Conservation and Divergence in Cellulosome Architecture between Two Strains of *Ruminococcus flavefaciens* 

Sadanari Jindou,1 Ilya Borovok,1 Marco T. Rincon,2 Harry J. Flint,2 Dionysios A. Antonopoulos,3 Margret E. Berg,3 Bryan A. White,3 Edward A. Bayer,4* and Raphael Lamed1

**Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv 69978, Israel;** 2Microbial Ecology Group, Rowett Research Institute, Bucksburn, Aberdeen AB21 9SB, United Kingdoms; 3Department of Animal Sciences, University of Illinois, Urbana, Illinois 61801; and Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

Received 4 July 2006/Accepted 10 September 2006

* Corresponding author. Mailing address: Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel. Phone: 972-8-934-2373. Fax: 972-8-946-8256. E-mail: ed.bayer@weizmann.ac.il.

**A 17-kb scaffoldin gene cluster in *Ruminococcus flavefaciens* strain FD-1 was compared with the homologous segment published for strain 17. Although the general design of the cluster is identical in the two strains, significant differences in the modular architecture of the scaffoldin proteins were discovered, implying strain-specific divergence in cellulosome organization.**

*Ruminococcus flavefaciens* is a gram-positive, anaerobic, cellulosome-producing, cellulytic bacterium which commonly inhabits the digestive tracts of ruminants, other herbivorous animals, and humans (11, 21). *R. flavefaciens* strain FD-1 and 17 are two commonly investigated strains of this species. Although the two strains were derived from different geographical locations and time frames (6, 12), they have been classified as the same species and, accordingly, show marked similarities in their properties.

The organization of cellulases into a multienzyme cellulosome complex is one of the major paradigms of efficient bacterial degradation of cellulose and related plant cell wall polysaccharides (5, 8, 10). The incorporation of the cellulosomal enzymes into scaffoldin components is the most defining feature of the cellulosome complex and is dictated by the high-affinity protein-protein interaction between two complementary modules, the cohesin and the dockerin. Knowledge of the types and specificities of cohesins and dockerins produced by a given bacterium and how they are arranged in the parent protein (i.e., scaffoldins and enzymes, respectively) provides insight into the overall architectural scheme of a given cellulosome system.

We have previously investigated the status of the cellulosome produced by strain 17 of *R. flavefaciens*. This was achieved by the sequencing of the genes encoding several enzymes (2, 13, 20) and scaffoldins (9, 18, 19) isolated from clone libraries. Individual cohesins and dockerins were subcloned and overexpressed, and the interactions among the expressed modules were evaluated biochemically. Using this approach, we proposed an architectural overview of the cellulosome in strain 17 (17).

Strains of *R. flavefaciens* from the rumen form a well-defined phylogenetic cluster based on 16S rRNA sequencing, but they show significant genetic variations (15). Significant functional diversity between strains, with respect to the breakdown of cellulose and of different types of plant material (7, 14) and the ability to utilize degradation products from xylans (3), has been reported. In the present report, we examine the architecture and component sequences of the cellulosome system in *R. flavefaciens* strain 17 and compare its characteristics with those of strain 17. The results indicate a general similarity in the cellulosome organization between the two strains, with a number of novel and very surprising properties.

The sequence of the *sca* gene cluster from *R. flavefaciens* strain 17 (accession number AJ278969) was compiled from several EMBL/GenBank entries with the following accession numbers: A1585075 (*scaC*), AJ278969 (*scaA-scaB*), AJ810898 (*cttA*), and AJ810899 (*scaE*). The corresponding gene cluster DNA sequence of strain FD-1 (accession number AM262974) was assembled from the draft genome (1) and PCR data, using the primers described in Table 1. As shown in Fig. 1A, the overall organization of the scaffoldin cluster of strain FD-1 matches well with that of strain 17 (17). The cluster consists of genes encoding four scaffoldins of different sizes and one additional gene, *cttA*, encoding a protein of currently unknown function.

The FD-1 *sca4* gene encodes a protein that contains two cohesins rather than three as in strain 17 (Fig. 1B). As a result, the FD-1 ScaA protein is markedly smaller in overall size than the orthologous protein of strain 17. The ScaA proteins in the two strains bear a similar type of N-terminal X module (53.9% sequence similarity) and similar dockerins (77.9% sequence similarity) at their C termini.

In the case of ScaB, however, the overall modular architecture is strikingly divergent for the two strains. ScaB of strain 17 carries seven contiguous cohesins, separated by linker sequences of various lengths. In contrast, the orthologous scaffoldin B of strain FD-1 bears nine cohesins. Both scaffoldins contain a characteristic X-module/dockerin pair (XDoc) at the C terminus. The integrity of the strain 17 ScaB XDoc has recently been shown to be critical for recognition by the ScaE cohesin (17). The differences between the two ScaB scaffoldins, however, are not only quantitative (i.e., in the respective number of cohesins) but also qualitative. Unlike the seven ScaB cohesins of strain 17, all of
which emanate from a single branch on the phylogenetic tree (9, 17), those of strain FD-1 are distinctively divergent: the first four N-terminal cohesins are related (46 to 78% sequence similarity) to those of ScaA (of both strains), whereas the five C-terminal cohesins are orthologous (53 to 79% sequence similarity) to the seven ScaB cohesins from strain 17 (Fig. 1B). Although both the ScaA and ScaB cohesins of strain 17 have been classified as type III cohesins, the first four N-terminal cohesins are related (46 to 78% sequence similarity) to the seven ScaB cohesins from strain 17, whereas ScaB cohesins 5 through 9 map together with all of the strain 17 ScaB cohesins (Fig. 1C). The FD-1 ScaB is but the second reported occurrence of the respective parent protein (17). The interstrain sequence due C-terminal LPXTG motifs for sortase-mediated anchoring of ScaE scaffoldins contain very similar (85.7% identity) 35-residue C-terminal dockerins (65.8% sequence similarity) and the C-terminal dockerins of strain FD-1 and amplified by PCR using appropriate primers (Table 1). Since all seven ScaB cohesins in strain 17 were shown previously (18) to bind selectively to ScaA, we employed in the present work selected cohesins from strain FD-1 to represent their conserved “paralogues.” Thus, FD-1 ScaA cohesin 1 represented cohesins 1 and 2 from this protein, ScaB cohesin 3 was representative of ScaB cohesins 1 to 4, and ScaB cohesin 6 represented ScaB cohesins 5 to 9. The resultant products were cloned into the pETCBMCoh cassette plasmid as described by Barak et al. (4), yielding C-terminally His-tagged cohesin modules, fused at their N termini with a CBM3a for improved display in the immobilized state. The DNA regions encoding the dockerin-bearing enzyme genes in strain FD-1 were amplified using an appropriate pair of primers (Table 1). The resultant PCR fragments were cloned into the pETXynDoc cassette (4), yielding the comparative architecture of the two R. flavefaciens strains. For this purpose, we employed a recently reported matching fusion-protein approach, combined with an enzyme-linked immunosorbent assay-like assay system (4). The gene fragments coding for five putative cohesins (two of them divergent cohesins from scaB) were selected from the four designated scaffolds of strain FD-1 and amplified by PCR using appropriate primers (Table 1). Since all seven ScaB cohesins in strain 17 were shown previously (18) to bind selectively to ScaA, we employed in the present work selected cohesins from strain FD-1 to represent their conserved “paralogues.” Thus, FD-1 ScaA cohesin 1 represented cohesins 1 and 2 from this protein, ScaB cohesin 3 was representative of ScaB cohesins 1 to 4, and ScaB cohesin 6 represented ScaB cohesins 5 to 9. The resultant products were cloned into the pETCBMCoh cassette plasmid as described by Barak et al. (4), yielding C-terminally His-tagged cohesin modules, fused at their N termini with a Clostridium thermocellum CBM3a for improved display in the immobilized state. The DNA regions encoding the dockerin modules derived from ScaA, ScaB, and ScaC of strain FD-1 were amplified using an appropriate pair of primers (Table 1). The resultant PCR fragments were cloned into the pETXynDoc cassette (4), yielding the desired His-tagged xylanase T6 fusion proteins bearing a dockerin at the C terminus. Orthologous dockerin-bearing enzyme genes in strain FD-1, encoding the equivalent of EndB (the family 44 enzyme Cel44A) and that of CesA-like dockerin (“Ce3B”-Doc) Homologous gene cloning

<table>
<thead>
<tr>
<th>Name</th>
<th>Nucleotide sequence</th>
<th>Location</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ScaA-Coh1-F</td>
<td>AGATCTTACAAGCTGCTAGTTGTTG</td>
<td>ScaA Coh1</td>
<td>Cohesin expression</td>
</tr>
<tr>
<td>ScaA-Coh1-R</td>
<td>CCTAGACGAGCTGCTAGTTGTTG</td>
<td>ScaB Coh3</td>
<td>Cohesin expression</td>
</tr>
<tr>
<td>ScaB-Coh3-F</td>
<td>GGATCCCAAGCTGCTAGTTGTTG</td>
<td>ScaB Coh6</td>
<td>Cohesin expression</td>
</tr>
<tr>
<td>ScaB-Coh3-R</td>
<td>GTACAGCTGCTAGTTGTTGTTG</td>
<td>ScaC cohesin</td>
<td>Cohesin expression</td>
</tr>
<tr>
<td>ScaC-Coh-F</td>
<td>AGATCTTACAAGCTGCTAGTTGTTG</td>
<td>ScaC dockerin</td>
<td>Dockerin expression</td>
</tr>
<tr>
<td>ScaC-Coh-R</td>
<td>CCTAGACGAGCTGCTAGTTGTTG</td>
<td>ScaC-XDoc</td>
<td>Dockerin expression</td>
</tr>
<tr>
<td>ScaE-Coh-F</td>
<td>AGATCTTACAAGCTGCTAGTTGTTG</td>
<td>ScaE dockerin</td>
<td>Cohesin expression</td>
</tr>
<tr>
<td>ScaE-Coh-R</td>
<td>CCTAGACGAGCTGCTAGTTGTTG</td>
<td>ScaE dockerin</td>
<td>Cohesin expression</td>
</tr>
<tr>
<td>ScaA-Doc-F</td>
<td>GGTACCTTACAAGCTGCTAGTTGTTG</td>
<td>ScaA dockerin</td>
<td>Dockerin expression</td>
</tr>
<tr>
<td>ScaA-Doc-R</td>
<td>GGATCCCAAGCTGCTAGTTGTTG</td>
<td>ScaA-like dockerin</td>
<td>Homologous gene cloning</td>
</tr>
<tr>
<td>ScaB-Doc-F</td>
<td>GGATCCCAAGCTGCTAGTTGTTG</td>
<td>ScaB-like dockerin</td>
<td>Homologous gene cloning</td>
</tr>
<tr>
<td>ScaB-Doc-R</td>
<td>GGATCCCAAGCTGCTAGTTGTTG</td>
<td>ScaB-like dockerin</td>
<td>Homologous gene cloning</td>
</tr>
<tr>
<td>CesA-Rf17-F</td>
<td>GAGCTCCAGAAGTGGTGGG</td>
<td>CesA-like dockerin</td>
<td>Homologous gene cloning</td>
</tr>
<tr>
<td>CesA-Rf17-R</td>
<td>GACAGAAGCTCAGGTTGTTG</td>
<td>CesA-like dockerin</td>
<td>Homologous gene cloning</td>
</tr>
</tbody>
</table>

a Restriction site sequences are underlined.

In previous studies (17), we formulated an architectural scheme of the cellulosome system in strain 17. In view of the divergent cohesin types in the FD-1 ScaB scaffoldin, it was of interest to determine the specificity characteristics of the cohesin-dockerin interactions in strain FD-1 and to determine the comparative architecture of the two R. flavefaciens strains.
between the two strains in their corresponding dockerin modules. The recombinant cohesin and dockerin modules were expressed as the desired fusion proteins in an *Escherichia coli* host cell system described previously (4). The cells were sonicated, the soluble proteins were applied batchwise to nickel-iminodiacetic acid agarose resin (Rimon Biotech, Rehovot, Israel), and the adsorbed proteins were eluted with 300 mM imidazole. The purified fusion proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (10%) with Coomassie brilliant blue staining. In all cases, a single major protein band was consistently obtained, as reported earlier (4).

To assess the binding properties among the putative cohesin and dockerin modules, a microtiter plate analysis was performed under standardized conditions, using an efficient, matching fusion-protein system (4). The expressed CBM3 cohesins were thus adsorbed onto microtiter plates and challenged by incremented concentrations of selected xylanase T6-

FIG. 1. The scaffoldin gene cluster in *R. flavefaciens* FD-1 and 17. (A) The scheme shows the organization on the genome of the *scaC*, *scaA*, *scaB*, and *scaE* genes. Scaffoldins of different sizes are indicated by the filled arrows. Numbers of amino acid residues for the given genes in the respective strain are shown. (B) Modular organization of the *sca* genes. The *scaA* and *scaB* genes contain multiple copies of cohesin modules (numbered). (C) Phylogenetic relationship among cohesin sequences from strains FD-1 and 17. The scale bar indicates the percentage of amino acid substitutions.
fused dockerins. Following extensive washing procedures, the amount of bound dockerin was assessed colorimetrically by primary anti-xylanase T6 antibody and secondary antibody-enzyme conjugate (4). The negative logarithms of the concentrations that produced the half-maximal chromogenic responses (pEC50 values) were determined from the respective binding curves (16).

As shown in Fig. 2, the ScaA-dockerin fusion protein interacts specifically with the conserved ScaB cohesins, i.e., cohesins 5 to 9 of ScaB of strain FD-1, which are orthologous to the ScaB cohesins of strain 17. The same ScaA dockerin fails to interact with the remaining ScaB cohesins (cohesins 1 to 4) of strain FD-1, which closely resemble in sequence those of ScaA in strain 17. As expected, the ScaA dockerin is unreactive toward its own cohesins, as well as those from ScaC and ScaE. The ScaB X-module/dockerin dyad also exhibited a high level of specificity for a different cohesin, the ScaE cohesin, and essentially failed to recognize the other cohesins. Conversely, the dockerins of ScaC and the enzyme equivalent of Cel44A both bound to the ScaA-like cohesins (i.e., the two reputed cohesins of ScaA and cohesins 1 to 4 of ScaB). No real preference was shown by either dockerin for the designated cohesins on the two scaffoldins, and the levels of recognition of the two dockerins were very similar. In contrast to the dockerin equivalent of Cel44A, that of the enzyme Ce3B displayed clear...
specificity for the cohesin of ScaC and essentially failed to recognize the cohesins of ScaA and cohesins 1 to 4 of ScaB.

The selective nature of the cohesin-dockerin interactions derived from this study allow us to propose an overall architectural scheme for R. flavefaciens strain FD-1 and to compare it with that proposed recently for strain 17 (17). As presented in Fig. 3, the same molecular players take part in cellulosome assembly in the two strains. The orthologous ScaE scaffolds are implanted into the respective cell surfaces via a similar sortase-mediated mechanism. The respective ScaE cohesins bind selectively to the XDoc dyad of ScaB, and ScaB cohesins serve to incorporate the primary ScaA scaffoldin. However, the major differences between the two strains are the number of confirmed cohesins in ScaA (two for strain FD-1 versus three for strain 17) and the number (nine versus seven, respectively) and nature (mixed versus homogeneous types, respectively) of the cohesins in ScaB. These differences are of apparent consequence to the respective architectures of the two cellulosome systems. In strain 17, for example, the combination of three ScaA cohesins and seven ScaB cohesins would imply the potential capacity to incorporate 21 different dockerin-containing proteins, either directly or mediated via ScaC. In strain FD-1, however, the two ScaA cohesins combined with the mixed cohesins in ScaB would presumably amount to a total of only 14 dockerin-containing proteins.

In both strains, the ScaA cohesins seem to recognize orthologous ScaC and homologous Cel44A-type dockers. The adaptor role of ScaC would thus appear to be similar for the two strains, i.e., a means of increasing the component repertoire of the respective cellulosome (19). The type of dockerin recognized by the ScaC cohesin has been identified for strain FD-1 but not for strain 17. The dockerin of the family 3 carbohydrate esterase equivalent, “Ce3B”-Doc, was shown in this work to bind strongly to the ScaC cohesin. The quotation marks signify that the enzyme itself remains unconfirmed, since only its dockerin was obtained by PCR amplification of appropriate primers derived from strain 17 sequences. Interestingly, the dockerin of the presumed orthologous enzyme in strain 17 failed to interact with its cognate ScaC cohesin (19). Although the two dockers show marked sequence similarity (59.0% identity and 74.4% similarity), the nature of the amino acid residues on homologous dockers that interact with ScaC in either or both species has yet to be elucidated.

In summary, the general design of the scaffoldin gene cluster in strains FD-1 and 17 of R. flavefaciens is essentially identical. The two clusters include the four scaffoldin orthologs in the same positions. Nevertheless, significant variations are evident in the modular architecture of the scaffoldin proteins, which implies a divergent organization of cellulosome architecture between the two strains. It appears that the two strains have undergone a considerable extent of evolutionary divergence, but it is currently unknown whether the differences observed here have functional consequences. In this context, the two strains may conceivably represent distinct biotypes that have become adapted to subtly different niches within the rumen ecosystem. The relationship between cellulosome architecture and the variations that occur in cell wall organization between plant species, or at different stages in the process of rumen digestion, is completely unknown.

We thank Alvaro Hernandez, W. Ryan Kim, Lei Liu, and Jyothi Thimmappar of the Keck Center for Comparative and Functional Genomics at the University of Illinois.

Parts of this research were supported by the National Research Initiative competitive grant no. 2002-53206-11634 from the USDA Cooperative State Research, Education, and Extension Service; by the Israel Science Foundation (grant no. 394/03 and 442/05); and by grants from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. The Rowett Research Institute is supported by the Scottish Executive Environmental and Rural Affairs Department. M.T.R. is supported by EU Grant GEMINI (QLRT-2001-02056). M.E.B. was supported by the Initiative for Future Agriculture and Food Systems grant no. 2001-52100-11527 from the USDA Cooperative State Research, Education, and Extension Service.

REFERENCES

