bind Ca^{2+}. R317 is hydrogen bonded to Glc(−2) O2, and a previously studied residue, Y318, appears to be hydrogen bonded to Glc(−2) O3, Glc(−3) O5, or Glc(−3) O6. All of the enzymes with mutations of the above residues, Y318A (21), Y318F (21), W209S, W256A, R317K, D261A, and D261N, have almost-normal activity on SWC and severalfold-higher activity on CMC, yet only 15% of wild-type activity on BC and reduced processivity. All of these residues are part of the Glc(−2) to Glc(−4) subsites, and it appears that each of them participates in binding crystalline substrates such as BC in the active cleft. The higher CMC activities could be due to the smaller side chains of the mutants allowing more room for the carboxymethyl groups to enter and leave the catalytic cleft.

Another interesting mutation is R378K, which makes two hydrogen bonds, NH1 and NH2, to Glc(+1) O2 (Fig. 2). The mutation does not change enzyme activity very much, but processivity is improved significantly. So far this is the only muta-

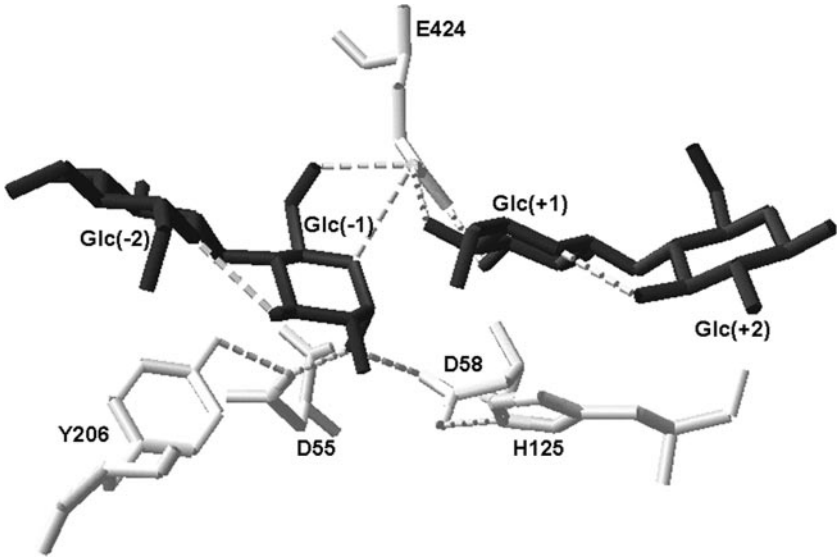

FIG. 4. Enzyme product complex (4TF4) showing the hydrogen bond network between Y206, D55, Glc(−1), D58, and H125. The hydrogen bonds from the catalytic acid E424 to Glc(−1) and Glc(+1) are also shown.

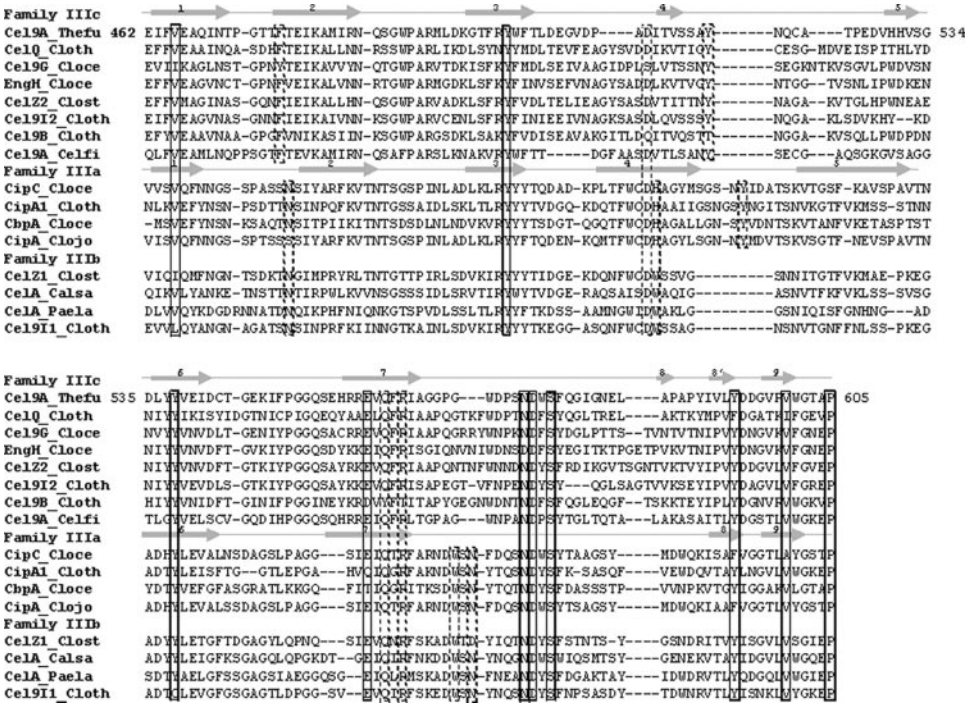

FIG. 5. Structure-based sequence alignment of family 3 CBMs. Secondary structural β-strands are indicated by arrows and enumerated. Proposed cellulose binding residues are shown in dashed boxes, and for *T. fusca* Cel9A-68, they are F476, D513, Y520, Q561, and R563. Residues shown in solid boxes are conserved, and those *T. fusca* residues are V465, Y501, Y538, E559, N574, D575, S577, Y594, V600, and P605. Sequences were obtained from the respective GenBank accession codes (included parenthetically) as follows: Cel9A_Thefu, from *Thermobifida fusca* (M73322); CelQ_Cloth, from *Clostridium thermocellum* (AB047845); Cel9G_Cloce, from *Clostridium cellulolyticum* (M87018); EngH_Cloce, from *Clostridium cellulovorans* (U34793); CelZ2_Clost, from *Clostridium stercorarium* (X55299); Cel9I2_Cloth, from *Clostridium thermocellum* (L04735); Cel9B_Cloth, from *Clostridium thermocellum* (X60545); Cel9A_Celfi, from *Cellulomonas fimi* (M64644); CipC_Cloce, from *Clostridium cellulolyticum* (U40345); CipA1_Cloth, from *Clostridium thermocellum* (X67406); CbpA_Cloce, from *Clostridium cellulovorans* (M73817); CipA_Clojo, from *Clostridium josui* (AB004845); CelZ1_Clost, from *Clostridium stercorarium* (X55299); CelA_Calsa, from *Caldicellulosiruptor saccharolyticus* (M36063); CelA_Paela, from *Paenibacillus lautus* (M76588); and Cel9I1_Cloth, from *Clostridium thermocellum* (L04735).

tion in the CD that results in obviously higher processivity. The D261A mutant enzyme has lower processivity than the wild type, and it is located near Glc(−4). An D261A/R378K double mutant enzyme was made, and its processivity decreased dramatically, even lower than that of the D261A mutant enzyme, indicating that processivity requires a precise balance between catalytic cleft binding on the two sides of the cleavage site.

The length of one cellobiose unit is about 10.3 Å (17), and the distance from the top of the catalytic cleft to the end of the CBM3c is about 70 to ~80 Å, so mathematically as many as 16 glucose units could span the CD and the CBM. Amino acids which interact with sugar residues other than those in the six known subsites could play important roles in activity and processivity. It is apparent from the structure that W128 (Fig. 2) could stack or interact with a possible Glc(+3) subsite. The W128A mutant enzyme lost activity on all substrates, and processivity dropped significantly. It is also interesting that this tryptophan side chain is predicted to have several energetically favorable positions based on the rotamer scores (three out of seven positions have scores of −1) reported by Swiss-PdbViewer.

Activity of CBM3c mutant enzymes. The initial step in cellulose degradation is the binding of the enzyme to the substrate, and this process clearly involves the CBM3c although Cel9A-68 binds only weakly to BC. There is no direct evidence for which CBM residues bind cellulose, but Tormo et al. (18) hypothesized that the flat face of the family 3a CBM, which is made up of four β-strands, could interact with three chains of cellulose: a strip of mostly aromatic residues, D, H, Y, R, and W, could interact with one chain; two well-conserved polar residues, N and Q, could interact with the middle chain; and nonconserved polar residues could interact with the third chain. The Cel9A CBM3c does not contain the strip of aromatic residues which exists in family 3a and 3b CBMs, even though it has the flat face which is made up of strands 1, 2, 4, and 7 (13). Six residues in the Cel9A-68 CBM that could potentially interact with a cellulose molecule were chosen for mutation based on sequence alignment (Fig. 5), a model (2), and consideration of the structure (Table 1). BC activity was lowered significantly in the R557A/E559A, Y520A, and R563A mutant enzymes. These amino acids are "in line" with the catalytic cleft (Fig. 1). SWC activity was only slightly lowered by these mutations, and CMC activity was not significantly affected. In spite of the lower BC activity, processivity was not decreased, which is an unexpected result. Mutation of the highly conserved F476 residue to A resulted in 21% decreased BC activity but increased processivity. The D513A mutation increased BC activity by 21%. D513 is located at the end of the CBM3c and appears to be not quite aligned with the catalytic cleft. Another interesting residue is I514, which is conserved

neither in family 3c nor on the β strands of the flat face. However, sequence alignment shows that this position is conserved, as an H in family 3a and a W in family 3b. In a family 3a CBM from *Clostridium cellulovorans*, the H is in a planar strip and plays an important role in cellulose binding (9). The I514H mutant enzyme has higher activity on BC and almost-wild-type activity on the other substrates. The BC binding of all of the tested CBM3c mutant enzymes was as good as or better than that of the wild type.

DISCUSSION

H125 was shown to be vital for activity and, together with Y206, forms a hydrogen bond network with Glc(−1) O1, D55, and D58. We suggest that this supporting hydrogen bond network is so important that it should be considered part of the catalytic mechanism. Addition of sodium azide to the reaction mix partially restored the activity of the D58A mutant enzyme and did not affect the activity of D55A or of the catalytic acid E424A mutant enzyme, which confirms D58 as the catalytic base. The azide rescue test did not fully restore the activity of the D58A mutant to the wild-type level as is usual for this procedure, probably due to the fact that although the azide ion can assist in hydrolysis, it could not fully imitate the role of the amino acid (16). Y429 does not appear to stabilize the transition state as suggested by Nerinckx et al. (11), as mutating it only partially reduced SWC activity.

An interesting finding is that all of the mutations that increase CMC activity are in the catalytic cleft from Glc(−2) to Glc(−4). The W209S, W256A, R317K, and D261A mutants all have smaller side chains than those in the wild type, which may allow the carboxymethyl-modified groups on CMC to enter and leave the catalytic cleft more easily to achieve more efficient cutting. This hypothesis does not explain why the D261N mutant enzyme has the highest CMC activity of any mutant enzyme studied so far, as the side chain is not changed in size by mutation. There is one hydrogen bond formed between D261 OD2 and Glc(−4) O2, and the distance between the two atoms is 2.76 Å. A hydrogen bond is also predicted between N261 ND2 and Glc(−4) O2, but the distance between the two atoms is 3.14 Å; therefore, this new hydrogen bond should be weaker than the original one so that the fragment of CMC after hydrolysis might leave the catalytic cleft more easily.

The ability of a cellulase to move along the substrate is critical for crystalline cellulose degradation. However, the mechanism of processivity is not yet completely understood. The R378K mutant is the first mutant with obviously higher processivity. R378 might have the function of moving the cellulose chain toward the catalytic center and into the right subsites. The R378K mutant enzyme loses a hydrogen bond to Glc(+1); thus, it might be easier for the enzyme to slide along the substrate. In contrast, the mutations in residues near subsites Glc(−1) to Glc(−4) (F205, Y429, W313, W256, R317, and D261) decrease processivity markedly. It is reasonable to propose that these mutations disturb the coordination between the release of the product and the subsequent binding of the cellulose molecule into these subsites as it moves from CBM3c.

Cel9A-51, which retains the linker but none of CBM3c, lost most of the activity on CMC and SWC and had barely measurable crystalline cellulose activity (Table 1). The CD alone has a ratio of soluble to insoluble reducing sugars of only 0.6 compared to 3.1 for Cel9A-68 (4). Clearly the CBM3c domain is essential for cellulose hydrolysis. Cel9A-68 D55C has 32% binding to BC compared to only 10% for Cel9A-51 D55C, indicating that the family 3c domain is also important for cellulose binding. Surprisingly, none of the CBM3c mutations reduced processivity very much; in fact most of them had improved processivity. Even the R557A/E559A double mutant, which has only 15% BC activity, has wild-type processivity. In contrast, all but one of the catalytic domain mutants had reduced processivity; some of them were drastically reduced. This strongly implies that it is the equilibrium binding into the catalytic cleft that controls processivity. It was proposed by Tormo et al. (18) that polar residues can interact with cellulose to contribute to affinity and may replace some of the interchain hydrogen bonds of cellulose so as to disrupt the tight cellulose structure. Four of the CBM3c mutants showed that BC activity was more severely affected than SWC activity, supporting this hypothesis. Based on these studies it appears that the function of CBM3c is to loosely anchor the enzyme on the substrate and to disrupt the hydrogen bond network in crystalline cellulose.

The detailed mechanism of how CBMs disrupt the hydrogen bond network in the crystalline substrate is still unclear. Preliminary data show that single mutations of N470, F476, K480, Y520, Q522, Q553, and R563, which are all suspected binding residues in the CBM3c, do not reduce the enzyme activity very much (the current data and data not shown). One reason could be that the pattern of interaction between the CBM and crystalline cellulose is so complicated, involving many residues simultaneously, that double or triple mutations are required to see changes in activity.

ACKNOWLEDGMENTS

We thank Edward Bayer of the Weizmann Institute and Colleen McGrath for valuable discussions.

This research was supported by grant DE-FG02-ER15356 from the U.S Department of Energy.

REFERENCES

1. **Barr, B. K., D. E. Wolfgang, K. Piens, M. Claeyssens, and D. B. Wilson.** 1998. Active-site binding of glycosides by *Thermomonospora fusca* endocellulase E2. Biochemistry **37:**9220–9229.
2. **Escovar-Kousen, J. M., D. Wilson, and D. Irwin.** 2004. Integration of computer modeling and initial studies of site-directed mutagenesis to improve cellulase activity on Cel9A from *Thermobifida fusca*. Appl. Biochem. Biotechnol. **113–116:**287–297.
3. **Ghose, T. K.** 1987. Measurement of cellulase activities. Pure Appl. Chem. **59:**257–268.
4. **Irwin, D., D. Shin, S. Zhang, B. K. Barr, J. Sakon, P. A. Karplus, and D. B. Wilson.** 1998. Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. J. Bacteriol. **180:**1709–1714.
5. **Irwin, D., L. Walker, M. Spezio, and D. Wilson.** 1993. Activity studies of eight purified cellulases: specificity, synergism, and binding domain effects. Biotech. Bioeng. **42:**1002–1013.
6. **Irwin, D. C., S. Zhang, and D. B. Wilson.** 2000. Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. Eur. J. Biochem. **267:**4988–4997.
7. **Jung, E. D., G. Lao, D. Irwin, B. K. Barr, A. Benjamin, and D. B. Wilson.** 1993. DNA sequences and expression in *Streptomyces lividans* of an exoglucanase gene and an endoglucanase gene from *Thermomonospora fusca*. Appl. Environ. Microbiol. **59:**3032–3043.
8. **Lao, G., G. S. Ghangas, E. D. Jung, and D. B. Wilson.** 1991. DNA sequences of three β-1,4-endoglucanase genes from *Thermomonospora fusca*. J. Bacteriol. **173:**3397–3407.
9. **Murashima, K., A. Kosugi, and R. H. Doi.** 2005. Site-directed mutagenesis and expression of the soluble form of the family IIIa cellulose binding domain from the cellulosomal scaffolding protein of *Clostridium cellulovorans*. J. Bacteriol. **187:**7146–7149.

10. **Nerinckx, W., T. Desmet, K. Piens, and M. Claeyssens.** 2005. An elaboration on the syn-anti proton donor concept of glycoside hydrolases: electrostatic stabilisation of the transition state as a general strategy. FEBS Lett. **579:** 302–312.
11. **Nerinckx, W., T. Desmet, and M. Claeyssens.** 2003. A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. FEBS Lett. **538:**1–7.
12. **Percival Zhang, Y. H., M. E. Himmel, and J. R. Mielenz.** 2006. Outlook for cellulase improvement: screening and selection strategies. Biotechnol. Adv. **24:**452–481.
13. **Sakon, J., D. Irwin, D. B. Wilson, and P. A. Karplus.** 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nat. Struct. Biol. **4:**810–817.
14. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
15. **Schülein, M.** 2000. Protein engineering of cellulases. Biochim. Biophys. Acta **1543:**239–252.
16. **Shallom, D., M. Leon, T. Bravman, A. Ben-David, G. Zaide, V. Belakhov, G. Shoham, D. Schomburg, T. Baasov, and Y. Shoham.** 2005. Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from *Geobacillus stearothermophilus* T-6. Biochemistry **44:** 387–397.
17. **Tarchevsky, I. A., and G. N. Marchenko.** 1991. Cellulose: biosynthesis and structure. Springer-Verlag, Berlin, Germany.
18. **Tormo, J., R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham, and T. A. Steitz.** 1996. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J. **15:** 5739–5751.
19. **Wilson, D. B.** 2004. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. Chem. Rec. **4:**72–82.
20. **Zhang, S., G. Lao, and D. B. Wilson.** 1995. Characterization of a *Thermomonospora fusca* exocellulase. Biochemistry **34:**3386–3395.
21. **Zhou, W., D. C. Irwin, J. Escovar-Kousen, and D. B. Wilson.** 2004. Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. Biochemistry **43:**9655–9663.