0273-2289/04/113/0287–0297/$25.00

Integration of Computer Modeling and Initial Studies of Site-Directed Mutagenesis to Improve Cellulase Activity on Cel9A from *Thermobifida fusca*

José M. Escovar-Kousen,*[,1] David Wilson,[2] and Diana Irwin,[2]

[1]Martek Biosciences, 555 Rolling Hills Lane, Winchester, KY 40391, E-mail: jescovar@martekbio.com; and [2]Department of Biochemistry, Cornell University, Ithaca, NY 14853, E-mail: dbw3@cornell.edu

Abstract

Cellulases are a complex group of enzymes that are fundamental for the degradation of amorphous and crystalline cellulose in lignocellulosic material. Unfortunately, cellulases have a low catalytic efficiency on their substrates when compared to similar enzymes such as amylases, which has led to a strong interest in improving their activities. *Thermobifida fusca* secretes six cellulose degrading enzymes: two exo- and three endocellulases and an endo/exocellulase Cel9A (formerly called E4). Cel9A shows unique properties because of its endo- and exocellulase characteristics, strong activity on crystalline cellulose, and good synergistic properties. Therefore, it is an excellent target for mutagenesis techniques to improve crystalline cellulose degradation. In this article, we describe research conducted to improve Cel9A catalytic efficiency using a rational design and computer modeling. A computer model of Cel9A was created using the program CHARMM plus its PDB structure and a cellohexose molecule attached to the catalytic site as a starting model. Initially molecular graphics and energy minimization were used to extend the cellulose chain to 18 glucose residues spanning the catalytic domain and cellulose-binding domain (CBD). The interaction between this cellulose chain and conserved CBD residues was determined in the model, and mutations likely to improve the binding properties of the CBD were selected. Site-directed mutations were carried out using the pET vector pET26b, *Escherichia coli* DH5-α, and the QuickChange mutagenesis method. *E. coli* BL21-DE3 was used for protein production and expression. The puri-

*Author to whom all correspondence and reprint requests should be addressed.

fied proteins were assayed for enzymatic activity on filter paper, swollen cellulose, bacterial microcrystalline cellulose, and carboxymethylcellulose (CMC). Mutation of the conserved residue F476 to Y476 gave a 40% improved activity in assays with soluble and amorphous cellulose such as CMC and swollen cellulose.

Index Entries: *Thermonospora fusca*; Cel9A; cellulases; protein engineering; computer modeling.

Introduction

Cellulose is the most widespread biologic material on Earth. It has been realistically estimated that 3.3×10^{11} tons of CO_2/yr is fixed on the world's surface and that approx 6.6% of this (e.g., 22,000 million t/yr will be cellulose (almost 4 t per person per year) *(1)*. Pure cellulose can be hydrolyzed by chemical or enzymatic methods to soluble sugars, which in turn may be used as the raw material for many different bioprocesses. Therefore, cellulose from both agriculture and forestry sources has the potential to become a major source of feedstock for the production of chemicals, single cell protein, and biofuels such as ethanol *(1)*.

However, cellulose is a complex biopolymer and invariably occurs in nature in close association with lignin, and to a lesser degree with hemicellulose, starch, proteins, and salts. The ability of lignocellulose to withstand degradative forces is witnessed by the longevity of trees and by the expensive energy-consuming pretreatment processes required to open up this complex structure to bioconversion. The high cost of converting cellulose into an easily metabolized raw material for industrial bioprocesses makes the use of cellulosic materials unattractive when compared with other sources of fermentable sugars such as corn starch or sugarcane. Therefore, important technological developments have to be achieved before economic use may be made of this plentiful compound *(2)*.

The enzymes responsible for cellulose degradation are a complex group of cellulases consisting of at least three classes: endo-1,4-β-glucanase (EC 3.2.1.4), also called endocellulases; exo-1,4-β-glucanase (EC 3.2.1.91), also called exocellulases; and β-1,4-glucosidase (EC 3.2.1.21) or cellobiase *(2–5)*. More than 200 different cellulase genes have been sequenced, and all cellulases and similar enzymes such as xylanases have been classified within 11 families that are sometimes referred to as cellulase families, with 7 of those families considered to contain true cellulases *(6–9)*. Structural studies show that the majority of cellulases are multidomain proteins and their structure resembles a tadpole *(7)* consisting of a catalytically competent core structure linked to a carbohydrate-binding domain (CBD) also known as a cellulose-binding domain (CBD) via a flexible, often highly glycosylated linker region *(6)*. Even though all of these structures share a similar acid/base catalytic mechanism involving two or more aspartate or glutamate residues, the structures of the cellulases from each family may be totally different.

Many microorganisms, particularly fungi and thermophilic bacteria, possess cellulolytic enzymes of different kinds *(10–21)*. Because of their

strong cellulose-degrading enzymes, extensive research has been performed on the cellulases of the following microorganisms: *Trichoderma reesei (11,17,18)*, *Clostridium thermocellum (8,22)*, *Thermomonospora fusca (2,4,5,19–21)*, *Cellulomonas fimi (23)*, and *Humicola insolens (9,12,14)*. *T. reesei* cellulases are particularly well known, degrade cellulose in a cooperative manner, and consist of cellobiohydrolases (CBH-I and CBH-II) releasing cellobiose from the reducing end and nonreducing end, respectively, of the cellulose chain and two endoglucanases (EG-I and EG-III) cleaving internal glucosidic bonds *(11)*. *T. reesei*'s cellulolytic mixture, coming from selected strains in industrial fermentation processes, has the strongest cellulose-degrading activity of all known microorganisms. However, the hydrolysis of cellulose using these microbial enzymes—or for that matter any other hydrolysis procedure—is a long and inefficient process.

The actinomycete *T. fusca*, a thermophilic, filamentous soil bacterium, produces six structurally and functionally distinct cellulases: two exocellulases—Cel48A (ex-E6), which attacks the cellulose chain from the reducing end, and Cel6B (ex-E3), which attacks from the nonreducing end; three endocellulases—Cel5A (ex-E5), Cel6A (ex-E2), and Cel9B (ex-E1); and an endo/exo cellulase—Cel9A or ex-E4 *(2,4,5)*. All are moderately heat stable and have a broad pH optimum centered at 6.5.

The enzyme Cel9A, the subject of this article, is a unique cellulase because it has characteristics of both an exocellulase and an endocellulase *(20,21)*. It is endocellulase-like in that it can hydrolyze and reduce the viscosity of carboxymethylcellulose (CMC), and it is exocellulase-like in that it produces mostly soluble oligosaccharides (87%) from insoluble cellulose *(20,21)*. Cel9A is also of interest because it possesses very high activity on bacterial microcrystalline cellulose (BMCC) and an unusual synergistic activity in cellulase mixtures. It is unique in that it shows synergism with endocellulases and both types of exocellulases *(20,21)*. It also retains more than 70% of its activity from pH 4.7 to 10.1.

Structurally, Cel9A consists of a catalytic domain (CD) and an adjacent CBD (CBM or CBD) (family IIIc), a linker, and another CBM (family II) (*see* Fig. 1). Cel9A CD is homologous to celD from *C. thermocellum* while the adjacent CBM is homologous to one in a scaffoldin from the same microorganism. Cel9A-68 (also E4-68) is generated by limited proteolysis and contains the CD and the adjacent CBM. Constructs containing the CD alone (Cel9A-51), and the CD plus the fibronectin linker and family II CBM (Cel9A-74), were also produced. The activities of all these enzymes were measured on BMCC, filter paper, swollen cellulose, and CMC. Removal of the internal CBD decreased activity substantially on every substrate. Cel9A-74 did bind to BMCC but had almost no hydrolytic activity, while Cel9A-68 retained 32% of the activity on BMCC even though it did not bind. Cel9A-68 maintains most of the activity toward all soluble substrates, suggesting that the family IIIc CBD is important for processivity while the second CBD is important for insoluble substrate binding *(20,21)*. This article focuses on Cel9A-68 in general and the adjacent CBM in particular.

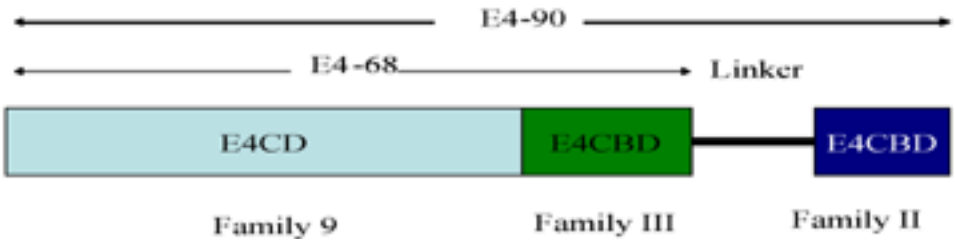

Fig. 1. Cel9A (E4) domain structure.

The crystal structure of Cel9A-68 has been determined by X-ray diffraction at 1.9 Å *(20)*. The structure reveals a CD from family 9 with an $(\alpha/\alpha)_6$ barrel fold directly linked to a CBD with an antiparallel β-sandwich fold. A model of the active-site configuration suggests that a water molecule is bound to Asp55 and Asp58, which are believed to function as the catalytic base, while Glu424 appears to be the catalytic acid. The observed cleavage product clearly shows the α configuration of the anomeric carbon, a fact expected from family 9 cellulases that cleave cellulose with inversion. Next to the active center there are six glucosyl-binding sites numbered for the nonreducing end, –4 to +2. Cel9A seems to processively cleave cellotetraose molecules from the nonreducing end of the cellulose chain until either they are completely hydrolyzed or it dissociates. The active-site cleft is blocked by a loop at the nonreducing end, and after each cleavage, the cellotetraose unit is expected to dissociate while the fragment that is bound to the catalytic site +1 and +2 and the family IIIc CBM is fed into the active site in a processive manner *(20,21)*.

To make biomass conversion more efficient, it will be necessary to either improve the catalytic activity of the enzymes (mainly cellulases) already known, discover new cellulases or cellulose-modifying proteins, increase the amount of protein secreted, modify the process, or combine these approaches. Cellulases show relatively low catalytic activity when compared with other hydrolases, and the low activities have called into question whether the genes encoding for the enzymes are under selection pressure to increase their catalytic efficiency *(24)*. Cellulase performance may be enhanced by the use of random mutagenesis techniques such as directed evolution *(25–29)* or rational protein design *(6,24,30)*. A major difficulty in making predictions for site-directed mutations for rational design in macromolecules is the lack of understanding of the forces that operate within proteins and between the protein and its substrate. Molecular mechanics calculations allow the theoretical positioning of substrate molecules into active sites through computer modeling and examination of the range of structural fluctuations that is possible for a protein-substrate complex *(31,32)*. Hence, computational studies usually facilitate the analysis of known protein structures as well as the design of proteins with novel

functional properties, reducing the cumbersome task of creating and studying countless mutants. Therefore, a good alternative for rational design is the use of computer modeling to determine individual changes followed by construction of the desired molecule by site-directed mutagenesis *(6,24,30)*. For our purposes, we concentrated on understanding the relationships among the structure of Cel9A-68, its catalytic activity, and particularly the interaction between the family IIIc CBM and cellulose. A molecular mechanics software and visualization programs *(31)* were used in combination with site-directed mutagenesis techniques.

Materials and Methods

Computer Modeling

All of the calculations were done using the CHARMM molecular mechanics program and the CHARMM22 parameter set for protein atoms *(33)*. All atoms from the glucose residues in the cellulose chain (initially a cellohexose) were modeled using parameters specifically developed for carbohydrates *(34)*. In all calculations, the lengths of chemical bonds to hydrogen atoms were kept fixed using the constraint algorithm SHAKE *(35)*.

A computer model of Cel9A-68 was created using the PDB file from the crystal structure published by Sakon et al. *(20)*. The protein structure plus the cellohexose molecule directly bound to the catalytic site were used as the starting point for all calculations. In the initial phase, molecular graphics visualization programs such as QUANTA and energy minimization calculations were used to place 12 additional glucose residues to extend the cellulose chain to 18 glucose residues spanning the CD and CBD. A number of possible conformations were evaluated using different starting positions and energy minimization-driven conformational refinement. The interactions between the cellulose chain and the conserved CBM residues (F476, D513, Y520, Q561, D513) *(36)* were determined in the model. Computational studies were conducted to determine the mutations most likely to create improved activity and were selected for site-directed mutagenesis.

Mutagenesis Procedure, Hosts, and Plasmids

The cloning and expression of recombinant proteins in *Escherichia coli* was done using the pET System developed by Novagen. The target gene from Cel9A-68 was ligated into pET26b to create plasmid pJE2 and transformed into *E. coli* DH5-α for mutagenesis. Plasmid pJE2 was constructed by ligating four DNA fragments: (1) a 4.3-kb *Sph*I-*Apa*I fragment from the pet vector, (2) a 1.15-kb *Apa*I-*Not*I fragment from pEJ containing the E2 signal peptide *(5)*, (3) a 1.0-kb *Not*I-*Acc*I polymerase chain reaction (PCR) fragment containing the mature Cel9A N-terminus plus a part of the gene obtained from pEJ2 (pD568) *(20)*, (4) a 0.7-kb *Acc*I-*Sph*I fragment containing the remaining part of Cel9A plus the C-terminus from pSZ46 (or pD687) *(21)*.

Plasmid pJE2 was used to construct the mutant plasmids (F476Y) using the QuickChange method for site-directed mutation. The forward and reverse primers were designed using the program DNAStar-Primer Select. The primers containing a silent mutation for threonine that created a unique restriction site for *Xma*I and *Sma*I, plus the desired codon change (underlined) for Tyr (TAC) instead of Phe (TTC), were as follows: Je1476tyr: TCA AC**ACC**CCGGGCACCACGTACACCGAGATC (forward); Je2476tyrrev: GATCTCGGTGTACGTGGTGCCCGG**GGT**GTTGA (reverse). The resulting PCR fragment was cloned into pET26b to produce pJE9. All coding regions in the new plasmids were sequenced by the Cornell Biotechnology Facility and only the desired mutations were found. All plasmids were transformed into *E. coli* BL21-DE3 for protein production and expression, and the modified enzymes were purified and characterized.

Protein Purification

Protein expression and production was scaled up to 2-L fermentors and a combination of hydrophobic interaction chromatography and ion-exchange chromatography proved very effective for purifying the enzymes *(20,21)*.

Enzyme Assays

The purified proteins were assayed for enzyme activity on filter paper, swollen cellulose, BMCC, and CMC following the procedures stated by Irwin et al *(5)*.

Results

Computer Modeling

Figures 2–4 indicate the CHARMM predictions for the structure and position of the cellulose chain (green) with respect to Cel9A (white and red) and the sequential manner in which the model was predicted by the program (one glucose residue at a time), starting from the original structure with 6 glucose residues *(20)* to 8 (*see* Fig. 2) to 12 (Fig. 3) and 16 (Fig. 4) with the CBD conserved residues in blue. Figure 5 shows the coordinates for the catalytic residues (yellow and red); the substrate (green); and the CBM conserved residues D513, R563, Y520, Q561, and F476 *(36)* (blue). As predicted in previous articles, based on the planar structure of the Cel9A-68 CBM, the cellodextrin chain is directly linked to the catalytic region through the CBD.

An important part of the present study was determining whether the residues suggested by Sakon et al. *(20)* (N470, CEL9A78, K480, R557, E559, Q561, and R563) and Bayer et al. *(36)* (D513, R563, Y520, Q561, and F476) interact with the cellulose substrate. The computational results indicate that all of these amino acids are well aligned and interact with the cellulose

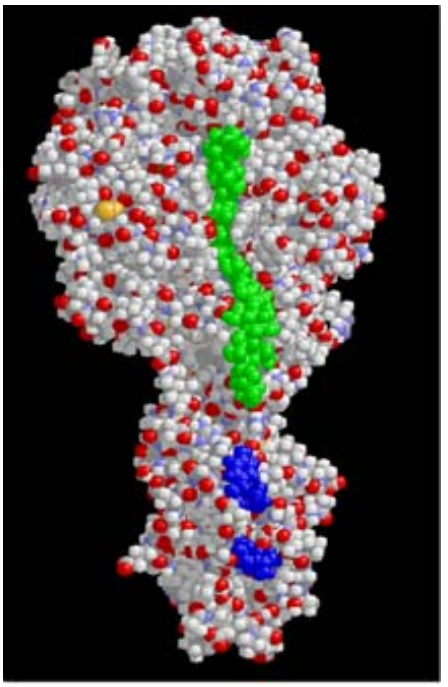

Fig. 2. CHARM progressive prediction for interaction between Cel9A and a cellulose chain with eight glucose residues.

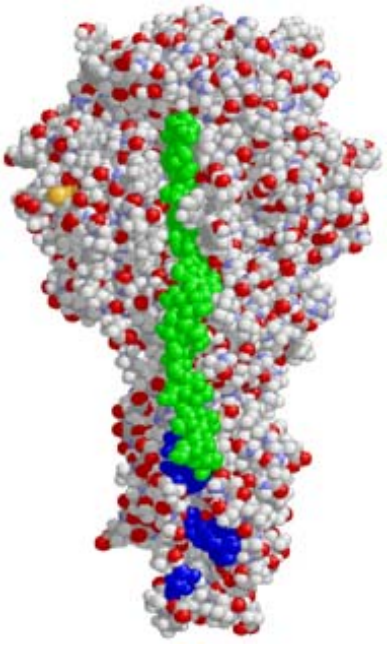

Fig. 3. CHARM progressive prediction for interaction between Cel9A and a cellulose chain with 12 glucose residues.

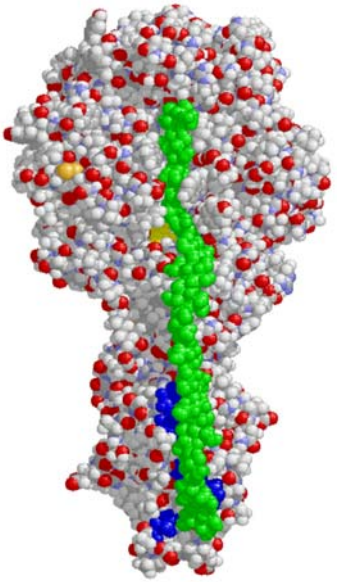

Fig. 4. CHARM progressive prediction for interaction between Cel9A and a cellulose chain with 16 glucose residues.

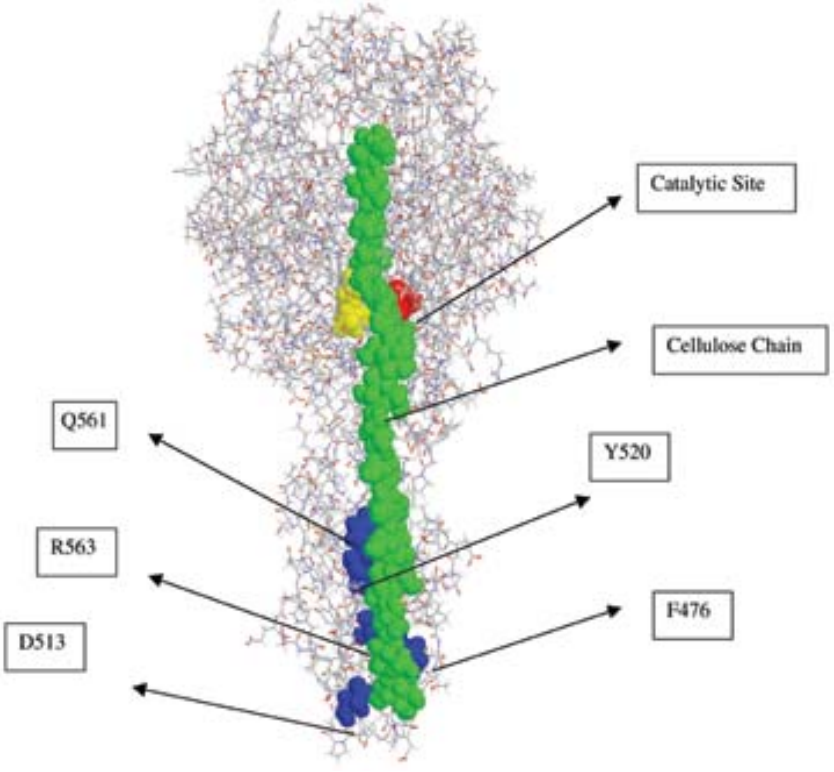

Fig. 5. CHARM predicted interaction between a 16 glucose cellulose chain and the CBM conserved residues.

Table 1
Cel9A-68 F476Y Activity [a]

Substrate	CMC (%)	Swollen cellulose (%)	Filter Paper (%)	BMCC (%)
Cel9A-68 F476Y	139	138	101	100

[a] With Cel9A-68 wild-type activity as 100%.

chain, with most of them forming hydrogen bonds or other interactions, i.e., a salt bridge and a planar interaction with the aromatic residue Y520 that are anchoring the surface of the CBM to the surface of the substrate.

The novel, conserved aromatic residue F476, which is characteristic of subfamily IIIc CBMs, was particularly interesting in the sense that it did not form a hydrogen bond with the cellodextrin chain, and as a consequence its interaction with the substrate seems to be weak. Therefore, computational studies were conducted on this residue, and the results indicated that mutation of F476 to Y created a hydrogen bond between the CBM and the cellulose chain. This was likely to improve binding, and thus it was selected for site-directed mutagenesis.

Mutagenesis

The enzyme from the F476Y mutant was purified and assayed on CMC, BMCC, filter paper, and swollen cellulose; the results are given in Table 1 as percentages of Cel9A-68 wild-type activity. An improvement in enzyme activity close to 40% was observed in soluble and amorphous cellulose (CMC and swollen cellulose), and no change or very small in crystalline cellulose (filter paper and BMCC). It has been found that the family IIIc CBM is important for processivity *(21)*. Processivity, as measured by the ratio of soluble reducing sugar ends to insoluble sugar ends, was found to be almost identical for the two enzymes: 3.1 for the wild type and 3.0 for the mutant.

Discussion

The combination of computational studies and site-directed mutagenesis was found to be an effective way to reduce the number of mutants to be tested. The enzyme activity assays of Cel9A-68 F476Y indicate an improvement of 40% in activity on CMC and swollen cellulose and no modification in activity on crystalline substrates (filter paper and BMCC). This result seems to confirm the computer model prediction that a new hydrogen bond would be the result of mutating the phenylalanine residue F476 to a tyrosine. This novel hydrogen bond in F476Y improved activity on soluble and amorphous cellulose probably owing to better binding properties that are important for placing the cellulose chain in the CBD for the mentioned substrates. The fact that no improvement was found for crystalline cellulose indicates a different kind of interaction between the CBD and crystalline cellulose, on one hand, and amorphous and soluble cellulose, on the other.

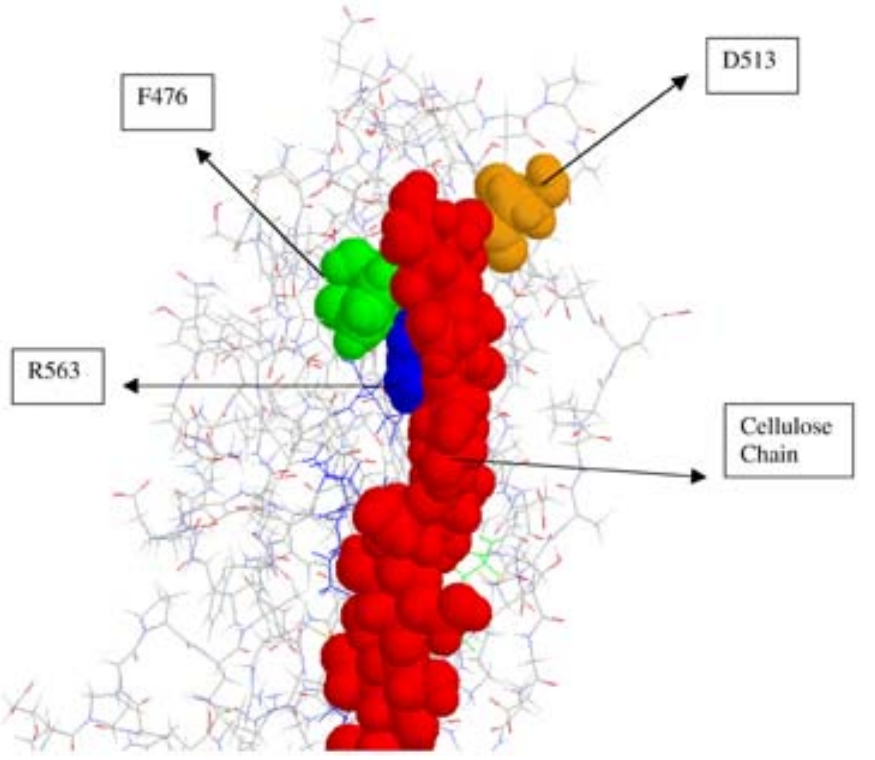

Fig. 6. Interaction between CBM residues R563, F476, and D513 with cellulose chain.

A possible explanation might be found in previous suggestions that Cel9A's CBD from family IIIc seems to bind transiently to only a single cellulose chain *(36)*, a factor that favors the interaction between the CBM and soluble and amorphous cellulose rather than crystalline cellulose.

Computer calculations indicate that the residues D513 and Y520 interact directly with the cellulose chain, a fact that is contrary to previous observations *(36)*. This result should be confirmed by future mutagenesis studies.

Because of the particular location at the end of the CBM (*see* Fig. 6) of residues D513 (orange), R563 (blue), and F476 (green), they seem to play an important role in the initial binding of the cellulose chain. By improving the binding properties of F476Y, the cellulose chain most likely binds faster to Cel9A, and as a result, there is an increase in the activity of the mutated enzyme. Future experiments will include other changes in these residues to prove or disprove the previous hypothesis.

Acknowledgment

We are indebted to Dr. John Brady at the Food Science Department at Cornell University for his invaluable collaboration and initial financial support.

References

1. Smith, J. E. (1996), in *Biotechnology*, Cambridge University Press, Cambridge, MA, pp. 22–23.
2. Walker, L. P., et al. (1993) *Biotech. Bioeng.* **42,** 1019–1028.
3. Teeri, T. T. (1997), *Trends Biotech.* **15,** 160.
4. Walker, L. P., et al. (1992), *Biotech. Bioeng.* **40,** 1019–1026.
5. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993), *Biotech. Bioeng.* **42,** 1002–1013.
6. Davies, G. J., et al. (1996), *Acta Crystallographica* **D52,** 7–17.
7. Henrissat, B., et al. (1989), *Gene (Amst.)* **81,** 83–95.
8. Juy, M., et al. (1992), *Nature* **357,** 89–91.
9. Davies, G. J., et al. (1993), *Nature* **365,** 362–364.
10. Takashima, S., Nakamura, A., Masaki, H., and Uozumi, T. (1996), *Biosci. Biotech. Biochem.* **60(1),** 77–82.
11. Kraulis, P. J., et al. (1989), *Biochemistry* **28,** 7241–7257.
12. Dalbøge, H. and Heldt-Hansen, H. P. (1994), *Mol. Gen. Genet.* **243,** 253–260.
13. Azevedo, Mde O., et al. (1990), *J. Gen. Microbiol.* **136,** 120–123.
14. Cui, Z., et al. (1992), *Biosci. Biotechnol. Biochem.* **56,** 1230–1235.
15. Huang, J., et al. (1992), *J. Bacteriol.* **174,** 1314–1323.
16. Ong, E., et al. (1989), *Trends Biotechnol.* **7,** 239–243.
17. Divne, C., et al. (1994), *Science* **265,** 524–528.
18. Rouvinen, J., et al. (1990), *Science* **249,** 380–386.
19. Spezio, M., et al. (1993), *Biochemistry* **32,** 9906–9916.
20. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P.,A. (1997), *Nat. Struct. Biol.* **4,** 810–818
21. Irwin, D. C., et al. (1998), *J. Bacteriol.* **180(7),** 1709–1714.
22. Sacco, M., Millet, J., and Aubert, J. P. (1984), *Ann. Microbiol. (Inst. Pasteur)* **135A,** 485–488.
23. Curry, C., et al. (1988), *Appl. Environ. Microbiol.* **54,** 476–484.
24. Konstantinidis, A. K., et al. (1993), *Biochem. J.* **291,** 883.
25. Stemmer, W. P. C. (1994), *Nature* **370,** 389–391.
26. Stemmer, W. P. C. (1996), *Nat. Biotechnol.* **14,** 315–319.
27. Stemmer, W. P. C. (1995), *Biotechnology* **13,** 549–553.
28. Matsumura, I. and Ellington, A. D. (1996), *Nat. Biotechnol.* **14,** 366.
29. Zhang, J. H., Dawes, G., and Stemmer, W. P. C (1997), *Proc. Natl. Acad. Sci. USA* **94,** 4504–4509.
30. Cleland, J. L. and Craik, C. (1996), in *Protein Engineering Principles and Practice*, Wyley-Liss, New York, NY, pp. 22.
31. Taylor, J. S., Teo, B., Wilson, D. B., and Brady, J. W. (1988), *Prot. Eng.* **8(11)**, 1145–1152.
32. Brooks, C. L., Karplus, M. and Pettit, B. M. (1988), in *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Adv. Chem. Phys.*, Vol. 71, Wiley-Interscience, New York, NY.
33. MacKerell, A. D., Bashford, D., et al (1998), *J. Physi. Chem.* **102,** 3586–3616.
34. Palma, R., Zuccato, P., et al. (2000), in *Glycosyl Hydrolases in Biomass Conversion*. Himmel, M. E., ed., American Chemical Society, Washington, DC.
35. Van Gunsteren, W. F., and Berendsen, H. J. C. (1977) *Mol. Physiol.* **34(5),** 1311–1327.
36. Bayer, A. E., et al. (1998) in *Carhohydrases from* Trichoderma reesei *and Other Microorganisms*. Claeyssens, M., Nerinckx, W., and Piens, K., eds., The Royal Society of Chemistry, pp. 39–65.