JOURNAL OF BACTERIOLOGY, Dec. 1994, p. 7328–7334
0021-9193/94/$04.00+0

Vol. 176, No. 23

Mutation Analysis of the Cellulose-Binding Domain of the *Clostridium cellulovorans* Cellulose-Binding Protein A

MARC A. GOLDSTEIN† AND ROY H. DOI*

Section of Molecular and Cellular Biology, University of California, Davis, Davis, California 95616

Received 5 July 1994/Accepted 4 October 1994

Cellulose-binding protein A (CbpA) has been previously shown to mediate the interaction between crystalline cellulose substrates and the cellulase enzyme complex of *Clostridium cellulovorans*. CbpA contains a family III cellulose-binding domain (CBD) which, when expressed independently, binds specifically to crystalline cellulose. A series of N- and C-terminal deletions and a series of small internal deletions of the CBD were created to determine whether the entire region previously described as a CBD is required for the cellulose-binding function. The N- and C-terminal deletions reduced binding affinity by 10- to 100-fold. Small internal deletions of the CBD resulted in substantial reduction of CBD function. Some, but not all, point mutations throughout the sequence had significant disruptive effects on the binding ability of the CBD. Thus, mutations in any region of the CBD had effects on the binding of the fragment to cellulose. The results indicate that the entire 163-amino-acid region of the CBD is required for maximal binding to crystalline cellulose.

Cellulose-binding domains (CBDs) from different CBD families have different characteristic lengths (5). While family I CBDs from many fungal enzymes are less than 40 amino acids long, many bacterial CBDs are considerably longer (greater than 100 amino acids), especially the family III CBD group, of which the CBD from cellulose-binding protein A (CbpA) is a member (5). The CBD region of CbpA, as determined by homology studies and binding assays, extends for more than 160 amino acids (8) and is larger than the other CBDs of this family. Because this CBD is about 30 amino acids longer than most of the others, we felt it important to determine whether all of this region or only a certain portion is required for the cellulose-binding function.

Once the full extent of the CBD was determined, we wished to determine the structure-function relationships in the CBD and to find out what types of amino acids are required and what types of interactions with cellulose are responsible for the binding of the CBD and, hence, the cellulase complex to crystalline cellulose. Although some mutagenesis work has been done on cellulases and endoglucanases (1, 2, 4, 14, 16, 17, 23), only one reference to a mutant CBD is available (17). As there are no references relating to family III CBD mutants, we began a mutagenesis study, deleting and altering amino acid codons, mostly those within highly conserved regions of the CBD. Two types of mutations were created: small internal deletions of three to eight amino acids and, later, missense mutations altering single residues. It was found that the CBD is highly sensitive to mutations involving deletions but that missense mutations generally have less of an effect.

MATERIALS AND METHODS

Bacterial strains and plasmids. *Escherichia coli* XL1-Blue was obtained from Stratagene, La Jolla, Calif., and was used for all cloning experiments. *E. coli* BL21 (DE3) and pET-3d were as described previously (22). Plasmid pET-f1-3d was derived by cloning a PCR-amplified copy of the f1 origin of plasmid pBluescript II KS+ between the *Sal*I and *Eag*I sites of plasmid pET-3d.

Materials. PC buffer (pH 7) contained 50 mM KH_2PO_4, 10 mM sodium citrate, and 1 mM NaN_3. TEDG buffer (3) contained 10 mM Tris (p H7), 0.1 mM EDTA, 0.1 mM dithiothreitol, and 5% (vol/vol) glycerol.

Creation of CBD fragments with N- or C-terminal deletions. The appropriate gene fragments were amplified by PCR with primers designed to place appropriate start and stop signals at the ends of the gene, as well as add restriction sites for cloning into a pET-f1-3d derivative of expression plasmid pET-3d. Table 1 contains a list of the PCR primer sequences used. PCR was performed with *Pfu* DNA polymerase (Stratagene) to take advantage of its 3′-to-5′ proofreading exonuclease activity, which results in a lower error rate than *Taq* polymerase. PCR was carried out in a total volume of 100 μl by following the manufacturer's instructions for buffer conditions, primer concentration, and source DNA concentration. Plasmid pCB1, containing the entire *cbpA* gene (21), was the DNA source. The PCR products were purified by GeneClean (Bio 101, La Jolla, Calif.), cleaved with the *Nco*I and *Bam*HI restriction enzymes, and run on a 3% NuSieve GTG agarose gel in TBE buffer (19). Major bands which corresponded to the expected sizes of the CBD gene fragments were cut from the gel, melted, and ligated to the pET-f1-3d vector DNA prepared by cutting with the same restriction enzymes by using standard techniques (18). Ligated DNA was used to transform *E. coli* XL1-Blue competent cells, which were then plated onto selective media and incubated overnight at 37°C. Several colonies from each plate were screened for the presence of the CBD insert. Positive clones were subjected to DNA sequencing by previously described techniques (21) to verify that the proper start and stop sequences were present and to eliminate the possibility of a PCR-induced mutation.

Creation of internal deletions and missense mutations. The portion of pET-f1-CBD (8) between the unique *Sal*I and *Bam*HI sites was subcloned into phage M13mp18. The subclone was sequenced to verify proper insertion of the CBD fragment. A large-scale single-strand preparation (~2 mg of DNA) was prepared as described previously (18). Primers corresponding to the DNA sequence of each mutation, flanked

* Corresponding author. Mailing address: Section of Molecular and Cellular Biology, University of California, Davis, Davis, CA 95616. Phone: (916) 752-3191. Fax: (916) 752-3085.

† Present address: Section of Plant Biology, University of California, Davis, Davis, California 95616.

TABLE 1. Primer sequences used to create site-directed CBD mutations

Primer	Sequence	Terminal deletion or mutant no.
MG3125	AGCTCCACGGATCCTATCCTGTAAC	ΔC12
MG3223	TCCTTGGATCCTTAAAATTGTCC	ΔC50
MG3324	AACTCTAACACCATGGCACAAACA	Δ17
MG3424	AACACATCTGCCATGGATTTAAAT	Δ35
MG3524	ATAACAAAATCAACTCAAACAAAT	35
MG3624	TACACTCAAACAGATGCAAGTAGT	36
MG3730	AACTCAATTACACCAAACACATCTGACAGT	37
MG3830	AATGACGTAAAAGTTAGTGATGGTACACAA	38
MG3930	AAAGTGACAGCAAACGAAACAGCAAGCCCA	39
MG4030	CCAACATCAACCTATTTTGGATTTGCAAGC	40
MG4124	CAGCAAGCCCAACATCAACCTATG	41[a]
MG4230	ACACAAGGACAAACTGCTGGTGCATTATTA	42
MG4330	ACTCTTAAAAAAGGACAAGGAAGAATAACA	43
MG4430	GTTGTAAATCCAAAAGGTGGAGCTAAAGTA	44
MG4536	TGTGACCATGCTGGTGCAAACACTAGCAAAGTGACA	45
MG4630	ACTCAAACAAATAAATATTCATTTGATGCA	46
MG4730	ACTCAAACAAATTCATATTCATTTGATGCA	47
MG4829	TTCTGGTGTGACAAAGCTGGTGCATTATT	48
MG4929	TTCTGGTGTGACGAAGCTGGTGCATTATT	49
MG5030	GATACATATGTTGATTTTGGATTTGCAAGC	50
MG51"30"	GATACATATGTTAAATTTGGATTTGCAAG	51
MG5230	TTTGCAAGCGGAAAAGCTACTCTTAAAAAA	52
MG5330	TTTGCAAGCGGAGAAGCTACTCTTAAAAAA	53
MG5428	TCAATGTCAGTTGACTTTTACAACTCTA	54
MG5528	TCAATGTCAGTTAAATTTTACAACTCTA	55
MG5630	GCTACTCTTAAAAGAGGACAATTTATAACT	56
MG5730	GCTACTCTTAAAGAAGGACAATTTATAACT	57
MG5827	CAAACAAATGACGGTTCATTTGATGCA	58
MG5927	ACAAATGACTATGGATTTGATGCAAGT	59
MG6027	AATGACTATTCAGGAGATGCAAGTAGT	60

[a] Sequencing primer 41 was designed to aid in sequencing some of the mutations.

by approximately 15 bases on each side, were synthesized (Table 1). An Amersham (Arlington Heights, Ill.) in vitro oligonucleotide-directed mutagenesis kit (catalog no. RPN1523) was used in accordance with the manufacturer's instructions to create the mutants. Each M13 mutant was sequenced to verify the mutation, and then the mutant *Sal*I-to-*Bam*HI fragments were cloned into vector pET-f1-3d.

Expression and purification of mutant CBD proteins. The mutant CBD gene fragments were expressed in the same fashion as wild-type CBD, as described previously (8). As with the wild type, the heterologously expressed CBD mutant proteins formed inclusion bodies within the *E. coli* BL21 (DE3) host cells and had to be purified through a complete denaturation-renaturation process. The insoluble fractions of the lysed cells from 200-ml NZYCM (18) cultures were solubilized by addition of 10 ml of 6 M guanidine hydrochloride, followed by incubation on ice for at least 1 h with periodic vigorous vortexing. The soluble phase was separated from the insoluble phase by centrifugation at 12,000 rpm in a Sorvall SS34 rotor for 30 min at 4°C. The soluble phase of each mutant preparation, containing the CBD protein, was then diluted 1:2 with TEDG renaturation buffer. This diluted fraction (30 ml) was then dialyzed against several changes of TEDG. Precipitated material was separated by centrifugation as described above, and the cleared supernatants were assayed for protein by UV spectroscopy. To aid in the solubilization of several mutant proteins which were insoluble at this step, inclusion of various detergents (purchased from Pierce, Rockford, Ill.) was tried. Although several detergents resulted in partial solubilization of the "insoluble" mutant proteins, it was determined that these detergents interfered with the binding of wild-type CBD to cellulose when present at concentrations appropriate for solubilization of the proteins.

Purification of the solubilized mutant CBDs was performed by one of two methods. Mutant CBDs that retained significant cellulose affinity were purified by an Avicel affinity method described previously (8), in which the protein was allowed to bind to Avicel and then several wash steps were performed to remove contaminating *E. coli* host proteins. Mutant CBDs were eluted with 6 M urea, which was then removed by dialysis against PC buffer. Mutant CBDs which did not bind significantly to Avicel were purified by fast protein liquid chromatography (FPLC) by using a Pharmacia FPLC apparatus equipped with a Resource Q anion-exchange column with a 1-ml bed volume. Samples containing up to 10 mg of total protein were applied to the column in 20 mM Tris buffer, pH 8.5. Proteins were eluted with a linear NaCl gradient ranging from 0 to 1 M across volumes of 20 to 40 ml. Samples, typically of 2 ml each, were collected and assayed for CBD protein by gel electrophoresis as described below.

Gel electrophoresis of mutant CBD proteins. Expression and purification of mutant CBD proteins were monitored by use of the Pharmacia PhastGel system. Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis was carried out with Homogeneous 20 gels with SDS buffer strips by using the manufacturer's recommended protocols. Low-molecular-weight standards were from Bethesda Research Laboratories, Inc., Gaithersburg, Md. Isoelectric focussing gels were run by using IEF 3-9 gels with the Pharmacia broad-range pI calibration kit (resuspended in 200 μl of water per vial) as standards. Silver staining of gels was accomplished in accordance with the manufacturer's protocols, except that the development times

were extended to approximately twice the recommended intervals, which resulted in better staining.

Analysis of the binding of mutant CBDs to cellulose. Two methods were used to assay the ability of CBDs to bind to cellulose. Those that were able to be purified by Avicel affinity, and thus exhibited measurable affinity for cellulose, were subjected to a quantitative binding assay. Precisely known amounts of protein, typically from 50 to 200 μg, were added to duplicate microcentrifuge tubes, one of which contained 10 mg of Avicel PH101 microcrystalline powdered cellulose in PC buffer. The total volume was 1 ml (final cellulose concentration, 10 mg/ml). The other tube contained buffer only. After 3 h of incubation at 37°C with slow rotational mixing, the cellulose was spun into a pellet and the free protein concentrations in both tubes were determined with a MicroBCA kit (Pierce). Bound protein was determined by subtracting the free protein concentration in the Avicel-containing tube from the (higher) protein concentration in the non-cellulose-containing tube. The cellulose added a small amount of reducing sugar to the soluble fractions; data from the MicroBCA assays were corrected for this background. These bound and free protein concentrations, in combination with the previously determined binding capacity of Avicel for wild-type CBD protein, were used to estimate the dissociation constant of the binding reaction between Avicel and each of the mutants.

Qualitative binding assays were attempted for proteins which did not bind tightly enough to Avicel to allow affinity purification. In these assays, a large amount of dry Avicel powder (20 mg) was added to tubes containing FPLC-purified protein samples (200 μl at 0.5 μg/μl), with a final cellulose concentration of 100 mg/ml. The tubes were incubated with rotational mixing for 3 h, and protein samples were removed from the cleared supernatants. These were applied to isoelectric focussing gels next to samples of the proteins which had not been exposed to the cellulose. Visual comparison of the lanes was used to score the samples as cellulose binding or lacking in cellulose binding.

Computer alignment. The PILEUP program, a component of the Wisconsin Genetics Computer Group package for the VAX, was used to create the alignment between the CBD from CbpA and other family III CBDs shown in Fig. 1. Included in this alignment were CBD sequences from CelZ (*Clostridium stercorarium* [12]), CipA (*Clostridium thermocellum* [6]), CelB (*Caldocellum saccharolyticum* [20]), and CelA (*Bacillus lautus* [9, 10, 20]) and the consensus sequence from several nearly identical endoglucanases from *Bacillus subtilis* (13, 15).

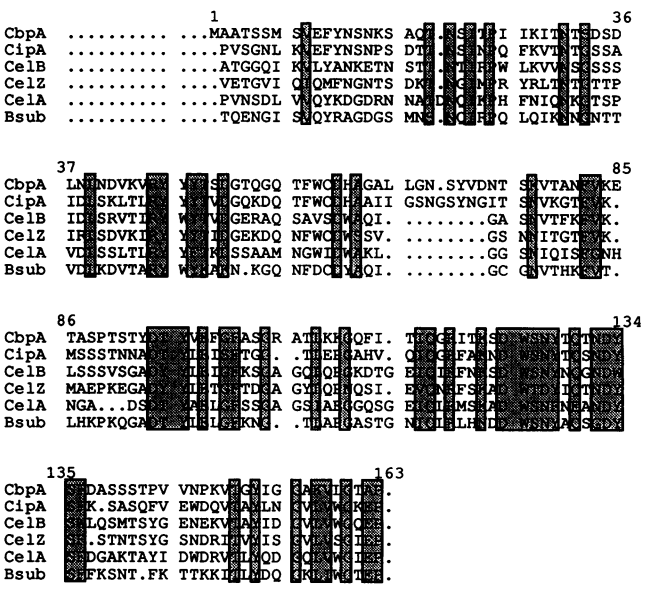

FIG. 1. Amino acid sequence alignment of several CBDs generated by the PILEUP program of the Wisconsin Genetics Computer Group package. Numbers 2 to 163 represent positions within the *C. cellulovorans* CBD sequence, corresponding to amino acids 28 to 190 from the CbpA sequence reported previously (21). Position 1 represents the methionine amino acid introduced to provide a translational start codon for the heterologously expressed CBD. Amino acids highly conserved are indicated by shaded boxes.

RESULTS AND DISCUSSION

Terminal deletions. To define the limits of the CBD region of CbpA, a series of truncated CBD-encoding clones were constructed. All of the CBD mutant proteins were expressed in *E. coli* at high levels. However, several of them did not renature into a buffer-soluble form. Of the six terminal deletions made, only three resulted in sufficient solubility to allow assays of binding ability. These were deletions ΔN17, ΔN71, and ΔC12, in which the first 17 and 71 N-terminal and the last 12 C-terminal amino acids, respectively, of the CBD region were deleted. Figure 2C is a diagram of the portions of CBD expressed in these clones. All three soluble proteins bound to cellulose with sufficient affinity to allow batch cellulose chromatographic purification of the mutant protein fragments, as shown in Fig. 3. Quantitative binding assays (Table 2) were performed with each of the three mutants, yielding estimations of the dissociation constants (K_ds) of the mutants. Each mutant was assayed in triplicate, and the results were averaged, producing K_ds for each mutant of 20 to 50 μM. To verify the validity of this technique, wild-type CBD protein, which binds Avicel tightly, and bovine serum albumin (BSA), which does not, were assayed as controls. When assayed in this fashion, the wild-type CBD protein yielded a K_d of about 1 μM, matching the results obtained during the previous characterization of the CBD-cellulose interaction (8). BSA, which presumably has only nonspecific interactions with Avicel, exhibited binding at

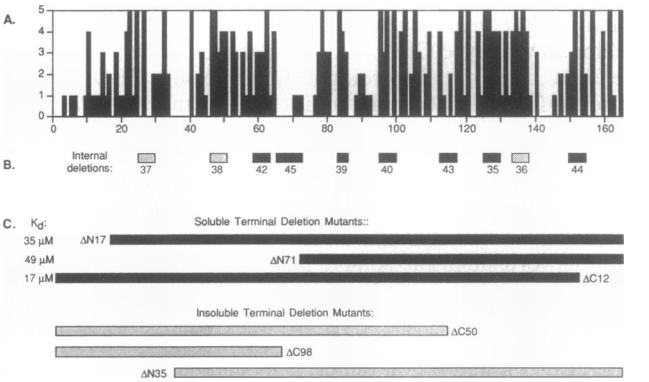

FIG. 2. Homology between CBDs of family III. (A) Histogram showing the numbers of residues matching between the CBD from CbpA and five aligned homologous CBDs, as shown in Fig. 1. (B) Positions and extents of internal deletions. Deletions resulting in insoluble proteins are shown as stippled bars, and those resulting in soluble proteins are black. (C) Positions of terminal deletions (bars indicate protein fragments produced). Soluble and insoluble proteins are patterned as in panel B. Approximate K_d values (averages of the values in Table 2) are given to the right of the soluble deletions.

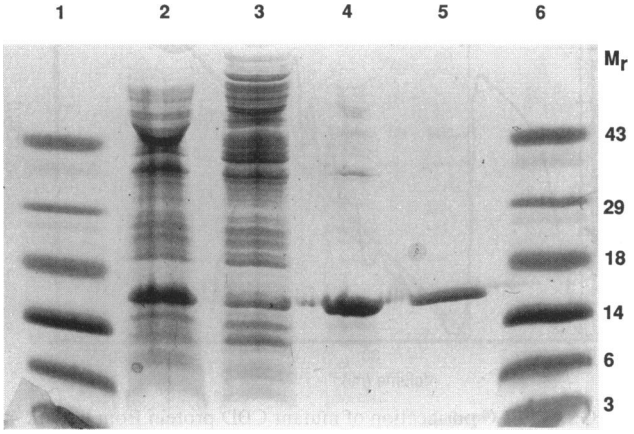

FIG. 3. Purification of mutant CBD from terminal deletion mutant ΔC12. Lanes: 1 and 6, Bethesda Research Laboratories low-molecular-weight markers (molecular masses [M_r] of the markers, in kilodaltons, are shown to the right of lane 6); 2, whole-cell fraction; 3, cell lysis supernatant fraction; 4, solubilized cell lysis pellet fraction; 5, Avicel affinity-purified CBD ΔC12 protein.

the threshold of detection of the assay, yielding a theoretical K_d on the order of 10^{-3} M, about 10^3 times greater than that of wild-type CBD and roughly 2 orders of magnitude greater than that of the CBD end deletion mutants. It should be noted that these numbers are only estimations based on the assumptions that (i) the binding capacity of cellulose is the same for the mutants as for wild-type CBD and (ii) the extent of the nonspecific binding of the mutants is relatively minor compared to the extent of the specific protein-cellulose interaction.

In support of the first assumption, no relationship between CBD size and binding capacity has ever been reported. Additionally, when a larger segment of the CbpA protein, containing the CBD and the first hydrophilic repeated domain (21), was subjected to a binding study, the result was that the binding capacity matched that of the CBD alone (data not shown). This suggests that cellulose samples have a set number of sites to which CBDs can bind, rather than the binding capacity being a function of the size of the CBD.

TABLE 2. Binding of CBD with terminal deletions to Avicel[a]

Protein or control (amt [μg] used)	[P] (μM)	[PC] (μM)	[C] (10^5 M)	K_d (μM)
Wild type (110)	0.77	5.47	8.53	1.2
Wild type (220)	2.20	10.3	3.73	0.8
ΔC12 (43)	1.53	1.07	1.29	18
ΔC12 (86)	3.47	1.74	1.23	25
ΔC12 (220)	11.2	2.13	1.19	63
ΔN17 (28)	1.07	0.75	1.32	19
ΔN17 (56)	2.67	0.92	1.30	36
ΔN17 (145)	8.27	1.15	1.29	93
ΔN71 (14)	0.65	0.75	1.32	11
ΔN71 (28)	1.68	1.13	1.29	19
ΔN71 (65)	4.20	2.32	1.17	21
BSA (100)	1.32	0.016	1.40	1,150
BSA (200)	2.64	0.026	1.40	1,420

[a] See Fig. 2 for descriptions of deletion mutants ΔC12, ΔN17, and ΔN71. [P], [PC], and [C] stand for free CBD protein, CBD-cellulose complexes, and unused binding sites on the cellulose, respectively. K_d is the dissociation constant of the protein-cellulose complex.

TABLE 3. Internal deletion mutations of CBD[a]

Mutant	Lesion(s)[b]	Sequence deleted	Predicted pI	Major-band pI	Binding[c]
None (wild type)	None		5.9	8.5	+++
35	Δ124–128	DWSNY	7.3	7	+
39	Δ82–84	FVK	5.1	4	−
40	Δ94–98	DTYVE	8.4	4–6	+
42	Δ58–62	FWCDH	6.8	7	−
43	Δ112–116	QFITI	5.9	4	−
44-1	Δ149–153, K148T	VTGYI	5.1	4.5	−
44-2	Δ149–153	VTGYI	5.9	6.5	−
45	Δ66–73	LLGNSYVD	7.3	7.5	+

[a] Assays used 100 μg of mutant CBD protein and 20 mg of Avicel powder.
[b] The mutations described here are diagrammed in Fig. 2B. The amino acid positions listed correspond to the reference numbers used in Fig. 1.
[c] Binding results are qualitative and based upon visual inspection of silver-stained gels as in Fig. 5. Scoring was as follows: −, no detectable binding; +, some binding; +++, strong (wild-type) binding. Although the assay does not give quantitative data, mutant CBDs designated ++ would be expected to have a K_d of <100 μM, and those designated + would have a K_d of <1,000 μM.

The second assumption, referring to the extent of nonspecific binding, is also likely a safe one for the CBDs measured. The molarity of the bound BSA was about 2 orders of magnitude lower than that of the CBD fragments measured under similar conditions. BSA was included as a measure of nonspecific binding; therefore, of the amount of bound CBD protein measured, only a small percentage was likely due to nonspecific interactions. This also indicates that nonspecific interactions become increasingly numerically significant in relation to specific binding as the K_d increases. Therefore, this assay was attempted only for proteins demonstrating substantial binding to cellulose during attempts at affinity purification.

The fact that all of these three proteins, in which the first 17 and 71 N-terminal and the last 12 C-terminal amino acids were deleted, were significantly impaired in affinity for cellulose indicated that those deleted regions were important for the binding function. However, the fact that they retained enough binding ability to allow affinity purification suggests that the central core of the CBD (from amino acids 72 to 151) contained a significant portion of the functional CBD as well. On the basis of these data, several models are possible. First, it could be that the core of the CBD is responsible for binding and that regions flanking the core are required for attaining the proper conformation. Alternatively, it has been suggested that CBDs contain several binding subsites which allow synergistic binding (7, 11).

Internal deletions. To investigate the contributions of individual regions of the CBD to its overall function, a set of 11 smaller internal deletions were created, each mutant lacking a three- to eight-amino-acid section of the CBD. Table 3 lists these deletions, which were generally in highly conserved regions of the CBD sequence (Fig. 2A). Although all of the mutants were well expressed in *E. coli*, as with the terminal deletions, several of them (no. 36 to 38) did not renature into a buffer-soluble form. Of the seven soluble mutants, only one, no. 35, showed any appreciable binding to Avicel during attempts to purify them by cellulose affinity, and even this was very weak. Interestingly, this mutant had a deletion in the most highly conserved region of the CBD. Because purification could not be accomplished by the previously described tech-

TABLE 4. Missense mutations of CBD[a]

Mutant	Lesion[b]	Δ Charge	Predicted pI	Major-band pI	Binding[c]
46	D133K	−→+	8.4	8	++
47	D133S	−→0	7.3	6–7	++
50	E98D	−→−	5.8	8.5	++
53	A105E	0→−	5.1	4	−
55	E10K	−→+	8.4	9	++
56	K110R	+→+	5.9	9	+
57	K110E	+→−	4.8	5	++

[a] Assays used 100 μg of mutant CBD protein and 20 mg of Avicel powder.
[b] The amino acid positions listed correspond to the reference numbers used in Fig. 1.
[c] Binding results are qualitative and based upon visual inspection of silver-stained gels as in Fig. 5. Scoring was as follows: −, no detectable binding; +, some binding; ++, significant binding. Although the assay does not give quantitative data, mutant CBDs designated as ++ would be expected to have a K_d of <100 μM and those designated as + would have a K_d of <1,000 μM.

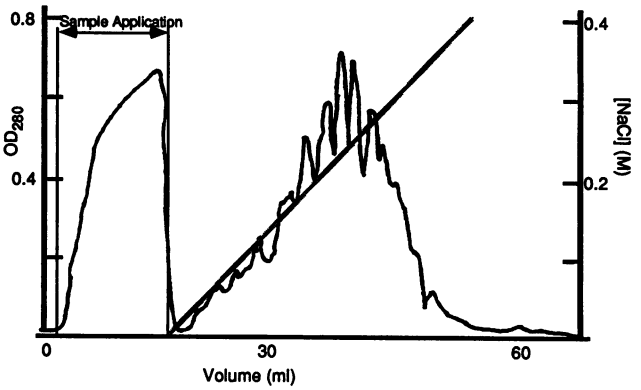

FIG. 4. FPLC purification of mutant CBD protein from mutant 40 (Table 3). Ten milliliters (approximately 15 mg) of the renatured protein preparation was applied to the column. Proteins bound to the column were eluted with a linear salt gradient (diagonal line). Elution of protein was monitored by UV absorbance (black line). Mutant CBD protein made up the majority (>90%) of the fractions collected between 30 and 50 ml. OD_{280}, optical density at 280 nm.

niques, an FPLC purification protocol was established. Computer calculations of the predicted isoelectric points of the mutants and wild-type CBDs were made (Tables 3 to 5). Predictions indicated that these proteins were, for the most part, neutral or somewhat acidic, so anion-exchange chromatography was performed. Tris buffer, pH 8.5, was selected as the loading buffer. Although this pH is out of the optimal buffering range of Tris, the purification protocol did not place heavy demands on the buffering capacity of the solution, so this choice was appropriate. In theory, all of the deletion mutants, as well as the wild-type protein, should have been negatively charged at this pH, so they should have all bound to the Pharmacia Resource Q column. However, when previously purified wild-type CBD was first used as a control, little binding to the column was seen. Most of the mutant CBDs, on the other hand, bound tightly to the column and were eluted by a salt gradient. An interesting phenomenon was noted: instead of eluting as a single, sharp peak, the CBD mutant proteins came off the column as several peaks spread over a large range of salt concentrations (Fig. 4). Attempts were made to "correct" this by altering the buffer and elution conditions and by substituting a Mono Q column for the Resource Q column. There was no improvement in the situation under any of the conditions tried. To determine if this was an artifact of the purification procedure or if the CBDs actually existed in several different forms with various isoelectric points, broad-range pI gels of crude mutant preparations, purified mutant proteins, and the wild-type CBD were run. All of the proteins appeared as multiple bands on the pI gels. Freshly Avicel affinity-purified wild-type CBD had the fewest minor bands, in addition to a major band in the pH ≈8.5 area of the gel. It was found that extended incubation (months) at 4°C increased the amount of protein banding at other pHs without significantly increasing the amount of degradation products seen on SDS-polyacrylamide gel electrophoresis gels. Extended incubation of the soluble terminal deletions resulted in protein that no longer bound cellulose (Fig. 5C). Also of interest was the fact that the mutants generally produced major bands or smears at pHs at or near their predicted isoelectric points, whereas the wild-type CBD major band was about 2.5 pH units above that predicted.

These results could be explained if the "actively binding" form of CBD was folded into a structure shielding some positively charged amino acids inside the protein, causing the protein to be effectively more negative (and hence have a higher gel pI) than the computer-predicted value. This hypothesis was tested by binding assays in which Avicel cellulose powder was added to protein samples, incubated for several hours to allow binding to occur, and then spun down. The supernatant containing nonbound protein was immediately run on pI gels next to protein samples that had not been exposed to Avicel. This would allow visual comparison of the binding to see which bands, if any, bound the cellulose. Wild-type CBD, included as a control, showed that the effective pI was not vital to binding, as bands with pIs ranging from 3.5 to 8.5 bound to the cellulose (Fig. 5). As determined by this method, only the CBD from mutant 35 of the internal deletion mutants bound the cellulose significantly although the high concentration of cellulose used (100 mg of Avicel per ml) would be expected to remove any cellulose-binding proteins from solution. Some marginal binding was observed for CBDs from mutants 40 and 45. These results from studies on solubility, activity, and apparent pI indicated that the CBD structure was very sensitive to disruption by small internal deletions, meaning either that formation of the actual binding site within the CBD was dependent on contributions from all of the regions altered or that the altered regions somehow interfered with the formation of the binding site.

TABLE 5. Insoluble mutant CBDs

Mutant	Lesion site[a]	Effect	Predicted pI
36	Δ132–136	NDYSF	7.3
37	Δ26–30	IIKIT	5.1
38	Δ45–49	RYYYT	5.1
51	E98K	−→+	8.4

[a] See Table 4, footnote *b*.

Missense mutations. The results of the internal deletions, which indicated generally low binding of the mutants, were surprising, since the larger terminal deletions resulted in proteins that retained some measure of affinity for cellulose. Somehow, the small deletions disrupted the binding face or surrounding protein structure to a greater extent than did the large deletions. To investigate this phenomenon further, a set of missense mutations in which only one amino acid was altered per mutant was created (Table 4). Because of the discrepancy between the wild type's predicted and gel-determined pIs, it was decided to begin with the charged amino

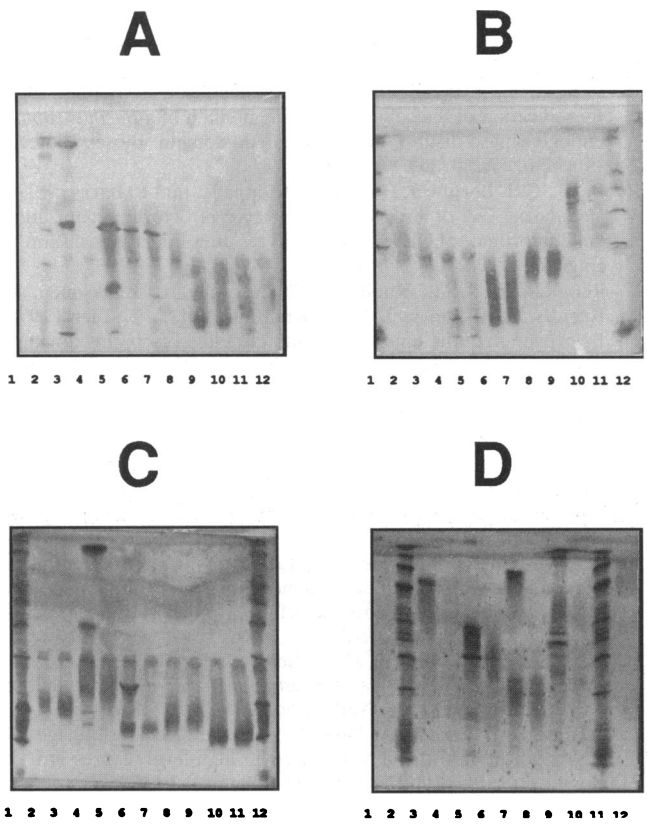

FIG. 5. Qualitative analysis of mutant CBD protein binding to Avicel. All gels were silver-stained isoelectric focussing gels (Pharmacia IEF 3-9). Protein samples (200 μl at 1 μg/μl) were incubated in the absence or presence of 20 mg of Avicel microcrystalline cellulose for 3 h at 37°C. Small aliquots (0.3 μl) of the supernatant fluid were then applied to the gels, which were run and stained in accordance with the manufacturer's protocols. The standards used were in the broad-range pI calibration kit from Pharmacia. The top of each gel is pH 9, and the bottom is pH 3, with a linear gradient in between. Lanes: (A) 1, control (double-distilled H_2O incubated with Avicel [no bands visible]); 2, pI standards; 3 and 4, wild-type CBD protein without and with Avicel, respectively; 5 and 6, mutant 35 (pre-FPLC) without and with Avicel, respectively; 7 and 8, mutant 35 (fraction 13) without and with Avicel, respectively; 9 and 10, mutant 39 without and with Avicel, respectively; 11 and 12, mutant 40 without and with Avicel, respectively; (B) 1 and 12, pI standards; 2 and 3, mutant 42 without and with Avicel, respectively; 4 and 5, mutant 43 without and with Avicel, respectively; 6 and 7, mutant 44-1 without and with Avicel, respectively; 8 and 9, mutant 44-2 without and with Avicel, respectively; 10 and 11, mutant 45 without and with Avicel, respectively; (C) 1 and 12, pI standards; 2 and 3, mutant 53 without and with Avicel, respectively; 4 and 5, mutant 56 without and with Avicel, respectively; 6 and 7, mutant 57 without and with Avicel, respectively; 8 and 9, CBD ΔN17 without and with Avicel, respectively (9-month-old sample); 10 and 11, CBD ΔC12 without and with Avicel, respectively (9-month-old sample); (D) 1 and 12, not used; 2 and 11, pI standards; 3 and 4, mutant 46 without and with Avicel, respectively; 5 and 6, mutant 47 without and with Avicel, respectively; 7 and 8, mutant 50 without and with Avicel, respectively; 9 and 10, mutant 55 without and with Avicel, respectively.

acids in the hope that a study of the relationship between the predicted and observed pIs of the mutants, correlated with the binding ability of these mutants, would reveal something of the structure-function relationships in the CBD. The results indicated that several of these mutations had little effect on cellulose binding, while a few mutant CBDs were dramatically impaired in binding. Alteration of the glutamic acid at position 98 to a lysine resulted in an insoluble mutant protein, the only one of the missense mutant proteins to be insoluble. However, alteration of this same residue to aspartic acid (E98D, mutant 50), resulted in a protein with the same charge as the wild type but substantially reduced binding ability. A similar effect was seen at position 110, where alteration of the lysine present in the wild type to an arginine, involving no change in charge, had a more drastic effect on the CBD function than alteration of that same lysine to glutamic acid, which resulted in a ΔpI of about 4 U. We plan to use these and additional data to create a map of critical amino acid sites within the CBD, which will be related to X-ray crystallographic data being prepared in the laboratory of Oded Shoseyov, at the Hebrew University in Jerusalem, Israel. This should allow determination of precise sites on the CBD likely to interact with the crystalline substrate.

Each of the types of mutations created has furthered our understanding of the function of the CBD. The N- and C-terminal deletions had drastically reduced but still measurable binding to cellulose. The C-terminal half of the CBD, which is the more highly conserved portion, was more critical for retention of any binding activity. This is supported by the fact that even the ΔN71 mutant protein retained some binding function, whereas both C-terminal deletions with more than 12 amino acids removed were insoluble. The ΔN71 mutation had the highest apparent K_d of any of the terminal deletions.

The small internal deletions, made predominantly in regions of high conservation, drastically reduced the binding ability of each of the mutant proteins. This is likely due to disruption of the protein's folding into proper secondary and tertiary structures, as evidenced by the fact that several of these mutations rendered the remaining CBD portion insoluble, even though only a small portion of the CBD was deleted. This set of deletions emphasized the essential nature of each region's contributions to the structure and function of the CBD.

The investigation of charge effects within the CBD indicated that individual charged amino acids often can be altered without drastically disrupting the function of the CBD and that substitution of like-charged amino acids had no less of an effect than a substitution in which the replacement had the opposite charge. These examples, in addition to our data indicating that wild-type CBD proteins banding at various isoelectric points all bind well, argue that the overall charge of the CBD is not a major factor in the binding of this protein to the cellulose.

Thus, our overall conclusion is that the entire CBD sequence (8) is necessary for efficient binding of CbpA to cellulose.

ACKNOWLEDGMENTS

This research was supported in part by Department of Energy grant DE-FG03-92ER20069.

REFERENCES

1. **Baird, S. D., M. A. Hefford, D. A. Johnson, W. L. Sung, M. Yaguchi, and V. L. Seligy.** 1990. The Glu residue in the conserved Asn-Glu-Pro sequence of two highly divergent endo-beta-1,4-glucanases is essential for enzymatic activity. Biochem. Biophys. Res. Commun. **169:**1035–1039.
2. **Belaich, A., H. P. Fierobe, D. Baty, B. Busetta, T. C. Bagnara, C. Gaudin, and J. P. Belaich.** 1992. The catalytic domain of endoglucanase A from *Clostridium cellulolyticum*: effects of arginine 79 and histidine 122 mutations on catalysis. J. Bacteriol. **174:**4677–4682.
3. **Chang, B.-Y., and R. H. Doi.** 1990. Overproduction, purification, and characterization of *Bacillus subtilis* RNA polymerase σA factor. J. Bacteriol. **172:**3257–3263.
4. **Chauvaux, S., P. Béguin, and J. P. Aubert.** 1992. Site-directed

mutagenesis of essential carboxylic residues in *Clostridium thermocellum* endoglucanase CelD. J. Biol. Chem. **267**:4472–4478.

5. **Coutinho, J. B., N. R. Gilkes, R. A. J. Warren, D. G. Kilburn, and R. C. Miller.** 1992. The binding of *Cellulomonas fimi* endoglucanase C (CenC) to cellulose and Sephadex is mediated by the N-terminal repeats. Mol. Microbiol. **6**:1243–1252.
6. **Gerngross, U. T., M. P. Romaniec, T. Kobayashi, N. S. Huskisson, and A. L. Demain.** 1993. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal SL-protein reveals an unusual degree of internal homology. Mol. Microbiol. **8**:325–334.
7. **Gilkes, N. R., E. Jervis, B. Henrissat, B. Tekant, R. C. J. Miller, R. A. J. Warren, and D. G. Kilburn.** 1992. The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. J. Biol. Chem. **267**:6743–6749.
8. **Goldstein, M. A., M. Takagi, S. Hashida, O. Shoseyov, R. H. Doi, and I. H. Segel.** 1993. Characterization of the cellulose-binding domain of the *Clostridium cellulovorans* cellulose-binding protein A. J. Bacteriol. **175**:5762–5768.
9. **Hansen, C. K.** 1992. Fibronectin type III-like sequences and a new domain type in prokaryotic depolymerases with insoluble substrates. FEBS Lett. **305**:91–96.
10. **Hansen, C. K., B. Diderichsen, and P. L. Jorgensen.** 1992. *celA* from *Bacillus lautus* PL236 encodes a novel cellulose-binding endo-beta-1,4-glucanase. J. Bacteriol. **174**:3522–3531.
11. **Henrissat, B., B. Vigny, A. Buleon, and S. Perez.** 1988. Possible adsorption sites of cellulases on crystalline cellulose. FEBS Lett. **231**:177–182.
12. **Jauris, S., K. P. Rucknagel, W. H. Schwarz, P. Kratzsch, K. Bronnenmeier, and W. L. Staudenbauer.** 1990. Sequence analysis of the *Clostridium stercorarium celZ* gene encoding a thermoactive cellulase (Avicelase I): identification of catalytic and cellulose-binding domains. Mol. Gen. Genet. **223**:258–267.
13. **MacKay, R. M., A. Lo, G. Willick, M. Zuker, S. Baird, M. Dove, F. Moranelli, and V. Seligy.** 1986. Structure of a *Bacillus subtilis* endo-beta-1,4-glucanase gene. Nucleic Acids Res. **14**:9159–9170.
14. **Mitsuishi, Y., S. Nitisinprasert, M. Saloheimo, I. Biese, T. Reinikainen, M. Claeyssens, S. Keranen, J. K. Knowles, and T. T. Teeri.** 1990. Site-directed mutagenesis of the putative catalytic residues of *Trichoderma reesei* cellobiohydrolase I and endoglucanase I. FEBS Lett. **275**:135–138.
15. **Park, J. S., A. Nakamura, S. Horinouchi, and T. Beppu.** 1993. Identification of the cellulose-binding domain of a *Bacillus subtilis* endoglucanase distinct from its catalytic domain. Biosci. Biotech. Biochem. **57**:260–264.
16. **Py, B., G. I. Bortoli, J. Haiech, M. Chippaux, and F. Barras.** 1991. Cellulase EGZ of *Erwinia chrysanthemi*: structural organization and importance of His98 and Glu133 residues for catalysis. Protein Eng. **4**:325–333.
17. **Reinikainen, T., L. Ruohonen, T. Nevanen, L. Laaksonen, P. Kraulis, T. A. Jones, J. K. C. Knowles, and T. T. Teeri.** 1992. Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase. Prot. Struct. Funct. Genet. **14**:475–482.
18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 1.5–1.52. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
19. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. B.23–B.24. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
20. **Saul, D. J., L. C. Williams, D. R. Love, L. W. Chamley, and P. L. Bergquist.** 1989. Nucleotide sequence of a gene from *Caldocellum saccharolyticum* encoding for exocellulase and endocellulase activity. Nucleic Acids Res. **17**:439.
21. **Shoseyov, O., M. Takagi, M. A. Goldstein, and R. H. Doi.** 1992. Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A. Proc. Natl. Acad. Sci. USA **89**:3483–3487.
22. **Studier, F. W., and B. A. Moffat.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**:113–130.
23. **Tomme, P., S. Chauvaux, P. Béguin, J. Millet, J. P. Aubert, and M. Claeyssens.** 1991. Identification of a histidyl residue in the active center of endoglucanase D from *Clostridium thermocellum*. J. Biol. Chem. **266**:10313–10318.