Artificial chromosome libraries of Streptomyces coelicolor A3(2) and Planobispora rosea

Rosa Alduina, Simona De Grazia, Luca Dolce, Paola Salerno, Margherita Sosioc, Stefano Donadio, Anna M. Puglia

Abstract

Using an Escherichia coli–Streptomyces shuttle vector derived from a bacterial artificial chromosome (BAC), we developed methodologies for the construction of BAC libraries of filamentous actinomycetes. Libraries of Streptomyces coelicolor, the model actinomycete, and Planobispora rosea, a genetically intractable strain, were constructed. Both libraries have an average insert size of 60 kb, with maximal insert larger than 150 kb. The S. coelicolor library was evaluated by selected hybridisations to DraI fragments and by end sequencing of a few clones. Hybridisation of the P. rosea library to selected probes indicates a good representation of the P. rosea genome and that the library can be used to facilitate the genomic analysis of this actinomycete.

Keywords: Bacterial artificial chromosome; Streptomyces coelicolor; Planobispora rosea

1. Introduction

Bacterial artificial chromosomes (BACs) were developed for cloning large fragments of DNA in Escherichia coli [1,2]. They can be easily isolated from chromosomal DNA and are very stably maintained in the cell, overcoming several disadvantages of yeast artificial chromosomes, such as chimaerism and instability. Consequently, BAC libraries of genomic DNA from animal, plant, fungal and bacterial species have been constructed and have become an important research tool in genome analysis [3–8]. Since genetic methodologies are not available for the vast majority of bacteria, the cloning of clustered genes as a single segment in a BAC may allow their genetic analysis in different backgrounds [6].

The order Actinomycetales, the high GC Gram-positive bacteria consisting of several genera, is industrially relevant for the production of bioactive secondary metabolites, some of which have found important applications in medicine and agriculture [9]. The genes for the production of secondary metabolites, comprising biosynthetic, resistance and regulatory determinants, are organised in clusters that can sometimes reach the 100-kb size [10]. Since efficient methodologies for genetic manipulation are available only for a few actinomycete strains, we have developed a method for the interspecies and intergeneric transfer of large DNA fragments within actinomycetes. This method uses E. coli–Streptomyces shuttle vectors derived from a BAC vector, which are designated with the generic name of ESAC (for E. coli–Streptomyces Artificial Chromosome) [11]. Here we describe methodologies for the construction of ESAC libraries of two actinomycetes, Streptomyces coelicolor and Planobispora rosea.

Received 11 May 2002; received in revised form 24 October 2002; accepted 6 November 2002

First published online 29 November 2002

Keywords: Bacterial artificial chromosome; Streptomyces coelicolor; Planobispora rosea

1 This work is dedicated to the memory of Prof. Franco Tato, former President of the Italian Society of General Microbiology and Microbial Biotechnologies.
2. Materials and methods

2.1. Bacterial strains and plasmids

The \textit{S. coelicolor} strain M145, a prototrophic SCP1\textsuperscript{−}SCP2\textsuperscript{−} derivative of the wild-type A3(2) strain, and \textit{P. rosea} ATCC 53733 \cite{12} were used in this work. Electro-competent \textit{E. coli} DH10B cells were obtained from Life Technologies, Gibco BRL. The ESAC vector, derived from a BAC, can be shuttled between \textit{E. coli}, where it replicates autonomously, and a suitable \textit{Streptomyces} host, where it integrates site-specifically into the chromosome \cite{11}.

2.2. Preparation of high molecular mass DNA

\textit{S. coelicolor} dense spore suspension was added to YEME medium and grown for 40 h at 30°C as described \cite{13}. \textit{P. rosea} was grown in TSB medium for 4 days at 30°C as described \cite{12}. The procedures described in \cite{13} were used to prepare plugs for pulsed field gel electrophoresis (PFGE) analysis.

Partial \textit{Sau}3AI digestion (25 U \text{ml}^{-1}) by limiting \text{Mg}^{2+} concentration was used to prepare large fragments of DNA from the plugs \cite{14}. The plugs were loaded onto a 0.8% SeaPlaque agarose gel, and the DNA was size-fractionated by PFGE on a KLBPulsaphor Electrophoresis Unit (33 s for 18 h at 160 V, 0.25 × TBE, 7°C). Agarose slices containing DNA of the appropriate size (100–300 kb) were cut out and digested with GELase (Epicentre Technologies, Madison, WI, USA) before ligation.

2.3. ESAC library preparation

The size-fractionated DNA was ligated to \textit{Bam}HI-digested and dephosphorylated ESAC vector (1:10 molar ratio) at 16°C overnight. The ligation mixture was desalted using Millipore filters (type VS, 0.025 mm) and then used to transform \textit{E. coli} DH10B electrocompetent cells with a Bio-Rad Gene Pulser instrument (2.5 kV, 100 \text{μF}, 25 \text{μA}). Recombinant clones derived from the \textit{S. coelicolor} and \textit{P. rosea} ligation were designated with Sc and Pr prefixes, respectively (i.e. Sc- and PrESACs). They were stored at −80°C in LB containing 20% glycerol in individual wells of microtiter plates.

2.4. Preparation and analysis of ESAC clones

The ESAC DNAs were prepared by the alkaline lysis method \cite{15}, digested with \textit{Dra}I and fractionated on a 0.8% agarose gel by PFGE (4 s for 4 h; 20 s for 14 h; at 160 V in 0.5 × TBE, 7°C).

DNA sequencing of selected ESACs, purified with the Plasmid Maxi kit (Qiagen) was performed by the SEQ-LAB Sequence Laboratories (Göttingen, Germany). End sequences of five PrESACs have been deposited in the GenBank Data Library under accession numbers BZ305591–BZ305600.

2.5. DNA manipulation and Southern blot analysis

DNA manipulations and Southern hybridisations were performed according to standard protocols \cite{15}. Total chromosomal DNA and DNA fragments were labelled with [α-\textit{32P}]CTP by using the Rediprime II kit (Amersham Pharmacia Biotech). Colony hybridisations were carried out according to the protocol for Hybond-N+ membrane kit (Amersham Pharmacia Biotech). Specific \textit{S. coelicolor} \textit{Dra}I fragments were obtained and recovered as described \cite{13}. Hybridisation with oligonucleotide Pep8 was performed as reported previously \cite{16}. The plasmid pGRP17 carrying three genes of \textit{S. coelicolor dnaK} locus was described in \cite{17}.

3. Results

3.1. Construction of a \textit{S. coelicolor} ESAC library

\textit{S. coelicolor} A3(2) is the best genetically characterised actinomycete \cite{9}. It has an 8.6-Mb linear chromosome whose DNA sequencing is now complete. A detailed genetic map is also available \cite{18}. Thus, we selected \textit{S. coelicolor} for setting up the methodologies for library construction. Initially, individual ScESACs were evaluated after \textit{Dra}I digestion and PFGE separation. \textit{Dra}I digestion is expected to release the insert as a single fragment in most cases, since only seven \textit{Dra}I sites are present in the \textit{S. coelicolor} genome \cite{13}. Thus, a total of three \textit{Dra}I fragments (two containing only vector sequences) are expected for most ESACs carrying \textit{S. coelicolor} DNA. Three relevant features were considered: the cloning efficiency; the fraction of ESACs carrying an insert larger than 40 kb; and the average insert size for the ESACs carrying inserts larger than 40 kb. Some steps revealed to be crucial for library construction; for example, the extent of partial \textit{Sau3AI} digestion of agarose-embedded DNA. Notwithstanding that the partially digested DNA was always fractionated by PFGE and the same interval of fragment sizes was recovered, results varied considerably depending on the extent of partial digestion. In one experiment, a 7-min \textit{Sau3AI} digestion yielded an 87% cloning efficiency with an average insert size of 70 kb. In another experiment, a 20-min digestion resulted in a 20% cloning efficiency and an average insert size of 24 kb (data not shown). These differences might be due to the variable amount of DNA recovered from the gel slice and to the extent of trapping of smaller fragments present in the excised fraction.

A total of 1200 ScESAC clones, resulting from four independent ligations, were individually analysed by \textit{Dra}I digestion and PFGE. Only the clones carrying an
insert larger than 40 kb were considered for further analyses. A total of 278 clones (or 23% of those analysed), with inserts ranging from 40 kb to 200 kb, were then collected and analysed as described below. The distribution of insert sizes of these 278 ScESACs is illustrated in Fig. 1. All together, these clones cover about 17 Mb of S. coelicolor DNA (a two-fold coverage of the genome), with an average insert size of 61 kb. A significant number of clones (65 out of this set of 278) contain inserts larger than 80 kb (Fig. 1). A first hint that the inserts derived from S. coelicolor DNA came from the low frequency of additional DraI sites observed: only four out of the 278 ScESACs contain a DraI site in the insert. This result was confirmed by observing that 70, randomly chosen ScESAC-carrying colonies were all positive upon hybridisation with labelled S. coelicolor genomic DNA (data not shown), indicating that most, if not all inserts derive from the DNA of this strain.

3.2. Quality control of the library

In order to determine the representativity of the library, we carried out hybridisations with selected S. coelicolor DraI fragments on 166 random clones. The DraI fragments, after separation by PFGE (Fig. 2A), were eluted from the gel and labelled. A typical hybridisation experiment is illustrated in Fig. 2B. A total of 11 ScESACs, representing all together 600 kb of insert DNA, hybridised with the 530-kb DraI-F fragment. Using the 820-kb DraI-D fragment as a probe, we identified 13 positive clones encompassing a total of 840 kb. Six clones, containing 781 kb of DNA, were positive to the 280-kb DraI-G fragment. The results of these hybridisations indicate a reasonable correspondence between probe size and the sum of inserts from the positive ScESACs, suggesting that no significant bias towards selected regions of the chromosome was introduced during the construction of the library. Furthermore, all clones but two hybridised with only one DraI fragment. We have not investigated whether these two clones, hybridising to two, non-adjacent fragments (DraI-D and -G in one case, DraI-F and -G in the other) resulted from chimaeric inserts or from spurious hybridisations. In any case, this result suggests that a high proportion of ScESACs contain an insert from a unique region of the S. coelicolor chromosome.

![Fig. 1. Size distribution of 278 ScESAC clones. Insert size was estimated by digestion of the plasmids with DraI, followed by analysis of the digestion products by PFGE.](image)

![Fig. 2. Hybridisation of ScESAC clones with the S. coelicolor DraI-F fragment. A: PFGE resolution of DraI fragments from S. coelicolor M145 (lane 1) and yeast chromosomes (lane 2). Size of DraI (A–H) fragments: 2100, 1770, 1650, 820, 690, 530, 280, 115 kb. B: An example of colony hybridisation with 67 colonies carrying ScESACs with inserts larger than 40 kb.](image)
Ten random ScESACs were subjected to sequencing from both insert ends. Except for one case, where no readable sequence was obtained from just one end, all sequences unambiguously identified a single region in the *S. coelicolor* genome (Table 1), confirming their origin from this strain. Furthermore, for the nine ScESACs for which it could be established, the inserts span a contiguous region of the chromosome. The insert sizes for each ScESAC, as deduced from DNA sequencing or as measured after PFGE, are reported in Table 1, together with their exact position in the *S. coelicolor* chromosome. The close correspondence between the measured and the deduced insert size (the highest difference is 15%) is consistent with each ScESAC carrying a contiguous, intact region of the chromosome. Finally, the large span of the genome covered by the nine, randomly selected ScESACs (from 2.7 to 7.7 Mb; Table 1) further supports the random origin of the inserts.

The results obtained thus far, from selected hybridisations to ScESAC sequencing, indicate that the methodology used for library construction can produce a representative library, with a reasonable fraction of clones carrying inserts larger than 40 kb and a small proportion, if any, of chimaeric clones. On this background, we applied the same approach to the construction of a *P. rosea* library.

### 3.3. Construction of a *P. rosea* ESAC library

*P. rosea* is an example of a genetically intractable strain, belonging to one of the less studied genera of actinomycetes. It produces the thiazolylpeptide antibiotic GE2270, a potent inhibitor of bacterial elongation factor Tu [19]. This strain thus represented a good candidate for applying the procedures for library construction described above. A preliminary analysis of *P. rosea* genomic DNA by PFGE showed at least 13 distinct DraI fragments, ranging in size from 180 kb to 2.0 Mb, summing up to 8.7 Mb (data not shown). This result indicated a genome size for *P. rosea* similar to that of *S. coelicolor* and that DraI digestion could also be used for analysis of the PrESAC clones. After electrottransformation of *E. coli* cells with four ligation mixtures, a total of 2620 kanamycin- and sucrose-resistant colonies were obtained. Approximately 860 clones (22% of the total) were analysed for insert size by PFGE. This analysis indicated that 262 PrESACs (33% of those analysed) had an insert larger than 40 kb. The insert distribution for these 262 PrESACs is illustrated in Fig. 3. The largest insert found in this set is about 150 kb, and the average insert size is 60 kb. All together these clones contain a total of 15.7 Mb. Seventy insert-carrying PrESACs were all positive upon hybridisation with labelled *P. rosea* genomic DNA, indicating that the inserts derive from the DNA of this strain. We also observed that 103 PrESACs (12% of those analysed) contain a DraI site within the insert. On the basis of the observed percentages, we can estimate that, among the 2620 kanamycin- and sucrose-resistant colonies, there are about 800 PrESACs containing inserts larger than 40 kb. Assuming an average insert size of 60 kb for all these clones, the *P. rosea* ESAC library should thus contain a total of 48 Mb of *P. rosea* DNA, or a 5.5-fold coverage, of its genome.

Very little information is available on *P. rosea* genes, so that an analysis comparable to that performed with the ScESACs could not be carried out on the *P. rosea* clones. However, recent data indicate that its genome contains multiple gene clusters for non-ribosomal peptide synthetases [16] and a single tuf gene encoding elongation factor Tu [19]. We thus screened the 262 colonies carrying the PrESACs of Fig. 3 with Pep8 [16], a consensus oligonucleotide derived from the gene segment encoding the thiolation domain of peptide synthases [20], with a 900-bp fragment containing the *P. rosea* tuf gene [19] and with a 3-kb fragment carrying three genes from the *S. coelicolor* *dnaK* locus, present as a single copy in actinomycete genomes [17]. These hybridisations resulted in the identification of 17, two and two clones positive to the Pep8, tuf and *dnaK* probes, respectively.

#### Table 1

End sequencing of 10 ScESACs

<table>
<thead>
<tr>
<th>Clone</th>
<th>Position on chromosome</th>
<th>Insert sizea (kb)</th>
<th>Sequence</th>
<th>PFGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A12</td>
<td>4100153</td>
<td>4168721</td>
<td>68.5</td>
<td>70</td>
</tr>
<tr>
<td>1C1</td>
<td>3383391</td>
<td>3441079</td>
<td>57.6</td>
<td>55</td>
</tr>
<tr>
<td>1H2</td>
<td>7577092</td>
<td>7654387</td>
<td>77.3</td>
<td>80</td>
</tr>
<tr>
<td>1B5</td>
<td>7018244</td>
<td>6890668</td>
<td>128.2</td>
<td>120</td>
</tr>
<tr>
<td>2F9</td>
<td>4793041</td>
<td>4880653</td>
<td>87.6</td>
<td>88</td>
</tr>
<tr>
<td>5A11</td>
<td>6638540</td>
<td>6717162</td>
<td>78.6</td>
<td>90</td>
</tr>
<tr>
<td>5F5</td>
<td>5647326</td>
<td>5555519</td>
<td>91.8</td>
<td>80</td>
</tr>
<tr>
<td>1C9</td>
<td>2842843</td>
<td>2747329</td>
<td>95.5</td>
<td>100</td>
</tr>
<tr>
<td>2F7</td>
<td>6344260</td>
<td>6440586</td>
<td>96.3</td>
<td>90</td>
</tr>
<tr>
<td>5C10</td>
<td>6407660</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

*Established by Blast searches of *S. coelicolor* database (www.sanger.ac.uk/Projects/S_coelicolor/blast_server.shtml). From refers to the insert end next to the T7 promoter, To to the insert end next to the SP6 promoter.

*Insert size calculated from the insert span as in the previous column, or measured by PFGE.
End-sequence analysis of five randomly chosen PrESACs revealed that seven out of 10 sequence stretches apparently encode proteins with homology in databases. As expected, most of the best matching sequences derive from *S. coelicolor*.

4. Discussion

In this paper we reported the construction of BAC libraries of GC-rich actinomycete DNA. During the development of procedures for library construction, we have considered three critical elements: the fraction of inserts larger than 40 kb, the average insert size and the largest insert. The first element served as a rough comparison between a BAC and a cosmids library. In fact, the latter has usually an upper cloning limit of about 40 kb. The second and third elements are normally employed to qualify BAC libraries. Judging from the equivalent number of clones analysed, the *S. coelicolor* and *P. rosea* libraries are comparable in terms of the three elements: 23 and 33% as the fraction of inserts larger than 40 kb, 61 and 60 kb as the average insert size, and 200 and 150 kb as the largest insert, respectively. We have not attempted to construct a high-coverage library of *S. coelicolor*. However, the methodologies developed and optimised for this strain yielded a 5.5-fold coverage of the *P. rosea* genome. So, we expect that similar results could be obtained with *S. coelicolor*.

The end sequencing of nine ScESACs clearly indicated the random origin of the inserts from the chromosome, the absence of major insertions or deletions, and the absence of insert concatemers. This is consistent with the hybridisations to selected fragments from the *S. coelicolor* chromosome. Since the PrESAC library was constructed with the same methodology, we can reasonably assume that it is of comparable quality. The only indications we have derive from the use of the Pep8, *tuf* and *dnaK* locus probes. Previous results indicate that *P. rosea* contains at least eight unlinked loci of Pep8-hybridising bands [16] and a single *tuf* gene [18]. The results reported here are quite consistent, since 262 clones, representing about 1.8 genomic equivalents, yielded 17 and two positive clones, respectively. Furthermore, just two *dnaK* locus positive clones were observed, as expected from a single-copy cluster [17]. To our knowledge, the PrESAC library is the first BAC library of an actinomycete DNA, with a reasonable genomic coverage and inserts ranging from 40 to 150 kb.

Analysing a *Mycobacterium tuberculosis* BAC library, Brosch et al. failed to obtain inserts larger than 110 kb [7]. These authors suggested that there could be an upper cloning limit for large inserts resulting either from lethal overexpression of certain genes or from the particular procedure necessary for the lysis of mycobacteria cell walls during chromosomal DNA preparation in agarose plugs. Considering the combined results of the ScESACs and the PrESACs, we observed that 34 out of 540 clones (about 6.3%) carry inserts larger than 100 kb (Figs. 1 and 3). However, inserts up to 150–200 kb could be observed, even if at a low frequency. The fact that inserts larger than 110 kb could be obtained from actinomycetes with a GC content higher than that of mycobacteria, suggests that the upper limit in insert size might be due to technical difficulties, and that improvements in the procedures for library construction may provide an increased fraction of large inserts.

Previous results [11] have shown that ESACs, carrying large inserts of actinomycete DNA, can be introduced into a genetically accessible strain such as *Streptomyces lividans*, where they are stably maintained in an integrated form in its chromosome. We also know that it is possible to construct *S. lividans* strains carrying selected PrESAC clones (R.A. and A.M.P., unpublished results). This should allow genetic analysis of *P. rosea* gene clusters, and eventually of other clusters from the many actinomycetes that are now genetically tractable. Since many soil microbes produce bioactive secondary metabolites, and genes required for their production are often clustered together with resistance and regulatory determinants, ESAC libraries can represent an additional tool for the search for novel bioactive metabolites from microbial sources.

Acknowledgements

We thank Benedetta Catalano and Valentina Di Caro for performing some experiments on the PrESAC library, Francesco Giusino for help at the initial stage of this work and Maurizio Noto for technical assistance. This work was supported by the Italian CNR, PF Biotecnologie and MIUR (Cofinanziamento programmi di rilevante interesse nazionale, 2000).

References

specific genome regions in which its transposons tend to cluster. Bio-