Cloning, Mapping, and Molecular Characterization of the rRNA Operons of Clostridium perfringens

THIERRY GARNIER, BRUNO CANARD, AND STEWART T. COLE*

Laboratoire de Génétique Moléculaire Bactérienne, Institut Pasteur, 28, rue du Docteur Roux, 75724 Paris Cedex 15, France

Received 12 April 1991/Accepted 30 June 1991

All 10 rRNA operons have been situated on the genome map of the anaerobic pathogen Clostridium perfringens. Four of these have been cloned and partially sequenced, and their transcriptional patterns in vivo and in vitro have been examined. Expression of rrnA, rrnB, and rrnE is directed by tandem promoters, P1 and P2, whereas rrnH is the only one to be expressed from a single promoter, which resembles P1. On inspection of the nucleotide sequences of the control regions, several sites which might be involved in the regulation of rrn expression were identified. These include a possible upstream activating region which could be recognized by the C. perfringens equivalent of the Escherichia coli Fis protein and a stringent response target site. Studies of maturation of 16S RNA identified two 5' cleavage sites and sequence analysis showed the dG+dC content of its gene, rrn, to be 52%, which is twice that of the genome.

The gram-positive, anaerobic bacterium Clostridium perfringens is responsible for several serious diseases in humans and domestic animals (20) but remains poorly characterized at the genetic level. Recently, Canard and Cole (5) have established a physical map of the chromosome of C. perfringens CPN50 (Fig. 1), which will serve as the focal point for future genetic studies about topics such as the localization of cloned genes, the genomic distribution of loci encoding virulence factors, and the characterization of deletion and transposon insertion mutants. During the construction of this map, a family of dispersed repeats was detected in a region representing about 30% of the chromosome, and these were shown to correspond to the rRNA operons, rrn, which account for about 1.5% of the organisms' coding capacity. Strikingly, the rrn genes appeared to contain clusters of cleavage sites for rare cutting restriction endonucleases with dG+dC-rich recognition sequences, such as Smal, whereas the genome as a whole has a very low dG+dC content (25%)

For some eubacteria, the rRNA operons (rrn) have been well studied, and a lot of information is available from both Escherichia coli and Bacillus subtilis. The E. coli genome comprises seven copies of the operon (1, 12), while B. subtilis has ten (11). In both organisms, the same canonical organization of the rrn operons is found, and they all consist of rrnD, rrl, and rrf genes, encoding the 16S, 23S, and 5S rRNAs, respectively (1). All three genes are transcribed as a polycistronic RNA from a common regulatory region which generally contains tandem promoters. The synthesis of rRNAs can represent up to 50% of total RNA production and is subject to several sophisticated control mechanisms such as antitermination and the stringent response as well as a complex maturation process (12). Consequently, we wished to establish the genomic organization of the rRNA operons in C. perfringens and to elucidate the nature of the transcriptional control signals, as this could provide insight into global regulatory mechanisms in this important pathogen.

Nucleotide sequence accession numbers. The nucleotide sequence of the 2.5-kb HindIII fragment comprising the rrnB gene encoding the 16S RNA (see Fig. 3) has been deposited at GenBank under accession number M69264. The sequences of the upstream regions of operons rrnH, rrnE, and rrnA (see Fig. 5) have been deposited at GenBank under accession numbers M69265, M69266, and M69277, respectively.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. The type A C. perfringens, strain CPN50 (9), also known as BP6K-N5, was used throughout. The strains of E. coli K-12 employed were MΔ15 [F− Δ(lac-pro) thi ara rpsL recA φ80 (lacZΔM15) (22)] for subcloning and plasmid preparation andJM101 [Δ(lac-pro) supE thi F' traD36 proAB lacPZΔM15 (29)] for M13 propagation.

Genomic banks of DNA from C. perfringens CPN50 were constructed in either plasmid pM1L22p (7) or cosmids pWE15 (26), and fragments were subcloned in pUC18 (29).

Nucleic acid techniques. Chromosomal DNA of C. perfringens CPN50 was prepared as described by Marmur (16), while RNA was extracted by the hot acid-phenol technique from mid-log-phase cultures grown in rich medium (8). The nucleotide sequence of the rrnB gene encoding the 16S rRNA was obtained by site-specific digestion of the rrnB subclone and ligation to appropriate M13 vectors. Recombinants were analyzed on an ABI370A sequencer from Applied Biosystems with a DNA sequencing kit. In some experiments, DNA sequences were obtained by the modified dyeodeoxy method (4).

All RNA analyses, i.e., S1 nuclease mapping, reverse transcriptase-mediated primer extension, and in vitro transcription, were performed exactly as described by Garnier and Cole (8) except that specific oligonucleotides, 32P-labelled with polynucleotide kinase, were used as primers in some experiments. Oligonucleotides were synthesized on an Applied Biosystems 391A synthesizer.

Southern blot hybridizations of genomic DNA were carried out as outlined by Canard and Cole (5).

* Corresponding author.
RESULTS AND DISCUSSION

Genomic organization of the ribosomal operons. While constructing the genetic map of C. perfringens, Canard and Cole (5) characterized a family of small HindIII-Smal fragments which were derived from the rRNA genes. Three clones (pBC15, pBC18, and pBC5) contained HindIII-Smal fragments of 1.5, 3.0, and 1.2 kb, respectively, bearing sequences homologous to the 5' end of the 16S rRNA, whereas others contained fragments of 400 bp from the intergenic region (pBC16) or the 5' end of the 23S rRNA gene (pBC14; see Fig. 2 for precise location). Southern blotting experiments of macrorestriction fragments, or smaller fragments, and hybridization with rRNA-specific probes revealed the existence of 10 copies of the rrn operons, rrnA to rrnJ, in strain CPN50 (Fig. 3A and B) (5) and in several other C. perfringens isolates (our unpublished results). Previously, rrnJ had been overlooked, as the corresponding restriction fragments had run out of the pulsed-field gels employed. The physical mapping studies described in reference 5 and shown in Fig. 3 allowed the 10 rrn operons to be positioned precisely on the genome map (Fig. 1).

Hybridization with specific probes indicated that the cloned DNAs in pBC15, pBC18, and pBC5 originated from rrnA, rrnE, and rrnH, respectively. This study was extended to include probes specific for the 23S and 5S RNAs and revealed that all 10 copies of the operon displayed the canonical organization 5'-rrs (16S), rrl (23S), rrf (5S)-3' (data not shown).

As indicated in Fig. 1, Smal and NruI sites are rare on the C. perfringens chromosome, most of them being clustered in the rRNA operons, each operon containing three sites for Smal (except rrnL, which lacks one) and two for NruI. To determine whether any of the intergenic regions between the 16S and 23S RNA genes contained additional sequences, pBC14 was used as a hybridization probe on conventional Southern blots of genomic DNA digested with Smal or NruI. Two bands were revealed in each case (data not shown), with the major species corresponding to an intergenic region such as that shown in Fig. 2 and 4. The minor band was about 200 bp larger and would have the capacity to code for two tRNAs in addition. This situation is very similar to that of B. subtilis, for which only two of the ten operons (rrnA and rrnO [15, 17]) contain tRNA genes between the 16S and 23S RNA genes. At present, we do not know which of the C. perfringens rrn operons contains the intervening sequences, but we suspect that it might be rrnJ, as its restriction profile is slightly different.
Cloning of \textit{rrnB}. To obtain a complete copy of an \textit{rrn} operon, pBC15 was used as a probe to screen a \textit{C. perfringens} cosmid library and several positively hybridizing clones obtained. One of these was characterized in detail and shown to carry \textit{rrnB}. The rRNA operon is carried by two \textit{HindIII} fragments of about 2.5 kb (pBC17) and 5.1 kb (pBC18), and a restriction map of this 7.6-kb segment is shown in Fig. 2.

Nucleotide sequence and maturation analysis. The complete nucleotide sequence of the 2,487-bp \textit{HindIII} fragment carrying the \textit{rrs} gene, encoding the 16S RNA, from the \textit{rrnB} operon was determined and is presented in Fig. 4. As the sequences of bacterial 16S RNAs are highly conserved, by homology with \textit{B. subtilis} and \textit{E. coli}, the 5’ nucleotide of the mature clostridial RNA was tentatively located at position 710 (Fig. 4). To obtain experimental support for this, two oligonucleotides were synthesized and used in transcript mapping as well as in sequencing the 5’ ends and regulatory regions of \textit{rrnA}, \textit{rrnE}, and \textit{rrnH}. The corresponding nucleotide sequences of \textit{rrnA}, \textit{rrnB}, \textit{rrnE}, and \textit{rrnH} from \textit{C. perfringens} are shown in Fig. 5, where they are aligned with \textit{rrnO} from \textit{B. subtilis}.

Primer 1, positions 733 to 717 (Fig. 5), was employed for precise mapping of the 5’ end of mature 16S RNA, and primer 2, positions 678 to 662, was employed for locating the sites of transcription initiation and processing events. Representative results obtained by using 32P-labelled primer 1 in reverse transcriptase mapping experiments are shown in Fig. 6. Three discrete species were detected and denoted M (mature; four bands), P (premature), and T2 (transcription start site 2). The relative abundance of transcripts M and P was estimated at about 80% and 15 to 20%, respectively, whereas T2 was barely detectable. The major bands of M correspond to nucleotides 709 and 710 in Fig. 5 and might represent the mature 5’ end of the 16S RNA. The two minor bands, positions 708 and 712, may be either due to inadequate processing or the result of nucleotide sequence variation in the different \textit{rrs} genes. Band T2 will be interpreted

FIG. 2. Genetic organization and restriction map of the \textit{rrnB} operon and clones used in this study. The positions of the \textit{rrs} (16S), \textit{rrf} (23S), and \textit{rrf} (5S, inferred) genes are shown, and the scale, in kilobases, is indicated. Abbreviations for restriction sites are as follows: C, \textit{SacII}; D, \textit{HindIII}; E, \textit{EcoRI}; H, \textit{HindIII}; K, \textit{KpnI}; N, \textit{NraI}; P, \textit{HpaI}; S, \textit{SmaI}.

FIG. 3. (A) Chromosomal DNA was double-digested with \textit{MluI} and \textit{SmaI} restriction endonucleases, with only the latter cutting in the rRNA operons. Fragments were then separated by pulsed-field gel electrophoresis, Southern blotted, and hybridized with probes for the 5’ end of \textit{rrs} (lane 1, pBC15, \textit{rrnA}) and the 3’ end of \textit{rrf} (lane 2, pBC18, \textit{rrnB}). In each case, 10 bands were seen, and their sizes were compared with the genome map of Fig. 1. When available, individual nonribosomal probes were then used to confirm the map positions of neighboring operons, and the corresponding fragments are indicated with asterisks. The letters A to J refer to the operons \textit{rrnA} to \textit{rrnJ}. Superimposing films from lanes 1 and 2 identified common bands indicating \textit{SmaI-MluI} fragments bearing both 5’-end (16S) and 3’-end (23S) sequences, e.g., bands G, E, F, D, H, and C from lane 1 and F, D, E, C, I, and B from lane 2, respectively. Finally, in addition to their ribosomal sequences, probes used in lanes 1 and 2 contained extra \textit{rrn}, unique flanking sequences which gave a stronger hybridization signal with their respective fragments. These are indicated with a filled circle in lanes 1 and 2. Sizes (in kilobases) are indicated. (B) Chromosomal DNA from strain CPN50 was digested with \textit{EcoRI} (lane 3) or \textit{HpaI} (lane 4), fractionated on an agarose gel (0.7%), and then hybridized with intragenic probes specific for \textit{rrs} (lane 3) or \textit{rrf} (lane 4). In each case, 10 fragments were detected.
FIG. 4. Nucleotide sequence of the 2.5-kb HindIII fragment comprising the rrsB gene encoding the 16S RNA. The 16S and 23S RNA genes are indicated by arrows above the sequence. The inverted-repeat sequences involved in maturation (Fig. 5, region III) are indicated by arrows below the sequence. The sequence was determined on both strands by the fluorescent dideoxy method.
FIG. 5. Nucleotide sequence alignment of the upstream regions of rnaA, rnaB, rnaE, and rnaI operons from C. perfringens and rno from B. subtilis (17). Arrows designated by T1, T2, and T3 represent the maturation, prematuration, and transcriptional sites found by reverse transcriptase mapping. The corresponding sites in B. subtilis are indicated below. When the homology between the clustidial operon is complete, the nucleotides are in capital letters. The boxed regions I, II, and III are discussed in the text. The -35 and -10 regions of the promoters P1 and P2 are indicated (when present), as well as the positions of oligonucleotides 1 and 2. The first nucleotide of the alignment corresponds to nucleotide 342 in Fig. 4.

Below. By nucleotide sequence homology with B. subtilis, the 3' end of the 16S RNA has been located at nucleotide 2224 (Fig. 4), delimiting a molecule 1,515 to 1,516 bases long. This is 35 and 27 bases shorter than the corresponding RNAs of B. subtilis and E. coli, respectively, and closer in size to the 16S RNA of Streptomyces coelicolor, which is 1,526 bases long (2).

**Processing signals.** In eubacteria, the rrn transcripts have the potential to form tripartite stem-loop structures with the 16S, 23S, and 5S RNAs inside each loop. Analysis of B. subtilis rrn operons (17, 23) led to the definition of a processing site, within the stem, which has been found in all the gram-positive bacteria studied, including mycoplasmas (10, 19, 24) and S. coelicolor (2). Examination of the rnbB sequence reveals one copy of this motif upstream and an inverted form downstream of rrs (underlined in Fig. 4). A second copy (with three mismatches) was found 5' to rrl.

When reverse transcriptase mapping experiments were performed, with either primer 1 (Fig. 6) or primer 2 (Fig. 7), a transcript (P) was found whose 5' end corresponds to this site and which might represent the premature form of the 16S RNA. It can be seen in Fig. 5 that the putative processing site (region III) and mature end points are identical for rnaA, rnb, rnaE, and rnh and the B. subtilis rno operon.

**Putative antitermination signals.** An intriguing feature of the E. coli rrr operons, described by Li et al. (14) and Berg et al. (3), is the presence of an antitermination system that is required to ensure complete transcription of the rRNA genes. Such a mechanism seems to be necessary for two reasons. Firstly, it is needed to avoid problems of polarity, as the rrr operons are transcribed but not translated, and translation is generally necessary for efficient transcription to occur. Secondly, it is well known that rRNAs and their genes have the potential to form complex secondary structures which could act as transcriptional terminators. In E. coli, the antitermination site, BoxA, has been identified and is located about 75 nucleotides upstream of the rrs gene. Comparison of the same region from organisms ranging from archaeabacteria to eubacteria, and especially the gram-positive bacterium B. subtilis, led to the definition of a consensus for BoxA, GCTCTTT. This motif is part of the processing site (Fig. 5, region III) upstream of the C. perfringens 16S and 23S RNA genes. Thus, it is possible that this region has a dual activity; during transcription initiation it would serve as a site for factors involved in antitermination, whereas when the gene has been transcribed, the stem-loop structure can form in the transcript and serve as a processing site.

**Further transcriptional studies.** As described above, when oligonucleotide 1 was used in primer extension experiments (Fig. 6), the 5' ends of three transcripts were detected and two of these, M and P, have been assigned to the maturation and processing sites. The third 5' end, T2, was much less abundant and might correspond to the unprocessed primary transcript. To confirm this hypothesis, in vitro transcription reactions were performed with purified C. perfringens RNA.
polymerase (8) and pBC5 (rrnH), pBC15 (rrnA), or pBC18 (rrnE) linearized with SacII as templates. In each case, a single run-off transcript was produced (Fig. 8), and although rrnA and rrnE both generated RNAs identical in length (about 341 nucleotides, which corresponds to the distance from T2 to the SacII site), rrnH initiated the synthesis of an RNA 18 bases longer.

As the latter species had not been visualized in the primer extension experiment shown in Fig. 6 and since the mature form of 16S RNA represents about 80% of the total transcripts, we increased the sensitivity of the primer extension analysis of rRNA synthesized in vivo by using primer 2, which is located upstream of the M site. As shown in Fig. 7, additional transcriptional initiation sites were now revealed and were named T1H, T1A, T1B, and T1X; these can be explained in two ways. Either they correspond to the uncharacterized rrr operons which have a completely different nucleotide sequence in the promoter region, or, as in B. subtilis and E. coli, C. perfringens rRNA operons are controlled by tandem or multiple promoters, only one of which was active in vitro. To discriminate between these two possibilities, S1 mapping experiments with probes specific for rrnA, rrnB, rrnE, and rrnH, produced with primer 2, were performed (Fig. 9). Two transcripts were detected with rrnA, rrnB, and rrnE, T2 and a longer species T1, whereas with the rrnH probe, only one RNA was found. These larger RNAs probably correspond to transcripts initiating at T1A, T1B, and T1E, whereas T1X (Fig. 7) could be explained by transcription initiating at one of the remaining six, as yet unstudied, rrr operons.

How many operons have tandem promoters? The data presented above are consistent with rrnA, rrnB, and rrnE being expressed from tandem promoters, while rrnH is transcribed from a single promoter. Analysis of the nucleotide sequence alignment of the four promoter regions (Fig. 5) confirmed this interpretation. All four are identical until the prematuration site (P), at which point rrnH diverges. The other three rrr control regions display a high degree of homology up to the transcriptional start T2, which is preceded by the proximal promoter, P2, composed of canonical −35 and −10 regions. From this point to the distal promoter, P1, preceding T1, there is no obvious conservation of sequence or spacing. The promoter responsible for transcription of rrnH appears to be of the distal, or P1, type.

To determine whether all of the five remaining rrr operons were like rrnB, a Southern blotting experiment was performed in which CPN50 DNA had been digested with HpaI or HindIII (which recognizes a degenerate version of the HpaI site) and hybridized with a probe specific for the 5' end of rrs. Ten different fragments were obtained with HpaI (Fig. 3B), which has one site in rrs (Fig. 2), the other being located in upstream sequences. In contrast, digestion with HindIII (data not shown) generated two bands, one of ≈1.3 kb which hybridized intensely and one of ≈3.2 kb which hybridized weakly and originated from rrnH. The HindIII site in the rrs coding sequence is the same as that recognized by HpaI (see Fig. 2), whereas the other site is located in the highly conserved −35 motif of P2. This analysis suggests that 9 of the 10 rrr operons in C. perfringens are probably expressed from tandem promoters, whereas only rrnH is transcribed from a single promoter.
Features of the promoter sites. All the ribosomal operons but one appear to possess dual promoters. It is striking that in vivo, at a fast growth rate (Fig. 7), or in vitro (Fig. 8) the proximal promoter, P2, is predominant. Both promoters display good −10 regions preceded by a TG motif 2 bp upstream. This sequence is known to increase the efficiency of promoters lacking a canonical −35 region (13, 18), but, although it may be required for P1, which lacks a recognizable −35 element, P2 possesses a canonical −35 motif, TGGACA, separated by the optimal distance (17 bp) from the −10 region. We do not know whether the TG motif in addition to a perfect −35 region can transform a strong promoter into a super promoter.

This could explain the extinction of the P1 promoter in vitro, in the presence of P2, although as illustrated by rrnH, P1 is functional on its own. These observations suggest that P1 may require an additional factor for activation. When the sequence alignment was inspected, a region of strong local homology (Fig. 5, region I) was found 120 bp upstream of P1. In the case of the E. coli rrnB operon (21), a similarly positioned site, known as the upstream activation region, which contains three sites for the Fis protein has been described. When present, this upstream activation region increases the efficiency of the promoter P1 20- to 30-fold in vitro. Although no homology can be found between the upstream activation region and region I, it is tempting to suggest that it represents another means of increasing the expression of rRNA operons.

Stringent control. The last points to be analyzed are the sites of transcription initiation. In all seven promoters characterized (four P1 and three P2) a motif was always found whose consensus sequence was GTCG(C/T)T (where the 5′ G is the transcriptional start point). This sequence, also present in the B. subtilis rrn and rna operons, is assumed to be the site required for stringent control (17). It should be noted that stringent promoters in E. coli are characterized by a GC-rich region located at the same place (12, 25), although this is not transcribed, and that such a sequence is not present in mycoplasmas (19, 24), which are intracellular organisms and probably not subjected to substrate limitation.

Features of the 16S RNA sequence. By using the secondary structure representation described by Woese et al. (28), all signature positions for a gram-positive eubacterium, with low dG+dC content (27), were found, but surprisingly, the main region of difference in length between C. perfringens and E. coli (or B. subtilis) 16S RNA is located in one of the stem-loop structures which, according to Woese (27), differentiate the eubacteria from the archaebacteria (data not shown).

The second interesting feature of the nucleotide sequence is the high dG+dC content (52%) of the rrs gene; this is in sharp contrast to that of the upstream and intergenic regions, which are close to the average genomic content of 25% (6). This point and the degree of sequence conservation reflect the high degree of evolutionary constraint imposed by both RNA function and interaction with the large number of other ribosomal components involved. This is strikingly exemplified when microorganisms such as C. perfringens (25%) and S. coelicolor (75%), from opposite ends of the dG−dC spectrum, are compared, as their 16S RNAs are composed of 52 and 59% guanine and cytosine (2), respectively, with an overall homology of 75%.

ACKNOWLEDGMENTS

Special thanks go to Brigitte Saint-Joanis and Anne-Marie Fargues for advice, encouragement, and editorial assistance.

This work was supported by funds from the Institut Pasteur and the Institut National de la Santé et de la Recherche Médicale (CRE 883003).

REFERENCES

