Patterns of Sequence Conservation in the S-Layer Proteins and Related Sequences in Clostridium difficile

Emanuela Calabi and Neil Fairweather*

Department of Biological Sciences, Centre for Molecular Microbiology and Infection, Imperial College of Science, Technology and Medicine, London SW7 2AY, United Kingdom

Received 1 February 2002/Accepted 2 April 2002

Clostridium difficile is the etiological agent of antibiotic-associated diarrhea. Among the factors that may play a role in infection are S-layer proteins (SLPs). Previous work has shown these to consist mainly of two components, resulting from the cleavage of a precursor encoded by the slpA gene. The high-molecular-weight (MW) subunit is related both to amidases from B. subtilis and to at least another 28 gene products in C. difficile strain 630. To gain insight into the functions of the SLPs and related proteins, we have further investigated the pattern of variability both at the slpA locus and at six nearby paralogs. Sequencing of the slpA gene from an S-layer group II strain and a variant S-layer group strain confirms a high degree of divergence in the low-MW SLP, which may result from diversifying selection. A highly conserved motif, however, is found at the C terminus in all low-MW subunits and may be essential for SlpA precursor cleavage. In strain 167, a variant cleavage product is present, suggesting a secondary processing site. Southern blotting analysis shows slpA-like open reading frames (ORFs) 2 to 7 to be conserved in all nine strains tested, with one exception: ORF2, which encodes a 66-kDa polypeptide coexpressed at low pH with the main SLPs in strain 630, may be partially deleted in strain 167. Polymorphism within the slpA-ORF7 cluster may be more pronounced in the region proximal to the slpA gene. Unexpectedly, a high-MW subunit probe cross hybridizes to sequences outside the slpA locus, which appear to vary in number in different strains.

E. Calabi and N. Fairweather*”}

* Corresponding author. Mailing address: Department of Biological Sciences, Centre for Molecular Microbiology and Infection, Imperial College of Science, Technology and Medicine, London SW7 2AY, United Kingdom. Phone: 44(0)20 7594 5247. Fax: 44(0)20 7594 3069. E-mail: n.fairweather@ic.ac.uk.
raise new questions as to the role of the low-MW SLP. In addition, the conservation of slpA paralogs supports the notion that they play a fundamental role in *C. difficile* physiology.

**MATERIALS AND METHODS**

**Bacterial strains and growth.** *C. difficile* strains 630, 17, 1, and Y were described previously (5). Strains 101, 167, 291, 371, and 959 are clinical isolates positive for cytotoxin and were obtained from C. Kelly, Harvard Medical School, Boston, Mass. The strains were cultured under anaerobic conditions in brain heart infusion broth and were maintained as described previously (5).

**PCR amplification, cloning, and DNA sequencing.** The *slpA* genes from strains 167 and Y were PCR amplified using the forward primer 5′-ATTCTATGTACATAATAAAGGATGTGTTGTA-3′ in combination with the reverse primer 5′-ATTAACTCCACAGCTAAATAAAC-3′ or the reverse primer 5′-ACCTTGACCAGTTTTCTATGCT-3′ under the conditions previously described (5). The products were cloned into *Escherichia coli* strain TG1 using the M13 vector sg131 (16) and were sequenced by MWG Biotech. Recombinant techniques were performed as described previously (23).

**N-terminal protein sequencing.** SLPs were acid extracted as previously described (5), fractionated on sodium dodecyl sulfate (SDS)-10% polyacrylamide gel electrophoresis (PAGE), and blotted onto a polyvinylidene difluoride membrane (Immobilon-P; Amersham) in 10 mM CAPS (3-[cyclohexylamino]-1-propanesulfonic acid; pH 11)-10% methanol for 1 h at 70°C. The blots were stained in 0.1% Ponceau S-1% acetic acid. N-terminal sequencing was carried out at the Protein and Nucleic Acid Chemistry Facility, Department of Biochemistry, University of Cambridge.

**Southern blotting analysis.** Probes were generated by PCR on genomic DNA from strain 630. The primers were as follows: (i) for the *slpA* low-MW subunit, 5′-TAAAGGCTATGGCAACTTGGGACAACA-3′ and 5′-AGGCCTGAGTGATTAGTTGCTTCAATC-3′; (ii) for the *slpA* high-MW subunit, 5′-TAAGGCCATGGCAAAATGATACAA-3′ and 5′-AGGCCTGAGGCTACATCATCTAAATAA-3′; (iii) for ORF2 to -7, previously described ORF-specific primers (5); (iv) for the 3′ region of ORF3, 5′-ATGCAGAAATAGAAGGTGGA-3′ and 5′-TTCAAGAAATGCTCTCTAT-3′.

PCRs were carried out in 30-μl volumes with ~1 μg of genomic DNA, 0.1 μM each primer, 2.5 mM MgCl₂, and 0.2 U of Taq DNA polymerase using the following cycle: 45 s at 95°C, 60 s at a temperature 5°C below the calculated melting temperature for the primers, and 60 s at 72°C, repeated 30 times. The expected PCR product was purified using the PCR purification kit from Qiagen according to the manufacturer’s instructions.

Probes were labeled using the ECL direct labeling and detection system (Amersham) according to the manufacturer’s instructions.

Approximately 5 μg of genomic DNA was digested in a 20-μl total reaction volume and fractionated on 0.7% agarose gels in Tris-acetate-EDTA buffer. The gels were acid treated, denatured, neutralized, and blotted to Hybond-N+ membranes. After UV cross-linking, they were prehybridized for 18 h at 70°C and followed by three washes in 1× SSC-0.1% SDS (20 min per wash). Detection was by the enhanced-chemiluminescence method.

**Nucleotide sequence accession numbers.** The nucleotide sequences of the *slpA* genes from strains 167 and Y were deposited in the GenBank database under accession numbers AF478570 and AF478571.

**RESULTS**

**Structural analysis of the SLPs from strains Y and 167.** Acid-extracted SLPs from *C. difficile* strain Y contain two main proteins with molecular masses of ~41 and 37 kDa in approximately equimolar amounts (Fig. 1A), which corresponds to a group II pattern (14). In contrast, strain 167 yields one major band at ~39 kDa and a minor band at ~43 kDa, in a ratio of approximately 4:1 based on Coomassie blue staining. Additional weaker bands are visible at 20 and 22 kDa and in the 35- to 35-kDa region (Fig. 1B). However, there is no substantial component with mobility comparable to those of the low-MW subunits from other strains (Fig. 1A). In order to clarify the nature of the variant SLPs, the *slpA* genes from strains 167 and Y were further characterized. PCR amplification was carried...
out using primers based on the sequence of strain 630 and spanning the whole slpA coding sequence, from 59 bp 5' of the ATG to 559 bp 3' of the stop codon. Products were obtained from both strains, indicating conservation of primer sequences.

Strain Y yielded a fragment of ~2,900 bp, and strain 167 yielded a fragment of ~2,400 bp (data not shown). These compare with a size of 2,778 bp for the PCR product from strain 630. In order to estimate the size of the low-MW subunit coding sequence, a second PCR was carried out using the same forward primer and a reverse primer (NF129) that maps immediately downstream of the sequence encoding the N terminus of the high-MW subunit. Fragments of ~1,600 and ~1,100 bp were obtained from strains Y and 167 compared to 1,523 bp from 630. These preliminary data are consistent with the observation that the size of the low-MW SLP is larger in strain Y than in S-layer group I strains whereas the high-MW SLPs are of similar sizes. In strain 167, a significant deletion is suggested within the low-MW SLP.

In order to determine the structures of the slpA genes from both strains in greater detail, PCR fragments were cloned and completely sequenced. The slpA ORF spans 2,301 bp in Y and 1,830 bp in 167. The predicted amino acid sequences are shown in Fig. 2, aligned with the previously reported SlpA precursor sequences (5, 13). In both strains, a signal peptide is evident by comparison to the previously characterized sequences. The sites of processing of the precursor were investigated by N-terminal sequencing of the 41- and 37-kDa proteins from strain Y and the major 39-kDa and minor 43-kDa proteins from 167. Unequivocal assignments were made for 13 to 15 residues in each case, which are shown in Fig. 2. In each case, a perfect match was found to predicted amino acid sequences. These data show the low-MW (N-terminal) and high-MW (C-terminal) SLPs from strain Y to be 351 and 392 amino acids, respectively, with estimated masses of 37,833 and 41,272 Da, in close agreement with the values determined by SDS-PAGE analysis (Fig. 1).

In strain 167, the N terminus of the predominant 39-kDa band was revealed to be at position 207, consistent with this protein being the high-MW SLP. The protein is predicted to contain 404 amino acids, with a mass of 42,390 Da. This would correspond to an appreciably slower mobility in SDS-PAGE than was determined experimentally, suggesting either abnormal electrophoretic behavior or a posttranslational modification. Cleavage of the SlpA precursor at residue 207 to release the 39-kDa protein would be expected to result in a small N-terminal protein of ~20 kDa. While a weak band of ~20 kDa which might correspond to this protein is visible in the SLP preparation from strain 167 (Fig. 1B), it is present in significantly smaller amounts than the 39-kDa protein. Thus, although the strain 167 SlpA precursor contains a sequence predicted to correspond to a low-MW subunit of 182 amino acids (approximately 20 kDa) and is processed to release the 39-kDa high-MW subunit, either the low-MW subunit is largely degraded or it is not efficiently extracted under the conditions that we have used.

N-terminal sequencing of the minor 43-kDa protein revealed it to match perfectly residues 25 to 37 of the SlpA precursor, showing this protein to be derived from the N terminus of the precursor protein after removal of the signal sequence. However, the 43-kDa protein cannot be the uncleaved precursor, which has a predicted MW of 62,312.

Based on the primary sequence alone, the C terminus of the 43-kDa component is estimated to lie around position 450 in the SlpA precursor. The sequences surrounding this region and the N termini of high-MW subunits were compared to identify patterns suggestive of a conserved cleavage site. Little conservation is apparent among the N termini of high-MW subunits, except that five out of six have Ala as the N-terminal residue. However, conservation is readily apparent in the sequences immediately upstream of the cleavage sites. A highly conserved motif is found in all low-MW subunits (positions 328 to 345 of strain 630 (Fig. 2 and 3A)). In strains 167 and Y, this motif extends for a further three residues, which are conserved between these two strains. These motifs may be recognized in an enzymatic-cleavage step. If this is the case, variant processing enzymes that have coevolved with the slpA gene may exist in different strains. No such motif is found at the predicted C terminus of the 43-kDa protein from strain 167. However, in strain 167, a different conserved motif is found ~50 residues C terminal to the predicted processing sites of both the 43-kDa and the 39-kDa proteins. This second, 17-amino-acid-long motif is part of the first and third of three ~120-amino-acid repeat units shared between the SLP high-MW subunit and the products of the B. subtilis CWLB and CWBA genes (Fig. 2). The copy found in the second repeat unit is more distantly related. Of all the other five sequenced SlpA precursors, only in Y is this motif similarly conserved between the first and third repeat (Fig. 3B).

**Identities of other surface-associated proteins.** Most S-layer preparations isolated from various strains by the low-pH method contain an ~66-kDa component (Fig. 1A), although this is not found in strain 167 (see below). For strain 630, we have determined the N-terminal sequence of this protein to be AETTQVKKET. A BLASTX search of the whole unfinished C. difficile genome showed an exact match to the gene product encoded by ORF2, an slpA paralog located ~3 kbp downstream of slpA and transcribed during vegetative growth (5). The gene product starts with a putative 23-amino-acid-long signal sequence that is highly related to that of the SlpA precursor. ORF2 contains 623 amino acids and, like the SlpA precursor, is 50% similar in amino acid sequence over its C-terminal region (residues 284 to 623) to a domain present in...
the B. subtilis N-acetylmuramoyl-\(\mathrm{L}\)-alanine amidase CWLB/LytC and its enhancer, CWBA/LytB (18) (Fig. 4A). Consistent with this, we have previously shown in a zymogram assay that this protein has amidase activity (5). A BLASTP search with the upstream, N-terminal region identifies significant homologies (39% amino acid similarity) to both the \(\sim 125\)-kDa SLP of \textit{Bacillus sphaericus} (3) and the transducer of rhodopsin (Htr) II from the archaeabacterium \textit{Natronobacterium pharaonis} (26) (Fig. 4B and C). Different sets of amino acids are conserved in the two pairwise comparisons. Of the 129 positions shared with \textit{N. pharaonis} Htr-II, 18 are also conserved in the signal domain of bacterial transducers like Tsr, including methylation sites.

**Pattern of variability in slpA.** The degree of relatedness between the SlpA protein sequences was estimated by calculating distance scores in each pairwise comparison with the program GeneDoc, using the Blosum 62 distance matrix (Table 1). Even excluding the hypothetical strain 167 component, low-MW subunits are little conserved, except over their COOH termini (data not shown). It has been argued that a 78-amino-acid sequence at the N termini of the low-MW subunits in strains C253 (SlpA sequence identical to that of 630 [13]) and 76-85 is related to the SLH (S-layer homology) domain found in SLPs from unrelated species (13). However, this sequence is not well conserved in different strains (Fig. 2).

In contrast, in the high-MW subunits, out of 263 positions at which all six different sequences can be aligned, 213 (81%) are highly conserved. Based on the close relatedness of high-MW subunit sequences, a dendrogram was constructed by the neighbor-joining method which suggests a tentative evolutionary relationship (Fig. 5). As expected, S-layer group I sequences are more related to each other than to Y or 167 sequences. The last strain is marginally more related than Y to S-layer group I sequences.

Although the SLP low-MW subunits are highly variable in amino acid sequence, they are conserved in overall length, ranging between 316 and 321 residues. The single exception is represented by the hypothetical low-MW subunit of 167, spanning only 182 amino acids. In contrast, high-MW subunits display greater variation in length, ranging from 373 (in strain 17) to 416 (in strain 1) residues. Most of the differences in sequence length among high-MW subunits can be mapped to one of four regions, i.e., between positions 446 and 449, 595 and 604, 684 and 701, and 732 and 761 in the multiple alignment shown in Fig. 2.

**Variability within the slpA gene cluster.** The genome of \textit{C. difficile} strain 630 contains at least 28 paralogs encoding polypeptides containing \(~45\)% amino acid similarity to the SlpA precursor (5, 13), which we have provisionally named ORFs 2 to 29, pending the annotation of the genome sequence (http://www.sanger.ac.uk/Projects/C_difficile). A number of these are closely linked to the slpA locus: slpA-like ORFs 2 to 7 are within 21 kbp of slpA, and slpA-like ORFs 8 to 12 are within 17 kbp 5’ of slpA. We have also shown that ORFs 2 to 7 are transcribed in cells during vegetative growth (5). The functions of these slpA-like genes are presently unknown. As a first step towards investigating it, we examined the extent of their variability by Southern blotting analysis.

ORF-specific probes were designed from the region of each ORF immediately upstream of the amidase homology domain (AHD), keeping to a minimum the overlap with the domain. DNA sequence comparison showed no significant cross homology among probes. Additional probes were made that spanned the whole slpA high-MW and low-MW subunit coding sequences, as well as from the sequence encoding the C-terminal region of ORF3.
Each probe was used on a panel of nine different strains, consisting of four strains in addition to 630, 17, 1, Y, and 167: strains 101 and 371 were isolated from patients with \textit{C. difficile}-associated diarrhea, whereas strains 291 and 959 were isolated from asymptomatic carriers (17). SDS-PAGE analysis showed all of these four strains to belong to S-layer group I based on their SLP patterns (Fig. 1C). For each strain, two different restriction enzyme combinations were used, chosen on the basis of the sequence from strain 630, to yield the best compromise between a convenient size range of fragments and maximum resolution of individual ORFs.

Figure 6A shows the positions of restriction sites within the \textit{slpA} locus of each strain based on the DNA sequence. Since, except for strain 630, the sequence has been determined only for the coding sequence and short flanking regions, in most cases the lengths of restriction fragments produced in each digestion are not known, although lower limits can be predicted. Regions showing >70% homology and therefore expected to hybridize to the strain 630 probes are indicated in the figure. Figure 7A shows a complete restriction map of the region spanning from \textit{slpA} to ORF7 in strain 630. The results of the Southern analysis are shown in Fig. 6B and 7B. In general, fragments hybridizing strongly with each probe closely match the sizes expected from the corresponding locus. There are, however, a few notable exceptions. (i) In strain 630, the low-MW subunit probe strongly hybridizes to an additional 9-kbp \textit{Hin}cII/\textit{Pvu}II fragment which is only very weakly positive with the high-MW subunit probe, as is visible on long exposures (data not shown). This fragment cannot be accounted for on the basis of the existing sequence. (ii) In strain 17, no hybridization to the predicted 0.4-kbp \textit{Hin}cII/\textit{Pvu}II fragment is observed with the low-MW subunit probe. This is most likely due to a modification of the upstream \textit{Pvu}II site (position 287; see below). (iii) In strain 17, a 5.5-kbp \textit{Hin}cII/\textit{Pvu}II fragment and a 9-kbp \textit{Hin}cII/\textit{Pvu}II fragment can be assigned to the 5' end of the \textit{slpA} gene (upstream of the \textit{Pvu}II site at position 287). While a fragment of the same size as the former is positive with the high-MW subunit probe, the 9-kbp \textit{Hin}cII/\textit{Pvu}II fragment is not. Thus, it is likely that the 5.5-kbp \textit{Hin}cII/\textit{Pvu}II fragments hybridizing to the high-MW and to the low-MW probes are derived from different loci. (iv) In strain Y, a 0.8-kbp \textit{Hin}cII/\textit{Pvu}II fragment is observed instead of the expected 0.6-kbp fragment. This may be explained by the upstream \textit{Pvu}II site (position 1497) being resistant to digestion under the high-salt conditions used for mixed \textit{Hin}cII/\textit{Pvu}II digestions. (v) In 630, larger-than-expected fragments are positive with the \textit{Pvu}II probe due to the \textit{Pvu}II site downstream of the probe being resistant to digestion, likely due to methylation. Thus, identical positive fragments are observed in digests from which \textit{Pvu}II has been omitted, and they match the size predicted for \textit{Hin}cII or \textit{Hin}cII fragments spanning the probe (Fig. 7B). (vi) Lack of detection of some fragments is explained by low homology to the strain 630 sequence used as a probe. This applies, in strain 1, to the fragment upstream of the \textit{Hin}cII site at position 847 (~50% homology), in strain Y to the 0.4-kbp \textit{Hin}cII/\textit{Pvu}II fragment (~45% homology), and in strain 167 to the 0.8-kbp \textit{Pvu}II/\textit{Pvu}II fragment (~43% homology). In addition to 630, fragments hybridizing to the low-MW subunit probe are only found in strains 17, 1, and 371, the last two giving identical patterns.

In the case of the high-MW subunit probe, strongly hybridizing fragments are found in all strains. With the few exceptions mentioned above, these fragments can be fully accounted for by the \textit{slpA} gene in the five strains for which sequences are available. By analogy, they must also be derived from the \textit{slpA} locus in the remaining four strains. Fragments hybridizing more weakly to the same probe must correspond to less homologous sequences mapping outside the \textit{slpA} locus. The number of these fragments varies between one and four. Six different patterns are observed among the nine strains, one consisting of three strains (101, 291, and 959), one of two strains (1 and 371), and the others of single strains. A similar grouping is also apparent from the results with probes for ORF2 to ORF6, the probe for ORF7 giving similar patterns in all strains. The nature of the sequences weakly homologous to the high-MW subunit probe remain to be established. In the case of 630, the 4.3-kbp \textit{Hin}cII/\textit{Pvu}II fragment and the 2.5-kbp \textit{Hin}cII/\textit{Pvu}II fragment may correspond to ORF2.

The probe for ORF2 does not detect any homologous sequence in 167. This may indicate a gene deletion, since the 66-kDa polypeptide encoded by this ORF is absent in SLP preparations from 167 (Fig. 1A).

The probe for the 5' region of ORF3 (ORF3-5') detects multiple fragments. A major band is detected in each digest or strain, which in 630 matches the size predicted for ORF3. The additional bands are stronger in 630, 101, 291, and 959 than in the other strains, and strain Y gives an overall weaker signal. The basis for this pattern is unclear, since a BLAST search of the whole unfinished 630 genome does not reveal any significant homology outside ORF3. Although the probe spans a \textit{Hin}cII site in 630, the portion upstream of the site extends for only 39 bp, and it is unlikely to contribute to any signal. When the same blot was hybridized to a probe corresponding to the 3' end of ORF3 (ORF3-3'), strong, unique signals were detected only in 630 and 17. These results are consistent with previously reported dot blot hybridization data showing the 3' end of the Cwp66/ORF3 gene to be less conserved than the 5' end in a panel of 36 strains (30).

We have identified polymorphic variants by using probes for each of the \textit{slpA}-like ORFs 2 to 6. Allelic variability seems to be restricted to the 5' end of the region under investigation. No polymorphism was mapped in or close to ORF7. In addition, probes for ORF5 and -6 give invariant \textit{Pvu}II/HindIII fragments in all strains. While size polymorphism is detected with these two probes on \textit{Pvu}II/HindIII digests, both probes hybridize to the same \textit{Pvu}II/HindIII fragment in each strain. This suggests that the \textit{Pvu}II/HindIII fragment polymorphism detected by the ORF5 and -6 probes is likely due to sequence variation at the 5' site.

In all ORFs, sequences immediately 5' of the AHD are conserved across strains. However, like the \textit{slpA} low-MW subunit coding sequence, the ORF3-3' sequence shows very limited DNA conservation. It remains to be seen whether this is also reflected in lower conservation at the amino acid sequence level.

DISCUSSION

Both the high- and low-MW \textit{C. difficile} SLPs show considerable variation in molecular mass. Initial comparison of eight
strains identified two patterns: group I, corresponding to masses of 45 to 47 and 32 kDa, and group II, corresponding to masses of 42 and 38 kDa for the high- and low-MW subunits. All of the slpA genes sequenced so far derive from group I strains (5, 13). In order to determine the structural constraints on the SLPs, we have extended the sequence analysis to two further strains: strain Y, a group II strain, and strain 167, which gives a variant pattern, consisting of a single major SLP component and other minor bands.

Our data show that the group II SLPs are less related to any of the previously sequenced group I SLPs than the latter are among themselves, which supports the hypothesis that group II belongs to a distinct class. Divergence is particularly striking over the low-MW SLPs. While sequence conservation is readily apparent in the high-MW SLPs and, to a lesser extent, over the C-terminal 50 amino acids of the low-MW SLPs, the remaining portions of the low-MW SLPs show little conservation. In pairwise comparisons between Y and group I strains, identities over the low-MW SLPs range from 34 to 40%. In contrast, the high-MW SLPs are 70 to 71% identical. This pattern of sequence variability in the low-MW SLPs is intriguing and shows impressive analogies with that of flagellins. In the flagellins, the central segment is highly variable, corresponding to surface-exposed regions and giving rise to a well-known class of antigenic determinants, the H antigens (11, 22, 28).

The sequence variability of the low-MW SLPs may simply reflect the lack of functional constraints. An alternative hypothesis is that it confers an evolutionary advantage, mediating escape from immune recognition and the ability to reinfect hosts. Consistent with the latter model, evidence has been presented that the variability of flagellins results from positive selection for amino acid replacements, i.e., diversifying selection, rather than from the absence of negative selection (19, 22). It is tempting to speculate that the high variability in the C. difficile low-MW SLPs reflects the operation of similar mechanisms. This subunit carries the dominant antigenic epitopes, as shown by Western blotting analysis of S-layer preparations using human sera (21). The high degree of divergence between the low-MW subunit coding regions of the group II and any of the group I strains prevents the calculation of meaningful rates of nucleotide change. However, within group I, rates of non-synonymous change are much higher over the low than over the high-MW subunit sequences and exceed the rates of

### Table 1: Distance scores in pairwise comparisons between SlpA signal sequences and low-MW and high-MW subunits from six C. difficile strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Distance score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>630</td>
</tr>
<tr>
<td>Signal sequence</td>
<td>209</td>
</tr>
<tr>
<td>Lower subunit</td>
<td>2,807</td>
</tr>
<tr>
<td>Upper subunit</td>
<td>3,301</td>
</tr>
</tbody>
</table>

**FIG. 4.** Sequence homology between the predicted product of slpA-like ORF2 and the SlpA high-MW subunits from strain 630, CWLB (N-acetylmuramoyl-L-alanine amidase) from B. subtilis (BS-Ami) (GenBank accession number Q02114) and CWBA (amidase enhancer) from B. subtilis (BS-Mod) (GenBank accession number P38537) (B), and Htr-II from N. pharaonis (GenBank accession number P42259) (C). The alignments in panel A were produced with ClustalX (version 1.81), using the Blosum 30 matrix, a gap opening penalty of 10, and a gap extension penalty of 0.1. The alignments in panels B and C were produced with pairwise BLASTP on the National Center for Biotechnology Information server (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html), using a gap opening penalty of 11, a gap extension penalty of 1, and no filter. Solid backgrounds highlight positions conserved in all sequences, and shaded backgrounds highlight other conserved positions. #, residues shared with the signal domain of bacterial Tsr transducers (26). Asterisks and numbers above the sequences mark every 10th residue.

**FIG. 5.** Dendrogram depicting the relationships among SlpA high-MW subunits. The diagram was constructed by the neighbor-joining method provided with ClustalX (version 1.81) based on the alignment shown in Fig. 2.
FIG. 6. (A) Deduced restriction maps of the \textit{slpA} locus from the five strains for which DNA sequences have been determined. The horizontal lines represent the DNA sequence, with the thick bar corresponding to the \textit{slpA} ORF. The vertical lines extending above the sequence indicate \textit{HindIII} sites, those below the sequence indicate \textit{Hin}cII sites, and those both above and below the sequence indicate \textit{Pvu}II sites. The solid horizontal bars indicate \( > 70\% \) homology to the high-MW subunit probe, and the shaded horizontal bars indicate \( > 70\% \) homology to the low-MW subunit probe. The numbers over the arrows indicate the size of the relevant restriction fragment in base pairs. (B) Southern blotting analysis of genomic DNA from \textit{C. difficile} strains with an \textit{slpA} low-MW subunit probe (left) and with an \textit{slpA} high-MW subunit probe (right). \textit{Pv}, \textit{Pvu}II; \textit{He}, \textit{Hin}cII; \textit{Hd}, \textit{HindIII}. 

\[ \text{DNA sequence} \]

\[ \text{Restriction maps} \]

\[ \text{Probes:} \textit{slpA} \text{ low MW subunit} \]

\[ \text{Probes:} \textit{slpA} \text{ high MW subunit} \]
FIG. 7. (A) Deduced restriction maps of the region spanning the slpA gene and slpA-like ORF2 to -7 in strain 630. The thin horizontal line represents the DNA sequence, with thick bars corresponding to ORFs. The solid segments indicate the AHD. The vertical lines extending above the sequence indicate HinfI sites, those below the sequence indicate Hincll sites, and those both above and below the sequence indicate PvuII sites. The solid horizontal bars under the sequence indicate the position of each probe. The numbers over the arrows indicate the size of the relevant restriction fragment in kilobase pairs. (B) Southern blotting analysis of genomic DNAs from C. difficile strains with probes from the indicated slpA-like ORF. Independent probes were used for each probe, except for ORF3-3/PvuII and ORF3-5/PvuII. Pv, PvuII; Hc, Hincll; Hd, HinfI.

VOL. 184, 2002
C. DIFFICILE
CELL WALL PROTEINS
3895
onymous change in some strain pairs. By analogy with flagellins, this may reflect diversifying selection. As to the mechanisms responsible for this high mutation rate, work with gram-negative enterobacteria has shown evidence for horizontal gene transfer and multiple recombination events (11, 19, 22). We have searched the genome of C. difficile strain 630 for evidence of intragenomic-recombination events that might give rise to one of the known SLP variants. While none was found, the possibility remains that sequences from other strains, or even other species, may act as donors.

The hypervariable sequence within the low-MW subunit spans a stretch of ~80 amino acids at the N terminus that in group I strains shows a distant relationship to the SLH domain (13). This domain has been implicated in anchoring S-layers to the cell wall in several species by binding either to the peptidoglycan or to secondary cell wall polymers (8, 12, 20). Since this sequence is not conserved in other C. difficile strains, it is unlikely to mediate binding of the low-MW SLPs to the cell surface.

The C-terminal region of the low-MW subunit does show a recognizable pattern of conserved residues. At least part of this pattern may be due to a requirement for proteolytic processing. Although the mechanism by which the high-MW and low-MW SLPs are released from the SlpA precursor has not been elucidated, cleavage is unlikely to represent an artifact of the isolation procedure, since identical subunit size, yield, and stoichiometry are obtained using different methods of extraction. Furthermore, cleavage does not occur when the SlpA precursor is expressed in a heterologous system, e.g., Lactococcus lactis or E. coli (E. Calabi and N. Fairweather, unpublished data). An amino acid motif N-terminal to the cleavage site is conserved in all strains. Interestingly, with respect to this motif, cleavage occurs in slightly different positions in group I strains than in Y and 167. Thus, the motif may reflect the binding constraints of the catalytic site of a specific peptidase, which may have evolved in different strains in parallel with the SLP sequences.

In strain 167, a different pattern of cleavage of the SlpA precursor is evident. The predominant protein produced is 39 kDa, representing the high-MW SLP. Very little, if any, of the predicted 20-kDa low-MW SLP is found at the cell surface. However, an additional component of 43 kDa is present which extends from the N terminus of the mature SlpA precursor to a putative secondary cleavage site within the precursor. The low levels of the 20- and 43-kDa proteins may be due either to instability and consequent rapid degradation of these polypeptides or to their inability to be anchored to the cell wall. In either case, their presence at the cell surface in the same molar ratio as the 39-kDa protein is clearly not required either for cellular viability or pathogenicity. It remains to be established whether one or both of the S-layer lattices identified in C. difficile are missing in this strain.

One of the most remarkable outcomes of the identification of the slpA gene has been the finding in strain 630 of a large family of related genes, many of which cluster around the slpA locus. As a prerequisite to further analysis of this family, we felt it important to establish whether its members are conserved across strains. To this end, we have used Southern blotting analysis of the slpA-like gene cluster downstream of slpA. While polymorphic variants have been identified for ORFs 2 to 6, allelic variability seems to be restricted to the 5' end of the region. Since most of the polymorphisms that we have examined are within coding sequences in strain 630, it is expected that these variants will be expressed at the protein level. The patterns of polymorphism seem to be linked across loci. Two additional significant conclusions emerge from our data. First, all seven slpA-like ORFs that we tested were found to be represented in every strain, with one exception. No signal was obtained with ORF2 in strain 167, consistent with the absence of a 66-kDa polypeptide in a low-pH extract. This suggests that the corresponding region is deleted. In all other cases, the signals obtained in different strains show minimal variation in strength, suggesting comparable degrees of homology with respect to 630. Second, a probe corresponding to the high-MW subunit coding sequence shows cross hybridization to a limited number of sequences outside the slpA locus. This result was unexpected, since while they are related at the protein level, homologies between slpA-like family members at the nucleotide level are below the threshold (~65%) required to result in significant hybridization under our experimental conditions. It remains to be established whether the sequences cross hybridizing to the slpA high-MW probe correspond to any of the known slpA-like family members or to yet-undiscovered chromosomal or extrachromosomal loci. However, it is interesting that their numbers appear to vary in different strains.

The occurrence of up to nine SLP-encoding ORFs sharing a short (~180-nucleotide) homology domain and arranged in a tight cluster has been reported in the genome of Campylobacter fetus (9). In this case, however, there is only a single slp promoter and only one sequence is expressed at any given time, the others serving as a reservoir for recombination. It remains to be seen whether recombination can occur between different slpA-like loci in C. difficile. Moreover, all six slpA homologues studied are transcribed, and the product of at least one is also present at the cell surface. Transcription of these genes must occur from separate promoters, since intergenic primers do not give positive results on reverse transcription-PCR (data not shown).

ACKNOWLEDGMENTS

We thank C. Kelly and L. Kyne for supplying some of the strains in this study. F. Calabi for helpful discussions, and the Sanger Centre for release of genome sequences prior to publication.

E.C. was supported by a fellowship from the Blancheflor Boncompagni Ludovisi Foundation.

REFERENCES

Lazarevic, V., P. Margot, B. Soldo, and D. Karamata.

Chauvaux, S., M. Matuschek, and P. Beguin.


