ScaC, an Adaptor Protein Carrying a Novel Cohesin That Expands the Dockerin-Binding Repertoire of the \textit{Ruminococcus flavefaciens} 17 Cellulosome

Marco T. Rincon, Jennifer C. Martin, Vincenzo Aurilia, Sheila I. McCrae, Garry J. Rucklidge, Martin D. Reid, Edward A. Bayer, Raphael Lamed, and Harry J. Flint

Microbial Genetics Group and Proteomics Unit, Rowett Research Institute, Aberdeen, United Kingdom; Institute of Protein Biochemistry, National Research Council, Naples, Italy; and Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, and Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv, Israel

Received 25 November 2003/Accepted 26 January 2004

A new gene, designated \textit{scaC} and encoding a protein carrying a single cohesin, was identified in the cellulolytic rumen anaerobe \textit{Ruminococcus flavefaciens} 17 as part of a gene cluster that also codes for the cellulosome structural components \textit{ScaA} and \textit{ScaB}. Phylogenetic analysis showed that the sequence of the \textit{scaC} cohesin is distinct from the sequences of other cohesins, including the sequences of \textit{R. flavefaciens} \textit{ScaA} and \textit{ScaB}. The \textit{scaC} gene product also includes at its C terminus a dockerin module that closely resembles those found in \textit{R. flavefaciens} enzymes that bind to the cohesins of the primary \textit{ScaA} scaffoldin. The putative cohesin domain and the C-terminal dockerin module were cloned and overexpressed in \textit{Escherichia coli} as His\textsubscript{6}-tagged products (\textit{ScaC-Coh} and \textit{ScaC-Doc}, respectively). Affinity probing of protein extracts of \textit{R. flavefaciens} 17 separated in one-dimensional and two-dimensional gels with recombinant cohesins from \textit{ScaC} and \textit{ScaA} revealed that two distinct subsets of native proteins interact with \textit{ScaC-Coh} and \textit{ScaA-Coh}. Furthermore, \textit{ScaC-Coh} failed to interact with the recombinant dockerin module from the enzyme \textit{EndB} that is recognized by \textit{ScaA} cohesins. On the other hand, \textit{ScaC-Doc} was shown to interact specifically with the recombinant cohesin domain from \textit{ScaA}, and the \textit{ScaA-Coh} probe was shown to interact with a native 29-kDa protein spot identified as \textit{ScaC} by matrix-assisted laser desorption ionization—time of flight mass spectrometry. These results suggest that \textit{ScaC} plays the role of an adaptor scaffoldin that is bound to \textit{ScaA} via the \textit{ScaC} dockerin module, which, via the distinctive \textit{ScaC} cohesin, expands the range of proteins that can bind to the \textit{ScaA}-based enzyme complex.

Since the first description of cellulosomes in the anaerobic thermophile \textit{Clostridium thermocellum} (3, 5), enzyme complexes responsible for plant cell wall breakdown have been found in many species of cellulolytic gram-positive bacteria (2). In general, different species of clostridia have fairly minor variations of the paradigm established with \textit{C. thermocellum}, in which a noncatalytic scaffold protein (the primary scaffoldin), which carries multiple cohesin domains and a cellulose-binding module, provides the framework for assembling a range of different enzymes. Enzyme subunits belonging to the cellulosome each carry a single type of dockerin that pairs specifically with the cohesin domains. Attachment of the complex to the bacterial cell surface occurs either via hydrophilic domains in the scaffolding proteins, as in mesophilic clostridia (6, 15, 18, 31), or via a C-terminal dockerin that links the primary scaffoldin to one of several cell wall anchoring proteins (anchoring scaffoldins) (16) via a second type of cohesin-dockerin specificity, as in \textit{C. thermocellum} (4, 21, 22).

Investigations with organisms that do not belong to the genus \textit{Clostridium} are, however, now revealing some potentially more fundamental variations in the organization of cellulose complexes. For example, the cellulosomal system of \textit{Bacteroides cellulosolvens} has a two-component scaffoldin arrangement similar to that of \textit{C. thermocellum}, except that the types of cohesins carried by the primary and anchoring scaffoldins are reversed (13, 33). In \textit{Acetivibrio cellulolyticus}, the number of enzymes incorporated into the cellulosome complex is amplified by the involvement of a multi-cohesin-bearing adaptor scaffoldin which mediates between the primary and anchoring scaffoldins (12, 34). \textit{Ruminococcus flavefaciens} is a major cellulolytic species in the rumen of herbivores (19, 25). This bacterium produces a complex in which many plant cell wall-degrading enzymes are attached to a structural protein, \textit{ScaA}, via specific dockerin interactions with the three conserved cohesin repeats of \textit{ScaA} (28). \textit{ScaA}, however, lacks an identifiable cellulose-binding module (CBM), although a CBM was identified in the cellulase \textit{EndB} (29). In addition, previous studies revealed at least two distinct classes of dockerins among enzymes from \textit{R. flavefaciens} 17 (1), and only one of these groups of dockerins, found in the \textit{EndB}, \textit{EndA}, \textit{XynB}, and \textit{XynD} enzymes, interacts with \textit{ScaA} (28). \textit{ScaA} interacts via its C-terminal dockerin with a larger protein, \textit{ScaB}, which contains seven distinctive cohesin repeats and is believed to mediate the anchoring of the cellulosome complex to the cell surface via an as-yet-unknown mechanism (14, 28). The \textit{scaA} and \textit{scaB} genes are located adjacent to each other in the chromosome.
In our quest for novel cohesin-dockerin interactions among proteins expressed by *R. flavefaciens* 17, we found a new structural protein encoded by a gene upstream of the *scaA* gene. This protein, designated ScaC, contains a unique cohesin domain with a binding specificity distinct from that of ScaA and ScaB cohesins, as well as a dockerin module that interacts with ScaA. We propose a role for the newly characterized ScaC protein as an adaptor molecule that greatly expands the repertoire of proteins able to participate in the *R. flavefaciens* cellulosome.

### MATERIALS AND METHODS

#### Strains and growth conditions. *R. flavefaciens* 17 was grown anaerobically (8) either in modified M2 medium (26) or in Hungate-Stack medium (17) with 0.4% microcrystalline cellulose (Avicel PH101; Honeywell & Stein, London, United Kingdom), 0.4% birchwood xylan (Sigma, St. Louis, Mo.), or 0.4% cellobiose (Sigma) as an energy source, as described previously (28). *E. coli* Solopack Gold XL10 and BL21(DE3) (supplied by Stratagene, La Jolla, Calif.) were used as hosts for transformation and protein expression, respectively, with constructs made in pET30 Ek/LIC (Novagen, Madison, Wis.). *E. coli* strains were routinely grown on Luria-Bertani medium with appropriate antibiotic selection.

#### Sequence of *scaC*. The *scaC* coding sequence was obtained from PCR extension of the *scaA* gene with the primers listed in Table 1 (Fig. 1). The reverse ScaAcoh9r and M13 forward (specific to the vector sequence) primers yielded a 1.8-kb DNA fragment from a pUC18 library of *R. flavefaciens* 17 chromosomal DNA partially digested with Sau3AI (28). Sequencing of the 1.8-kb DNA fragment was then purified and concentrated with a QIAquick PCR purification kit (Qiagen). Ligation was carried out by using the pET30 hardvector as a template vector by the manufacturer, and the resulting recombinant plasmid was transformed in *E. coli* SoloPack Gold XL10 chemically competent cells, plated on Luria-Bertani agar supplemented with 30 μg of kanamycin ml⁻¹ and incubated at 37°C for 16 h.

Double His₆-S-tagged proteins from pET30 Ek/LIC constructs were overexpressed following transformation in *E. coli* BL21(DE3). Cells were recovered and lysed by sonication (Soniprep; Sanyo) after growth at 37°C on Luria-Bertani medium supplemented with 1% (vol/vol) glycerol, 1% (wt/vol) glucose, and 30 μg of kanamycin ml⁻¹. Induction was carried out by adding 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside) to a culture at an optical density of 600 nm of 0.8 to 1.0 and then incubating the culture either at 37°C for 4 h or at 16°C for 16 h in a refrigerated incubator. Protein constructs were purified by nickel affinity chromatography as described previously (14).

#### Cell surface-associated protein extraction from *R. flavefaciens*. Protein extracts used in this study were prepared by the method of Devilard et al. (11), with some modifications. Briefly, bacterial cell pellets were harvested from a 200-ml culture of *R. flavefaciens* at the stationary phase by centrifugation at 10,000 × g for 10 min at 4°C. The pellet was washed twice with 200 ml of TBS buffer (25 mM Tris-HCl buffer [pH 7.0], 150 mM NaCl) and resuspended in 10 ml of TBS buffer containing 2% (wt/vol) skim milk, 0.05% (vol/vol) Tween 20, and 1 mM CaCl₂). Incubation with the pET30 Ek/LIC vector as a stop codon for the cloning of internal domains. Residues in boldface type were added in addition to the ligation-independent cloning sequence as a stop codon for the cloning of internal domains.

#### Western blotting and detection of cohesion-dockerin interaction. Protein extracts were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) by the Laemmli method (20). After electrophoresis, proteins were blotted onto a polyvinylidene difluoride (PVDF) membrane (Immobilon-P; Millipore, Bedford, United Kingdom) following the manufacturer’s instructions and using a semidry blotting system (Biometra, Göttingen, Germany) at 0.8 mA/cm². Membranes were blocked for 1 h at room temperature by rocking in 50 ml of blocking buffer (TBS buffer [pH 7.0] containing 3% [wt/vol] skim milk, 0.05% [vol/vol] Tween 20, and 1 mM CaCl₂). Incubation with the His₆-S-tagged protein probes was performed in the same blocking buffer at 4°C with rocking overnight. The following day the membranes were washed three times in CaTTBS buffer (TBS buffer with 1% [wt/vol] skim milk, 0.05% [vol/vol] Tween 20 and 1 mM CaCl₂) and then incubated for 1 h with a 1:5,000 dilution of nickel-activated horseradish peroxidase (HRP) (India HisProbe HRP; Pierce, Rockford, Ill.) in CaTTBS buffer containing 1% skim milk. The bands were visualized by enhanced chemiluminescence (Supersignal West Pico chemiluminescent substrate; Pierce) by following the manufacturer’s instructions after the membranes were washed five times at 5-min intervals in CaTTBS buffer. Spot tests were done following the manufacturer’s instructions after the membranes were washed five times at 5-min intervals in CaTTBS buffer.
carried out in the same way as the Western blotting analysis on a nylon membrane spotted with 1 μl of His$_{6}$-tagged recombinant proteins that were previously purified (14, 28, 29) and probed either with an HRP-conjugated S-protein (Novagen) or biotinylated His$_{6}$-tagged ScaC-Coh recombinant protein. Sample preparation for two-dimensional (2D) gel electrophoresis. A total of 300 μg of protein extracts was precipitated by using Perfect-FOCUS (Geno Technology, St. Louis, Mo.) as recommended by the manufacturer. Each protein sample was suspended in a solution containing 9 M urea, 4% 3-(3-cholamidopropyl)-dimethylammonio)-1-propanesulfonate (CHAPS), 0.5% BioLyte (pH 3 to 10; Bio-Rad), and 3% DTT and was homogenized gently with a plastic pestle. The sample was sonicated twice for 2 min with cooling for 5 min between the cycles and was allowed to sit for 20 min to facilitate protein solubilization. Finally, the sample was cleared of debris by centrifugation at 14,000 × g before 340 μl (total volume) was loaded into the sample tray. Immobilized pH gradient (IPG) strips were carefully placed face down on the sample so that there were no trapped air bubbles beneath the strip.

Isoelectric focusing. Bio-Rad IPG strips (pH 3 to 6) were used for separation of proteins in the first dimension. Rehydration was performed at 20°C for 1 h without an applied voltage. After 1 h, each lane was overlaid with mineral oil to prevent the strips from drying out. The rehydration procedure was switched to the active mode for a further 16 h (50 V/strip). A wick wetted with 10 μl of H$_2$O and another wick wetted with 10 μl of 3.5% DTT were placed at the anode and cathode, respectively, in a fresh focusing tray before an IPG strip was transferred from the rehydration tray. The strip was overlaid with mineral oil, and the initial start-up and ramping protocol used was the protocol described in the Bio-Rad manual; after this the strips were subjected to a total of 60 kV h. After 1 h of focusing the strips were moved to a tray containing fresh wicks and overlaid with mineral oil, and the focusing was continued.

Reduction and alklylation of IPG strips. After focusing, the strips were removed from the focusing tray and were placed face up and equilibrated in a solution containing 6 M urea, 2% SDS, 3.075 M Tris-HCl (pH 8.8), 20% glycerol, and 130 mM DTT for 10 min, after which each strip was drained and placed into a solution containing 0.375 M Tris-HCl (pH 8.8), 20% glycerol, and 135 mM iodoacetamide for 10 min.

Second dimension (SDS-PAGE). The second-dimension procedure was performed by using an IsoDalt tank apparatus (Hoeffer Scientific) essentially as described by Laemmli (20), with the following modifications. The strips were rinsed in Laemmli electrode buffer, drained, and placed on top of an 8 to 16% SDS-polyacrylamide gradient gel (18 by 18 cm), and each strip was sealed in position with 2% agarose in Laemmli electrode buffer. Standard M$_{r}$ proteins (Bio-Rad Protein Plus standards) were loaded onto paper wicks (5 by 5 mm), dried, and loaded on top of the gel alongside the IPG strips, and each standard strip was also sealed in position with agarose. The gels (up to 10 gels per run) were electrophoresed at 200 V for 9.5 h before removal from the cassettes and staining.

Colloidal Coomassie blue staining. Fixation and staining steps were performed in 200-ml solutions in Nuncion plastic dishes (23 by 23 cm; Fisher Scientific). The gels were fixed in 50% ethanol−2% p-aminophenolic acid for at least 3 h, after which they were rigorously washed in distilled H$_2$O for at least 1 h. The gels were placed in 34% methanol−2% p-aminophenolic acid containing 17% (NH$_4$)$_2$SO$_4$ and 0.05% Coomassie blue G (Sigma) for 3 days, after which they were rinsed thoroughly in distilled H$_2$O to remove adherent particulate Coomassie blue before imaging. The gels were imaged with a Bio-Rad GS-710 scanner and were analyzed by using the Bio-Rad PD Quest image analysis software.

Spot cutting and acquisition of MALDI spectra. Spots of interest were excised from a gel manually, placed in a 96-well V-bottom plate, and trypsinized by using a MALDI-TOF spectrometer. Tryptic peptides were spotted onto a matrix-assisted laser desorption ionization (MALDI) target plate (Teflon 2 × 96 plate; Applied Biosystems), mixed with recrystallized α-cyano-4-hydroxycinnamic acid in aceto-nitrile-water-trifluoroacetic acid (70:30:0.1), and dried before acquisition of spectra with an Applied Biosystems Voyager DE MALDI-time of flight (TOF) instrument; the machine was calibrated with a peptide standard for capillary electrophoresis (Sigma).

Phylogenetic analysis. Phylogenetic trees and a sequencing alignment were generated by using the ClustalW program (http://www2.ebi.ac.uk/clustalw/). Dockerin sequences were obtained from the GenBank website (http://www.ncbi.nlm.nih.gov/) or the Carbohydrate-Active Enzymes server (CAZy website; http://afmb.cnmrs-mrs.fr/~pedro/CAZy/dh.html) designed by Coutinho and Henrissat (7, 10). Sources of the cohesins and enzyme-borne dockerin sequences used in this work have been described previously (28, 33, 34).

Miscellaneous. Protein concentrations were determined by the Bradford method. Biotinylation of the His$_{6}$-tagged ScaC-Coh recombinant protein for the spot test was carried out as described previously (28).

Nucleotide sequence accession numbers. The DNA sequence of scaC and a partial sequence of synX have been deposited in the EMBL nucleotide sequence database under accession numbers AJ198075 and AJ1986790, respectively.

RESULTS

Sequencing and domain structure of scaC. scaC was discovered by genome walking upstream of the previously reported scaA and scaB genes (Fig. 1A) (14, 28). The scaC start codon (ATG) is preceded by a Shine-Delgarno sequence (AGAAA GGAGG) (Fig. 1C). The intergenic spacer region between scaC and scaA is only 18 bp long and lacks any obvious rho-independent hairpin-like secondary structure, suggesting that the two genes might be cotranscribed.

scaC encodes a 308-amino-acid protein with an estimated molecular mass of 31.8 kDa (Fig. 1C). Sequencing analysis of the gene product of scaC with the program SignalP (http://www.cbs.dtu.dk/services/SignalP/) resulted in prediction of a typical prokaryote gram-positive signal-sorting peptide sequence consisting of 29 amino acids with the estimated cleavage site at AA-GE. A 160-amino-acid region follows the sorting signal, and sequence analysis with the SIB BLAST network service (http://ca.expasy.org/tools/blast/) revealed significant homology to cohesin domains from scaffolding and anchoring proteins previously identified in C. thermocellum (CipA, CipB, AncA), Clostridium cellulovorans (CbpA), Clostridium acetobutylicum, Clostridium cellulolyticum (CipC, CipCCA), Clostridium josui (CipA), B. cellulosolvens (CipBC), A. cellulolyticus (CipV), and R. flavefaciens (ScaA, ScaB). Furthermore, a search analysis of the Pfam database (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00963) also confirmed that the N-terminal 160-residue region of ScaC is probably a cohesin domain. Phylogenetic analysis (Fig. 2A) and multiple-sequence alignment (Fig. 3) revealed that the ScaC sequence is novel. Like cohesins from A. cellulolyticus ScaC (34), the cohesin from R. flavefaciens ScaC does not fall into any of the three currently recognized groups of bacterial cohesins (Fig. 2A).

At the C terminus of ScaC there are two recognizable 22- to 2579 NOVEL CELLULOSOMAL ADAPTOR IN R. FLAVEFACIENS

FIG. 1. Diagrams showing the disposition on the genome, the domain organization, and the sequence of the R. flavefaciens scaC gene. (A) Gene organization of the structural scaffoldin proteins in R. flavefaciens 17. (B) Sequence strategy and domain architecture. scaC was discovered to be upstream of scaA by PCR walking extension with primers ScaAcoh5′r (9r) and M13f (see Table 1 and Materials and Methods for details) from a pUC18 R. flavefaciens 17 plasmid library. The sequence was completed by using the additional internal primers ScaAcoh11r (11r), ScaAcoh10f (10f), and ScaAcoh10r (10r). Domains within ScaC are indicated. The solid box indicates a linker sequence, and the striped boxes indicate N-terminal signal peptides (SP). The positions of the ScaC cohesin and dockerin are indicated. (C) Nucleotide sequence and sequence of the gene product (top) of scaC. Shine-Dalgarino sequences (SD) and the predicted cleavage site that removes the signal peptide (▿) are indicated. The F-hand repeat motifs typical of the Ca$^{2+}$-binding loop of dockerin domains are indicated by open boxes. The putative cohesin domain is indicated by a solid box. The initial sequence coding for ScaA is indicated by a striped box.
(A) Cohesin Tree

(A) Cohesin Tree

(B) Dockerin Tree

(B) Dockerin Tree
24-residue F-hand variations of the EF-hand motif that are characteristic of loops involved in the binding of calcium (Fig. 1C), which is typical of dockerin modules (9, 27). Phylogenetic analysis (Fig. 2B) placed the ScaC sequence on a branch of the dockerin tree together with the EndB-like dockerins of *R. flavefaciens* 17 glycoside hydrolases EndA, EndB, XynB, and XynD, which have previously been shown to interact with the ScaA type III cohesins (14, 28). Moreover, the predicted recognition residues (23, 24, 27) of the ScaC dockerin are remarkably similar to those of the EndB-like enzymes. On this basis

**FIG. 2.** Relationship of *R. flavefaciens* ScaC-borne cohesin and dockerin domains to previously described cellulosomal domains. (A) Phylogenetic analysis of the ScaC cohesin relative to the known type I, II, and III cohesins. The ScaC cohesin maps on a separate branch, distinct from all other cohesins. The ScaC cohesin emanates from the central branch close to the branching point of the three *A. cellulolyticus* ScaC cohesins and approximately equidistant from the type I cohesins and the point of detection that separates the type II cohesins from the type III cohesins. (B) Phylogenetic analysis of the ScaC dockerin relative to other enzyme and scaffoldin-borne dockerins. The ScaC dockerin (solid square) maps among the EndB-like dockerins from *R. flavefaciens* enzymes (solid circles) but distinct from the dockerins of CesA, XynE, and XynX. The other known *R. flavefaciens* scaffoldin-borne dockerin (ScaA) and the dockerins of *C. thermocellum* (Cltom), *A. cellulolyticus* (Acece), and *B. cellulosolvens* (Bacce) are indicated by open squares. Other dockerin-borne enzymes include enzymes from *Ruminococcus albus* (Rumal-EgV, Rumal-EgVI, Rumal-EgVII) and *Ruminococcus* sp. (Rumsp-Xyn1) and a selection of enzymes from *C. thermocellum* and mesophilic clostridia (*C. cellulolyticum*, *C. cellulovorans*, and *C. josui*). For a list of the proteins and their accession numbers, see references 28, 33, and 34. Scale bars = 0.1% amino acid substitutions.

**FIG. 3.** Multiple-sequence alignment of the cohesin domain from *R. flavefaciens* ScaC with cohesin domains from other cellulolytic bacteria. RUMFL-ScaC, *R. flavefaciens* ScaC (AJ585075); RUMFL-ScaB, *R. flavefaciens* ScaB (accession no. tr:Q9AE52); RUMFL-ScaA, *R. flavefaciens* ScaA (tr:Q9AE53); ACECE-ScaC, *A. cellulolyticus* ScaC (tr:Q7WYN2); CLOTH-CipA, *C. thermocellum* CipA (sp:Q06851); ACECE-CipV, *A. cellulolyticus* CipV (tr:Q9RPLO); CLOCE-CipC, *C. cellulolyticum* CipC (sp:Q45996); CLOJO-CipA, *C. josui* CipA (sp:O82830); CLOCL-CbpA, *C. cellulovorans* CbpA (sp:P38058); CLOAC-0910, *C. acetobutylicum* 0910 (tr:Q977Y4). Amino acids that are conserved in sequences are highlighted. The alignment was constructed by using ClustalW (http://www2.ebi.ac.uk/clustalw/) and was edited by using GeneDoc (http://www.psc.edu/biomed/genedoc).
proteins were overexpressed in *E. coli* (see Materials and Methods for details). The recombinant product was purified by nickel affinity chromatography and was separated by SDS-PAGE. Proteins were stained with Coomassie blue. Lane M contained a protein molecular size marker. The predicted molecular masses of the protein constructs are 16.7, 23.2, and 33.9 kDa for His6-ScaC-Doc, His6-ScaC-Coh, and His6-ScaC-mat, respectively. The faint bands above the protein constructs appear to be products of dimerization. The numbers on the left indicate the positions of the protein molecular mass markers (in kilodaltons).

we predicted that ScaC dockerin should interact specifically with cohesin domains from ScaA.

**Cloning and Western blot analysis of recombinant protein from ScaC.** The mature peptide of ScaC (full-length sequence excluding the putative signal peptide), as well as the cohesin domain and the dockerin module, were all cloned as double His6-S-tagged proteins. Recombinant ScaC-mat and ScaC-Coh cohesin were overexpressed at 37°C after IPTG induction. The protein yield obtained from the soluble fraction after Ni-nitrilotriacetic acid affinity purification (Fig. 4) revealed that both recombinant proteins are highly soluble and fairly stable under the expression conditions used. In contrast, the ScaC dockerin was overexpressed at 16°C. The latter conditions were chosen because of the low yields obtained at 37°C, apparently due to the formation of inclusion bodies when the protein was expressed at 37°C.

The purified His6-S-tagged ScaC-Coh recombinant protein was used as a probe for a cell surface-associated protein fraction of *R. flavefaciens* 17 grown on birchwood xylan or crystalline cellulose (Avicel) following one-dimensional SDS-PAGE separation and Western blotting. This approach revealed positive interactions with several proteins whose sizes ranged from 45 to 90 kDa (Fig. 5). For comparison, the same protein fraction was probed with His6-tagged ScaA-Coh. ScaA-Coh recognized a different set of proteins in xylan-grown cells than in cellulose-grown cells, but in both cases the proteins differed from the proteins recognized by ScaC-Coh. On average, the polypeptides recognized by ScaA-Coh were larger than those recognized by ScaC-Coh (Fig. 5). To further ascertain the difference in protein recognition between cohesins from ScaC and ScaA, a similar Western blot analysis was carried out with the cell surface-associated protein fraction of *R. flavefaciens* 17 grown on birchwood xylan after separation by 2D SDS-PAGE (Fig. 6A). One protein preparation was divided into three aliquots, and each aliquot was resolved by 2D gel electrophoresis. One gel was subjected to Coomassie blue staining in order to reveal the total number of spots. Proteins from the two other gels were transferred to a blotting membrane; the proteins from one of these gels were subsequently probed with ScaA-Coh, and the proteins from the other gel were probed with ScaC-Coh. The results revealed ~20 positive spots that interacted with ScaA-Coh with a pI range of 4 to 5 and a molecular weight range consistent with the one-dimensional Western blot analysis (Fig. 5). ScaC cohesin, on the other hand, was able to recognize 11 different spots with a much wider pI range and an average molecular mass of 37 to 90 kDa (Fig. 6C), as predicted from the one-dimensional Western blot analysis (Fig. 5).

Among the *R. flavefaciens* native proteins recognized by His6-tagged ScaA-Coh was a 29-kDa protein, which is consistent with the estimated molecular mass of ScaC. Although not shown in Fig. 5, this protein was clearly one of the prominent spots in the 2D Western blot analysis (Fig. 6B). A protein profile analysis of the 29-kDa protein spot recovered from 2D gel separation (Fig. 6A) was carried out by trypsinization followed by MALDI-TOF analysis, and the resulting peptide fingerprint was compared with the theoretical trypsinized pattern for ScaC generated by using the computer-assisted program.
A total of seven masses analyzed as specific enzymatic cleavage products of ScaC, which together contained 139 amino acids (45% of the protein sequence), were found to match the theoretical pattern of ScaC. Similarly, MALDI-TOF analysis of the cell-associated protein fraction shown in Fig. 6A also revealed the presence of other components of the \textit{R. flavefaciens} cellulosome, such as the structural proteins ScaA (predicted pI 4.42; molecular mass, 87.7 kDa) and the enzyme EndB (pI 4.60; molecular mass, 84.4 kDa). Interestingly, in the 2D Western blot analysis the latter protein was one of the major spots interacting with ScaA-Coh. Although the 2D profile of these proteins was not consistent with the theoretical molecular weight, the difference in the case of ScaA can be attributed to the presence of covalently linked glycosyl groups (14, 28). A string of spots having the same molecular size but slightly different pIs is commonly observed and is likely to be due to posttranslational modifications that affect charged groups (32).

**Spot test analysis.** Interactions of the ScaC dockerin and cohesin were investigated further by using His$_6$-tagged recombinant proteins obtained from cloned \textit{R. flavefaciens} 17 genes available from previous work (14, 28). Recombinant proteins were spotted onto nylon membranes and incubated with His$_6$-S-tagged ScaC-Doc or biotinylated ScaC-Coh. The ScaC recombinant proteins used as probes were cloned into the pET30 Ek/LIC system so that they had an S tag in addition to the His$_6$ tag. The S tag could therefore be used as a differential label in the spot test with an HRP-conjugated S-protein as a second label. However, some of the other protein constructs used in this work were also cloned by using the same pET30 Ek/LIC system, and the expressed proteins thus had both tags. In such cases, the recombinant protein probe was biotinylated in order to provide a distinctive tracking system for Western blot analysis. The results revealed that the His$_6$-S-tagged ScaC dockerin interacted specifically with His$_6$-tagged ScaA-Coh but failed to interact with recombinant His$_6$-tagged cohesins from ScaB.

![FIG. 6. 2D gel electrophoresis of cell surface-associated protein fraction of \textit{R. flavefaciens} 17 grown on birchwood xylan. (A) Colloidal Coomassie blue-stained gel. The positions of the enzyme EndB and the structural cellulosomal proteins ScaA and ScaC are indicated. The numbers at bottom indicate the pH gradient of the first separation. The numbers on the right indicate the positions of the protein markers for the second separation. (B) 2D Western blot analysis performed with a similar gel after blotting onto a PVDF membrane and probing with His$_6$-ScaA-Coh (28). The positions of EndB and ScaC are indicated. (C) 2D Western blot analysis performed with a similar gel after the protein spots were blotted onto a PVDF membrane and probed with His$_6$-ScaC-Coh. Cohesin-dockerin interactions were exposed as described in the legend to Fig. 5.](http://jb.asm.org/2004/186/2583/fig6.png)
Furthermore, there was no detectable interaction between biotinylated ScaC-Coh and either of the two previously characterized types of dockerins derived from known *R. flavefaciens* proteins, including EndB, XynB, XynX, XynE, CesA, and ScaA. (Fig. 7)

**DISCUSSION**

The results of the present study established that a newly identified protein, ScaC, interacts with the primary ScaA scaffoldin via its C-terminal dockerin module. ScaC is thus the first noncatalytic protein in *R. flavefaciens* that has been shown to pair with ScaA cohesins, which otherwise recognize dockerins borne by the enzymes EndA, EndB, XynD, and XynB. The unique cohesin domain of ScaC appears to mediate interactions with a range of native *R. flavefaciens* 17 proteins that exhibit no overlap with the subset of proteins that is recognized by the ScaA cohesins. This conclusion is supported by the failure of the ScaC cohesin to recognize representative recombinant dockerins from the enzymes EndB and XynB or from the scaffolding protein ScaA. It appears, therefore, that ScaC acts as a type of adaptor, allowing many proteins that are not recognized directly by the ScaA cohesins to be bridged into the cellulosome assembly.

We assume that the subset of proteins that is recognized by the ScaC cohesin must have a distinct type of dockerin domain that has yet to be identified. In previous work it has been found that the dockerins present in the enzymes CesA and XynE, whose sequences differ from the sequences of the dockerins in EndA, EndB, XynB, and XynD (1), failed to recognize ScaA cohesins (28). However, we found no evidence in this study that these dockerins were recognized by ScaC. This argues that there should be at least the following four distinct cohesin-dockerin binding specificities in *R. flavefaciens* 17: (i) between the ScaB cohesins and ScaA-Doc; (ii) between the ScaA cohesins and the EndB-type enzyme dockerins; (iii) between CesA-XynE dockerins and an unknown cohesin(s); and (iv) between ScaC-Coh and an unknown type of dockerin.

The role proposed here for ScaC further emphasizes the differences in cellulosome organization between *R. flavefaciens* and cellulosolytic *Clostridium* spp. In the cellulosome model described for *C. thermocellum* the catalytic subunits are organized along the primary scaffoldin protein, CipA. CipA in turn interacts with at least three other anchoring scaffoldins (SdbA, OlpA, and OlpB) (4, 30) through its dockerin domain. Moreover, CipA possesses a family 3a CBM with high affinity toward cellulose fibrils and is thought to mediate the specific interactions of the bacterium with its substrate. A similar overall arrangement, with distinctive variations on the theme, exists in the cellulosomes systems of *A. cellulolyticus* and *B. cellulolovorans* (12, 13, 33, 34). In contrast, in *R. flavefaciens* the catalytic subunits are organized in a different way. The dockerin of the primary scaffoldin ScaA interacts with any of the seven cohesin repeats in the putative anchoring scaffoldin ScaB, and certain (EndB-like) catalytic subunits are then able to interact with the three cohesin repeats present in ScaA. ScaC, however, introduces another dimension into the interaction with ScaA, whereby an additional subset of proteins, which we assume possess distinctive dockerin modules, are recruited into the cellulosome. The single ScaC cohesin could thus have a true adaptor function, which is distinct from the proposed adaptor function described for *A. cellulolyticus* cellulosomal ScaB (34). In the latter case, the major function of the ScaB scaffoldin seems to be to amplify the number of enzymes that can be incorporated into the *A. cellulolyticus* ScaC anchoring scaffoldin (34).

The functional significance of the *R. flavefaciens* ScaC adaptor mechanism has yet to be fully established. If production of the ScaC protein were subject to regulation independent of ScaA, then it might provide an effective mechanism for regulating the subunit composition of the cellulosome. We can also speculate that the additional cohesin specificity provided by ScaC could allow the incorporation of plant cell wall-degrading enzymes acquired by horizontal gene transfer from phylogenetically distant organisms into the *R. flavefaciens* cellulosome.

One interesting observation from this work is that ScaC-Coh seems to interact with proteins whose molecular sizes are smaller than those of the range of proteins recognized by ScaA-Coh. The enzymes so far known to interact with ScaA cohesins all have complex multimodular structures, but it seems likely, judging by their small molecular sizes, that proteins interacting with ScaC have a simpler modular organization. This might reflect their evolutionary origin (e.g., acquisition from different source organisms via horizontal transfer) or
might indicate entirely different roles (e.g., nonenzymatic) on the cell surface. The functions of the proteins that bind ScoA, including the question of whether they are in fact enzymes or are structural components, have yet to be established.

The combination of Western blot analysis and 2D gel electrophoresis proved to be a valuable tool for the study of protein interactions in the cellulosome organization of *R. flavefaciens* 17 cell wall-degrading enzymes. The main priority for future research on this system is to use proteomic analysis to identify proteins separated by 2D gel electrophoresis that can be pinpointed as structural components of the cellulosome by protein affinity analysis.

ACKNOWLEDGMENTS

R.R.I. is supported by the Scottish Executive Environment and Rural Affairs Department. M.T.R. is employed under the EU framework V program GEMINI. This research was supported by the Israel Science Foundation (grants 394/03 and 771/01), by the U.S.-Israel Binational Agricultural Research and Development Fund (BARD) research grant 3106–99C), and by a grant from the United States-Israel Binational Science Foundation, Jerusalem, Israel.

REFERENCES