JOURNAL OF BACTERIOLOGY, Apr. 2004, p. 2508–2510 Vol. 186, No. 8
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.8.2508–2510.2004

Clostridium difficile IStron Cd*ISt1*: Discovery of a Variant Encoding Two Complete Transposase-Like Proteins

Oliver Hasselmayer,[1]* Veit Braun,[1] Claudia Nitsche,[1] Michael Moos,[1] Maja Rupnik,[2] and Christoph von Eichel-Streiber[1]

Institute of Medical Microbiology and Hygiene, Johannes Gutenberg University of Mainz, D-55101 Mainz,[1] and Department of Bacteriology, Georg-August-University Goettingen, D-37075 Goettingen,[2] Germany

Received 30 June 2003/Accepted 21 November 2003

Screening a *Clostridium difficile* strain collection for the chimeric element Cd*ISt1*, we identified two additional variants, designated Cd*ISt1*-0 and Cd*ISt1*-III. In in vitro assays, we could prove the self-splicing ribozyme activity of these variants. Structural comparison of all known Cd*ISt1* variants led us to define four types of IStrons that we designated Cd*ISt1*-0 through Cd*ISt1*-III. Since Cd*ISt1*-0 encodes two complete transposase-like proteins (TlpA and TlpB), we suggest that it represents the original genetic element, hypothesized before to have originated by fusion of a group I intron and an insertion sequence element.

Recently, the genetic element Cd*ISt1* was identified in the genome of *Clostridium difficile* strain C34 (1, 5) and was shown to combine features of group I introns and insertion sequence (IS) elements. To distinguish between individual Cd*ISt1* variants from different strains, the strain C34 element is herein designated Cd*ISt1*-C34, since the designation Cd*ISt1* includes all variants combining features of the group I intron and the IS elements of the IS*605* type. The 434 bp at the 5′ end of Cd*ISt1*-C34 show the typical structures and key features of group I introns, while the 3′ part harbors two open reading frames (ORFs) coding for a truncated (*tlpA*) and a complete (*tlpB*) protein. The putative proteins TlpA and TlpB show high homology to transposases, which are characteristic products of composite IS elements of the IS*605* type (Fig. 1) (4). In contrast to classical group I introns, Cd*ISt1*-C34 is found in protein-encoding genes of a bacterial chromosome (1, 2). Sequence analysis of Cd*ISt1*-C34, several of its variants (Cd*ISt1*a-C34 through Cd*ISt1*j-C34), and their integration sites revealed that these genetic elements show all features characteristic of an IS*605*-like mobility mechanism (1, 3). Therefore, we assumed that the IS element component mediates the spread of Cd*ISt1* while the intron component is responsible for efficient splicing of Cd*ISt1* from precursor mRNA. Due to their chimeric nature, we called these elements IStrons.

Previously, two types of IStrons (Cd*ISt1*-I and Cd*ISt1*-II) were identified in *C. difficile*, and these types differ in the IS element component (Fig. 1) (1). We have grouped all characterized Cd*ISt1* IStron variants into four types (Cd*ISt1*-0 to Cd*ISt1*-III) based on their structural features. Within each type, the variants differ in their genomic integration sites and the strains in which they are located. Compared to Cd*ISt1*-I IStrons, Cd*ISt1*-II IStrons show a deletion of about 100 bp covering nearly the entire *tlpA* gene (Fig. 1). To verify the proposal that Cd*ISt1* was generated by fusion of a group I intron and an IS element from the IS*605* family, we searched for additional Cd*ISt1* types without deletion of the IS element component. *C. difficile* isolates were screened for Cd*ISt1*-related sequences by using two internal primers (P1, CGACAA CCTCAAAAATGATAAA, and P2, TCTTAATCCTTCTTT TAATATATTT) that amplify a major part of the Cd*ISt1* sequence. PCR conditions were as follows: 30 cycles of 10 s at 94°C, 30 s at 42°C, and 4 min at 68°C. Using this PCR assay, we detected Cd*ISt1*-related sequences in the genomes of all *C. difficile* isolates tested but not in those of other clostridial species such as *C. acetobutylicum*, *C. bifermentans*, *C. botulinum*, and *C. perfringens*.

Six different patterns of PCR amplification products were obtained with *C. difficile* chromosomal DNA as the template (Fig. 2). PCR amplicons of 1,397 bp identified type I variants, while amplicons of 1,318 bp identified type II variants in the bacterial genome. The amplicon sizes in four of the six patterns differed remarkably from those corresponding to Cd*ISt1* types I and II (Fig. 2). This indicated the existence of yet unknown Cd*ISt1* variants in the genomes of the respective *C. difficile* strains. One of these PCR products (Fig. 2, lanes 5 and 6) was approximately 250 bp larger than the PCR product representing Cd*ISt1*-I, and the other (Fig. 2, lanes 1, 4, and 6) was more than 1.3 kb smaller than the PCR product representing Cd*ISt1*-II. Sequencing of the two unusual PCR products verified that they actually represent two novel Cd*ISt1* types. The larger was designated type 0 (Cd*ISt1*-0), and the smaller was designated type III (Cd*ISt1*-III) (Fig. 1).

As described by Braun et al. (1), we then used primers P3 (CGGAGCTTACCTGCTGATTG) and P4 (GCCCCTACCG GACACCTCTT) for inverse PCR and determined the integration sites and the complete sequences of Cd*ISt1*l-C34 (Cd*ISt1*-III variant) and Cd*ISt1*a-SE918 (Cd*ISt1*-0 variant) in strain SE918. Cd*ISt1*l-C34 has a size of 734 bp and is located in an ORF homologous to one from *C. perfringens* encoding a protein tyrosine phosphatase. Cd*ISt1*l-C34 contains the complete ribozyme component showing 97.5% identity to Cd*ISt1*-I and Cd*ISt1*-II copies. Downstream of the intron component, we found a deletion in Cd*ISt1*l-C34 of 1,247 bp covering *tlpA*

* Corresponding author. Mailing address: Institute of Medical Microbiology and Hygiene, Johannes Gutenberg University of Mainz, Obere Zahlbacher Strasse 67, D-55101 Mainz, Germany. Phone: 49 6131 39-32204. Fax: 49 6131 39-33364. E-mail: hasselm@mail.uni-mainz.de.

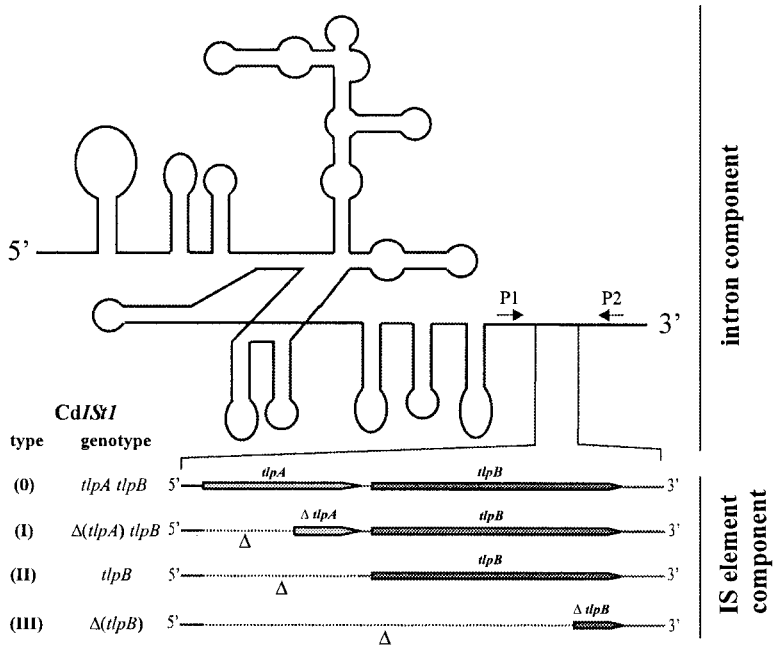

FIG. 1. Schematic drawing of Cd*ISt1* IStron variants types 0 to III. The four Cd*ISt1* types are depicted; the intron component (upper part) shows little variation, whereas the IS element component (lower part) varies as indicated. All variants have conserved ribozyme activities, and deletions of the putative TlpA and TlpB proteins are found in all but type 0 variants. Arrows labeled P1 (5′-CGACAACCTCAAAAATGATA AA-3′) and P2 (5′-TCTTAATCCTTCTTTTAATATATTT-3′) indicate the primers for PCR amplification of the four IStron types.

and the major part of *tlpB* (positions 558 to 1804 of Cd*ISt1*-C34). The 3′ ends of Cd*ISt1*l-C34 and Cd*ISt1*-I IStrons are identical and contain the 3′ splice site. Using reverse transcription-PCR and an in vitro splice assay, we verified that Cd*ISt1*l-C34 is precisely and efficiently excised from precursor mRNA in vitro and in vivo (data not shown). This demonstrates that all structures associated with the ribozyme activity of the IStron are preserved in Cd*ISt1*l-C34.

Cd*ISt1*a-SE918 is integrated into an ORF homologous to one from *Thermogota maritima* encoding a 16S pseudouridylate synthase. Strikingly, in *C. difficile* strain C34, the same ORF is interrupted by a Cd*ISt1*-II variant and not by a Cd*ISt1*-0 variant (1). Sequence analysis revealed that Cd*ISt1*a-SE918 shows very high homology to Cd*ISt1*-I IStrons, but Cd*ISt1*a-SE918 contains an additional DNA stretch of 239 bp located upstream of the truncated *tlpA* of type I IStrons (Fig. 1). Except for a TAA stop codon at position 40, this additional sequence is in frame with the remnant of *tlpA*. Analyzing *C. difficile* strains 55767 and R9304, which also contain large Cd*ISt1*-0 variants (Cd*ISt1*a-55767 and Cd*ISt1*a-R9304), we found a CAA sequence instead of this TAA stop codon. Within the three Cd*ISt1*-0 IStron variants, we found a potential Shine-Dalgarno sequence 7 bp upstream of the start codon of the *tlpA* ORF. Thus, *tlpA* of Cd*ISt1*-0 variants seems to code for a complete transposase-like protein. The calculated TlpA size (135 amino acids) and molecular mass (16 kDa) are typical of IS*200* homolog transposases. A database search using the translated amino acid sequence of the putative TlpA encoded by Cd*ISt1*-0 variants of strains SE918, 55767, and R9304 verified high similarity to IS*200* homologous transposases from *C. perfringens* (accession number, NP_562804), *Deinococcus ra-*

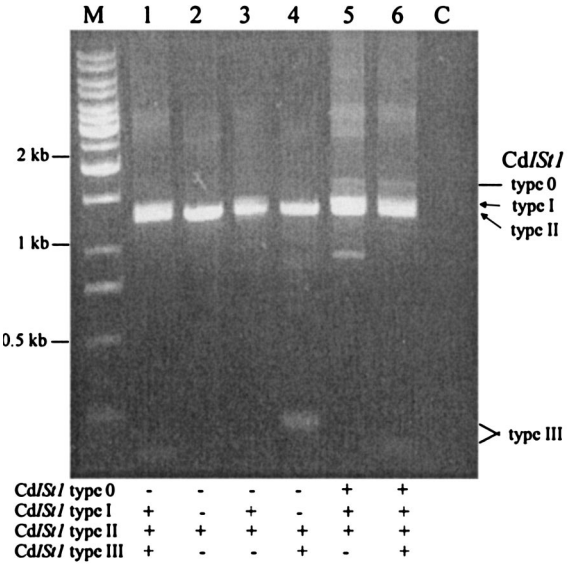

FIG. 2. PCR patterns representing Cd*ISt1* variants from different *C. difficile* isolates. Primers P1 and P2 were used for PCR analysis of genomic DNA prepared from different *C. difficile* strains, as indicated in the text. Lanes are as follows: M, 1-kb ladder (MBI Fermentas); C, negative control (PCR run without template DNA); 1 to 6, chromosomal DNA template derived from strains C34 (1), ST1322/96 (2), 37444 (3), 46307 (4), SE918 (5), and CH3859 (6). The double band at 1,397 and 1,318 bp identifies Cd*ISt1* type I and type II variants, respectively. Bands at approximately 1,650 bp in lanes 5 and 6 correspond to PCR products derived from Cd*ISt1* type 0 variants; bands with sizes of 150 bp (lanes 1 and 6) and ~210 bp (lane 4) correspond to Cd*ISt1* type III PCR products. The size difference results from deletions of different portions of *tlpB* in various isolates.

diodurans (accession number, BAA32389), *Methanosarcina mazei* strain Goel (MM1729; accession number, NP_633753), and *Helicobacter pylori* (TnpA; accession number, AAD11513). The ribozyme activity of Cd*ISt1*a-SE918 was again checked in an in vitro splice assay. This assay demonstrated that Cd*ISt1*a-SE918 is able to carry out self-splicing (data not shown). Obviously, the huge spacing of more than 1,750 bp between the intron core structures and the 3′ splice site in Cd*ISt1*a-SE918 does not interfere with the ribozyme activity of this chimeric element.

To conclude, more than 20 Cd*ISt1* copies from different *C. difficile* strains have been characterized by now (1; this study), and four types have been found: Cd*ISt1*-0 (TlpA/TlpB), Cd*ISt1*-I (ΔTlpA/TlpB), Cd*ISt1*-II (TlpB), and Cd*ISt1*-III (ΔTlpB) (Fig. 1). All Cd*ISt1* types share a complete and functional group I intron component but differ in their IS element components. The Cd*ISt1*-0 variants alone seem to contain a nearly complete IS element of the IS*605* family coding for two transposase-like proteins. We assume that Cd*ISt1*-0 variants may represent the original genetic element that—according to a recent proposal—originated by the fusion of a group I intron and an IS element.

Nucleotide sequence accession numbers. Sequences of Cd*ISt1*l-C34, Cd*ISt1*a-SE918, Cd*ISt1*a-55767, and Cd*ISt1*a-R9304 have been deposited in the EMBL database under accession numbers AJ440753, AJ579718, AJ579716, and AJ579717, respectively.

This work was supported by grant Ei 206/10-1 and 10-2 from the Deutsche Forschungsgemeinschaft.

C.V.E.-S. expresses his special gratitude to the Johannes Gutenberg University in Mainz for providing his group with laboratory space in the Verfügungsgebäude für Forschung und Entwicklung.

REFERENCES

1. **Braun, V., M. Mehlig, M. Moos, M. Rupnik, B. Kalt, D. E. Mahony, and C. von Eichel-Streiber.** 2000. A chimeric ribozyme in Clostridium difficile combines features of group I introns and insertion elements. Mol. Microbiol. **36:**1447–1459.
2. **Edgell, D. R., M. Belfort, and D. A. Shub.** 2000. Barriers to intron promiscuity in bacteria. J. Bacteriol. **182:**5281–5289.
3. **Kersulyte, D., N. S. Akopyants, S. W. Clifton, B. A. Roe, and D. E. Berg.** 1998. Novel sequence organization and insertion specificity of IS605 and IS606: chimaeric transposable elements of Helicobacter pylori. Gene **223:**175–186.
4. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
5. **Mehlig, M., M. Moos, V. Braun, B. Kalt, D. E. Mahony, and C. von Eichel-Streiber.** 2001. Variant toxin B and a functional toxin A produced by Clostridium difficile C34. FEMS Microbiol. Lett. **198:**171–176.