Cloning and Analysis of the Telomere and Terminal Inverted Repeat of the Linear Chromosome of *Streptomyces griseus*

Kohei Goshi, Tetsuya Uchida, Alexander Lezhava, Masayuki Yamasaki, Keiichiro Hiratsu, Hidenori Shinkawa, and Haruyasu Kinashi*

Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter, Hiroshima University, Higashi-Hiroshima 739-8530, Japan

Received 27 December 2001/Accepted 22 March 2002

*Corresponding author. Mailing address: Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8530, Japan. Phone and fax: 81-824-24-7869. E-mail: kinashi@hiroshima-u.ac.jp.

*Streptomyces* species are gram-positive soil bacteria with high G+C base compositions (70 to 74%) in their DNAs. They display complex morphological differentiations and produce many secondary metabolites, including medically useful antibotics. Moreover, all of the *Streptomyces* species studied hitherto carry an approximately 8-Mb linear chromosome (14, 16, 18, 19, 23). *Streptomyces* are also known to frequently possess large linear plasmids (11). Both the linear chromosomes and plasmids contain principally similar structural features, namely, terminal proteins bound to the 5' ends and terminal inverted repeats (TIRs) bound at both ends.

Cohen and his colleagues (1, 25) studied the replication mechanisms of *Streptomyces* linear plasmids pSLA2 and pSCL. They showed that the plasmid DNA replicates bidirectionally from a replication origin near the center of the molecule. Replication in the leading strand is completed at the 3' end, but the lagging strand does not reach the 5' end and leaves a 280-bp gap (1), which was speculated to be filled in by protein-primed DNA synthesis (24). Chen (2) proposed that a foldback secondary structure formed at the 3' protruding end is involved in the patching of the single-stranded gap. Huang et al. (8) sequenced the telomeric regions of a total of six linear chromosomes and plasmids from *Streptomyces* species. All of the first 166 to 168 nucleotide (nt) sequences are similar to each other and can form a foldback structure with Y-shaped loops or a hairpin loop at the 3' ends. Qin and Cohen (21) showed experimentally that replication at the telomere of pSLA2 was consistent with this model and was abolished by the elimination of even 4 nucleotides (nt) from the end.

We have been studying the linear chromosome of *Streptomyces griseus*, including its physical structure, deletion, and circularization (10, 16, 17). However, the nucleotide sequence of the telomere has not been clarified. To answer the question of whether the foldback secondary structure is always present in *Streptomyces* telomeres, we cloned and sequenced the telomere of *Streptomyces griseus*. Both inside ends of the TIR were also cloned and sequenced, which confirmed the size of the TIR and allowed us to formulate a hypothesis on its formation mechanism.

**DNA preparation.** The protease-treated total DNA was prepared as follows. The mycelium from a 50-ml culture was suspended in 10 ml of TSE-sucrose medium (30 mM Tris, 50 mM NaCl, 5 mM EDTA [pH 8.0], 10.3% sucrose) to which 3 ml of lysozyme (5 mg/ml in TSE-sucrose) was added, and the mixture was incubated at 37°C for 30 min. Two milliliters of Actinase E (5 mg/ml in TSE medium; Kaken Pharmaceuticals, Tokyo, Japan) was added and the mixture was incubated at 37°C for 1 h, and then 2 ml of 10% sodium dodecyl sulfate was added and the mixture was shaken for an additional 30 min. The mixture was extracted with phenol, incubated with RNase (100 µg/ml) at 37°C for 1 h, and extracted again with phenol-chloroform. For the DNA not treated with protease, the Actinase E digestion step was omitted and phenylmethylsulfonyl fluoride (final concentration, 2 µg/ml) was added before the sodium dodecyl sulfate treatment to inhibit the strong intrinsic protease activity of *S. griseus*.

**Cloning and analysis of the telomere.** We previously constructed a cosmid library for *S. griseus* strain 2247 (16) and aligned cosmids in some hundreds of kilobases at both chromosomal ends (17). In the terminal cosmid maps, cosmids 9D2 and 6E12 were located at the leftmost and rightmost positions, respectively, and the latter was found to be about 300 bp from the right end (17). The *Sal*I end fragment was subcloned from cosmid 6E12 and named p6E12R (Fig. 1A). After subjecting the total DNA digest to probing using p6E12R, a 2.7-kb *Sal*I fragment was deduced to be the extreme end fragment of the chromosome. Hybridization was carried out using the DIG system (Roche Diagnostics GmbH, Mannheim, Germany) overnight at 70°C in standard buffer.

To clone the extreme end of the chromosome, we first isolated a 11-kb *Afl*II fragment, which was deduced to be the end fragment and was completely separated from bigger fragments in conventional agarose gel electrophoresis (16). Since the 11-kb *Afl*II fragment is located inside of the 24-kb TIR region,
it was impossible to distinguish the right and left end fragments. The AfIII fragment was digested with SalI and then the 2.7-kb end fragment was isolated and force cloned into pUC19 which had been predigested with SalI plus SmaI. In the force cloning (7, 12, 15), the 2.7-kb SalI end fragment, which after protease treatment should still have been carrying a small peptide covalently bound to the 5' end, was directly ligated to the vector.

To eliminate the risk that the extreme end was cut with SalI or AfIII before cloning, protease-treated and nontreated total DNAs were digested with SalI plus AfIII and compared on agarose gel electrophoresis. When probed by p6E12R, the nontreated sample gave a retarded hybridizing signal compared with that of the protease-treated sample (Fig. 1B). Protein-bound DNAs remain at the loading well (16, 18) or move slowly (9) in agarose gel electrophoresis. The assay result shown in Fig. 1B confirmed that the sample used for cloning contained the protein-bound extreme end. Among about 1,000 ampicillin-resistant transformants, three colonies were selected by positive hybridization to p6E12R and named pSGE1, pSGE2, and pSGE3. Nucleotide sequencing showed that pSGE1 and pSGE2 had identical sequences beginning at the SalI cloning site, while the corresponding sequence of pSGE3 was 10 nucleotides (nt) shorter.

To prove that the extreme end was cloned in pSGE1 and pSGE2, end fragments of the chromosome and pSGE1 were compared as follows. Total strain 2247 DNA and pSGE1 were each digested with SacI, KpnI, ApaI, and Eco52I. When pSGE1 was digested with ApaI and Eco52I, SacI and KpnI were added, respectively, because ApaI and Eco52I cannot cut the multicloning site of the vector. When probed by the KpnI end fragment of pSGE1 (Fig. 1A), each pair of 2247 and pSGE1 digests gave hybridizing signals at the same position (Fig. 1C).

This result suggested two possibilities: (i) the chromosomal end is located at the left end of these fragments (Fig. 1A) or (ii) the recognition sites for all of the four restriction endonucleases are located there. There is no chance that the latter possibility is the case, because four recognition sites cannot be accommodated within the detection limit of size differences (about 10 bp in Fig. 1C). Therefore, we concluded that the chromosomal end was cloned in pSGE1 and pSGE2 and that pSGE3 lost a 10-bp end DNA during the cloning process.

The 600-nt sequence of pSGE1 and pSGE2 is shown in Fig. 2A. Unexpectedly, a recognition site for a rare cutter DraI was found close to the end (nt 71 to 76), which added two 116-nt terminal sequence, all of which can form a hairpin loop structure together with palindromes II and III (8). This secondary structure has been suggested to play an important role in the initiation of terminal replication (8, 21). In contrast to these examples, the end sequence of the 2247 chromosome cannot form a foldback secondary structure.
We previously determined the terminal sequence of another linear plasmid, SCP1 (12), where a foldback secondary structure was likewise not possible. The end sequences of SCP1 and the 2247 chromosomes do not show any homology to the typical *Streptomyces* telomere sequences. We have also determined the telomere sequence of the linear plasmid pSLA2-L (7), which coexists with pSLA2-M and pSLA2-S (also known as pSLA2) in *S. rochei* 7434AN4 (13). It showed a similarity to the typical telomere sequences and can make a terminal foldback structure. Therefore, the foldback secondary structure is not absolutely necessary for terminal replication of *Streptomyces* linear genomes. The original protein-priming model reported by Salas (24) does not require it.

**Cloning and analysis of the TIR.** Lezhava et al. (16) showed by restriction analysis that the *S. griseus* 2247 chromosome has a 24-kb TIR. In order to confirm the size of the TIR and to get a hint about its formation mechanism, cloning of both inside ends of the TIR was carried out. Previous studies revealed that the inside ends of the TIR were located on the 2.8-kb *Pst*I fragment of the left end cosmid 9D2 and the 1.4-kb *Pst*I fragment of the right end cosmid 6E12 (17). These two *Pst*I fragments were subcloned to give plasmids pTL and pTR. Comparison of their restriction maps (Fig. 3A) located the left and right inside ends of the TIR on the 350-bp *Sac*II- *Sma*I fragment and the 240-bp *Sac*II fragment, respectively.

The nucleotide sequences of these fragments were determined and are presented for comparison in Fig. 3B. The TIR regions extended 205 nt from the *Sac*II site, which confirmed the size of the TIR to be 24 kb. Frame analysis revealed an open reading frame (ORF) in both TIR regions which started before the leftmost *Pst*I site and stopped at 2 nt from the inside end of the TIR (Fig. 3B). Two nucleotides (nt 179 and 185) differed between these ORFs; this, however, does not affect their amino acid sequences. Homology searches of databases failed to find any proteins with significant homologies to these proteins.

Most of the linear chromosomes and plasmids so far isolated from *Streptomyces* have a TIR at both ends. The sizes of TIRs of linear plasmids are quite different, ranging from 44 bp for SLP2 in *S. lividans* (3) to 95 kb for pPZG101 in *S. rimosus* (6). However, Kalkus et al. (9) reported that pHG201 in *Rhodococcus opacus* has a TIR of only 3 bp, if a sequence of so short a length can be called a TIR. Pandza et al. (20) showed that recombination between pPZG101 and the linear chromosome of *S. rimosus* led to exchange of their ends. The hybrid plasmid generated seemed not to have a TIR based on its restriction map, although the possibility that an extremely short TIR was present could not be excluded. The sizes of *Streptomyces* linear chromosomes are also different, ranging from 24 kb for *S. griseus* (16) to 210 kb for *S. ambofaciens* (14).

The biological role and formation mechanism of TIRs have not been clarified. Fischer et al. (5) reported that recombination of two sigma factor-like genes caused chromosomal arm replacement and generated an extremely long TIR in *S. am-
bofaciens. We also found chromosomal arm replacement in one of the deletion mutants of S. griseus (unpublished result). In the present study, two almost identical ORFs have been located at the inside ends of the TIR. These results suggest that chromosomal arm replacement frequently occurs in Streptomyces by recombination of two homologous ORFs and that this may be a general mechanism by which a TIR is produced.

This hypothesis was strongly supported by analysis of the available S. coelicolor genome sequence (Streptomyces coelicolor genome project; http://www.sanger.ac.uk/Projects/S_coelicolor/). Identical insertion sequence elements are located at the left and right inside ends of the TIR (TIR-L and TIR-R, respectively) of the S. coelicolor chromosome. The one at TIR-L is flanked by direct repeat sequences (CTTTTC) while that at TIR-R is flanked by different sequences (CTTTTC and CTTAAA). This fact indicates that TIR-R was formed by recombination of two identical insertion sequence elements which, however, carried different direct repeats, CTTTTTC and CTTAAA.

Accumulated data suggest that TIR itself is not but the extreme end sequence is essential for terminal replication of Streptomyces linear genomes. The 144-nt telomeric sequence was sufficient for replication of the plasmids artificially constructed from pSLA2 (21). However, it may be possible that replication of a linear plasmid with a very short TIR is supported by another factor coded on the chromosome, as suggested for SLP2 by Redenbach et al. (22). Therefore, a final determination of the essential telomeric structure of Streptomyces linear genomes must wait for more comprehensive studies.

This work was supported by a grant-in-aid for scientific research from the Ministry of Education, Science, Sports and Culture of Japan.

REFERENCES

structures of linear plasmids from *Rhodococcus opacus*. Microbiology 144: 1271–1279.


