Cellulosome from *Clostridium cellulolyticum*: Molecular Study of the Dockerin/Cohesin Interaction†

Henri-Pierre Fierobe,*,‡ Sandrine Pagès,‡ Anne Bélaïch,‡ Stéphanie Champ,‡ Doris Lexa,‡ and Jean-Pierre Bélaïch‡,§

Laboratoire de Bioénergétique et Ingénierie des Protéines, Centre National de la Recherche Scientifique, 31, Chemin Joseph-Aiguier, F-13402 Marseille, France, and Université de Provence, 3, place Victor Hugo, F-13331 Marseille, France

Received May 21, 1999; Revised Manuscript Received July 7, 1999

ABSTRACT: *Clostridium cellulolyticum* produces cellulolytic complexes (cellulosomes) made of 10−13 cell wall degrading enzymes tightly bound to a scaffolding protein (CipC) by means of their dockerin domain. It has previously been shown that the receptor domains in CipC are the cohesin domains and that the cohesin/dockerin interaction is calcium-dependent. In the present study, surface plasmon resonance was used to demonstrate that the free cohesin1 from CipC and dockerin from CelA have the same K_D (2.5×10^{-10} M) as that of the entire CelA and a larger fragment of CipC, the latter of which contains, in addition to cohesin1, a cellulose binding domain and a hydrophilic domain of unknown function. This demonstrates that neither the catalytic domain of CelA nor the noncohesin domains of CipC have any influence on the interaction. Dockerin domains are composed of two conserved segments of 22 residues: removal of the second segment abolishes the affinity for cohesin1, whereas modified dockerins having twice the first segment, twice the second, or both segments but in a reverse order have K_D values for cohesin1 in the same range as that observed for wild-type dockerin. These data indicate that if two segments are required for the complexation with the cohesin, segments 1 and 2 are similar enough to replace each other. Calcium overlay experiments revealed that the dockerin domain has one calcium binding site per conserved segment. Circular dichroism performed on wild-type and mutant dockerins indicates that this domain is well structured and that removal of calcium only weakly affects the secondary structure, which remains 40−45% helical.

Clostridium cellulolyticum has been studied over the past 10 years due to its ability to produce large enzymic complexes, called cellulosomes, able to degrade crystalline cellulose (*1*). These complexes are also synthesized by other cellulolytic *clostridia* such as *Clostridium thermocellum* and *Clostridium cellulovorans* (*2*−*4*). In all cases, it has been observed that these complexes are composed of 10−20 cell wall degrading enzymes (mainly cellulases) tightly bound to a large scaffolding protein (*5*−*7*), devoid of enzymic activity, which binds crystalline cellulose (*8*−*10*). So far, 10 genes encoding cellulosomal polypeptides from *C. cellulolyticum* have been cloned and sequenced (*11*, *12*; A. Bélaïch, unpublished results). Most of these encode cellulases belonging to various families of the classification of glycosyl hydrolases (*11*). Sequence analysis indicates that these enzymes all contain a noncatalytic domain located at their C-terminal extremity named the dockerin domain (*11*), which is found in many other clostridial enzymes involved in cellulosome complexes (*13*). This peculiar domain is made of two homologous but not identical segments of 22 residues spaced by a linker of 8−10 residues (*13*).

So far, five cellulases, CelA, C, D, F, and G, from *C. cellulolyticum* have been purified from recombinant *Escherichia coli* strains overexpressing their genes (*14*−*18*). The enzymic characterization of these cellulases has revealed that they differ in both their substrate specificities and their ability to degrade crystalline cellulose (*11*).

Very recently, the gene *cipC*, which encodes the scaffolding protein CipC of the cellulosome, was completely sequenced (*12*). The corresponding 1519 residue mature protein is composed of a N-terminal type III cellulose binding domain (CBD),[1] followed by a first hydrophilic domain (called HD1), seven hydrophobic domains of 150 residues called cohesin domains, a second hydrophilic domain (HD2), and finally an eighth cohesin domain (*12*). Previous studies have shown, by using truncated forms of the scaffoldin, that the cohesin domain is required for the docking of the cellulases onto the scaffolding protein and that only cellulases containing a dockerin domain are able to anchor the scaffolding protein CipC. It was also observed that this interaction is inhibited by EDTA (*13*). In addition, sequence comparisons with EF-hand calcium binding proteins strongly suggest that the dockerin domain contain two calcium binding sites, one per conserved segment (*13*).

Similar observations have been made for the cellulosomes from *C. thermocellum* and *C. cellulovorans* (*13*, *19*, *20*), in

†This work was supported by grants from the Région Provence Alpes Côte d'Azur and the EU biotechnology program (BIOTECH Contract CT 97−2303).

* Address correspondence to this author: Telephone + 33 (0)4 91 16 45 47; Facsimile + 33 (0)4 91 71 33 21; E-mail hpfierob@ibsm.cnrs-mrs.fr.

‡ CNRS.

§ Université de Provence.

[1] Abbreviations: CBD, cellulose binding domain; HD, hydrophilic domain; SPR, surface plasmon resonance; CD, circular dichroism; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight.

10.1021/bi9911740 CCC: $18.00
Published on Web 09/03/1999

terms of both calcium dependence and the necessity of cohesin and dockerin domains for the docking of cellulases onto the scaffoldin.

The three-dimensional structures of two cohesin domains from *C. thermocellum* have been solved, as has the structure of the CBD of the scaffoldin from the same bacterium. These domains were found to share the same overall jellyroll topology composed of nine β-strands connected by loops (*21*−*23*). To date, however, no structural data concerning the dockerin domain of any cellulosome-producing clostridia are available. A recent study has shown that, in the case of *C. thermocellum*, both duplicated segments are required to bind the cohesin domain (*24*). It has also been demonstrated that, despite strong sequence homologies, the cohesin/dockerin interaction is species-specific, since the *C. thermocellum* cellulases fail to interact with the *C. cellulolyticum* scaffolding protein and vice versa (*13*). Based on these data, a model was proposed for the cohesin/dockerin interaction, and it was hypothesized that some conserved residues at specific positions either are involved in calcium binding or serve as specificity determinants during the complexation with the cohesin partner (*13*).

In the present study, by means of site-directed mutagenesis, surface plasmon resonance (SPR), calcium overlay, and circular dichroism (CD), several questions concerning the assembly of the cellulosomes from *C. cellulolyticum* were addressed: (i) Are cohesin and dockerin solely responsible for the docking of cellulases to the scaffolding protein? (ii) What role does calcium play in the folding and complexation processes? (iii) Is the dockerin domain structured when it is not bound to the cohesin domain? (iv) Are the upstream flanking regions of both cohesin and dockerin domains required? (v) Are the two conserved segments of the dockerin domain required for the complexation with the cohesin partner, and if so, do they have the same importance? Finally, the validity of various models describing the cohesin/dockerin complex formation is discussed.

EXPERIMENTAL PROCEDURE

Strains and Plasmids. DH5α (Bethesda Research Laboratories) *Escherichia coli* strain was used as host for the pJF118EH derivative vectors pA2 (*14*), pA3 (*25*), and pA4-H. BL21(DE3) *E. coli* strain was used as host for the following pET22b(+) (Novagen) derivative vectors: pETCip1, pETCip2 (*5*), pET-HD1coh1, pET-coh1A, pET-coh1B, pET-dokeA, pET-dokeB, pET-doke11, pET-doke22, and pET-doke21.

Construction of the Various Expression Vectors. The DNA fragment encoding modified CelA, having only the first duplicated segment of the dockerin domain and a C-terminal His tag, was obtained from pA2, by PCR using the synthetic forward primer A4f (5′-GCATATTGTCCATGGAATTTTGCAGGG-3′), which is homologous to an internal sequence of the *celA* gene, which contains a *Nco*I site (underlined). The reverse primer was A9r (5′-CCCCCAAGCTTAGTGGTGGTGGTGGTGGTG**CAAATTCTTTAC**-3′) having partial homology (in boldface type) with the region encoding the first duplicated segment of the dockerin domain and introducing six histidine codons, a new stop codon, and a *Hin*dIII site (underlined). The amplified fragment was purified (Qiaquick, Qiagen) and digested by *Nco*I and *Hin*dIII. The resulting DNA fragment was purified by electrophoresis and prep-A-gene (Bio-Rad) and ligated into a *Nco*I−*Hin*dIII linearized plasmid pA2. The resulting plasmid was called pA4-H, and the inserted fragment was sequenced (Genome Express Society) by using a Perkin-Elmer 373 fluorescent sequencing apparatus (Applied Biosystems Dye terminator method).

The DNA fragment encoding HD1, the first cohesin domain of CipC, and a His tag at the C-terminus was prepared by PCR from pETCip1. The forward primer was cip17f (5′-GGCGGGCCATATG**GACGGTGGAAACCCACCT**-3′), which has partial homology with the DNA region encoding the linker region located at the N-terminus of HD1 (in boldface type) and introduces an ATG codon and a *Nde*I site (underlined). The reverse primer used was cip21r (5′-CCAGCTCGAGCTAGTGGTGGTGGTGGTG**TACTGCTACTTTAAGTTCCTTTGTAG**-3′), which has partial homology with the region encoding the C-terminus of cohesin1 (in boldface type) and introduces five histidine codons, a stop codon, and a *Xho*I site (underlined). After purification, the DNA fragment was digested by *Nde*I and *Xho*I and ligated into pET22b(+) linearized with the same endonucleases, as described above. The resulting plasmid was named pET-HD1coh1, and the insert was sequenced as previously described.

Construction of pET-coh1A and pET-coh1B encoding truncated and entire cohesin1, respectively, was performed as described above with the exception that the primers used for PCR amplification were cip18f (5′-GGTGGGTCATATG-**ACTGTTTTACCAAAAGATATCCCT**-3′) and cip21r in the case of pET-coh1A, whereas cip19f (5′-CGCCACGCATATG**AAAGTTACAGTAGGA**-3′) (*Nde*I site and ATG codons are underlined, and homologous regions are in boldface type) and cip21r were used for the construction of pET-coh1B.

Construction of pET-dokeB and pET-dokeA, designed for the production of the dockerin domain of *celA* commencing at the first conserved residue (pET-dokeB) or eleven residues upstream of the first conserved residue (pET-dokeA), was carried out using the same procedure as for pET-coh1A. The primers used for the amplication by PCR from pA2 were A5f (5′-GGTGGGTCATATG**GGAGATTATAACAATGATGGA**-3′) and A10r (5′-CCCCTCGAG**CTTACTTACCATACCAAG**-3′) in the case of pET-dokeA and A6f (5′-CCGCCCGCATATG**GCCAAGACAGATCCTGAC**3′) and A10r in the case of pET-dokeB (shown in boldface type, homologous areas; underlined, *Nde*I site in the case of A5f and A6f primers and *Xho*I site in the case of the primer A10r).

Modified Dockerins. pET-doke11 encoding a dockerin containing twice the first duplicated segment was constructed as follows: the region coding for the first segment was amplified by PCR from pET-dokeB by use of the forward primer A6f and the reverse primer A11r (5′-TACATAAGCATGGTCAGC-3′), which is homologous to DNA encoding the linker region between the two conserved segments and contains an *Nla*III site (underlined). The amplified fragment was called Seg1A. A second amplification of the DNA encoding the first duplicated segment was performed from pA4-H using the forward primer A7f (5′-*GCTGAC*CATGCTTATGTAAAGAAT**GGAGATTATAACAATGAT**-3′), which has homologies with both the region encoding

the N-terminus of the first segment (in boldface type), and the region encoding the linker region (in italic type with *Nla*III site underlined). The reverse primer was A12r (5′-CCCCTCGAG*CTTACTTACCAT***AGCCATAATATATTT-CTT**-3′), which has homologies with both the region encoding the C-terminus part of the first segment (in boldface type) and the region encoding the C-terminus extremity of the dockerin domain (in italic type), and which introduces a *Xho*I site (underlined). This amplified fragment was named Seg1B. Seg1A was digested by *Nde*I and *Nla*III, whereas Seg1B was digested by *Nla*III and *Xho*I. The digested fragments were purified and ligated into a *Nde*I−*Xho*I-linearized pET22b(+). The resulting plasmid was named pET-doke11.

A similar strategy was used to construct pET-doke22: the region encoding the second duplicated segment was amplified from pET-dokeB by use of the forward primer A8f (5′-C G C C C G C A T A T G *G C C A A G A C A G A T C C T-GACCCAGTAATTGTATAT***TTGGATGTTATTCTCGAC**-3′), which is homologous to both the region encoding the N-terminus of the second duplicated segment (in boldface type) and the DNA region encoding the 11 residues located upstream of the first segment (in italic type) and which introduces a *Nde*I site (underlined). The reverse primer was A13r (5′-*TACATAAGCATGGTCAGC***ACCAAGCAGATA-TTTTTT**-3′), homologous to both the DNA encoding the linker region between the duplicated segments (in italic type with *Nla*III site underlined), and the C-terminus part of the second conserved segment (in boldface type). The amplified fragment obtained was called Seg2A. A second amplification of the region encoding the second segment was carried out from pET-dokeA with the forward primer A9f (5′-GCT-GACCATGCTTATGTAAAGAAT-3′), homologous to the region encoding the linker region between the two duplicated segments and containing a *Nla*III site (underlined), and the reverse primer A10r. The amplified fragment was called Seg2B. Seg2A was digested by *Nde*I and *Nla*III, whereas Seg2B was digested by *Nla*III and *Xho*I. After purification, the digested fragments were ligated into a *Nde*I−*Xho*I-linearized pET22b(+). The resulting plasmid was named pET-doke22.

The vector pET-doke21, encoding a modified dockerin domain containing segment 2 followed by the segment 1, was obtained by ligation of Seg2A (digested by *Nde*I and *Nla*III) and Seg1B (digested by *Nla*III and *Xho*I) into a *Nde*I−*Xho*I-linearized pET22b(+).

Expression and Purification of the Recombinant Proteins. *E. coli* DH5α or BL21(DE3) was grown at 37 °C in Luria−Bertani medium supplemented with ampicillin (200 μg/mL) to an $OD_{600} = 1.5$. The culture was then cooled to 25 °C, and isopropyl thio-β-D-galactoside (Appligene) was added to a final concentration of 250 μM. After 12−16 h, the cells were centrifuged (10 min, 3000*g*), resuspended in 30 mM Tris-HCl, pH 8, and broken in a French press. The purification of entire CelA (CelA2) and truncated CelA (CelA3) was performed as previously described by Fierobe et al. (*14*), and Bélaïch et al. (*25*), respectively. MiniCipC1 was purified according to Pagès et al. (*5*). All other recombinant proteins (the present study) contain a His tag at their C-termini and were purified at 4 °C as follows: NaCl (0.2 M) and DNase I (1 g/L) were added to the crude extract. The suspension was then loaded on 2−5 mL of Ni−NTA resin (Qiagen), equilibrated in 30 mM Tris-HCl, pH 8.0, and 0.2 M NaCl (TN buffer). The resin was subsequently washed with 10 mM imidazole in TN buffer. The protein of interest having the C-terminal His tag was eluted by a linear gradient from 0.01 to 0.15 M imidazole in TN buffer. Analysis by SDS−PAGE (Phast System, Pharmacia Biotech Inc.) of the recombinant protein-containing fractions indicated that no further purification was required. The fractions were pooled, concentrated by ultrafiltration (Millipore, cutoff 10 kDa), and dialyzed overnight at 4 °C against 10 mM Tris-HCl, pH 8.0.

Determination of Protein Concentration and Other Analytical Procedures. The concentration of purified proteins was routinely estimated at 280 nm in 6 M guanidine hydrochloride and 25 mM sodium phosphate, pH 6.5, by using the program ProtParam tool (http://expasy.hcuge.ch/sprot/protparam.html), and also by quantitative amino acid analysis on a Waters Pico-tag amino acid system. N-Terminal amino acid sequences were determined by stepwise Edman degradation on a gas-phase sequencer (Model 470 A, Applied Biosystems). Mass spectrometry was performed by MALDI-TOF on a Voyager DE-RP (Perseptive Biosysteme), with apomyoglobin as the standard.

Biotinylation of the Proteins. MiniCipC1, HD1-cohesin1, cohesin1A, and cohesin1B were biotinylated with biotinyl *N*-hydroxysuccinimide ester as described by Bayer and Wilcheck (*26*), using the Biotin labeling kit (Boehringer-Mannheim) according to the manufacturer's instructions, except that the biotinyl *N*-hydroxysuccinimide ester/protein ratio was kept between 0.7 and 1.

Surface Plasmon Resonance. The kinetic parameters of the interaction between cohesin-containing proteins and the various dockerins were determined on a BIAcore apparatus (BIAcore), at 25 °C. The biotinylated MiniCipC1, HD1-cohesin1, cohesin1A, or cohesin1B was coupled to a streptavidin−dextran layer on the sensor chip's surface. Biotinylated recombinant proteins were injected for 50−200 s, resulting in approximately 600−800 resonance units (RU) of immobilized protein. The flow cell was routinely equilibrated with 10 mM $CaCl_2$, 0.005% surfactant P20 (BIAcore), and 50 mM Tris−maleate buffer, pH 6.5, except for the pH dependence experiments, for which 50 mM glycine hydrochloride (pH 3.0−4.0), sodium acetate (pH 4.0−5.5), Tris−maleate (pH 5.8−7.5), Tris-HCl (pH 7.5−8.8), or glycine−NaOH (pH 8.5−9.8) were used instead of Tris−maleate buffer, pH 6.5. The ligands (CelA2, CelA3, CelA4-H, DokeA, DokeB, Doke11, Doke22, or Doke21) were diluted in the same buffer and allowed to interact with the sensor surface by a 300 s injection. In all cases, at least three different concentrations of ligand were injected. The obtained sensograms were evaluated with the biomolecular interaction analysis evaluation software (BIAcore) to calculate the kinetic constants. The data were interpreted on the basis of the simple model $L + A \leftrightarrow LA$, where L denotes the mobile ligand and A the immobilized receptor. For all tested ligands, control experiments were performed by direct injection onto the streptavidin−dextran layer. In addition, with the exception of CelA2, the sensor chip's surface was regenerated by injection of 10 mM EDTA. For CelA2, complete removal of the ligand was not possible by such treatment, and a new surface had to be used for each concentration of CelA2.

Calcium Overlay Experiments. Calcium overlay was performed essentially as has previously been described (*27*).

Briefly, proteins were subjected to native PAGE (12.5% or 20%) and transferred onto nitrocellulose (Schleicher and Schüll) membrane. The membrane was incubated for 15 min in isoamyl alcohol (20%) and subsequently washed twice with 60 mM KCl, 5 mM $MgCl_2$, and 10 mM imidazolium, pH 7.0 (KMI buffer), during 10 min. The membrane was then incubated for 45 min at room temperature in KMI buffer supplemented with ^{45}Ca (10 μCi/mL). The nitrocellulose sheet was again washed twice with deionized water for another 10 min. The membrane was dried at 37 °C for 3 h, and bound ^{45}Ca was detected with a Storm 820 PhosphorImager (Molecular Dynamics). The nitrocellulose membrane was then stained with Ponceau red (Sigma).

Circular Dichroism Spectroscopy. CD spectra were acquired with a CD6 Jobin-Yvon apparatus (ISA, Division d'Instruments, S.A.). The temperature was 10 °C, unless stated otherwise. The bandwidth was 1 nm with an integration time of 30 s. The spectra were measured in a 50 μm optical path cuvette. The protein concentrations were 0.1−0.3 mM in 2.5 mM Tris-HCl buffer, pH 8.0. Spectra were corrected for background contributions from the buffer by omitting the protein. Ellipticities were normalized to residue concentration by using the relationship $[\Theta] = \Theta_o M_r/lc$, where Θ_o is the observed ellipticity in degrees, M_r is the average molecular mass of an amino acid in the polypeptide, l is the path length in centimeters and c is the protein concentration in grams per liter. Deconvolution of the far-UV CD profiles was carried out by using standard computer procedures (DICHROPROT V2.4 by G. Deleage) including neural network (*28*).

RESULTS

Design of Truncated CipC and Various Dockerin Domains. In a former study it has been shown that miniCipC1 (Figure 1) containing the CBD, first hydrophilic domain (HD1), and first cohesin domain was able to bind the dockerin-containing cellulases of *C. cellulolyticum* (*5*). In the present study, based on sequence homologies and the available crystal structures of cohesin from *C. thermocellum* (*21*, *22*), three new miniCips, possessing a C-terminal His tag, were designed (Figure 1): HD1-cohesin1, which lacks the CBD; cohesin1A, which contains the entire cohesin 1 domain flanked by the linker regions, and finally cohesin1B, which starts at the second conserved residue of the cohesin domain and has the same C-terminal extremity as cohesin1A.

A similar strategy was used in the design of the dockerin domains (Figure 1), although no structural data are available concerning this domain, which is typical of cellulosomal enzymes. It was shown by Pagès et al. (*13*) that the entire CelA (CelA2) from *C. cellulolyticum* binds miniCipC1, conversely to CelA3, a truncated form of the cellulase lacking the C-terminal dockerin domain. In the present study, an attempt was made to evaluate the importance of each conserved segment and the flanking regions. The various recombinant dockerins engineered are summarized in Figure 1. Entire CelA (CelA2) was used as the reference. A first modified CelA, CelA4-H, that lacks the second duplicated segment was built. Second, two forms of the free dockerin domain of CelA (without the catalytic domain) were engineered: DokeA, commencing 11 residues upstream of the first conserved residue of the dockerin domain, and DokeB, starting at the first conserved residue. Finally, to determine the relative importance of each segment, three additional modified dockerins were constructed: Doke11, which contains twice the first segment; Doke22, which contains twice the second segment; and Doke21, having both segments but in reverse order compared to the wild-type dockerin domain. The three latter mutant dockerins also contain the 11 residues located upstream of the first segment.

Production and Purification of the Various Recombinant Proteins. All recombinant proteins were produced in *E. coli*. Despite high production yields (usually between 0.1 and 0.2 g/L of culture), all recombinant proteins were found in the soluble fraction of the bacterial lysate. The new recombinant proteins designed in the present study were all His-tagged at the C-terminus to allow rapid purification on Ni−NTA resin. In all cases, the quantity of resin was kept far below the amount necessary to bind all recombinant protein present in the crude extract. This induced important losses of the protein of interest. The complete saturation of the Ni−NTA resin by the recombinant protein, however, rendered further purification steps unnecessary, as indicated by SDS−PAGE analysis (data not shown) of the fractions containing the His-tagged proteins. MiniCipC1, CelA2, and CelA3, devoid of His tag, were purified as previously described (*5*, *14*, *15*).

Mass spectrometry was performed on the purified recombinant proteins. As shown in Table 1, the measured masses were in very good agreement with the masses calculated from the amino acid sequences, indicating that no proteolytic cleavage, decarboxylation, or deamination occurred during the production and purification.

SPR. Pagès et al. (*13*) showed that calcium was required for the binding of cellulases to the scaffoldin CipC. Most of the experiments were thus performed in the presence of 10 mM $CaCl_2$, in 50 mM Tris−maleate, pH 6.5, and 0.005% P20 surfactant, to avoid nonspecific interactions with the sensor chip surface. In a first series of experiments, CelA2 was used as the ligand, and various biotinylated fragments of CipC containing the first cohesin domain were immobilized on the sensor chip surface. The results reported in Table 2 clearly demonstrate that neither the removal of the CBD (HD1-cohesin1) nor that of the CBD and the first hydrophilic domain (cohesin1A) affects the affinity of CelA for this cohesin domain, since the obtained K_D values remain in the $(2-2.5) \times 10^{-10}$ M range. When cohesin1B, lacking both the N-terminal linker region and first highly conserved residue of the cohesin domain, was used as the immobilized receptor, the affinity of CelA2 for this "truncated cohesin" was only reduced by a factor of 4, indicating that this linker region located upstream of the cohesin domain is not essential for complexation of the cellulases.

In the case of CelA4-H, which lacks the second duplicated segment, no binding was detected (see Table 2) even at high ligand concentrations (up to 100 μM), thus demonstrating that two segments of the dockerin domain are required for the interaction. To confirm this absence of interaction, gel retardation with an equimolar mixture of CelA4-H and cohesin1A as well as western blots using biotinylated cohesin1A were attempted. Both of these tests also generated negative results (data not shown).

To determine the influence of the catalytic domain on the binding parameters to cohesin, DokeA, which commences 11 residues upstream of the first conserved residue of the

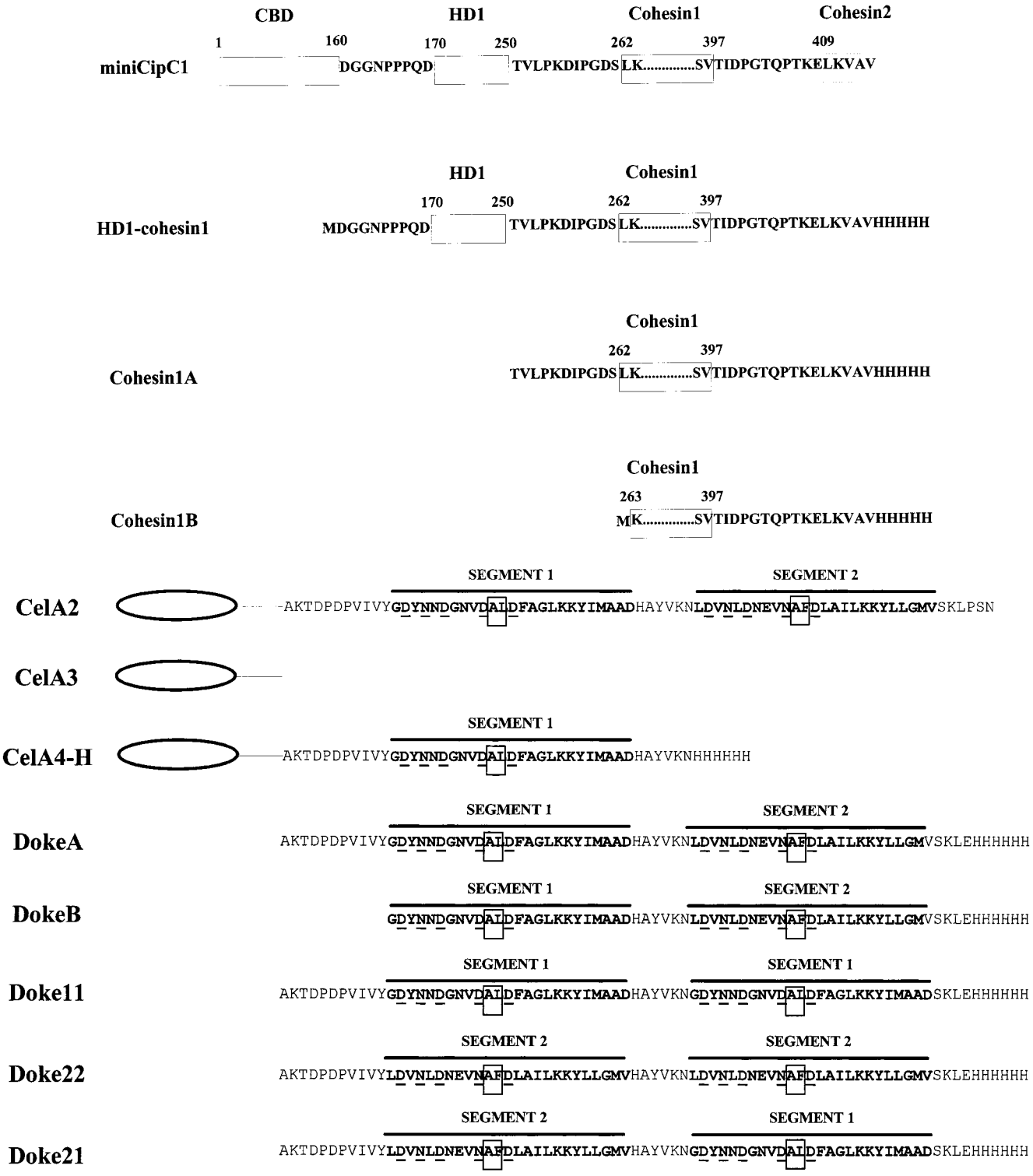

FIGURE 1: Diagram of the recombinant proteins used. (Top) Cohesin-containing proteins: boxes represent the domains of CipC (CBD, HD, cohesin), and numbers refer to the residue number in the mature protein. The primary sequences between the boxes are the linker regions defined on the basis of sequence comparisons. (Bottom) Dockerin-containing proteins: the white oval is the catalytic domain of CelA, and residues in boldface type are the conserved segments of the dockerin domain. Residues participating in the calcium binding site are underlined, and boxes represent those residues suspected of serving as selectivity determinants (*13*).

dockerin domain (Figure 1), was used as the ligand in the BIAcore experiments. The binding parameters (Table 2) were almost identical to those of CelA2. As was observed in the case of the entire enzyme, these parameters remained unchanged from miniCipC1 to cohesin1A (Table 2). These results strongly suggest that the catalytic domain of CelA has no influence on the binding to cohesin1. Furthermore, they indicate that the His tag grafted to the C-terminus of the dockerin domain does not disturb the binding to the cohesin partner. As observed in the case of CelA2, a small but significant decrease in the affinity (5-fold) was observed when cohesin1B was immobilized on the sensor chip surface. A noticeable difference was, however, observed between CelA2 and DokeA: in both cases, the dissociation was very

Table 1: Molecular Masses of Recombinant Cohesin- and Dockerin-Containing Proteins

proteins	theorotical (Da)	measured (Da)
HD1-cohesin1	26 972.5	26 974.9 ± 2.1[a]
cohesin1A	17 483.9	17 483.1 ± 1.5
cohesin1B	16 378.6	16 376.7 ± 2.4
DokeA	8 687.8	8 688.2 ± 1.0
DokeB	7 488.4	7 487.5 ± 0.3
Doke11	8 601.6	8 600.9 ± 0.5
Doke22	8 774.0	8 773.3 ± 0.3
Doke21	8 687.8	8 687.6 ± 0.6

[a] Standard deviation.

long (k_{off} values are around 3×10^{-4} s^{-1}; Table 2), preventing new injections of ligand. To accelerate this dissociation and regenerate the sensor chip surface, EDTA was assayed, since it is known (*13*) that this compound inhibits the cellulase/scaffoldin complex formation. Thus, the dissociation kinetic was followed for 5−10 min to determine the k_{off}, and then short injections of 10 mM EDTA were performed to remove all residual ligand. DokeA could easily be removed from the sensor chip surface by injection of 10 mM EDTA, whereas such a procedure was inefficient with CelA2. Even at concentrations up to 100 mM EDTA, only partial desorption of CelA2 was observed. One possible explanation for this difference is that the calcium atoms remain exposed to EDTA in DokeA, whereas they are protected from EDTA by the large catalytic domain in the case of CelA2. More drastic treatments, such as the short injection of 0.05% SDS and 10 mM NaOH, were able to completely remove residual CelA2 from the immobilized protein. Such procedures, however, lead to an irreversible denaturation of the immobilized recombinant protein. Quite unexpectedly, injection of nonbiotinylated cohesin1A or miniCipC1 up to 50 μM did not desorb CelA2 or DokeA but led to an increase in the RU. One possible explanation for this is that the cohesin domain has affinity for itself. It was observed, for example, that at high protein concentrations, purified cohesin1A generated a single band on SDS−PAGE, whereas a second faster-migrating minor band appears when nondenaturing conditions are used during electrophoresis (data not shown). It is also worth noting that the cohesin domains 3 and 7 of CipA from *C. thermocellum*, the crystal structures of which have been solved, both formed dimers in the crystal (*21*, *22*).

Since the sensor chip surface was easily regenerated with DokeA, and the binding parameters were very similar to those of CelA2, the cohesin1A/DokeA couple was used to study pH dependence of the binding parameters (Figure 2) in the presence of calcium. It appeared that both k_{on} and k_{off} vary very little between pH 5 and 9. Large variations were only observed at acidic pH ($K_D = 6.0\ 10^{-7}$ M at pH 3.0, versus 2.21×10^{-10} M at pH 6.5), whereas at pH 9.85 the affinity was barely 5-fold reduced ($K_D = 1.0 \times 10^{-9}$ M at pH 9.85, versus 2.21×10^{-10} M at pH 6.5). It was unfortunately not possible to investigate the binding parameters at a pH lower than 3.0 or higher than 10, since the sensor chip surface was irreversibly damaged at these pHs. Nevertheless, the results available indicate that this interaction remains strong over a wide range of pH.

DokeB (Figure 1) was engineered to determine whether the 11 residues located upstream of the first duplicated

Table 2: Association and Dissociation Constants[a] for the Binding of Dockerin Domains to Immobilized Cohesin Domains

	immobilized receptor											
	miniCipC1			HD1-cohesin1			cohesin1A			cohesin1B		
ligand	k_{on} (s^{-1} M^{-1})	k_{off} (s^{-1})	K_D[b] (M)	k_{on} (s^{-1} M^{-1})	k_{off} (s^{-1})	K_D (M)	k_{on} (s^{-1} M^{-1})	k_{off} (s^{-1})	K_D (M)	k_{on} (s^{-1} M^{-1})	k_{off} (s^{-1})	K_D (M)
CelA2	1.5×10^6 ±6.5×10^4[c]	3.1×10^{-4} ±1.1×10^{-5}	2.1×10^{-10}	1.7×10^6 ±3.4×10^4	3.5×10^{-4} ±2.9×10^{-5}	2.1×10^{-10}	1.2×10^6 ±9.4×10^4	3.0×10^{-4} ±2.5×10^{-5}	2.5×10^{-10}	9.0×10^5 ±8.0×10^4	9.1×10^{-4} ±4.0×10^{-5}	1.0×10^{-9}
CelA3	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
CelA4-H	*d*	*d*	*d*	*d*	*d*		*d*	*d*	*d*			
DokeA	1.3×10^6 ±1.1×10^5	3.2×10^{-4} ±8.1×10^{-6}	2.5×10^{-10}	1.4×10^6 ±8.1×10^4	3.4×10^{-4} ±2.1×10^{-5}	2.4×10^{-10}	1.4×10^6 ±2.3×10^4	3.1×10^{-4} ±1.9×10^{-5}	2.2×10^{-10}	7.8×10^5 ±7.5×10^4	8.6×10^{-4} ±4.5×10^{-5}	1.1×10^{-9}
DokeB							7.1×10^4 ±5.5×10^3	9.7×10^{-3} ±5.1×10^{-4}	1.4×10^{-7}			
Doke11							6.3×10^5 ±2.5×10^4	8.3×10^{-4} ±3.9×10^{-5}	1.3×10^{-9}			
Doke22							1.0×10^6 ±3.1×10^4	6.6×10^{-4} ±2.5×10^{-5}	6.6×10^{-10}			
Doke21							7.7×10^5 ±1.7×10^4	2.8×10^{-3} ±1.5×10^{-4}	3.6×10^{-9}			

[a] Determined at 25 °C in 50 mM Tris−maleate, pH 6.5, 10 mM $CaCl_2$, and 0.005% P20. [b] The apparent equilibrium dissociation constant K_D was directly calculated from the ratio k_{off}/k_{on}. [c] Standard deviation. [d] No binding detected. Blank spaces indicate not determined.

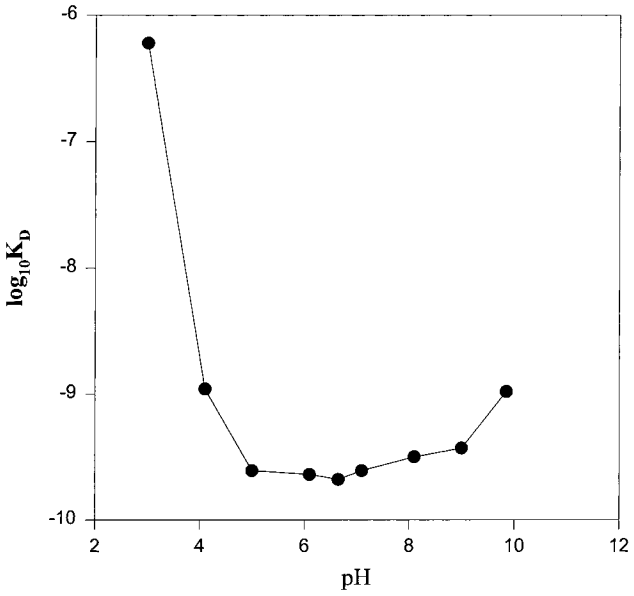

FIGURE 2: pH dependence of the DokeA/cohesin1A interaction. The K_D values were determined in the presence of 10 mM $CaCl_2$, 0.005% P20 surfactant, and 50 mM buffer (see Experimental Procedures).

segment have any influence on the binding to cohesin. SPR experiments with DokeB as the injected ligand show that the affinity for cohesin1A is approximately 600-fold reduced compared to CelA2 or DokeA. Both k_{on} (20-fold reduced) and k_{off} (30-fold increased) were affected by the removal of these residues. This suggests that these 11 residues, or at least some of them, are directly or indirectly involved in the complex formation with the cohesin domain. This point is further discussed in the context of circular dichroism experiments below.

Because of the lowered affinity of DokeB for cohesin1A, the 11 upstream residues were maintained in the modified dockerins having various combinations of segments 1 and 2. The binding parameters for cohesin1A obtained with these mutant dockerins are summarized in Table 2. In all cases, the affinity for cohesin1A was reduced compared to either wild-type DokeA or CelA2 but remained in the same range nonetheless. Doke21 was found to have the lowest affinity, with a 16-fold increase in its K_D value compared to DokeA, whereas Doke22 had only a 3-fold increase in K_D. These variations are almost exclusively due to an increased k_{off}, whereas the k_{on} values observed are at most reduced by a factor of 2, compared to wild-type dockerin. The use of CelA4-H as the ligand has shown that two segments are required for the binding to cohesin1A. The experiments performed with the mutant dockerins indicate that the type of segment or the order in the dockerin domain can be changed without drastically affecting the dockerin's function.

Calcium Binding. Pagès et al. (*13*) have shown that, as in the case of *C. thermocellum* (*19*), calcium is required for the interaction to occur. SPR experiments performed in the present study also confirm that EDTA inhibits the complex formation and can even dissociate the complex in the case of DokeB. Sequence comparisons strongly suggest that the dockerin domain contains two EF-hand calcium binding sites, located at the beginning of each conserved segment (*13*, *29*) (Figure 1). It has never been demonstrated, however, that the dockerin domain of *C. cellulolyticum* binds calcium. To

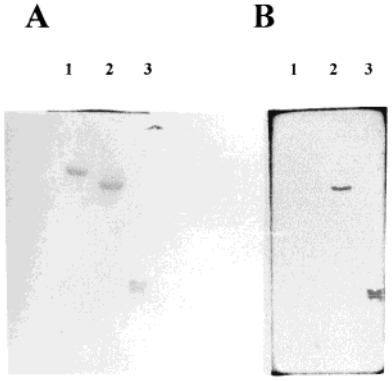

FIGURE 3: Calcium overlay experiment. (A) Nitrocellulose membrane colored with Ponceau red, and (B) detection of bound ^{45}Ca with a Phosphorimager. The amounts of proteins used are as follows: Lane 1, 5 μg of cohesin1A; lane 2, 5 μg of cohesin1A and 2.7 μg of DokeA; lane 3, 2.7 μg of DokeA.

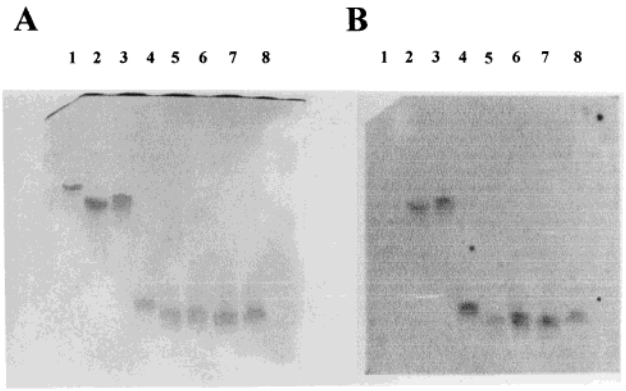

FIGURE 4: Calcium overlay experiment. (A) Nitrocellulose membrane colored with Ponceau red, and (B) detection of bound ^{45}Ca with a Phosphorimager. The amounts of proteins used are as follows: Lane 1, 5 μg of CelA3; lane 2, 6 μg of CelA2; lane 3, 6 μg of CelA4-H, lane 4, 2.7 μg of DokeA; lane 5, 3.1 μg of DokeB; lane 6:2.9 μg of Doke11; lane 7:3.5 μg of Doke22; lane 8, 3.1 μg of Doke21.

check this hypothesis, the calcium overlay technique was used (see Experimental Procedures). The results shown in Figure 3 indicate that cohesin1A domain does not bind calcium, contrary to the DokeA or DokeA/cohesin1A complex, thus confirming that the dockerin does contain calcium binding site(s). In a second set of experiments (Figure 4), the calcium overlay technique was applied to different forms of CelA or its dockerin domain. With the exception of CelA3, which lacks the entire dockerin domain, all proteins were found to bind calcium. The fact that CelA4-H, Doke11, and Doke22 were able to bind the divalent cation strongly supports the hypothesis that each duplicated segment contains a calcium binding site.

Circular Dichroism. To date, no structural data are available concerning the dockerin domain. Crystallization attempts are currently underway, but suitable crystals for X-ray diffraction have not been obtained to date. To gain some insight on the structure of this key domain, the CD approach was thus tried. The far-UV spectrum of DokeA in the presence of 2 mM $CaCl_2$, recorded at 10 °C, is shown in Figure 5A. This spectrum is characteristic of a α-helix-rich domain. A stepwise increase in temperature on the same protein sample, to 25, 37, and 50 °C, had no effect on the spectrum of the dockerin. At higher temperatures (55−60

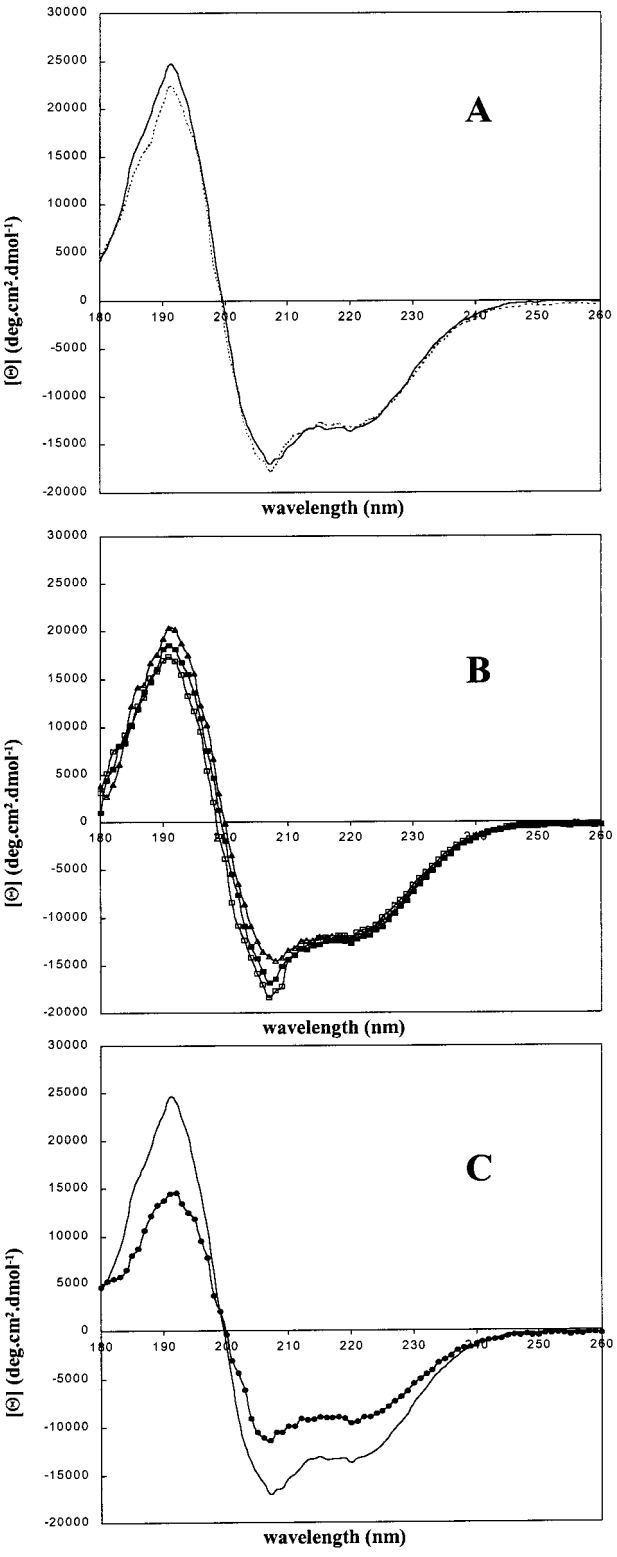

FIGURE 5: Far-UV CD spectra of the wild-type and modified dockerin domains recorded at 10 °C, in 2.5 mM Tris-HCl pH, 8.0. (A) DokeA in the presence of 2 mM $CaCl_2$ (—), or 2 mM EDTA (···). (B) Doke11 (△), Doke21 (■), and Doke22 (□) in the presence of 2 mM $CaCl_2$. (C) DokeA (—) and DokeB (●) in the presence of 2 mM $CaCl_2$.

°C) the protein sample started to precipitate, which made all ellipticity measurements impossible.

The profile of the metal-free form of DokeA (in the presence of 2 mM EDTA) was very similar (see Figure 5A) to the spectrum of DokeA recorded in the presence of 2 mM $CaCl_2$. This result indicates that the removal of calcium has very minor effects on the secondary structure of this domain. The CD spectra of Doke11, Doke21, and Doke22, recorded in the presence of 2 mM $CaCl_2$, were also quite similar to that of DokeA (Figure 5B). Some small but significant differences in the height of the peaks were, however, observed at 191 and 207 nm. The molar ellipticity values of the mutants at 191 nm were reduced by 15−20% compared to DokeA. Analysis of the peak centered at 207 nm reveals a more interesting feature since "signature" for each type of segment emerges from the signal at this wavelength. The molar ellipticity values at 207 nm in the case of DokeA and Doke21, which both contain the two types of segments, are −17 000 and −16 800 deg·cm^2·dmol^{-1}, respectively. In the case of Doke11, the molar ellipticity value at this wavelength is −14 100 deg·cm^2·dmol^{-1}, whereas in the case of Doke22 the ellipticity value is decreased to −18 400 deg·cm^2·dmol^{-1}. Conversely, the spectrum of the "truncated" dockerin, DokeB, displays marked differences compared to that of DokeA (Figure 5C). Although the overall shape is quite similar, the molar ellipiticity values were significantly reduced compared to the other dockerins. Such a decrease cannot be attributed to an important error in the determination of the protein concentration, since great care was taken to determine this crucial parameter. Indeed, two different procedures were used (amino acid composition and absorption at 280 nm in 6 M guanidine hydrochloride), which generated very similar protein concentrations (within 5% deviation) for all proteins. The large variations in the CD spectrum of DokeB versus DokeA are most probably due to important changes in the secondary structure of the truncated dockerin. As for DokeA, spectra of the various dockerin mutants were also recorded in the presence of 2 mM EDTA and vary very little compared to those in the presence of 2 mM $CaCl_2$.

Table 3: Secondary Structure Content in Percentage and in Number of Residues As Determined from the Far-UV Region of the CD Spectra of the Recombinant Wild-Type and Mutant Dockerin Domains of CelA[a]

protein	α-helix (%)	β-strand (%)	α-helix (res)	β-strand (res)
DokeA	40−45	10−15	31−35	8−12
DokeB	23−28	15−20	15−19	10−13
Doke11	37−42	10−15	29−32	8−12
Doke22	37−44	10−15	29−34	8−12
Doke21	40−44	10−15	29−34	8−12

[a] See Figure 5 in the presence of 2 mM $CaCl_2$.

At this point, the secondary structure of the various dockerin domains was inferred from the far-UV CD spectra. The results are summarized in Table 3. It is worth noting that, as expected in view of their spectra, DokeA, Doke11, Doke21, and Doke22 seem to share very similar secondary structures, whereas in the case of the truncated dockerin, DokeB, the percentage of α-helices drops to approximately 25%, and a higher score for random coils is observed (35%). However, caution needs to be exerted regarding these percentages and the absolute number of α-helical and β-strand residues, since these values may be biased by a number of factors: deconvolution methods, amino acid composition (occurrence of aromatic residues), etc. (*30*). In addition, since no other structural data concerning the dockerin from CelA

or any homologous dockerin from *C. cellulolyticum*, *C. thermocellum*, or *C. cellulovorans* are available to date, it was not possible to correlate the values given in Table 3 with a dockerin of known structure. These numbers therefore must not be considered as absolute values. These secondary structure percentages, however, are well correlated to the SPR data (see above), which indicate, in the case of DokeB, a drastic decrease in the affinity for cohesin1A, whereas Doke11, Doke21, and Doke22 have K_D values for cohesin1A in the same range as that of DokeA.

DISCUSSION

It has been shown over the past few years that the dockerin/cohesin interaction plays a key role in the assembly of the cellulosomes from *C. thermocellum*, *C. cellulolyticum*, or *C. cellulovorans* (*13*, *19*, *20*) and presumably in other cellulosome-producing clostridia such as *Clostridium josui* (*31*) or *Clostridium papyrosolvens* (*32*, *33*). Recently it was demonstrated that, despite sequence homologies between the dockerin and cohesin from *C. thermocellum* and *C. cellulolyticum*, the cellulases from *C. cellulolyticum* were unable to bind the cohesin domains from *C. thermocellum* and vice versa (*13*). The latter study opens new fields of investigation, since hybrid miniscaffoldin containing both types of cohesin can be designed to bind cellulases or other enzymes harboring the appropriate dockerins, at specific positions. However, investigations to determine the boundaries of the domains involved, the independence of these domains, the strength of the interaction, and, in the case of the dockerin domain, the role of each conserved segment, are still required. It is these questions that have motivated the present study.

It is clear from the SPR data that neither the CBD nor the hydrophilic domain, located upstream of the first cohesin domain of CipC, has any influence on the binding parameters of the cohesin1/CelA interaction. This result strongly suggests that the first cohesin domain is functionally independent. It was also shown that the first cohesin domain from CipC is unable to bind calcium.

The cohesin domains of CipC are spaced by very short linkers, of about 8−10 residues (*12*), which is contrary to CipA from *C. thermocellum*, for which spacers are 25 residues long (*34*). To investigate the importance of these linkers, the upstream linker and the first conserved residue of the cohesin domain (L262), which by comparison to the available crystal structures of *C. thermocellum* cohesin domains is expected to participate to the first β-strand, were removed. The obtained truncated cohesin, cohesin1B, had only 4-fold reduced affinity for CelA2, indicating that these residues are not critical for binding. The role of the hydrophilic domain was not clarified by our experiments since it appears that it has no influence on the binding of CelA2 to the cohesin domain. It has been previously shown (*8*) that this domain, which is also found in the scaffoldins from *C. thermocellum*, and *C. cellulovorans* (*34*), does not interfere with the binding of the CBD of CipC to crystalline cellulose. Since this domain is not involved in the main functions of the scaffoldin, i.e., binding of cell wall degrading enzymes and binding to cellulose, its role might be at a different level. It has been shown in the present and former studies (*5*, *8*) that the free CBD, the free first cohesin domain, the miniCipC1, and HD1-cohesin1 are produced in large amounts in the cytoplasmic compartment of *E. coli* in a soluble form. An attempt was also made in the present study to produce a protein containing the CBD immediately followed by the first cohesin domain in *E. coli*, but this recombinant protein formed inclusion bodies when induction was performed at 25 °C, and no recombinant protein was detected in the soluble fraction of the crude extract. This aggregation was partly avoided by decreasing the temperature of induction to 15 °C. This observation suggests that the hydrophilic domains may prevent the aggregation of the scaffoldin due to the large number of hydrophobic cohesin domains.

The SPR data obtained when DokeA was used as the ligand also proves that the dockerin domain is independent and that the catalytic domain of CelA does not interfere with the binding to cohesin. The present study also demonstrates that, as expected due to sequence comparisons, the dockerin domain of CelA alone is able to bind calcium. Tavares et al. (*35*) suggested a model according to which complexation with the cohesin may increase the affinity for Ca^{2+} by stabilizing the structure of the EF-hand motifs of the dockerin or by contributing directly to the coordination of Ca^{2+}. The calcium overlay experiments, however, do not indicate whether the affinity for calcium increases when the dockerin is in complex with the cohesin partner. The same technique, applied to the various dockerin mutants, also indicates that each conserved segment can bind calcium, as suggested by the sequence comparisons between dockerins and calcium binding loops in the EF-hand motif (*13*, *29*). It has been observed in the case of the dockerin of CelS from *C. thermocellum*, however, that only the first segment was able to chelate the divalent cation (*27*). Our results with Doke22 clearly show that the second segment of the dockerin of CelA can also bind Ca^{2+}. In a recent study, Lytle and Wu (*24*) have separately overproduced each duplicated segment of the dockerin from CelS and shown by gel retardation that a stable complex with cohesin 3 from CipA was formed only when the two types of segments were present. The experiments performed in the present study with CelA4, which only contains the first segment of the dockerin and the linker region, also indicate that, in the case of *C. cellulolyticum*, two segments are also required to interact with the cohesin domain.

The SPR data for the dockerin mutants containing twice the first segment (Doke11), twice the second segment (Doke22), or both types of segment but in a reverse order compared to the wild type (Doke21) show that although two segments are required to interact with the cohesin domain, the type of segment is not really important since the K_D values obtained are 3−16-fold higher compared to the wild-type dockerin domain, DokeA. This observation was quite unexpected since the dockerin domain from CelA is quite atypical for a *C. cellulolyticum* dockerin: among the residues suspected of serving as selectivity determinants, according to Pagès et al. (*13*), the characteristic couple AL is found in the first segment, whereas the sequence AF is found in the second segment, while in all other known dockerin domains from *C. cellulolyticum* the sequences AL, AI, or AV are found at this specific position. The replacement of an aliphatic residue by an aromatic one does not seem to alter the binding to cohesin since Doke22 had the best binding parameters compared to either Doke11 or Doke21. The latter

result suggests that a certain amount of flexibility is permitted at this position, which was identified as critical by Pagès et al. (*13*). Doke21 was found to be the mutant having the lowest affinity for the cohesin domain despite the fact that it contains, like the wild-type dockerin, the two types of segments. This observation is not very surprising since in the two other dockerin mutants, one segment out of two remains in a wild-type configuration, whereas in Doke21, the flanking regions of both segments have been changed.

Analysis of the properties of DokeB, which commences at the first conserved residue of the dockerin domain, i.e., one residue upstream of the first putative calcium binding loop, also provides additional information concerning this domain: This truncated dockerin still binds calcium, thus indicating that at least one of the calcium binding sites remains operational, but it interacts with the cohesin domain with a 600-fold lower affinity compared to DokeA. Removal of the 11 residues located upstream of the first conserved amino acid of the dockerin thus has an important effect on the binding to cohesin, although these residues are not conserved among the cellulases from *C. cellulolyticum* or *C. thermocellum*. The most probable explanation for this drop in affinity is that these residues, or at least some of them, are involved in the correct folding of the first segment.

Whether the dockerin domain is structured or adopts a folded structure only upon calcium binding or upon interaction with the cohesin domain has become a major issue in our understanding of the cellulosome assembly. It has been observed that the free dockerin domain of CelD (without the catalytic domain) from *C. thermocellum* is insoluble in the absence of detergents or chaotropic agents (*36*), and it was suggested that complexation with the cohesin partner is required for high-affinity calcium binding and protein stability and/or structure (*35*, *36*). Pagès et al. (*13*) proposed a structural model for the dockerin domain based on sequence comparisons and secondary structure predictions, in which the calcium binding loop of each conserved segment is followed by an α-helix. These last authors do not mention, however, if the dockerin domain has this type of structure in a free or cohesin-bound state. In the present study, CD spectroscopy gave unambiguous answers concerning the structure of the free dockerin domain of CelA from *C. cellulolyticum*. The CD data demonstrate for the first time that this domain (DokeA) has a well-defined secondary structure in the free state. As many as 40−45% of the residues would belong to α-helices, whereas only 10−15% of the residues would be folded in β-strands. These data are in good agreement with the recently published secondary structure predictions based on primary sequence analysis (*13*). The presence of 2 mM calcium, or 2 mM EDTA, does not affect the secondary structure of the polypeptide at concentrations ranging from 0.1 to 0.3 mM. One could argue that, in the case of structural calcium, 2 mM EDTA might not be sufficient to trap all proteic calcium. The concentration of EDTA had to be kept as low as possible during the CD experiments, however, to lower the noise due to that compound. Electrophoresis under nondenaturing conditions was thus used to verify that under the same conditions (2 mM EDTA, 0.2 mM DokeA, and 2.5 mM Tris-HCl, pH 8.0) no complex with cohesin1A (0.2 mM) was formed. Furthermore, it was observed that 1 mM EDTA was enough to completely inhibit the binding of DokeA to cohesin1A during the SPR experiments. We are therefore confident that the CD spectrum recorded in the presence of 2 mM EDTA represents the calcium-free state of DokeA. A secondary structure independent of the calcium binding seems quite common in EF-hand domains or proteins. For example, this characteristic has been observed for the integrin α5β1 EF-hand domain (*30*) and in the S100 family proteins (*37*, *38*). It is of interest to note that although the secondary structure of these EF-hand proteins does not significantly change upon Ca^{2+} binding, these proteins were found to have very different apparent binding constants for this metal, varying from 10^3 M^{-1} (*37*) to 3×10^7 M^{-1} (*38*). Therefore, the absence of structural changes upon Ca^{2+} binding gives no indication about the affinity of DokeA for the divalent cation.

The CD data concerning Doke11, 22, and 21, show that duplication or permutation of the conserved segments led to some minor but significant changes in the secondary structure of the dockerin domain. These small changes concern the yield of α-helix residues, which, according to the deconvolution methods used, is at most 5% lower than that of wild type. The overall secondary structure of these mutants, however, can be considered to be very close to that of the wild type. This is supported by their strong affinity for the cohesin domain, which remains within the range of that observed for the wild-type dockerin. In contrast, the 600-fold lower affinity of DokeB for cohesin1A is best explained by the large variations observed in the CD profile as compared to DokeA. The analysis of the spectrum indicates a decrease in α-helix content to approximately 25% of the residues. In our opinion, this reflects considerable modification in the secondary structure of the first segment, or even an absence of structure since the percentage of random coil residues increases to 35% (versus 20−25% for the other dockerins). As suggested above, the nonconserved upstream residues, which are lacking in DokeB, therefore appear necessary to allow the first segment to adopt a correct folding.

In conclusion, it was shown that both cohesin and dockerin are functionally independent domains. As suggested by primary sequence analysis, each conserved segment of the dockerin domain can bind calcium. Two segments of the dockerin are required for an interaction with the cohesin to occur, but the kind of segment, first or second, is not important in the complexation with the cohesin partner. It was also demonstrated that the free dockerin has a well-defined α-helix-rich structure and that filling or emptying the calcium binding sites affects the secondary structure of this domain very little.

ACKNOWLEDGMENT

We are grateful to Drs. C. Cambillau and E. Loret for their helpful discussions concerning the structural data. We are indebted to Drs. C. Tardif and C. Gaudin for helpful discussions and critical evaluation of the manuscript. O. Valette is thanked for expert technical assistance. We thank Dr. M. T. Giudici-Orticoni for her valuable help with the use of the BIAcore apparatus. J. Bonicel is thanked for MALDI-TOF and N-terminal microsequencing, and N. Zylber for performing amino acid compositions. We are also grateful to P. Sauve for providing the oligonucleotides used in this study and to Dr. M. Johnson for correcting the English manuscript.

REFERENCES

1. Gal, L., Pagès, S., Gaudin, C., Bélaïch, A., Reverbel-Leroy, C., Tardif, C., and Bélaïch, J.-P. (1997) *Appl. Environ. Microbiol. 63*, 903−909.
2. Bayer, E. A., Setter, E., and Lamed, R. (1985) *J. Bacteriol. 163*, 552−559.
3. Béguin, P., and Lemaire, M. (1996) *Crit. Rev. Biochem. Mol. Biol. 31*, 201−236.
4. Shoseyov, O., and Doi, R. H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3483−3487.
5. Pagès, S., Bélaïch, A., Tardif, C., Reverbel-Leroy, C., Gaudin, C., and Bélaïch, J.-P. (1996) *J. Bacteriol. 178*, 2279−2286.
6. Salamitou, S., Raynaud, O., Lemaire, M., Coughlan, M., Béguin, P.; and Aubert, J.-P. (1994) *J. Bacteriol. 176*, 2822−2827.
7. Takagi, M., Hashida, S., Goldstein, M. A., and Doi, R. H. (1993) *J. Bacteriol. 175*, 7119−7122.
8. Pagès, S., Gal, L., Bélaïch, A., Gaudin, C., Tardif, C., and Bélaïch, J.-P. (1997) *J. Bacteriol. 179*, 2810−2816.
9. Goldstein, M. A., Takagi, M., Hashida, S., Shoseyov, O., Doi, R. H., and Segel, I. R. (1993) *J. Bacteriol. 175*, 5762−5768.
10. Morag, E., Lapidot, A., Govorko, D., Lamed, R., Wilchek, M., Bayer, E. A., and Shoham, Y. (1995) *Appl. Environ. Microbiol. 61*, 1980−1986.
11. Bélaïch, J.-P., Tardif, C., Bélaïch, A., and Gaudin, C. (1997) *J. Biotechnol. 57*, 3−14.
12. Pagès, S., Bélaïch, A., Fierobe, H.-P., Tardif, C., Gaudin, C., and Bélaïch, J.-P. (1999) *J. Bacteriol. 181*, 1801−1810.
13. Pagès, S., Bélaïch, A., Bélaïch, J.-P., Morag, E., Lamed, R., Shoham, Y., and Bayer, E. A. (1997) *Proteins*: *Struct., Funct., Genet. 29*, 517−527.
14. Fierobe, H.-P., Gaudin, C., Bélaïch, A., Loutfi, M., Faure, E., Bagnara, C., Baty, D., and Bélaïch, J.-P. (1991) *J. Bacteriol. 173*, 7956−7962.
15. Fierobe, H.-P., Bagnara-Tardif, C., Gaudin, C., Guerlesquin, F., Sauve, P., Bélaïch, A., and Bélaïch, J.-P. (1993) *Eur. J. Biochem. 217*, 557−565.
16. Shima, S., Igarashi, Y., and Kodama, T. (1993) *Appl. Microbiol. Biotechnol. 38*, 750−754.
17. Reverbel-Leroy, C., Pagès, S., Bélaïch, A., Bélaïch, J.-P., and Tardif, C. (1997) *J. Bacteriol. 179*, 46−52.
18. Gal, L., Gaudin, C., Bélaïch, A., Pagès, S., Tardif, C., and Bélaïch, J.-P. (1997) *J. Bacteriol. 179*, 6595−6601.
19. Yaron, S., Morag, E., Bayer, E. A., Lamed, R., and Shoham, Y. (1995) *FEBS Lett. 360*, 121−124.
20. Tamaru, Y., Liu, C., Malburg, L., and Doi, R. H. (1999) in *Cellulose degradation* (Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S., and Kinura, T., Eds.) pp 488−494, Uni Publishers Co., Ltd.
21. Tavares, G. A., Béguin, P., and Alzari, P. M. (1997) *J. Mol. Biol. 273*, 701−713.
22. Shimon, L. J. W., Bayer, E. A., Morag, E., Lamed, R., Yaron, S., Shoham, Y., and Frolow, F. (1997) *Structure 5*, 381−390.
23. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J. 15*, 5739−5751.
24. Lytle, B., and Wu, J. H. D. (1998) *J. Bacteriol. 180*, 6581−6585.
25. Bélaïch, A., Fierobe, H.-P., Baty, D., Busetta, B., Bagnara-Tardif, C., Gaudin, C., and Bélaïch, J.-P. (1992) *J. Bacteriol. 174*, 4677−4682.
26. Bayer, E. A., and Wilchek, M. (1990) *Methods Enzymol. 184*, 138−160.
27. Choi, S. K., and Ljungdahl, L. G. (1996) *Biochemistry 35*, 4906−4910.
28. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) *Protein Eng. 6*, 383−390.
29. Chauvaux, S., Béguin, P., Aubert, J.-P., Bhat, K. M., Gow, L. A., Wood, T. M., and Bairoch, A. (1990) *Biochem. J. 265*, 261−265.
30. Banères, J.-L., Roquet, F., Green, M., LeCalvez, H., and Parello, J. (1998) *J. Biol. Chem. 273*, 24744−24753.
31. Kakiushi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., Karita, S., Sakka, K., and Ohmiya, K. (1998) *J. Bacteriol. 180*, 4303−4308.
32. Pohlschröder, M., Leschine, S. B., and Canale-Parola, E. (1994) *J. Bacteriol. 176*, 70−76.
33. Pohlschröder, M., Canale-Parola, E., and Leschine, S. B. (1995) *J. Bacteriol. 177*, 6625−6629.
34. Bayer, E. A., Morag, E., Lamed, R., Yaron, S., and Shoham, Y. (1997) in *Carbohydrases from Trichoderma reesei* and other microorganisms (Claeyssens, M., Nerinckx, W., and Piens, K., Eds.) pp 39−66, The Royal Society of Chemistry, London.
35. Tavares, G. A., Souchon, H., Guérin, D. M., Lascombe, M.-B., and Alzari, P. M. (1997) in *Carbohydrases from Trichoderma reesei* and other microorganisms (Claeyssens, M., Nerinckx, W., and Piens, K., Eds.) pp 174−181, Royal Society of Chemistry, London.
36. Béguin, P., Raynaud, O., Chaveroche, M.-K., Dridi, A., and Alzari, P. (1996) *Protein Sci. 5*, 1192−1194.
37. Franz, C., Durussel, I., Cox, J. A., Schäfer, B. W., and Heizmann, C. W. (1998) *J. Biol. Chem. 273*, 18826−18834.
38. Gribenko, A. V., and Makhatadze, G. I. (1998) *J. Mol. Biol. 283*, 679−694.

BI9911740