National Library of Canada — Bibliothèque nationale du Canada

Canadian Theses Service — Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

UNIVERSITY OF ALBERTA

MOLECULAR STUDIES OF A LINEAR PLASMID FROM *STREPTOMYCES CLAVULIGERUS*

BY

XIAONING WU

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MICROBIOLOGY

EDMONTON, ALBERTA

FALL, 1990

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-64855-4

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: XIAONING WU

TITLE OF THESIS: MOLECULAR STUDIES OF A LINEAR PLASMID FROM *STREPTOMYCES CLAVULIGERUS*

DEGREE: DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED: 1990

PERMISSION IS GRANTED TO THE UNIVERSITY OF ALBERTA LIBRARY TO REPRODUCE SINGLE COPIES OF THIS THESIS AND TO LEND OR SELL SUCH COPIES FOR PRIVATE, SCHOLARLY OR SCIENTIFIC RESEARCH PURPOSES ONLY.

THE AUTHOR RESERVES OTHER PUBLICATION RIGHTS, AND NEITHER THE THESIS NOR EXTENSIVE EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT THE AUTHOR'S WRITTEN PERMISSION.

Wu Xiaoning

Biotechnology Laboratory

Room 237-Wesbrook Building

6174 University Boulevard

University of British Columbia

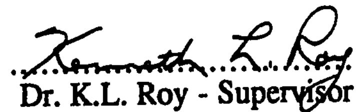
Vancouver, B.C. Canada V6T 1W5

Date: July 11, 1990.

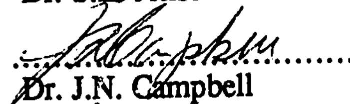
UNIVERSITY OF ALBERTA

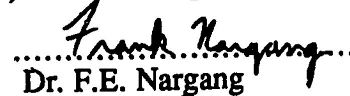
FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled MOLECULAR STUDIES OF A LINEAR PLASMID FROM *STREPTOMYCES CLAVULIGERUS* submitted by XIAONING WU in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

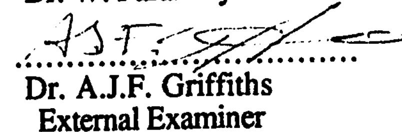

Dr. K.L. Roy - Supervisor

Dr. S.E. Jensen


Dr. J.N. Campbell


Dr. F.E. Nargang

Dr. W. Paranchych


Dr. A.J.F. Griffiths
External Examiner

Date: June 8, 1990

Dedicated to my wife Weixing,

to my parents and my two younger brothers.

ABSTRACT

Linear plasmids are unusual genetic elements which exist in a variety of organisms. Almost all of the linear plasmids in prokaryotes are found in the genus *Streptomyces*. To better understand gene organization and regulation of linear plasmids in *Streptomyces*, the complete nucleotide sequence of a small linear plasmid (pSCL) from *Streptomyces clavuligerus* has been determined. The sequence data reveal that this plasmid is 11, 696 base pairs (bp) in length with a 71.9% G+C content and has 969 bp inverted terminal repeats. Comparison of the inverted terminal sequence of pSCL with that of a linear plasmid from *Streptomyces rochei* shows that the two terminal sequences have a rather high degree of similarity (65.4%). Moreover, several small inverted repeats found in the long terminal sequences of both plasmids are highly conserved. Open reading frame (ORF) prediction and G+C percentage analysis indicate that pSCL has 5 highly probable protein-coding ORFs. Two RNA transcripts from pSCL were detected in *S. clavuligerus* and they do not correspond to any of the 5 highly probable ORFs. The larger transcript (980 nucleotides) corresponds to an ORF, indicating that it is likely to be a mRNA for a protein. On the other hand, the smaller transcript (450 nucleotides) does not correspond to any ORF. However, its 5' end is complementary to the 5' end of a predicted ORF, suggesting that it may function as an antisense RNA. Interestingly, it was shown that the larger RNA was transcribed at a high level during the early stage (12 hours) of cell growth, then decreased and remained at a lower level through the later stages of growth, while the transcriptional level of the smaller RNA remained relatively constant through all stages of cell growth. Since pSCL is the first prokaryotic linear plasmid to be completely sequenced, the information provided by this study may enable us to further understand the genetic behavior of other linear plasmids in the genus *Streptomyces*, particularly the giant linear plasmids.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor, Dr. K.L. Roy, for his excellent supervision, enthusiastic encouragement as well as endless support and faith during my entire graduate work in his lab. Undoubtedly, I have learned so much, not only from his invaluable expertise and experience, but also from his rigorous attitude in science.

I also sincerely appreciate all members of my supervisory committee, Drs. S.E. Jensen, J.N. Campbell, F.E. Nargang, D.W.S. Westlake, W. Paranchych, for their continuous guidance and support through these years. In particular, without the helpful and friendly cooperation with Dr. S.E. Jensen and the people in her lab, this work would not be accomplished. I still remember the encouraging and faithful words from Dr. J.N. Campbell, the first person who introduced me to this department.

I am grateful to Dr. J. Doran for his valuable suggestions and technical assistance in my first two years. Also, I have really enjoyed working with the other graduate students in this lab, W. Henry and D. Spodafora. There is no doubt that the useful discussions between us and the enthusiastic atmosphere in the lab made the hard work much easier. Special thanks go to J. Hoyem for his expertise in the computer analysis. In addition, as the first graduate student from China in this department, I thank all people in the department, who helped and supported me during the years. The warm hospitality will never be forgotten.

Of course, I deeply appreciate my family, my parents and especially my wife Weixing, for their endless encouragement and understanding. I also thank one special friend, Gary Ritzel, for his continuous help and encouragement since I came to Canada.

Finally, I acknowledge the Department of Microbiology for the graduate studentship, the Alberta Heritage Foundation for Medical Research for the full-time studentship and the University of Alberta for the dissertation fellowship.

TABLE OF CONTENTS

Page

Abstract v

Acknowledgements vi

List of Tables ix

List of Figures x

List of Abbreviations xi

1. INTRODUCTION 1

    1.1 Chromosomal and Extrachromosomal Elements in *Streptomyces* 2

    1.2 Interactions Between Extrachromosomal Elements and Host Chromosomal DNA 11

    1.3 Linear Plasmids in *Streptomyces* 13

    1.4 Linear Plasmids in Eukaryotes 20

    1.5 DNA Viruses with Protein Linked to the 5' Termini of Their Genomes 25

    1.6 Objectives of This Study 32

2. MATERIALS AND METHODS 35

    2.1 Materials 35

    2.2 DNA Isolation and Manipulations 37

    2.3 Cloning of Internal Fragments and Terminal Fragments of pSCL 39

    2.4 Cloning of 3 Small *Sal* I Fragments by Polymerase Chain Reaction 41

    2.5 DNA Sequencing 42

    2.6 RNA Isolation, Northern Transfer and Hybridization 43

    2.7 Mapping of 5'-Termini of Two RNA Transcripts 45

    2.8 *In Vitro* Transcription and Translation Using *S. lividans* Extracts 46

3. RESULTS 47

3.1 Cloning Strategies 47

3.2 Sequencing Strategies 52

3.3 Complete Sequence of pSCL 58

3.4 Open Reading Frame (ORF) Analysis 77

3.5 Properties of the 5 Predicted Polypeptides Deduced from the 5 Most Probable ORFs 84

3.6 RNA Transcripts from pSCL 86

3.7 Characterization of Two RNA Transcripts 93

3.8 *In Vitro* Coupled Transcription and Translation Using *S. lividans* Cell Extracts 103

4. DISCUSSION 106

5. BIBLIOGRAPHY 127

6. APPENDIX 140

LIST OF TABLES

Table	Description	Page
Table 1	List of Some Linear Plasmids Found in *Streptomyces*	18
Table 2	List of Some Linear Plasmids Found in Eukaryotes	21
Table 3	List of Some DNA Viruses with Protein Attached to the Ends of DNA Genomes	26
Table 4	Summary of Positions of the Most Probable ORFs in pSCL	82
Table 5	Summary of the Anticipated Promoter Regions of the Two RNA Transcripts from pSCL	96

LIST OF FIGURES

Figure	Description	Page
Figure 1	Autoradiograph of a 6% Polyacrylamide Denaturing Gel Demonstrating the Terminal Sequences of pSCL in pUC119, Using Sanger's Chain Termination Method	54
Figure 2	Autoradiograph of a 20% Polyacrylamide Denaturing Gel Demonstrating the Sequence of the 3' Termini of pSCL DNA, Using the Maxam and Gilbert Chemical Cleavage Method	56
Figure 3	Complete Double Stranded Nucleotide Sequence of pSCL	59
Figure 4	Alignment of the 969 bp Left Terminal Sequence of pSCL with the 969 bp Right Terminal Sequence of pSCL	72
Figure 5	Comparison of the 969 bp Left Terminal Sequence of pSCL with the 838 bp Left Terminal Sequence of pSLA2	75
Figure 6	Graphs Illustrating the %G+C in the Three Codon Positions for the Five Most Probable ORFs	78
Figure 7	Northern Transfer and Hybridization Analysis of Total RNA Isolated from *S. lividans* and *S. clavuligerus*, Using pSCL DNA as a Probe	87
Figure 8	Autoradiograph of a 6% Polyacrylamide Denaturing Gel Used for Mapping the 5' Terminus of the Large RNA Transcript of pSCL	89
Figure 9	Autoradiograph of a 6% Polyacrylamide Denaturing Gel Used for Mapping the 5' Terminus of the Small RNA Transcript of pSCL	91
Figure 10	Graphs Illustrating the %G+C in the Three Codon Positions for the Two RNA Molecules Shown to be Transcripts of pSCL	94
Figure 11	Potential Stem-Loop Structures Close to the 3' Termini of the two RNA Transcripts from pSCL	99
Figure 12	Northern Transfer and Hybridization Analysis of Total RNA Preparations from Different Stages of Cell Growth	101
Figure 13	Incorporation of ^{35}S-Methionine into Hot TCA-Precipitable Material, Using Various DNA Templates with *S. lividans* Cell Extracts	104

LIST OF ABBREVIATIONS

AMV	Avian myeloblastosis virus
ATP	Adenosine-5'-triphosphate
bp	Base pair(s)
BSA	Bovine serum albumin
cccDNA	Covalently closed circular DNA
Ci	Curie
dATP	2'-deoxyadenosine-5'-triphosphate
dC	2'-deoxycytidine
dCTP	2'-deoxycytidine-5'-triphosphate
DEAE	Diethylaminoethyl
dG	2'-deoxyguanosine
dGTP	2'-deoxyguanosine-5'-triphosphate
DMSO	Dimethylsulfoxide
ddATP	2', 3'-deoxyadenosine-5'-triphosphate
ddCTP	2', 3'-deoxycytidine-5'-triphosphate
ddGTP	2', 3'-deoxyguanosine-5'-triphosphate
ddTTP	2', 3'-deoxythymidine-5'-triphosphate
DNase	Deoxyribonuclease
DTT	Dithiothreitol
dTTP	2'-deoxythymidine-5'-triphosphate
EDTA	Ethylenediaminetetraacetic acid
kb	1000 base(s) or base pairs
kcal	1000 calorie
Mb	1,000,000 base(s) or base pairs

mRNA	Messenger RNA
nt	Nucleotide(s)
ORF	Open reading frame
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFG	Pulsed field gel electrophoresis
rDNA	Ribosomal DNA
RNase	Ribonuclease
SDS	Sodium dodecylsulfate
7-deaza dGTP	2'-deoxy-7-deazaguanosine-5'-triphosphate
TCA	Trichloroacetic acid
TE	Tris-EDTA buffer
TEA	Tris-EDTA acetate buffer
TEB	Tris-EDTA borate buffer
tRNA	Transfer RNA
UV	Ultraviolet

1. INTRODUCTION

The genus *Streptomyces* comprises a group of Gram-positive soil microorganisms belonging to the family Streptomycetaceae. These organisms have a fairly complex life cycle. When they undergo vegetative growth after spore germination, they produce an extensively branching primary or substrate mycelium. The vegetative mycelia further develop aerial mycelia or aerial hyphae. These aerial mycelia then differentiate to form chains of spores (reviewed by Kutzner, 1982). Many members of this genus are capable of producing antibiotics. Over 60% of known, naturally-occurring antibiotics come from the genus *Streptomyces*. Some species of *Streptomyces* are even able to produce more than one antibiotic. Many of these antibiotics have been extensively used in medicine for control of infectious diseases. Because of the medical and industrial importance of these organisms, much attention has been focused on their genetics, particularly the organization and regulation of genes involved in antibiotic biosynthesis. With the development of recombinant DNA technology, considerable progress has been made. Many genes involved in antibiotic biosynthesis have been cloned and characterized (reviewed by Seno and Baltz, 1989). New recombinant DNA techniques even permit us to create new antibiotics which do not exist in nature (Hopwood *et al.*, 1985a). However, genetic studies of *Streptomyces* are complicated by the unusual genetic behavior of these organisms, such as differentiation and genome rearrangements, which frequently results in change or loss of phenotypic traits (Hutter and Hintermann, 1985; Schrempf, 1985; Cullum *et al.*, 1986). Therefore, extrachromosomal elements in *Streptomyces*, including plasmids, transposons, insertion sequences (IS) and bacteriophages, have become attractive for further detailed studies of gene regulation and expression in these organisms. Many extrachromosomal elements are widely spread, structurally small and simple, and relatively easy to obtain. Some of them actually encode structural or regulatory genes involved in antibiotic biosynthesis. Moreover, some extrachromosomal elements are

believed to be involved in the genetic instability of some *Streptomyces*. The genetic behavior of these extrachromosomal elements could also serve as model systems for further studies of genomic instability in *Streptomyces* spp. Finally, a better understanding of these extrachromosomal elements will undoubtedly enable us to construct new cloning vectors for genetic engineering of these organisms.

1.1 Chromosomal and Extrachromosomal Elements in *Streptomyces*

A. Genomes of *Streptomyces*

Streptomyces, like other typical prokaryotes, contains one major circular chromosome. The genome size was estimated (Benigni *et al.*, 1975) to be approximately 1.5 to 2 times as large as that of *E. coli* (4.7 Mb, Kohara *et al.*, 1987; Smith *et al.*, 1987). However, at the present time there has been no reliable estimate published in which pulsed field gel electrophoresis (PFG) methods were employed. Genome complexity has been analyzed by Antonov *et al.*, (1977) in *S. coelicolor* A3(2), the only species of the genus *Streptomyces* with well-established genetics. It has been estimated that 2% of the genome consists of "foldback" DNA and that about 5% represents repetitive sequences (largely rDNA sequences and possibly tRNA gene sequences, Keen, C.L. and K.L. Roy, unpublished results) in about four copies per genome. A chromosomal linkage map of *S. coelicolor* A3(2) has been established by using conjugation and protoplast fusion. More than 100 genes, including auxotrophic, drug-resistance, temperature-sensitive, morphogenic and some antibiotic biosynthesis genes, have been mapped on the genome (Chater and Hopwood, 1983).

An unusual characteristic of *Streptomyces* genomes is their high G+C content of about 70% to 74%. This high G+C content results in a strong, biased distribution of G+C composition at the third position of codons within a protein-coding region (Bibb *et al.*, 1984). Normally, for a given protein-coding region, G+C percentages at the 3 positions in

codons are about 70% at the first position, 50% at the second and over 90% at the third. This characteristic distribution has been used as a criterion for structural genes isolated from *Streptomyces*. Because of the high G+C content, it has often been observed that restriction endonucleases recognizing sites low in G+C cut *Streptomyces* DNA at lower frequencies than those enzymes recognizing sites high in G+C. Moreover, it is has been suggested that the high G+C content may reduce the possibility of UV damage.

B. Plasmids

The first *Streptomyces* plasmid which was isolated physically was SCP2 from *S. coelicolor*. With the development of efficient and reliable procedures for the isolation of plasmids it was shown that many *Streptomyces* strains isolated from nature carry plasmids (Hopwood *et al.*, 1986b). In most cases, plasmids were detected as covalently closed circular DNA molecules. Occasionally, some plasmids were found to exist as linear molecules (Hayakawa *et al.*, 1979; Kinashi and Shimaji, 1987). The sizes of *Streptomyces* plasmids vary from several kilobases to several hundred kilobases. They have copy numbers in the range of less than one to several hundred per chromosome. The G+C content in *Streptomyces* plasmids is believed to be similar to the host DNA. Recently, the base composition of the plasmid pIJ101 from *S. coelicolor* has been reported to be 73% G+C, based on its nucleotide sequence (Kendall and Cohen, 1988).

Most of these plasmids are not known to confer a distinctive phenotype on their hosts. However, a considerable number of the plasmids can give a pock formation phenotype; in other words, they can transfer from one host to another. *Streptomyces* plasmids have also been shown to promote fertility, i.e. host chromosomal transfer and recombination. In some cases, plasmids do carry antibiotic resistance and biosynthesis genes. For instance, SCP1 from *S. coelicolor* A3(2) (Kirby and Hopwood, 1977), and pSV1 from *S. violaceus-ruber* SANK 95570 (Aguilar and Hopwood, 1982) encode genes

for the production of, and resistance to, methylenomycin. Some plasmids have been suggested to play a regulatory role in antibiotic production (Kirby, 1978).

Replication and copy number control of *Streptomyces* plasmids are not well understood. Generally speaking, a minimal size less than 2 kb is sufficient for replication of *Streptomyces* plasmids (Kieser *et al.*, 1982). Nucleotide sequence data and genetic analysis together will be required to reveal the mechanisms of *Streptomyces* plasmid replication and copy number control.

Most *Streptomyces* plasmids are stable. It has been reported, however, that some of them undergo structural changes. Bibb and Hopwood (1981) reported that a 9 kb segment of SCP2* was duplicated to give four tandemly repeated copies when this plasmid was transferred from *S. coelicolor* A3(2) to *S. parvulus* ATCC12434. Jaurin and Cohen (1984) also reported a 1.1 kb duplication of a region in pSLP101. Deletions of plasmid DNA to yield small derivatives were also observed in pIJ101 (Kieser *et al.*, 1982). Interestingly, some *Streptomyces* plasmids normally integrate into the host chromosome in one strain, but exist as autonomously replicating plasmids in another strain. For instance, the plasmids SLP1, pIJ101 and pIJ408 were found to be chromosomal in *S. coelicolor* A3(2). When crossed with other, plasmid-free strains, such as *S. lividans* 66, *S. parvulus* ATCC12434 and *S. glaucescens* ETH 22794, respectively, the integrated sequences were excised and detected as free plasmids in the recipients (Hopwood *et al.*, 1984a). Like other prokaryotic plasmids, *Streptomyces* plasmids can also be cured by treatment with acridine dyes or ethidium bromide. It has also been reported that the process of protoplast formation and regeneration can successfully eliminate plasmids from *Streptomyces* (Hopwood, 1981).

In several cases, genetic evidence has suggested that certain phenotypes might be carried by extrachromosomal elements. But, some of these elements have never been isolated. The elements SCP1 in *S. coelicolor*, SLP2, SLP3 and SLP4 in *S. lividans* (Hopwood *et al.*, 1984a), SRP1 and SRP2 in *S. rimosus* (Rhodes *et al.*, 1984), SQP1 in

S. qingfengmyceticus (Zheng *et al.*, 1982) are good examples. Recent evidence has indicated that SCP1 is a giant linear plasmid in *S. coelicolor* with a size of 350 kb (this will be discussed later).

The most remarkable characteristic consequence of *Streptomyces* plasmids is pock formation. When one plasmid-containing *Streptomyces* species is growing on lawns of another, plasmid-free, *Streptomyces* strain, patches of the resulting plasmid-containing individuals are surrounded by narrow zones in which growth of the plasmid-free culture is retarded. These narrow zones are known as pocks. Since the majority of *Streptomyces* plasmids have this property, the pock formation phenotype has become a valuable tool for detection of plasmids in *Streptomyces*. Moreover, this type of phenotype is analogous to lethal zygosis in *E. coli* (Skurray and Reeves, 1973) and therefore, was designated as Ltz^+. If some strains harbor a particular plasmid, they are resistant (Ltz^R) to pock formation by other cultures carrying the same plasmid or its Ltz^+ derivatives, but they are sensitive (Ltz^S) to cultures carrying most other plasmids (Kieser *et al.*, 1982). The pock phenomenon has been shown to be associated with plasmid transfer, rather than diffusible agents (Bibb *et al.*, 1977). In the case of SCP2* plasmid, it has been demonstrated that when an SCP2* harboring strain was grown on an SCP2 harboring lawn, the SCP2* plasmid was able to be detected in the zone up to the recipient mycelium at the edge of the pock, while the genotypes of the SCP2* parent were found to be confined to the central area of the pock (Bibb and Hopwood, 1981). Moreover, a transfer-defective derivative of the plasmid SCP1 failed to give pock formation (Kieser *et al.*, 1982), suggesting that the genes on the plasmids which are responsible for pock formation are closely associated with those responsible for plasmid transfer. Recent studies (Kendall and Cohen, 1987) of pIJ101 from *S. lividans* have shown that the functional genes for pock formation (*kilA*) and plasmid transfer (*tra*) overlapped in one region. Insertions into the region abolished the pocking phenotype. In fact, sequencing data indicated that the *kilA* and *tra* genes were in the same open reading frame (Kendall and Cohen, 1988), suggesting that a single

polypeptide may be responsible for both functions. In addition, it has been proposed that the function of resistance to pock formation may be independent of the functions of plasmid transfer (Hopwood *et al.*, 1986b). In another case, the function of resistance to pock formation (*por* 1 and *por* 2) by pSK3* from *S. kasugaensis* has been localized in the regions responsible for plasmid transfer (*tra* 1 and *tra* 2) (Akagawa, 1987). It seems that the functions of pock formation and resistance are closely associated with those of plasmid transfer. The complete picture of the molecular mechanisms of these functions remains unclear.

Another interesting phenotype of pock formation is $Ltz^{+/-}$ (Kieser *et al.*, 1982). When strains harboring some pIJ101 derivatives, which lacked a specific region of the plasmid, grew on lawns of plasmid-free *S. lividans*, pocks appeared in a characteristic and reduced size. Like the normal Ltz^{+} phenotype, the plasmids were indeed transferred to the recipient. This phenomenon gave rise to the concept of "spread" function. It has been hypothesized that besides intermycelial transfer of a plasmid, intramycelial migration of the transferred plasmid, i.e. intercellular spread along a mycelium, is also required to give a Ltz^{+} phenotype. Plasmids with the $Ltz^{+/-}$ phenotype are believed to be spreading-defective. The genes responsible for this "spread" function have been mapped on the plasmids pIJ101 and pSK3* (Kieser *et al.*, 1982; Akagawa, 1987).

Most of the self-transmissible *Streptomyces* plasmids, except SLP3 and SLP4, have been shown to promote host chromosomal transfer and recombination, known as fertility. Non-transferable plasmids failed to promote such transfer and recombination. When one plasmid-carrying strain is crossed with a plasmid-free strain, among the spores of the new generations, new genotypes coming from both parents can be detected. The level of fertility varies from one plasmid to another, indicating that plasmids play an active role in fertility. However, the level of fertility is decreased when both parents carried the same plasmid (Hopwood *et al.*, 1984b). The reasons for this are not obvious. In crosses between a strain harboring the SCP2 plasmid and an SCP2-free strain, the progeny showed

no predilection to inherit genetic markers from the plasmid-carrying parent. In contrast, in crosses between SCP1$^+$ strains of *S. coelicolor* and SCP1$^-$ strains, the progeny had a tendency to inherit genetic markers preferentially from the SCP1$^+$ parent. Interestingly, it was strongly suggested by genetic evidence that the SCP1 plasmid from *S. coelicolor* A3(2) integrates into the host chromosome at various locations (Chater and Hopwood, 1983). Thus, it was assumed that the preferential appearance of genetic markers from the SCP1$^+$ parent in the new progeny could be due to the integration of SCP1 in the chromosome, as seen with the fertility of the F$^+$ plasmid in *Escherichia coli* (Curtiss and Stallious, 1969). Unlike the plasmid SCP1, the unisolated plasmids SLP3 and SLP4 (they may also be giant linear plasmids) are transferable, but do not promote fertility.

Nevertheless, fertility in *Streptomyces* is mediated by specific plasmids. The mechanism of this phenomenon is not understood and a specific plasmid region promoting such fertility has not yet been identified. But, plasmid-mediated fertility in *Streptomyces* is certainly different from that demonstrated in *E. coli*, mediated by the F plasmid, simply because the former leads to bidirectional transfer of chromosomal markers, while the latter is unidirectional.

Although most plasmids in *Streptomyces* are of the conventional cccDNA type, the genus *Streptomyces* does contain a number of species which harbor linear plasmids. Many of them are very large in size. Their structural and biological features will be discussed in a later section.

C. Transposable Elements

Searches for insertion sequences (IS) in *Streptomyces* have been ongoing for a number of years. The first physical evidence of an IS-like element was found in *S. coelicolor* by Chater *et al.* (1985). This element, designated as IS*110*, has a length of 1.6 kb and was capable of transposing from the *S. coelicolor* chromosome into a preferred site in ϕC31 phage derivatives, with the deletion of the *attP* site. The defective phage with

IS*110* allowed integration into the chromosome by recombination with IS*110* copies in the chromosome. The insertion site on the phage has been sequenced, with and without IS*110* (Bruton and Chater, 1987). It consists of a GC rich sequence and is associated with an imperfect direct repeat. The complete nucleotide sequence of IS*110* has also been determined (Bruton and Chater, 1987). It has been shown that, like other transposable elements, it contains short terminal inverted repeats, although the inverted repeats are not perfectly matched. Moreover, one large open reading frame (ORF) on one strand and a small one on the other strand were found. The polypeptides deduced from the two ORFs showed no homology to proteins from other IS elements. Although IS*110* was the first IS-like element in *Streptomyces* to be physically isolated, some similar IS-like elements have since been observed, such as IS*281* in the phage ϕC43 from *S. lividans* 803 (Sladkova, 1986) and a 1.2 kb insertion element in the phage SH10, as well as in *S. chrysomallus* 40341 (Chater *et al.*, 1988).

Despite the fact that all IS-like elements exist as linear forms, a transposable element with a circular intermediate has been reported in *S. coelicolor* (Lydiate *et al.*, 1986, 1987). This element exhibits a very low copy number in *S. coelicolor* and has a size of 2.6 kb. It exists in a cccDNA form in addition to two linear integrated copies in the *S. coelicolor* chromosome. Interestingly, when the element was transformed into *S. lividans* which lacks the mini-circle sequences, it integrated into the host chromosome at the preferred sites. Moreover, the 2.6 kb mini-circle is able to carry foreign DNA and to integrate the resulting hybrid into the *S. lividans* chromosome. This element has been sequenced and it has one large ORF and two small ORFs on opposite strands in the same DNA region (Chater *et al.*, 1988). The integration site of the element was located within a non-coding region. The sites on the chromosome at which the element inserted were shown to be associated with an imperfect inverted repeat. There was no apparent sequence homology between the integration site on the element and the chromosomal DNA target.

The only transposon reported in any species of *Streptomyces* is Tn*4556* from a neomycin-producing strain of *S. fradiae* (Chung, 1987). It is 6.8 kb in length. This transposon was discovered in a plasmid isolated from *S. fradiae*. When this plasmid was introduced into another *S. fradiae* host with the ϕSF1 phage, a 6.8 kb fragment on the plasmid was deleted. After the plasmid with the deletion was reintroduced back into the origin host, however, the 6.8 kb element transposed back from the host chromosome to the plasmid, but at many new locations and in two orientations. Thus, the transposon was isolated and many new transposons with antibiotic resistance were able to be constructed *in vitro*. It has since been shown that transposons with a viomycin resistance gene (*vph*) were able to transpose to various locations in the *S. lividans* chromosome (Chung, 1987). Sequence data revealed that Tn*4556* has 38 bp terminal inverted repeats, and the sequence of the terminal inverted repeats has some similarity to the Tn*3* transposon family in *E. coli* (Olson and Chung, 1988). Tn*4556* will probably be a very useful tool for transposon mutagenesis in *Streptomyces*.

D. Bacteriophages in *Streptomyces*

In principle, it is not appropriate to discuss *Streptomyces* phages in the section on extrachromosomal elements. However, it has been postulated that linear plasmids in *Streptomyces* may have evolved originally from bacteriophages since structural analysis has shown that linear plasmids share several similar structural features with many bacteriophages and eukaryotic viruses (Esser and Kempken, 1986; Keen *et al.*, 1988). DNA polymerases from these bacteriophages and viruses contain conserved domains which are also found in certain genes in almost all known eukaryotic linear plasmids. Therefore, it is certainly worthwhile to examine briefly some features of *Streptomyces* phages.

Bacteriophages, temperate and virulent, can be found for many *Streptomyces* species. Unfortunately, genetic and molecular information is not available at the present

time for most of them. The best-studied *Streptomyces* phage is φC31 from *S. coelicolor* A3(2) (reviewed by Chater, 1986). Phage φC31 has a broad host range. The viral genome is 41.5 kb long with a 63% G+C content. The DNA has cohesive ends that result from short, single-stranded complementary termini, just like λ phage in *E. coli*. When the phage infects the host, it undergoes either a lytic or a lysogenic life cycle. In the lysogenic pathway, phage φC31 DNA integrates into the host chromosome, probably by a mechanism similar to that used in the integration of phage λ DNA into the *E. coli* chromosome. The gene (*attP*) responsible for the integration of φC31 has been mapped on the phage genome, and the region into which the phage DNA inserts has been located on the *S. coelicolor* chromosome. Other temperate phages, such as R4, VP5, SH10, etc., have also been studied for their potential use as cloning vectors (Chater, 1986). These phages are closely related to each other in terms of genomic structures. It is interesting to note that a temperate phage of *S. griseus* (Pg2) with a 42 kb genome has been reported to have terminally repeated sequences of about 1 kb (Sladkova, 1982). Another phage, φSF1 (82 kb) from *S. fradiae*, also has DNA with terminally redundant ends (Chung, 1982). This indicates that Pg2 and φSF1 may have a different replication mechanism from that of φC31.

Transduction mediated by the temperate phage SV1 in *S. venezuelae* is the only example which has been well-established. Other phages, such as SH10 and φSF1, have also been reported to mediate transduction (Suss and Klaus, 1981; Chung and Molnar, 1983)

E. Cloning Vectors Derived from Plasmids and Bacteriophages

An understanding of the molecular biology of *Streptomyces* plasmids and phages has allowed the construction of various vectors useful for gene cloning, manipulation and expression in *Streptomyces*. Many recently developed plasmid cloning vectors and their

hosts have been reviewed by Hopwood *et al.* (1986b, 1987). Phage ϕC31 derived vectors have also been developed (reviewed by Chater, 1986).

1.2 Interactions Between Extrachromosomal Elements and Host Chromosomal DNA

The first evidence that *Streptomyces* plasmids integrate into the host chromosome came from genetic analysis of the plasmid SCP1. This plasmid was originally found in *S. coelicolor* by its pock formation and methylenomycin production phenotypes, but it was not physically isolated (Vivian, 1971). Like some other *Streptomyces* plasmids, SCP1 also mediated host chromosomal transfer and recombination, i.e. fertility. Based on genetic analysis, it was found that SCP1 interacted with the *S. coelicolor* chromosome (Hopwood and Wright, 1973) and gave a so-called NF (normal fertility) strain in which a chromosomal gene, *dagA* , at the 9 o'clock position was inactivated (Hodgson and Chater, 1981). When the NF strain was crossed with an SCP1$^-$ strain, genetic markers on both sides of the position into which SCP1 integrated on the NF chromosome were donated to the recipient. The greater the distance the genetic markers are located from the point of SCP1 integration, the lower the frequencies with which they were transferred to the recipient. All progeny carried the SCP1$^+$ phenotype, indicating that the SCP1 plasmid was transferred to the SCP1$^-$ strain. Moreover, in some cases, SCP1 was found to carry other genetic markers from different locations on the chromosome, producing SCP1' plasmids, which gave a high frequency of genetic marker donation in crosses with SCP1$^-$ strains (Hopwood and Wright, 1976). Taken all together, it strongly suggests that SCP1 somehow interacts with the host chromosome. Kendall and Cullum (1986) proposed that SCP1 may integrate via about 1 kb DNA sequences on both plasmid and the chromosome. Since SCP1 was never isolated physically, it was impossible to locate precisely the regions directly involved in the interaction. The recent discovery that SCP1 is a giant linear

plasmid (Kinashi *et al.*, 1987) may help to solve the mystery because some eukaryotic linear plasmids are known to integrate into host chromosomes (to be discussed later).

Further evidence of interactions between plasmids and host chromosomes came from the study of the SLP1 series of plasmids. As mentioned earlier, SLP1 was identified as a plasmid only when *S. coelicolor* A3(2) was crossed with *S. lividans* 66 (Bibb *et al.*, 1981). Free SLP1 plasmid was not detectable in *S. coelicolor* A3(2). However, when the two strains were crossed with each other, SLP1 was excised from the *S. coelicolor* chromosome and transferred into *S. lividans* in various sizes with deletions. The majority of the transferred plasmids replicated autonomously, but about 10% of recipients showed a plasmid-free phenotype (Omer and Cohen, 1984). This indicated that SLP1 was indeed transferred, but integrated into the *S. lividans* chromosome. The size of the SLP1 integrated form was determined to be 17 kb, which was designed as SLP1^{int}. The regions involved in SLP1^{int} integration have been determined: the *attP* site on SLP1^{int} and the *attB* site on the *S. lividans* chromosome. Integration also requires two additional regions on SLP1^{int}, *intA* and *intB*. These regions were all sequenced. The sequencing data suggest that homologous recombination between the plasmid and the chromosome leads to the integration (Omer and Cohen, 1986). This explains why some of the transferred plasmids only replicated autonomously rather than integrating, because they lacked the *attP* site. Another plasmid, pSAM2 from *S. ambofaciens* , also exists in a free and a chromosomally integrated form, but in the same host (Pernodet *et al.*, 1984). In addition, pSAM2 can integrate into the *S. lividans* chromosome at a preferred site. Besides the plasmids mentioned above, a transposon-like 2.6 kb mini-circle is also able to integrate at a specific site on the host chromosome, as described in the previous section. Furthermore, the IS-like element IS*110* in the *S. coelicolor* chromosome is able to insert into a preferred site in the DNA of ϕC31 phage derivatives (Chater *et al.*, 1985). The locations of IS*110* in the chromosome remain unknown.

In summary, a number of *Streptomyces* extrachromosomal elements were found to exist in both integrated and free forms. They always have a specific site in their host's chromosomes for their integrations. It seems very likely that integration take place by homologous recombination between specific sequences on the plasmids and the host chromosomes. The mechanism of *Streptomyces* plasmid-mediated fertility is not clear. The different levels of fertility associated with different plasmids suggest that plasmids play an active role in fertility. It seems probable that plasmids have to have direct contact with the host chromosomes. A direct interaction between the SCP1 plasmid and the *S. coelicolor* chromosome has been demonstrated by genetic evidence, and is the only example so far. There is no information available for other *Streptomyces* plasmids at present.

It is interesting to note that both the *S. coelicolor* linear plasmid (SCP1) and the circular plasmid (SLP1) exhibit two forms, free and integrated, within the cell. They should rely on different mechanisms to integrate themselves into the host chromosome, because a circular form needs only one cross-over, while a linear form requires two.

1.3 Linear Plasmids in *Streptomyces*

Although most plasmids found in *Streptomyces* exist as cccDNA, and in a few cases some of them are able to integrate into host chromosomes, autonomously replicating linear plasmids have also been found in this genus. Hayakawa *et al.* (1979) first reported that two linear plasmid-like DNAs (designated as pSLA1 and pSLA2) could be isolated from *S. rochei* 7434-AN4, a lankacidin- and lankamycin-producer. One of them (pSLA2) has been extensively studied. The copy number of pSLA2 was estimated to be 60 per chromosome (Hirochika and Sakaguchi, 1982). Further structural analysis has revealed that the plasmid pSLA2 has a length of 17 kb. It was found to have proteins covalently linked to the 5' termini since pSLA2 DNA was detectable only after Pronase treatment of the cell lysates and it was also resistant to λ exonuclease. It was shown that the linkage

between the terminal protein and pSLA2 DNA was alkali-labile, as are those in adenoviruses (Rekosh *et al.*, 1977) and the *Bacillus* phage ϕ29 (Salas *et al.*, 1978). The terminal fragments of pSLA2 DNA have been cloned. Subsequent sequencing indicated that pSLA2 contains a 614 bp long inverted terminal repeat (Hirochika *et al.*, 1984). Several small inverted repeats were also found in the terminal half of the long repeated sequences, suggesting that specific DNA-binding proteins may be associated with these terminal sequences. Based on these observations, Hirochika *et al.* (1984) proposed that pSLA2 DNA might form a racket frame-like structure *in vivo*, similar to the structure of adenoviruses suggested by Challberg *et al.* (1980). Because of the long inverted terminal repeats, which are a common structural feature in transposable elements and in some eukaryotic viruses, it has been postulated that pSLA2 may have the ability to insert itself into the host chromosome (Hirochika *et al.*, 1984).

After the first report of linear plasmids from *Streptomyces*, several small linear plasmids were also found in other *Streptomyces* species, such as pSRM (43 kb in length) from *S. rimosus* (Chardon-Loriaux *et al.*, 1986) and the 12 kb pSCL from *S. clavuligerus* (Keen *et al.*, 1988). These linear plasmids in *Streptomyces* share some common structural characteristics. They were all found to have terminal proteins covalently linked to the 5' ends of the DNA and to have long inverted terminal repeats. In the case of pSCL in *S. clavuligerus*, a 900 bp terminal repeat was demonstrated by cross hybridization (Keen *et al.*, 1988). These characteristic features have been reported in other linear plasmids found in eukaryotes and in some eukaryotic and prokaryotic viruses (they will be discussed in the next section).

In some cases, there is evidence which suggests that the linear plasmids carry the genes responsible for some phenotypes of their hosts. For instance, curing experiments using ethidium bromide suggested that pSLA2 was involved in the production of lankacidin. The pSRM-free mutants of *S. rimosus* obtained by treatment with ethidium bromide or acridine orange were found to have lost their ability to form aerial mycelia. It

should be stressed that many plasmid curing agents can cause chromosome mutations in *Streptomyces* (Hopwood *et al.*, 1986b). Thus, the loss of a particular phenotype of the host may not result from the loss of the plasmid.

Another efficient method for curing *Streptomyces* plasmids is protoplast formation and regeneration (Hopwood, 1981). The linear plasmid SCP1 in *S. coelicolor* was cured in this way. But, preliminary experiments with this technique failed to eliminate pSCL from *S. clavuligerus* (Keen *et al.*, 1988). It has been suggested that this method may be plasmid and strain specific.

Recently, in a study of lasalocid biosynthesis in *S. lasaliensis*, Kinashi and Shimaji (1987), using orthogonal-field-alternation gel electrophoresis (OFAGE, a PFG variant), were able to show that giant linear plasmids exist in many antibiotic-producing *Streptomyces* strains. The sizes of these linear plasmids are quite large, in the range of several hundred kb, compared with linear plasmids found in eukaryotes. Interestingly, for a long time there has been genetic evidence suggesting that the plasmid SCP1 of *S. coelicolor* A3(2) was involved in methylenomycin production (Kirby and Hopwood, 1977), but it was never isolated physically. Using the OFAGE technique, it was shown that SCP1 is a giant linear plasmid with a size of 350 kb in *S. coelicolor* M138 and M146 strains. In addition, a series of giant linear plasmids differing in size by 30 kb increments were found in a wild-type strain of *S. coelicolor* A3(2), but not in SCP1$^-$ strains (Kinashi *et al.*, 1987). It has been proposed that these plasmids in the wild-type strain may derive from a 350 kb progenitor of the original SCP1 plasmid by multiplication of a 30 kb segment.

The detection of these giant linear plasmids allowed establishment of a possible relationship between the plasmids and antibiotic production. In the case of *S. lasaliensis*, which is able to produce lasalocid and echinomycin, protoplast fusion of an antibiotic producer (harboring a linear plasmid pKSL) and a non-producer always resulted in a high frequency transfer of antibiotic biosynthesis capability to the non-producer. A 520 kb

linear plasmid, pKSL, was consistently detected in all producer strains, but not in the non-producer strains. Moreover, pKSL DNA from the original producer strain was used as a probe which hybridized with all the plasmid bands in the producer strains resulting from protoplast fusion. This indicated that pKSL was transferred to the non-producer strains and gave the phenotype of production of lasalocid and echinomycin (Kinashi and Shimaji, 1987). Similarly, the cloned methylenomycin biosynthetic (*mmy*) and resistance (*mmr*) genes were also used as a probe. The several giant linear plasmids found in the *S. coelicolor* A3(2) strain with the SCP1^{+} phenotype were found to contain the antibiotic biosynthetic and resistance genes, suggesting that these plasmids were closely related to SCP1 (Kinashi *et al.*, 1987). Taking all of this evidence together, it implies that the linear plasmids pKSL and SCP1 are directly involved in the production of these antibiotics. Previously, genetic studies suggested that the plasmid SCP1 integrated into the *S. coelicolor* chromosome. However, the hybridization results indicate that SCP1 seems to exist in an autonomously replicating form rather than an integrated form.

To verify the relationship between the giant linear plasmids and antibiotic production, a number of antibiotic producing strains of *Streptomyces* were tested for the presence of giant linear plasmids. Some of them indeed contain giant linear plasmids, but some do not (Kinashi and Shimaji, 1987). The involvement of these linear plasmids in antibiotic biosynthesis requires further investigation.

It is interesting to note that since SCP1 was known to have the ability to induce pock formation, another linear plasmid, pSRM from *S. rimosus* , was also tested for pock formation. A Ltz^{+} phenotype was observed with some *Streptomyces* strains with this plasmid. There is no evidence for pock formation available for other linear plasmids in other *Streptomyces*. Despite the fact that both linear and circular *Streptomyces* plasmids bring about pock formation, the question remains unanswered whether the mechanisms utilized for plasmid transfer are the same or different.

Another unusual linear plasmid, pSA1, was found in *S. azureus*, a thiostrepton producer. This plasmid was believed to be responsible for spontaneous pock formation in the organism. Initially, the plasmid was detected in plate cultures by sucrose gradient sedimentation of cleared lysates, followed by agarose gel electrophoresis. It was shown that it had a linear form since CsCl buoyant density centrifugation failed to demonstrate any cccDNA species in the same cleared lysates (Ogata *et al.*, 1983). However, after growing in liquid culture, the plasmid was found to have a circular form (Miyoshi *et al.*, 1986). Hybridization evidence indicated that the circular plasmid was derived from the linear one. Moreover, the linear plasmid had a low copy number, less than 1 per chromosome. After UV irradiation, the circular plasmid had a copy number of about 20-30 copies per genome. These observations imply that the linear plasmid may replicate through a circular intermediate. However, further information has not been published despite statements suggesting the presence of such a circular intermediate in the replication of the linear plasmid pSA1.

Nevertheless, linear plasmids are unusual genetic elements in prokaryotes, especially in that they are found predominantly in the genus *Streptomyces*. The linear plasmids so far reported in *Streptomyces* are summarized in Table 1. It can be seen that giant linear plasmids appear to be more common in *Streptomyces* species than small ones. Because of the large size of some of the linear plasmids, it will be difficult to study their genetic behavior at the molecular level by conventional techniques. But, the smaller plasmids may be used as models for further studies of gene organization and regulation in the linear plasmids, including their possible involvement in antibiotic production.

List of Some Linear Plasmids Found in *Streptomyces*

Strain	Size (kb)	Antibiotics the strain produces	Reference
S. lasaliensis	520	Echinomycin Lasalocid A	Kinashi and Shimaji, 1987
S. fradiae	420	Tylosin	Kinashi and Shimaji, 1987
S. violaceoruber (*S. coelicolor*)	410, 440, 470, 500, 530, 560, 590	Methylenomycin	Kinashi and Shimaji, 1987
S. parvulus	520, 560, 580	Actinomycin D	Kinashi and Shimaji, 1987
S. venezuelae	130	Chloramphenicol	Kinashi and Shimaji, 1987
S. rochei	17, 90, 180	Lankacidins	Hayakawa *et al.*, 1979; Kinashi and Shimaji, 1987
S. rimosus	43	Oxytetracycline	Chardon-Loriaux *et al.*, 1986
S. azureus	9, 14.5	Thiostrepton	Ogata *et al.*, 1983
S. clavuligerus	12	Cephamycins	Keen *et al.*, 1988

Table 1. List of some linear plasmids found in *Streptomyces*.

List of Some Linear Plasmids Found in *Streptomyces*

Strain	Size (kb)	Antibiotics the strain produces	Reference
S. lasaliensis	520	Echinomycin Lasalocid A	Kinashi and Shimaji, 1987
S. fradiae	420	Tylosin	Kinashi and Shimaji, 1987
S. violaceoruber (*S. coelicolor*)	410, 440, 470, 500, 530, 560, 590	Methylenomycin	Kinashi and Shimaji, 1987
S. parvulus	520, 560, 580	Actinomycin D	Kinashi and Shimaji, 1987
S. venezuelae	130	Chloramphenicol	Kinashi and Shimaji, 1987
S. rochei	17, 90, 180	Lankacidins	Hayakawa *et al.*, 1979; Kinashi and Shimaji, 1987
S. rimosus	43	Oxytetracycline	Chardon-Loriaux *et al.*, 1986
S. azureus	9, 14.5	Thiostrepton	Ogata *et al.*, 1983
S. clavuligerus	12	Cephamycins	Keen *et al.*, 1988

1.4 Linear Plasmids in Eukaryotes

Although linear plasmids are found almost exclusively in *Streptomyces,* among the prokaryotes, they are found in a wide variety of eukaryotic organisms, from animals and plants to fungi. Table 2 summarizes some of the linear plasmids reported in eukaryotes.

Pring *et al.* (1977) first observed that two linear plasmids exist in the mitochondria of a higher plant, *Zea mays.* The two plasmids were present in equal amount and were about 5 times more abundant than mitochondrial DNA, suggesting that the plasmid DNA replicated independently. After the first discovery, linear plasmids were also reported in the nuclei of the nematode *Caenorhabditis elegans*, and in many species of fungi. From Table 2, it is clear that the linear plasmids found in eukaryotes are relatively small (less than 15 kb) compared with those found in *Streptomyces.* In addition, many species contain more than one linear plasmid, and of different sizes. Most of the eukaryotic linear plasmids were found associated with mitochondria; only one was reported to be in chloroplasts (Turmel *et al.*, 1986).

In some cases, linear plasmids were found to be responsible for certain phenotypes of their hosts. For instance, the linear plasmids S1 and S2, with sizes of 5.45 kb and 6.4 kb respectively, were believed to be correlated with male sterility in *Zea mays* (Levings *et al.*, 1980). In addition, investigations on cytoplasmic revertants to fertility, using restriction analysis, Southern transfer and hybridization analysis, indicated that the linear plasmids S1 and S2 integrated into the mitochondrial genome in these revertants (Kemble and Mans, 1983). This evidence suggested that excision of these S plasmids may lead to sterility, whereas reintegration may result in reversion to fertility. In contrast, the mitochondrial linear plasmid (mtAr-kalDNA) in *Neurospora intermedia* is believed to be responsible for the senescence phenotype which results from its integration in the host mitochondrial genome (Bertrand and Griffiths, 1989; Myers *et al.*, 1989).

Table 2. List of some linear plasmids found in eukaryotes.

List of Some Linear Plasmids Found in Eukaryotes

Source	Size (kb)	Location	Possible Function	Reference
Animal				
Caenorhabditis elegans	1.6	nuclear	unknown	Ruan and Emmons, 1984
Higher plant				
Zea mays	2-7	mitochondria	male sterility	Pring *et al.*, 1977
Green alga				
Chlamydomonas moewusii	5.9	chloroplast	unknown	Turmel *et al.*, 1986
Fungi				
Kluyveromyces lactis	13.4, 8.9	cytoplasm	killer toxins	Gunge *et al.*, 1981
Agaricus bitorquis	2.6	mitochondria	unknown	Mohan *et al.*, 1984
Claviceps purpurea	1.1-6.7	mitochondria	unknown	Tudzynski *et al.*, 1983

Moreover, the two linear plasmids (pGKl1 of 8.9 kb and pGKl2 of 13.4 kb) of the yeast *Kluyveromyces lactis* have been shown to carry the genes for the production of killer toxins (Gunge *et al.*, 1982). pGKl2 also contains the gene(s) for maintenance of the two plasmids because without pGKl2, pGKl1 cannot survive (Niwa *et al.*, 1981). Possible functions for other linear plasmids in eukaryotes are not known at the present time.

Further analysis of these linear plasmids has revealed that, like the linear plasmids in *Streptomyces,* they all have covalently bound proteins at the 5' ends of their DNA genomes. Moreover, all of the linear plasmids so far investigated contain long inverted terminal repeat (Esser and Kempken, 1986). The sizes of these long inverted terminal repeats vary from one plasmid to another. The sequences of the terminal repeats of similar plasmids from a given organism may be identical or different. The two linear plasmids S1 and S2 from *Zea mays* have identical terminal sequences up to 208 bp (Paillard *et al.*, 1985). On the other hand, the long terminal inverted repeat sequences of the two linear plasmids pGKl1 and pGKl2 (202 bp and 184 bp, respectively), isolated from the same organism, *K. lactis* , share no homology at all, except that both of them are AT rich (Hishinuma *et al.*, 1984). Surprisingly, a linear plasmid (pSKL), from another fungus, *Saccharomyces kluyveri*, contains a 16 bp terminal sequence identical to pGKl2 (Kitada and Hishinuma, 1987). pSKL has its own long terminal inverted repeats of 483 bp. Moreover, alignment of the terminal nucleotide sequences of the S2 plasmid and five *Bacillus* phages which have a similar structure indicates a high degree of homology. A 6 bp sequence, AAAGTA, was found in all terminal sequences of the phages and also in the terminal sequences of S2 at the second nucleotide from the 5' ends. This suggests that linear plasmids in higher plants may have a viral origin (Levings and Sederoff, 1983).

Although many eukaryotic linear plasmids have been identified from a variety of organisms, only 5 linear plasmids have been completely sequenced so far, S1 and S2 from *Zea mays*, pGKl1 and pGKl2 from *Kluyveromyces lactis*, and pClKl from *Claviceps purpurea*. Jung *et al.* (1987a) first observed that there are three conserved amino acid

domains among DNA polymerases from several viruses and bacteriophages. These viruses and phages share some structural similarities, such as *Bacillus* phage ϕ29, *E. coli* phage PRD1, adenoviruses, etc. Interestingly, these conserved DNA polymerase domains were also found in almost all linear plasmids which have been sequenced; for example, URF 3 of S1 (Kuzmin and Levchenko, 1987), ORF 1 of pGKl1 (Jung *et al.*, 1987b), ORF 3 of pGKl2 (Tommasino *et al.*, 1988), and ORF 1 of pClK1 (Oeser and Tudzynski, 1989). It obviously suggests that the linear plasmids encode their own DNA polymerase for their replication. Moreover, URF 1 of S2 and ORF 2 of pClK1 have considerable homology to the mitochondrial RNA polymerase of *S. cerevisiae* and RNA polymerases of bacteriophages T3 and SP6 (Kuzmin *et al.*, 1988; Oeser, 1988; Oeser and Tudzynski, 1989). The ORF 974 of pGKl2 also shows homology to subunits of several RNA polymerases, particularly the β and β' subunits of the *E. coli* RNA polymerase (Wilson and Meacock, 1988). This evidence supports the idea that some linear plasmids may have a plasmid-specific gene expression system, which is partially independent of the host apparatus. This might also explain why pGKl1 could be maintained in its host only when pGKl2 is present. That is, pGKl2 probably encodes an RNA polymerase which is essential for expression of the genes on both plasmids. In fact, genes from the plasmids failed to be transcribed by the nuclear RNA polymerases I, II or III (Wilson and Meacock, 1988). In the case of S1 and S2, the plasmid S1 encodes a putative DNA polymerase while S2 encodes a putative RNA polymerase. The two plasmids may complement each other and thus co-exist in the same host.

Replication of the unique structure of linear plasmids seems to require a specific DNA polymerase similar to certain viral DNA polymerases, but different from that of the host. This implies that linear plasmids in eukaryotes may have a replication mechanism similar to that of some bacteriophages. Based on the considerable similarities between eukaryotic linear plasmids and linear DNA viruses, it is reasonable to speculate that these linear plasmids may be derived from a common type of ancestor, probably bacteriophages.

It is interesting to note that a different type of linear plasmid has recently been reported in the the bacterium *Borrelia burgdorferi* (49-kb plasmid, Barbour and Garon, 1987) as well as in the fungus *Rhizoctonia solani* (pRS64-1, -2 and -3 plasmids, Miyashita *et al.*, 1990). Unlike linear plasmids in *Streptomyces* and in the eukaryotic organisms mentioned earlier, these linear plasmids have quite different terminal structures. Instead of having terminal proteins covalently linked to the 5' termini of plasmid DNA, they have hairpin loops at their termini, in other words, the ends of the double-stranded DNA are covalently closed. In addition, they do not have long inverted terminal repeats. The unique terminal structures of these linear plasmids suggest that they may have a replication mechanism similar to the one proposed for vaccinia virus DNA replication (Baroudy *et al.*, 1983) since similar terminal structures have also been found in the genomic DNA of vaccinia virus (Gesheli and Berns, 1974).

1.5 DNA Viruses with Protein Linked to the 5' Termini of Their Genomes

As mentioned earlier, a unique structural feature, i.e. a linear double-stranded DNA genome with terminal proteins covalently linked to the 5' termini, is found not only in linear plasmids but also in a number of eukaryotic viruses and bacteriophages (summarized in Table 3). Because of this structural similarity, it is particularly interesting to examine some of their biological functions, especially the replication of these viruses, since at present there is no information on how linear plasmids replicate and what cellular components are involved in their replication. Since the replication of some animal viruses and bacteriophages, such as the adenoviruses and *Bacillus* phage ϕ29, has been studied extensively, a general mechanism proposed for their replication may be applied to the case of linear plasmids.

The genomic structure of the adenoviruses has been well-documented. These viruses are linear, double-stranded DNA viruses and have about 36 kb genomes with

Table 3. List of some DNA viruses with protein attached to the ends of DNA genomes.

List of Some DNA Viruses with Protein Attached to the DNA Ends

Virus	Host	Reference
<u>Animal Viruses</u>		
Adenoviruses	Human, Simian, Rodent, etc.	Robinson *et al*, 1973
<u>Bacteriophages</u>		
ϕ29	*B. subtilis*	Ito *et al.*, 1978
Cp-1	*S. pneumoniae*	Garcia *et al.*, 1983
PRD1	*E. coli* and *S. typhimurium*	Bamford *et al.*, 1983

terminal proteins covalently attached to the 5' ends. All adenoviruses have inverted terminal repeat sequences in the range of 63 bp to 165 bp, depending on the serotypes and strains. The nucleotide sequences at the termini of many adenovirus DNAs have been determined (reviewed by Tamanoi, 1986). Almost all terminal sequences contain a conserved core sequence, ATAATATACC very close to the termini. This sequence has been shown to play an important role in the initiation of viral DNA replication. The last nucleotide of the 5' terminal sequences in almost all adenoviruses is a deoxycytidine residue. The terminal protein is encoded by the virus, and has a molecular weight of about 55 kd and is processed from a precursor protein of 80 kd. This protein is believed to serve as a primer to initiate the viral DNA replication. The linkage between the terminal protein and the DNA is via a phosphodiester bond between the β-hydroxyl group of a serine residue in the protein and the 5' phosphate group of the deoxycytidine residue.

The replication of adenoviruses has been extensively investigated by electron microscopy. Also, an *in vitro* replication system has been established (Challberg and Kelly, 1979). The initiation of replication requires three viral encoded proteins: the terminal protein precursor, a DNA polymerase and a single-stranded DNA binding protein, and two cellular proteins: nuclear factor I and nuclear factor III. It has been shown that these five proteins are sufficient for the synthesis of full-length adenovirus DNA *in vitro* (Nagata *et al.*, 1983). *In vitro* and *in vivo* experiments, using constructs with the terminal sequences of an adenovirus, revealed that the replication origin of the virus is located at the termini (first 20 bp) of the genome, and a linear DNA containing adenovirus terminal sequences at the ends was capable of initiating the replication (Tamanoi and Stillman, 1982; Hay *et al.*, 1984). The viral DNA binding protein and the two nuclear factors have been shown to make direct contact with the terminal sequences. *In vitro* viral DNA replication (Tamanoi and Stillman, 1982) demonstrated that the viral genome, with proteins at the ends, functioned as a good template for initiation of replication, but proteinase-treated viral DNA did not. However, if the terminal protein was completely removed by piperidine

treatment, the naked viral DNA still functioned, albeit poorly, as a template. This indicates that the terminal protein on the parental DNA is important in viral DNA replication, but it is not absolutely essential for initiation of replication. In fact, plasmid constructs with two copies of the adenovirus terminal sequence can replicate *in vivo* along with intact adenovirus DNA as a helper (Hay *et al.*, 1984).

The information obtained from *in vitro* and *in vivo* biochemical studies, and observations using electron microscopy, seem to support the model for adenovirus DNA replication proposed by Lechner and Kelly (1977). A basic characteristic of this model is that the terminal protein is used as a primer to initiate replication at the 5' ends of the viral genome, and the replication is carried out by a viral DNA polymerase in the 5' to 3' direction, in a continuous manner. However, many aspects of this model are still unclear. For instance, a proposed panhandle structure has not been demonstrated *in vivo* or reassembled *in vitro.*

Another interesting feature of adenovirus DNA replication is the finding of covalently closed circles of adenovirus 5 DNA *in vivo* (Ruben *et al.*, 1983). The circular viral DNA has been cloned and the resulting plasmid DNA can generate linear infectious viruses when transfected into human cell lines, suggesting that circular viral DNA may play some role in viral replication (Graham, 1984; Graham *et al.*, 1989). Besides, under certain conditions of viral infection, the entire viral DNA has been found to integrate into the host chromosome (Ruben *et al.*, 1982).

In prokaryotes, some bacteriophages have the same genomic structure as the adenoviruses. These bacteriophages can be divided into three groups based on their hosts: the *Bacillus subtilis* phage φ29 family, the *Streptococcus pneumoniae* phage Cp-1 family and the *E. coli* and *Salmonella typhimurium* phage PRD1 family. They all have a double-stranded DNA genome with terminal proteins covalently linked to their 5' ends. The linkages between the DNA and terminal proteins for phages φ29, Cp-1 and PRD1 have been determined. In the case of φ29, the linkage is a phosphodiester bond between the OH

group of a serine residue in the terminal protein and a 5' phosphate group on deoxyadenosine (Hermoso and Salas, 1980), while in phage Cp-1, it is between a threonine residue and a 5' phosphate on deoxyadenosine (Garcia *et al.*, 1986); in phage PRD1, a hydroxyl group of a tyrosine residue is esterified to a 5' phosphate on deoxygaunosine (Bamford and Mindich, 1984).

Like adenoviruses, all of these bacteriophages have inverted terminal repeats. The phages in the φ29 family have 6-8 bp long terminal repeats and all contain a core sequence of AAAGTA except GA-1 which has AAATAGA at the termini. On the other hand, the phages of the Cp-1 family and the PRD1 family have fairly long inverted terminal repeats, for instance, Cp-1 has a 236 bp terminal repeat, while PRD1 has a 109 bp terminal repeat. In addition, the terminal sequences of phages within each family all contain a high degree of homology (reviewed by Salas, 1988). It seems that the terminal sequences found in phages from the different families are not related to each other.

The genomic organization and replication of the phage φ29 have been extensively characterized. The complete nucleotide sequence of the φ29 genome (19,285 bp) was determined (Garvey *et al.*, 1985; Vlcek and Paces, 1986). It is known that gene 2 of φ29 encodes a DNA polymerase and gene 3 encodes the terminal protein (Garcia *et al.*, 1983; Blanco and Salas, 1984). At least two genes (5 and 6) encode proteins which are directly involved in the initiation of viral DNA replication. For example, a protein (p6) from gene 6 stimulated viral replication *in vitro* (Pastrana *et al.*, 1985), while the protein (p5) from gene 5 shows the ability to bind to single-stranded DNA (Salas, 1988). An *in vitro* replication system for phage φ29 has been developed, using the purified terminal protein and DNA polymerase in the presence of φ29 DNA as a template (Blanco and Salas, 1985). By using this system, template requirements for initiation of the viral replication have been investigated. It was demonstrated that the terminal sequence of the viral DNA is absolutely required (Garcia *et al.*, 1984). No initiation could be observed when a plasmid containing the terminal sequence was circular or when the terminal sequence was not placed at the

ends of the linearized plasmid. Interestingly, as found with adenoviruses, when the $\phi29$ DNA-protein complex was treated with proteinase K, the resulting DNA did not function as a template. However, the piperidine-treated $\phi29$ DNA was active, but the activity was about 5-10 times lower than the native $\phi29$ DNA-protein complex (Gutierrez *et al.*, 1986). This suggests that the terminal protein is important for initiation of replication, but it is not essential.

Based on studies *in vitro*, Salas (1988) has proposed a model for the protein-primed replication of $\phi29$. In this model, a free terminal protein and DNA polymerase first contact the terminal protein bound to the viral genome by protein-protein interaction to form an initiation complex. Then, in the presence of other viral proteins such as p6 and p5 or other factors, the DNA polymerase catalyzes the formation of a covalent bond between a serine in the free terminal protein and dATP, followed by elongation of a daughter DNA strand. This model does not propose a circular or panhandle form as an intermediate like the one for adenoviruses, probably because the 6 bp inverted terminal sequences of $\phi29$ may be too short to allow the formation of such structures. This does not eliminate the possibility of such a structure for other phages, like Cp-1 or PRD1, which have long inverted terminal repeats.

Generally, the models of replication for adenoviruses and for phage $\phi29$ are quite similar. Both of them are based on the same principle, that a specific DNA polymerase encoded by the virus is capable of initiating DNA replication using a specific terminal protein. As a general mechanism this proposal can also be applied to linear plasmids in prokaryotes and eukaryotes, not only because linear plasmids have the same structure as these viruses, but also because all linear plasmids that have been sequenced so far seem to encode their own DNA polymerases. The putative DNA polymerases from these linear plasmids all share conserved domains found in some viral DNA polymerases (Jung *et al.*, 1987a; Oeser and Tudzynski, 1989). Although at this time nothing is known about the terminal proteins of linear plasmids, it is very likely that they are also encoded by the linear

plasmids. DNA binding proteins required for replication of linear plasmids, similar to those used by adenoviruses and phage ϕ29, may be encoded by the plasmids or by cellular genomes.

1.6 Objectives of This Study

About five year ago, in a unrelated study, a linear plasmid (pSCL) was discovered in *S. clavuligerus* by C. Keen and K.L. Roy (Keen *et al.*, 1988). Like linear plasmids found in eukaryotes and in other *Streptomyces* species, this plasmid shared certain common structural characteristics. It was shown that it had a linear double-stranded DNA form and contained terminal proteins covalently linked to the 5' termini of the DNA. Cross-hybridization experiments demonstrated that a long inverted terminal repeat sequence (about 900 bp) existed at its termini. A restriction endonuclease cleavage map was established, which revealed that the linear plasmid was 12 kb in length.

The major objectives of this study were to determine the complete nucleotide sequence of pSCL and to identify the possible protein-coding regions on pSCL. In addition, the analysis of messenger RNA transcribed from the plasmid would allow me to confirm whether the genes detected by sequencing were actually active *in vivo*. This would also enable me to locate important genetic elements on the plasmid, such as promoters and terminators. Also, it was anticipated that the protein products of the plasmid could be identified by using an *in vitro* coupled transcription and translation system established in *S. lividans*. The linear plasmid pSCL was chosen to conduct this research for a number of reasons.

1. The organism *Streptomyces clavuligerus* is of medical and industrial importance because of its ability to produce clavulanic acid and several β-lactam antibiotics such as cephamycins (Jensen, 1986). Based on observations that extrachromosomal elements in *Streptomyces* may be involved in antibiotic biosynthesis, for example, SCP1 in methylenomycin production and pSLA2 in

lankacidin production, a considerable effort has been made to find extrachromosomal elements in this organism. Kirby (1978) observed that an unstable genetic element may exist in *S. clavuligerus*, but the extrachromosomal element could not be identified physically. Furthermore, Saunders *et al.* (1984) also reported a possible plasmid in *S. clavuligerus*. There has been no further evidence to support the observation since then. All attempts to isolate covalently closed circular plasmids have failed (Jensen, S.E., personal communication). The linear plasmid pSCL was the first and only extrachromosomal element identified in the organism *S. clavuligerus*. Obviously, it would be a good candidate for exploration of the possible relationship between an extrachromosomal element and antibiotic production in this organism.

2. Despite the fact that a considerable amount of knowledge about eukaryotic linear plasmids has been accumulated, for instance, several linear plasmids in higher plants and fungi have been sequenced, almost nothing is known about linear plasmids in *Streptomyces* in terms of gene organization and regulation. The only sequence data available for prokaryotic linear plasmids comes from the analysis of a 614 bp inverted terminal repeat in the linear plasmid pSLA2 from *S. rochei* (Hirochika *et al.*, 1984). In addition, most linear plasmids found in *Streptomyces* are quite large (Kinashi and Shimaji, 1987). It is almost impossible to investigate their detailed gene organizations and functions by using current gene manipulation techniques. Among the known linear plasmids found in the genus *Streptomyces*, pSCL is one of the smallest isolated so far. For this reason, pSCL has advantages for the study of its gene organization and regulation. Information obtained from molecular studies of pSCL should provide us with a model for the further understanding of genetic behavior of other linear plasmids, including the giant linear plasmids of *Streptomyces*.

3. Previous research work on pSCL has established conditions suitable for carrying on this study. A detailed restriction endonuclease cleavage map of pSCL has been determined, using more than 20 different restriction endonucleases. More importantly, methods for the preparative isolation of pSCL by sucrose gradient sedimentation were established. As a result, a relatively large amount of fairly pure pSCL DNA could be obtained for subsequent cloning and sequencing.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and Enzymes

Dimethylsulfate, piperidine, and sodium 4-amino salicylate were purchased from Aldrich Chemical Co. Hydrazine and sodium-triisopropylnaphthalene sulfonate were from Eastman Kodak Co. Glyoxal was from MCB Chemical. Dimethylsulfoxide was from Alfa Products. Nitrocellulose and nylon filters were from Amersham. Kodak XAR5 X-ray film was obtained from Eastman Kodak Co., through Innomed Imaging.

All nucleotides and 2'-deoxyadenosine-5'-O-(1-thiotriphosphate) were purchased from Pharmacia; 2'-deoxy-7-deazaguanosine-5'-triphosphate (7-deaza dGTP) was from Boehringer Mannheim (BM). Radioisotopically labeled compounds, including ^{35}S-methionine (1200 Ci/mmole), α-^{32}P-dATP and α-^{32}P-ATP (3000 Ci/mmole), and γ-^{32}P-ATP (4500 Ci/mmole), were obtained from New England Nuclear and ICN Biochemicals Inc.

M13 universal sequencing primer, hybridization primer and random primers were purchased from the Regional DNA Synthesis Laboratory, University of Calgary, Canada. All other oligonucleotides used in this study were made in this laboratory using a model 381A DNA synthesizer from Applied Biosystems.

All restriction enzymes were purchased from Bethesda Research Laboratories Inc. (BRL), Boehringer Mannheim (BM) and New England Biolabs Inc. (NEB). T4 DNA ligase, the Klenow fragment of *E. coli* DNA polymerase I, and exonuclease III from *E. coli* were purchased from BRL; avian myeloblastosis virus (AMV) reverse transcriptase, S1 nuclease, micrococcal nuclease and polynucleotide kinase from Pharmacia; Sequenase (chemically modified T7 DNA polymerase) from United States Biochemicals (USB); *Thermus aquaticus* (*Taq*) DNA polymerase from NEB; terminal deoxynucleotidyl

transferase from BM; RNase-free DNase from Promega Corp. All enzymes were used as recommended by the manufacturers.

2.1.2 Bacterial Strains, Culture Conditions, Plasmids and Bacteriophages

S. lividans 1326 was obtained from S.E. Jensen, Department of Microbiology, University of Alberta, Canada. Growth conditions and spore preparations for this strain were as previously described (Hopwood *et al.*, 1985b). *S. clavuligerus* (NRRL 3585, Peoria, Illinois) was also from S.E. Jensen. Preparation of spores and culture conditions for *S. clavuligerus* were according to Hopwood *et al.* (1985b) and Keen *et al.* (1988).

Streptomyces plasmid pIJ702 was obtained from S.E. Jensen. *E. coli* strains JM83 and JM109 and the plasmids pUC18 and pUC19 were purchased from Pharmacia. The M13 vectors mp18 and mp19 were from BM. *E. coli* strain MV1193, M13 phage M13KO7, and plasmids pUC118 and pUC119 were a gift from J. Vieira, Department of Biochemistry, University of Minnesota, USA. Growth conditions for these strains were according to Yanisch-Perron *et al.* (1985) and Vieira and Messing (1987). Preparation of competent cells from these *E. coli* strains was done according to Morrison (1979) and Hanahan (1985).

2.1.3 Computer Software

The FRAME computer program (Bibb *et al.*, 1984) adapted for the Apple Macintosh computer was kindly provided by S.E. Jensen. All other computer analysis was performed on PCGENE software from IntelliGenetics Inc. and on DNA Strider (C. Marck, 1988) from C. Marck, Department of Biology, Institute of Fundamental Research, France.

2.2 DNA Isolation and Manipulations

2.2.1 DNA Isolation

The linear plasmid pSCL was isolated as described previously (Keen *et al.*, 1988). The *Streptomyces* plasmid pIJ702 was prepared according to Kieser (1984). *E. coli* plasmids were prepared as described by Birnboim (1983) and Maniatis *et al.* (1982). The recombinant M13 phages were isolated according to Messing (1983). Single-stranded plasmid DNA was prepared as described by Vieira and Messing (1987) with the following modifications.

The recombinant plasmid was transformed into *E. coli* MV1193. A single colony plus 0.1 ml concentrated helper phage M13KO7 stock were grown with shaking for 1 hour at 37°C in 10 ml 2 x YT broth (Messing, 1983) containing 150 µg/ml ampicillin, then kanamycin was added to a final concentration of 70 µg/ml. The culture was incubated with aeration at 37°C overnight. Single-stranded DNA was isolated and extracted as described by Messing (1983).

The M13KO7 helper phage stock inoculum was prepared as follows. A single plaque was inoculated into 10 ml of *E. coli* MV1193 overnight culture in 2 x YT liquid medium. The culture was incubated at 37°C for 1 hour with shaking. Kanamycin was added to a final concentration of 70 µg/ml. The culture was grown at 37°C overnight. Then, 10 ml of the culture was used to inoculate 500 ml of 2 x YT broth containing 70 µg/ml kanamycin. After growing at 37°C overnight, the cells were removed by centrifugation. The phages in the supernatant were precipitated by dissolving 20 g (4% w/v) polyethylene glycol 8000 (BDH), 14.6 g (0.5 M) NaCl and keeping at 4°C overnight. The mixture was centrifuged at 12,000 x *g* in a Beckman centrifuge (model J2-21) with a JA20 rotor. The pellet was collected and resuspended in 50 ml SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl, pH 8.0 and 2% [w/v] gelatin). The phage stock was incubated at 60°C for 30 minutes, then stored at 4°C.

2.2.2 Unidirectional Deletions

Unidirectional deletions of recombinant M13 phage or pUC plasmids using exonuclease III were performed as described by Henikoff (1987). The resulting phages or plasmids with various deletions were screened by agarose gel electrophoresis and selection was based on their sizes. Plasmids with appropriate deletions were used for production of single-stranded DNA, to be used for sequencing. Additionally, M13 phages with different deletions were screened by using a so-called "T track" method, also described by Henikoff (1984, 1987).

2.2.3 Preparation of DNA Probes

DNA probes used in hybridizations were made by nick-translation, as described by Maniatis *et al.* (1982). Oligonucleotide probes were labeled by T4 polynucleotide kinase and γ-^{32}P-ATP as described by Maxam and Gilbert (1980). DNA fragments were labeled in some cases by using random primers according to Feinberg and Vogelstein (1983, 1984). M13 hybridization probes were made according to Hu and Messing (1982). All probes were passed through a 3 ml Sephadex G-50 or G-25 column after labeling to remove unincorporated nucleotides.

DNA probes used in screening of recombinant plasmids containing the terminal fragments of pSCL were made as follows. The linear plasmid pSCL was first tailed with α-^{32}P-dATP using terminal deoxynucleotidyl transferase. Then, the labeled pSCL was subjected to *Sal* I restriction enzyme digestion. The resulting DNA was passed through a 3 ml Sephadex G-50 column.

2.2.4 Southern Transfer and Hybridization

DNA fragments were first separated electrophoretically on an agarose gel, stained with ethidium bromide, then photographed. The DNA was transferred to a nitrocellulose or nylon filter as described by Southern (1979), and modified for nylon membranes by Reed and Mann (1985) and Rigaud *et al.* (1987). After baking at 80°C for 2 hours (for nitrocellulose membranes), the filter was prehybridized at the appropriate temperature (normally 60°C) in a heat-sealable plastic bag (Dazey Corp.) with a prehybridization solution containing 6 x SSPE (1 x SSPE is 0.15 M NaCl, 20 mM sodium phosphate, pH 7.0 and 2 mM EDTA), 5 x Denhardt's solution (Denhardt, 1966), 0.5% SDS and 100 μg/ml denatured, sheared salmon sperm DNA (Sigma) for at least 2 hours. Then, the ^{32}P-labeled DNA probe was denatured and added. Hybridization was at 60°C overnight. After hybridization, the filter was removed and washed twice with two litres of 2 x SSPE at 60°C for 30 minutes, then one litre of 0.2 x SSPE at 60°C for 30-60 minutes, if necessary. The filter was wrapped with Saran Wrap and autoradiographed using Kodak XAR5 X-ray film, with a Dupont Cronex Lightning Plus intensifying screen at -80°C, if necessary.

Plaque hybridization for screening of the recombinant M13 phages and colony hybridization for screening of the recombinant plasmids were performed according to Benton and Davis (1977) and Grunstein and Hogness (1975), respectively. Hybridization and washing conditions for these procedures were the same as described above.

2.3 Cloning of Internal Fragments and Terminal Fragments of pSCL

To clone the internal fragments of pSCL, the plasmid was digested with *Sal* I restriction enzyme. Then, the 2 large internal fragments (3 kb and 4.3 kb) of pSCL were ligated directly into pUC119 and pUC118 vectors at *Sal* I cleavage sites. Initially, because the 5' termini of pSCL were blocked by one or more amino acids, the terminal fragments were cloned by a tailing method. The plasmid pSCL was first tailed with dGTP by using terminal deoxynucleotidyl transferase. Then, the tailed pSCL was digested with *Sal* I. At

the same time, a portion of the vector pUC19 was digested with *Kpn* I and then tailed with dCTP, followed by digestion with *Sal* I. The digested vector was precipitated by addition of PEG 8000 to 6% and NaCl to 0.5 M to eliminate the small polylinker fragment (Lis, 1980). Therefore, the vector should have had one terminus tailed with dC while the other had a *Sal* I cohesive end. The tailed terminal fragments and the vector were mixed together in an annealing buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl and 1 mM EDTA). The mixture was incubated at 68°C for 1 hour, then cooled slowly to room temperature to allow GC base pairing. The DNA was precipitated by the addition of ethanol. The resulting DNA pellet was redissolved in a ligation buffer containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 5% (w/v) PEG 8000. Ligation was performed at 12°C overnight. The resulting open circular DNA molecules were used to transform competent *E. coli* JM83 cells. The recombinant clones were identified by colony hybridization and sizing on agarose gels. The cloned end fragments were then subcloned into the vectors pUC118 and pUC119 for production of single-stranded DNA.

To confirm the sequence of the last several nucleotides at the ends, the end fragments of pSCL were cloned again by the following approach, based on the observation that the linkage between the terminal proteins and the 5' end of the *Streptomyces rochei* linear plasmid DNA was alkali-labile (Hirochika *et al.*, 1984). The plasmid pSCL was first digested with *Sal* I and *Hinc*II. Then, the 0.9 kb left end fragment and the 1 kb right end fragment were purified from low melting agarose gels, separately (Maniatis *et al.*, 1982). Both fragments were treated with 0.2 N NaOH for 30 minutes at 37°C. The DNA was recovered by ethanol precipitation. The denatured DNA was renatured by incubation in an annealing buffer at 68°C for 1 hour, followed by cooling slowly to room temperature. The renatured DNA fragments were recovered by ethanol precipitation. The left end fragment was ligated into a sample of the pUC119 vector which had been digested with *Sal* I and *Sma* I. The right end fragment was ligated into *Hinc*II digested pUC119 vector. Each ligation mixture was used to transform competent *E. coli* MV1193 cells. The resulting

colonies were screened by colony hybridization and sizing on agarose gels. A number of positive clones containing left or right end fragments were selected for sequencing.

2.4 Cloning of 3 Small *Sal* I Fragments by Polymerase Chain Reaction

To obtain the 3 small *Sal* I fragments of pSCL (78 bp, 267 bp and 575 bp, respectively) in their correct orientations, an approach using the polymerase chain reaction (PCR, Kleppe *et al.*, 1971; Saiki *et al.*, 1988) was performed manually. Two oligonucleotides (CCATCGGCGAACACAAA) and (CTGGACCATGTTCACCCT) were synthesized, based on known sequences flanking these fragments. One is located in the left end fragment, the other in the 3 kb *Sal* I fragment. The plasmid pSCL was first denatured by treatment with 0.2 N NaOH and recovered by ethanol precipitation. Then, the DNA was dissolved in a reaction buffer containing 65 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 15 mM $(NH_4)_2SO_4$, 10 mM β-mercaptoethanol, 100 μg/ml BSA, 10% DMSO, 100 μM each of dATP, dGTP, dCTP and dTTP, and 1 μg each of the two primers. The reaction solution was covered with mineral oil and heated at 90°C for 3 minutes, followed by incubating at 60°C for 5 minutes. Two units of *Thermus aquaticus* (*Taq*) DNA polymerase was added. Twenty five to thirty cycles were performed. The temperature and time period used in each reaction cycle were 93°C for 30 seconds, 60°C for 1 minute and 70°C for 3 minutes. The amplified DNA fragment (1.1 kb) was purified from low melting agarose gels. The purified DNA fragment was denatured by heating at 90°C for 3 minutes followed by phosphorylation of its 5' termini using T4 polynucleotide kinase. The phosphorylated DNA was renatured by heating at 90°C for 3 minutes, then cooling slowly to room temperature. The resulting DNA fragment was cloned into the pUC119 vector at a *Hinc*II cleavage site. The original 3 small *Sal* I fragments were also cloned directly from pSCL, separately. To detect any misincorporation during the PCR, the fragments cloned from PCR and directly from pSCL were all subjected to sequencing. No misincorporation of nucleotides during the PCR was found.

2.5 DNA Sequencing

The complete sequence of pSCL was determined using Sanger's chain termination method (Sanger *et al.*, 1977) modified by Tabor and Richardson (1987) for use with Sequenase. Sequenase was used in all sequencing reactions. In addition, an analogue of dGTP, 7-deaza dGTP (Mizusawa *et al.*, 1986), was used to replace dGTP in sequencing mixes to help resolve compressions resulting from the high G+C content. Also, DMSO was used (10% final concentration) to help minimize non-specific stopping during elongation steps in the sequencing reactions (Winship, 1989). Sequencing procedures were performed according to Tabor and Richardson (1987). Double stranded DNA sequencing was performed as described by Chen and Seeburg (1985), except using Sequenase, 7-deaza dGTP and DMSO. Sequencing reaction products were separated electrophoretically with a BRL sequencing apparatus (model S1) on 6% polyacrylamide (38:2, acrylamide:N, N'-methylene bisacrylamide) gels containing 8.3 M urea, using a TEB buffer system (60 mM Tris, 60 mM boric acid and 1.2 mM EDTA). After electrophoresis, the labeled DNA bands were visualized by autoradiography, usually with Kodak XAR5 X-ray film.

To determine the sequence of the last few nucleotides of pSCL, the Maxam and Gilbert chemical sequencing method (Maxam and Gilbert, 1980) was used. The 3' ends of the plasmid pSCL were first labeled by using α-^{32}P-ATP and terminal deoxynucleotidyl transferase (Maxam and Gilbert, 1980). Then, the labeled plasmid was subjected to *Sal* I digestion. The labeled left and right terminal fragments were purified separately from low melting agarose gels. The sequencing reactions and gel electrophoresis were carried out according to Maxam and Gilbert (1980) and Rubin and Schmid (1980).

Sequencing was done on both strands of the pSCL double-stranded DNA. Any ambiguous area has been resequenced at least twice, either by selecting new clones overlapping the difficult region or by making an oligonucleotide as a primer close to the ambiguous region. All regions containing sites (e.g. *Sal* I sites) used for cloning have been

sequenced using clones (usually generated by PCR) which extend through the sites to avoid missing multiple restriction enzyme cleavage sites which may be too close to each other to be detected by electrophoresis. The sequencing data have been read independently by two people to ensure accuracy.

2.6 RNA Isolation, Northern Transfer and Hybridization

Transfer RNA from *S. lividans* was prepared as follows. *S. lividans* 1326 was cultured as described by Thompson *et al.* (1984). The mycelia were harvested by centrifugation at 12,000 x *g* in a Beckman centrifuge (model J2-21) with a JA20 rotor. The cell pellet was resuspended in 10 ml of TE buffer (10 mM Tris-HCl and 1 mM EDTA). The resulting cell suspension was extracted twice with phenol. The total nucleic acids in the aqueous phase were recovered by ethanol precipitation. After redissolving the nucleic acid pellet in a buffer containing 0.1 M NaCl and 0.1 M Tris-HCl (pH 8.0), the solution was loaded on a DEAE-cellulose (Whatman International Ltd.) column, previously equilibrated with the same buffer. The column was washed several times with the buffer, then the tRNA was eluted from the column with an elution buffer containing 1 M NaCl and 0.5 M Tris-HCl (pH 8.0). The eluted solution containing tRNA was incubated at 37°C for 1 hour to hydrolyze aminoacyl esters, followed by the addition of 2 volumes of ethanol. The tRNA was finally recovered from ethanol suspension by centrifugation.

Total RNA was isolated from *S. lividans* and *S. clavuligerus* according to Hopwood *et al.* (1985b) with the following modifications.

When the culture reached the appropriate stage of growth, it was cooled to 0°C immediately by the addition of crushed ice. The mycelia were collected by vacuum filtration on Whatman No. 1 filters, followed by washing several times with pre-cooled water. The collected mycelia were quickly resuspended in Kirby's mixture (Hopwood *et al.*, 1985b) which contains 1% (w/v) sodium triisopropylnaphthalene sulphonate, 6% (w/v) sodium 4-amino salicylate and 6% (v/v) liquified phenol. The cell suspension was

ground for a few minutes in the presence of alumina (ALCOA Chemicals, bacteriological grade) and liquid N_2. The homogenate was transferred to a fresh tube and extracted several times with phenol/chloroform mixture (phenol:chloroform:isoamyl alcohol, 50:50:1). The aqueous phase was transferred to another fresh tube followed by the addition of 1/10 volume of 3 M sodium acetate (NaOAc) and an equal volume of isopropyl alcohol. After standing at room temperature for a few minutes, the precipitated chromosomal DNA was spooled out. The remaining DNA and RNA were recovered by centrifugation. The nucleic acid pellet was resuspended in a DNase digestion buffer containing 40 mM Tris-HCl, pH 8.0, 10 mM NaCl, 6 mM $MgCl_2$ and 20 units of RNase-free DNase. After incubation at 37°C for 30 minutes the mixture was extracted with phenol/chloroform mixture twice and with chloroform once. A small portion of the solution was used to assess the quantity of the RNA by measuring the absorbance at 260 nm (1 A_{260} unit = 40 µg RNA). The rest of the RNA was precipitated by the addition of 1/10 volume of 3 M NaOAc and 2.5 volumes of ethanol. The precipitated RNA was stored at -80°C.

Northern transfer and hybridization were performed according to Williams and Mason (1985) with the following modifications.

The RNA (40 µg) was denatured by using deionized glyoxal (Thomas, 1983) and was then fractionated by electrophoresis on an agarose gel containing 10 mM sodium phosphate buffer (pH 6.5). After electrophoresis, the RNA was transferred to a nylon membrane with 20 x SSPE. The filter was baked at 80°C for 2 hours and prehybridized at 50°C in a hybridization solution (see the section describing Southern hybridization) containing 50% deionized formamide for at least 2 hours. After prehybridization, the ^{32}P-labeled DNA probe was denatured and added. Hybridization was at 50°C overnight. The washing conditions were the same as for Southern hybridization.

2.7 Mapping of 5'-Termini of Two RNA Transcripts

The 5' ends of two RNA transcripts from pSCL were mapped by using the method of 5'-primer extension (Calzon *et al.*, 1987). Two oligonucleotides (GATGTTTCGGTCGGACG for the large RNA, AGCAGGTCGAGGAGTG for the small RNA) were made, based on the known DNA sequences. The experimental procedure was performed as follows.

Samples containing 80 μg of total RNA preparation from *S. lividans* (used as a control), 80 μg of total RNA preparation from *S. clavuligerus* at the culture stage of 12 hours (used for mapping of the large RNA), and 160 μg of total RNA preparation from *S. clavuligerus* at the culture stage of 48 hours (used for mapping of the small RNA) were precipitated from ethanol suspension separately. Each RNA pellet was redissolved in a total of 10 μl of buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 40 mM KCl, 1 mM DTT, 0.5 μg actinomycin D, 100 μM each of dATP, dGTP, dCTP and dTTP, and 15 pmoles of the γ-^{32}P-ATP labeled oligonucleotide primer. This mixture was incubated at 70°C for 5 minutes, and then kept at 45°C for 30 minutes. After adding 9 units of AMV reverse transcriptase, the mixture was incubated at 45°C for an additional hour. When the reaction was finished, the sample was heated at 90°C for 1 minute, then kept on ice. A total of 40 μl of TE buffer and 1 μl RNase (1 mg/ml, Sigma) were added to the mixture. After the sample was incubated at 37°C for 30 minutes, it was extracted with phenol/chloroform mixture twice. The remaining nucleic acids were precipitated by the addition of 5 μg tRNA, 5 μl of 3 M NaOAc and 150 μl ethanol. The resulting pellet was redissolved in 10 μl of sequencing loading buffer containing 80% deionized formamide, 1 mM EDTA, 0.1% (w/v) xylene cyanol and 0.1% (w/v) bromphenol blue. Two μl of the final mixture was loaded on a 6% sequencing gel, along with sequencing reactions conducted using the same oligonucleotide as a primer. The DNA template used in the sequencing reactions was the single-stranded, cloned pSCL 4.3 kb DNA in M13, which is complementary to the RNA transcripts.

2.8 *In Vitro* Transcription and Translation Using *S. lividans* Extracts

Preparation of tRNA from *S. lividans* was carried out as described previously. Growth of the culture, preparation of S30 cell extracts, assays of coupled transcription-translation and protein gel electrophoresis were all performed according to Thompson *et al.* (1984), except that the plasmids pIJ702 (Katz *et al.*, 1983; Hopwood *et al.*, 1985b) and pAT153 (Twigg and Sherratt, 1980) were used as control templates. Various assay conditions, such as different amounts of *S. lividans* tRNA, different amounts of S30 cell extract, different concentrations of magnesium ions and different amounts of input DNA were examined, as described by Chen and Zubay (1983).

3. RESULTS

3.1 Cloning Strategies

To accomplish the major objective of the study, i.e. determination of the total nucleotide sequence of pSCL, it was essential to obtain a large amount of pure pSCL DNA. Since the linear plasmid could not be purified by conventional methods, such as CsCl gradient centrifugation, pSCL DNA was initially obtained by sucrose gradient sedimentation. Contamination with *Streptomyces* chromosomal DNA was recognized to be a problem. Therefore, the cloning of pSCL into an *E. coli* host/vector system was a first important step to achieve the major objective. Besides, Sanger's chain termination sequencing techniques were to be employed. Cloning of the pSCL into sequencing vectors, such as M13 vectors mp18 and mp19 or the single-stranded DNA producing plasmids pUC118 and pUC119, was therefore obviously demanded. Based on these considerations, the first part of this study was to clone the linear plasmid pSCL into *E. coli* vectors.

Initially, a random subcloning approach was employed (Messing *et al.*, 1981). Basically, the intact linear plasmid pSCL was first digested with restriction endonucleases, such as *Taq* I or *Mbo* I, which gave DNA fragments in the range of less than 2 kb. The resulting DNA fragments were then directly cloned into M13 mp18 and mp19 vectors at *Acc* I or *Bam* HI cleavage sites. Plaque hybridization revealed that many recombinant M13 phages indeed contained DNA fragments from pSCL. However, subsequent sequencing analysis indicated that spontaneous recombination and deletion events within the cloned pSCL DNA fragments occurred in the *E. coli* host. It was assumed that the high G+C content in *Streptomyces* DNA might have caused these recombinations or deletions in *E. coli* since almost all regions involved in these events had clusters of GC rich sequences.

To avoid the problem mentioned above, another cloning approach was used. In this approach, the large pSCL DNA fragments were cloned into *E. coli* vectors, using the

known restriction endonuclease cleavage sites on pSCL. From a physical map of pSCL (Keen *et al.*, 1988), 6 *Sal* I cleavage sites have been identified, which gave the sizes of 7 DNA fragments ranging from 4.3 kb to 78 bp, including the two terminal fragments. These *Sal* I cleavage sites were used for cloning of the 7 pSCL DNA fragments into *E. coli* cloning vectors.

3.1.1 Cloning of Internal *Sal* I Fragments

Cloning of the 5 internal *Sal* I fragments into *E. coli* vectors was straightforward. Initially, these fragments were cloned into the pUC19 vector without any problem. When the 3 kb and the 4.3 kb *Sal* I fragments were subcloned into M13 mp18 and mp19 vectors with the *E. coli* host JM103, spontaneous deletions within the cloned DNA fragments also occurred. Apparently, the sizes of the cloned DNA fragments were not beyond the practical limits of the M13 cloning system (Yanisch-Perron *et al.*, 1985). It was believed that these spontaneous deletions observed in the pSCL DNA fragments were more likely due to the properties of the cloned DNA fragments and the recombination systems of the *E. coli* host. When a new *E. coli* host, JM109, with a *recA*1 mutation (Yanisch-Perron *et al.*, 1985), was used instead of JM103, the two large pSCL DNA fragments were stably cloned into M13 vectors in both orientations. It appears likely that the *E. coli* host recombination system, particularly the RecA protein, was responsible for the deletions we observed.

There are three small fragments in pSCL released by digestion with *Sal* I. They were directly cloned into M13 vectors. However, because there are few available restriction endonuclease cleavage sites within these fragments, it was difficult to determine their orientations, even after the complete sequences of the fragments had been obtained. Because these three small *Sal* I fragments are adjacent, it was clear that their correct orientations could be determined if they could be cloned directly on a single fragment of DNA. Cloning of a such DNA fragment was achieved by using the polymerase chain

reaction (PCR, Kleppe *et al.*, 1971; Saiki *et al.*, 1988). PCR reactions were performed manually in the presence of 10% DMSO. Thus, a DNA fragment containing the 3 small *Sal* I fragments was successfully obtained by using PCR, and it was cloned into pUC119. A number of reports have noted the artifacts caused by misincorporation of nucleotides during PCR (reviewed by Vosberg, 1989). To obtain accurate sequence information, the DNA fragments cloned from PCR amplification reactions and directly from pSCL were all subjected to sequence analysis. Fortunately, no such misincorporation was found.

3.1.2 Cloning of the Terminal Fragments of pSCL

The two terminal fragments of pSCL were obtained by *Sal* I digestion of pSCL, the 0.9 kb on the left and the 2.8 kb on the right. The 5' termini of pSCL are blocked by terminal proteins, and even after treatment with proteinase, one or more amino acid residues remain (Keen *et al.*, 1988). Since this amino acid-DNA complex at the 5' termini prevents ligation of the terminal fragments into cloning vectors, direct cloning of these two fragments was impossible. It has been reported that the linkage between terminal proteins and the 5' ends of a *Streptomyces rochei* linear plasmid DNA (pSLA2) was alkali-labile (Hirochika *et al.*, 1984). This is also true for adenoviruses (Rekosh *et al.*, 1977) and bacteriophage ϕ29 (Salas *et al.*, 1978). In the case of pSLA2, the intact plasmid was treated with NaOH and then later renatured so that the terminal proteins were removed from the 5' termini of the DNA. After digestion of the renatured plasmid with restriction endonucleases, the terminal fragments were purified and ligated into cloning vectors. This approach was also attempted with pSCL, but very little intact plasmid was recovered after the alkali treatment. It is possible that the denatured plasmid DNA formed a stable, large stem-loop structure due to the long inverted terminal repeat sequences.

A different approach was used to overcome this problem. This approach took advantage of the protein-free 3' termini of pSCL and the *E. coli* host repair system. The 3' termini of pSCL were first tailed with dGTP and then digested with *Sal* I. The two

terminal fragments should then have had one dG tailed terminus and one *Sal* I cohesive terminus. Meanwhile, the vector pUC19 was digested with *Kpn* I and then tailed with dCTP. The tailed vector was then cleaved with *Sal* I. Thus, the resulting vector contained one end tailed with dC and also another with a *Sal* I cohesive end. After mixing the tailed vector with the tailed terminal fragments and incubating at an annealing temperature, the tailed ends were annealed together by dG:dC base pairing. In a subsequent ligation reaction, the *Sal* I cohesive ends on the terminal fragments and on the vector were covalently joined by DNA ligase. The resulting open circular DNA molecules were finally transformed into the *E. coli* host JM83. The host repair system removed the amino acids linked to the 5' ends of the terminal fragments. In fact, the principle of this cloning approach has been used previously by Okayama and Berg (1982) to clone cDNA into vectors. A major difference is that in the latter there were no amino acids covalently linked to the 5' end of the cDNA. By using this tailing method, the two terminal fragments of pSCL were cloned into pUC19 vectors quite successfully.

Initially, dG:dC base pairing was used because of the relatively higher thermostability than dA:dT base pairing. An adverse consequence of dG and dC tailing was the formation of a long GC homopolymer at one end of each of the two terminal fragments. It became very difficult to identify the last few nucleotides at the termini because the last 7 nucleotides at the termini of pSCL consist only of dG and dC. To obtain the correct sequence from the termini of pSCL, it was necessary to reclone the terminal fragments by a different method.

The second approach used to clone the terminal fragments of pSCL was based on the observation that the linkage between terminal proteins and the 5' ends of linear plasmid DNA is alkali-labile (mentioned earlier). Instead of using the intact plasmid DNA, the two terminal fragments of pSCL were first purified from restriction endonuclease digests and then treated with NaOH. The denatured terminal fragments were renatured under appropriate conditions. The renatured left terminal fragment (0.9 kb), generated by *Sal* I

digestion of pSCL, was inserted into a pUC119 vector between its *Sal* I and *Sma* I sites, while the right terminal fragment (1.1 kb), generated by *Hinc*II digestion of pSCL, was inserted into a pUC119 vector cut at its *Hinc*II site. The right end fragment generated by *Hinc*II was chosen in preference to the larger *Sal* I fragment to improve the recovery of the denatured DNA fragment after treatment with NaOH. This approach allowed the two terminal fragments to be cloned again. Subsequent sequence analysis of the new clones provided complete sequence information on the two te[illegible]ents, and confirmed the previously determined sequences.

The fact that the termini of pSCL can be [illegible]y ligated (after alkaline hydrolysis) to the blunt ends of cloning vectors suggests that these termini are blunt. This conclusion was further confirmed by sequencing (see the next section for details).

It is interesting to note that an attempt had been made to subclone the tailed right end fragment from the pUC19 into M13 vectors, but without success. Spontaneous deletions within the right end fragment occurred in the M13 system, even when *E. coli* JM109 (RecA$^-$ strain) was used. The right end fragment was subjected to digestion with restriction enzymes. The three smaller DNA fragments (each about 1 kb) resulting from digestion with *Sst* II and *Hinc*II were subcloned again into M13 vectors. Surprisingly, they were all unstable. When the right end fragment was cloned into pUC119, a vector capable of producing single-stranded DNA, the recombinant plasmid was stable and single-stranded plasmid DNA was produced without any problem. The reason for the instability of this fragment in the M13 system is not clear, but M13 recombinants are known to be less stable than plasmid derivatives (Messing, 1983). To avoid any loss of sequence information which might occur with the M13 system, most sequencing of the right terminal fragment was performed using the single-stranded plasmid approach.

3.2 Sequencing Strategies

Because *Streptomyces* DNA has an unusually high G+C content, only a few restriction endonucleases are available to generate suitable DNA fragments for subsequent sequencing. To allow rapid sequence analysis of the fragments of pSCL cloned into *E. coli*, the exonuclease III unidirectional deletion method (Henikoff, 1984, 1987) was employed to generate overlapping DNA clones for sequencing. The main advantage of this technique is that the generation of overlapping clones is independent of the arrangement of restriction enzyme sites on a DNA fragment of interest. Using this method, a large number of overlapping clones generated from each fragment of pSCL was quickly collected. The resulting recombinant plasmids with various deletions were screened by electrophoresis on 20 cm x 20 cm agarose gels. As little as 100 bp difference in size between these plasmids can be distinguished. Clones with the appropriate deletions were selected for production of single-stranded plasmid DNA to be used for sequencing.

Most sequencing of pSCL was performed using single-stranded M13 phage or single-stranded plasmid DNA. Double stranded DNA sequencing has been used occasionally but only gave reliable results up to about 150 bp. The reason for the limited usefulness of double-stranded DNA sequencing is believed to be the high G+C content of *Streptomyces* DNA. The denatured double-stranded DNA reassociates rapidly during the sequencing reactions, so that further elongation of the newly synthesized DNA fragments is prevented. In addition, direct sequencing using the linear plasmid pSCL as a template and a pSCL-specific oligonucleotide as a primer has failed to give any meaningful data.

We have found that Sequenase (from USB) gave more clear and reliable sequencing results than the Klenow fragment of *E. coli* DNA polymerase I on pSCL DNA. Also, "compressions" resulting from the high G+C content can usually be resolved by using 7-deaza dGTP, which replaces the normal dGTP in sequencing mixes (Mizusawa *et al.*, 1986). A problem associated with the use of 7-deaza dGTP is that DNA bands beyond 250 bp on sequencing gels were not as sharp as when dGTP was used. The reason for this

phenomenon is not clear. DMSO was included in all sequencing reactions (at a final concentration of 10%). Its ability to prevent the formation of secondary structures minimized non-specific pausing during elongation steps.

To determine the sequences of the last few nucleotides on the left and right termini of pSCL, the two alkali-treated, cloned terminal fragments were subjected to sequencing using Sanger's chain termination method. The result is presented in Fig. 1. The sequencing data showed that the left (1) and right (2) terminal sequences are identical and the sites on the vectors into which the terminal fragments were inserted are exactly what were expected. In order to confirm that all sequence information was retained during the cloning procedures, the 3' termini of the native pSCL were labeled with α-^{32}P-ATP using terminal deoxynucleotidyl transferase and subjected to sequencing using Maxam and Gilbert's chemical method. The sequencing data are shown in Fig. 2 (1 indicates the left end of pSCL, 2 indicates the right end of pSCL). The results clearly show that the sequences are identical to those determined by the chain termination method, and demonstrate that the complete terminal sequences of pSCL have been successfully cloned. These results further confirm that the termini of pSCL are indeed blunt.

Combining unidirectional deletion and PCR techniques with the modifications to the normal sequencing methodology mentioned above, I believe that the data collected from sequencing both strands of pSCL are reliable and accurate. In fact, as pointed out in Materials and Methods, the sequencing of any ambiguous area has been repeated at least twice, either by selection of new clones overlapping the region or by making an oligonucleotide for use as a primer closer to the region. More than 20 oligonucleotides have been made to help sequence those difficult regions. The sequencing data were read independently by two people before they were entered into the computer.

Fig. 1. Autoradiograph of a 6% polyacrylamide denaturing gel demonstrating the terminal sequences of pSCL in pUC119, using Sanger's chain termination method. (1). The sequence of the left terminal fragment cloned in pUC119 between the *Sma* I and *Sal* I sites. (2). The sequence of the right terminal fragment cloned in pUC119 at the *Hinc*II site. Lanes G, A, T and C indicate the specific dideoxynucleotides used in the sequencing reactions (G: ddGTP; A: ddATP; T: ddTTP; C: ddCTP). Arrows pointing upward indicate the vector sequences. Arrows pointing downward indicate the terminal sequences of pSCL.

GATCGATC

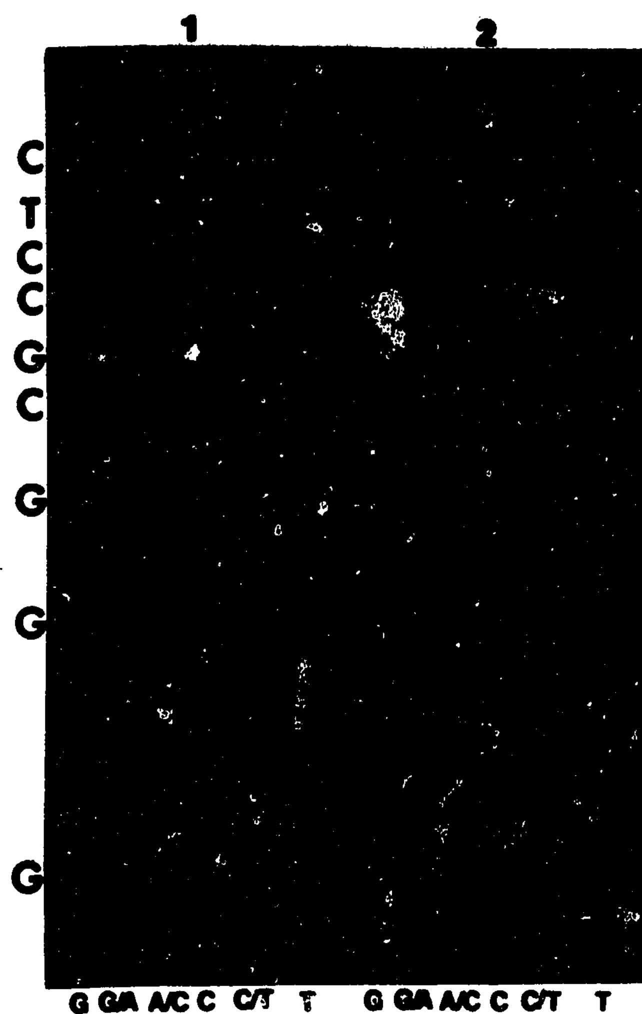
AGCCATGGGGGCGCCTCGCC

AGATCTCAGGGCGCCTCGCC

1

2

Fig. 2. Autoradiograph of a 20% polyacrylamide denaturing gel demonstrating the sequence of the 3' termini of pSCL DNA, using the Maxam and Gilbert chemical cleavage method. The native pSCL DNA was labeled with α-^{32}P-ATP at the 3' termini using terminal deoxynucleotidyl transferase. Then, after cleavage with *Sal* I, the terminal fragments of pSCL were purified from low melting agarose gels. The purified terminal fragments were partially cleaved at guanines (G), guanines and adenines (G/A), adenines and cytosines (A/C), cytosines (C), cytosines and thymines (C/T), and thymines (T). (1). The sequence of the left terminal fragment. (2). The sequence of the right terminal fragment. The small arrows indicate the positions of the ^{32}P labeled and cleaved DNA fragments.


1
2
C
T
C
C
G
C
G
G
G
G G/A A/C C C/T T G G/A A/C C C/T T

3.3 Complete Sequence of pSCL

The complete nucleotide sequence of pSCL is presented in Fig. 3. The total length of pSCL is 11,696 base pairs, which is in close agreement with the estimate in the previous study (Keen *et al.*, 1988). The total percentage of G+C is 71.9%, which is in the range of the G+C content of typical *Streptomyces* DNA. Most of the restriction endonuclease cleavage sites described previously (Keen *et al.*, 1988) have been confirmed to be correct. However, two new cleavage sites for the enzymes *Kpn* I and *Sst* II have been found at the extreme termini of pSCL. The reason for missing these sites in the previous studies is most likely the locations of these sites very close to the ends. The fragments generated by the enzymes were much too small to be detected by agarose gel electrophoresis.

3.3.1 Long Inverted Terminal Repeats

As result of cross-hybridization studies, Keen *et al.* (1988) proposed that pSCL has long inverted terminal repeat sequences (about 900 bp). The sequencing data clearly indicate that pSCL indeed has a 969 bp long inverted terminal repeat sequence. Figure 4 presents the alignment of the 969 bp left sequence (PSCLLEFT) with the 969 bp right sequence (PSCLRIGHT) of pSCL. From the alignment of the two sequences, it is clear that the left terminal sequence of pSCL is completely identical to the right terminal sequence up to 690 bp. Then, after a 3 bp difference between the two sequences, ACA on the left sequence, CAC on the right, the identical sequences extend to 894 bp. Even beyond the 894 bp, the two terminal sequences are still highly homologous up to 969 bp. The length of the long inverted terminal repeats is in good agreement with the previous results (Keen *et al.*, 1988). Long terminally repeated sequences are common characteristics of all known linear plasmids. In the case of a linear plasmid (pSLA2) from *S. rochei* , 614 bp inverted terminal repeats have been found. The long inverted terminal repeats of pSCL are about

Fig. 3. Complete double-stranded nucleotide sequence of pSCL, from 5' terminus to 3' terminus and from left end to right end. The arrows between the double strands indicate inverted repeats. Single lines above or below the double-stranded sequence indicate the predicted ORFs. Double lines below the double-stranded sequence indicate the demonstrated RNA transcripts. Small thicker arrows indicate the orientations of the ORFs and the transcriptional directions of the two RNA transcripts.

```
                   10        20        30        40        50        60
                    |         |         |         |         |         |
  1  CCCGCGGAGCGGGTACCTAGGCGCTGCGCGCCTAGCGAGTGCATGCCCCCGCTGCGCGGG
     GGGCGCCTCGCCCATGGATCCGCGACGCGCGGATCGCTCACGTACGGGGGCGACGCGCCC

 61  GGCAGTGACGGAGTGTCACGGAGCCCGCAAGCGGGCTCCGTTGGGCCCGCTGCGCGGGCC
     CCGTCACTGCCTCACAGTGCCTCGGGCGTTCGCCCGAGGCAACCCGGGCGACGCGCCCGG

121  TGTCCTCTCCGCTGCGCTGCGAGGGGCACTCTTCCGCTGTGCTTCAGAGTGCTGGCCTGG
     ACAGGAGAGGCGACGCGACGCTCCCCGTGAGAAGGCGACACGAAGTCTCACGACCGGACC

181  ATGAACTCTGGAATGGGCCCCGGGCTGCGCCCGGGGCCGAGGCCGCTCCGCTGCGCTCCG
     TACTTGAGACCTTACCCGGGGCCCGACGCGGGCCCCGGCTCCGGCGAGGCGACGCGAGGC

241  GGCCCTGGCCGCTTCGCGGCGGCTGGCTACGGGGAGTGGGGGCGGGGGTGGTGGAGTTGG
     CCGGGACCGGCGAAGCGCCGCCGACCGATGCCCCTCACCCCCGCCCCCACCACCTCAACC

301  GTGCGGGTTCTGAGCCTGCATCAAAAGATGCGGCATGTACGGTTGGCAGATCCGGGCCAC
     CACGCCCAAGACTCGGACGTAGTTTTCTACGCCGTACATGCCAACCGTCTAGGCCCGGTG

361  ACCCCCCGACGAGGCGCTGTGCGCCGATCTGACAGCCGTTCACCCCAGGATGGACACAGA
     TGGGGGGCTGCTCCGCGACACGCGGCTAGACTGTCGGCAAGTGGGGTCCTACCTGTGTCT

421  GACCCACTCCGGGGGTCGGGAGACGCTGACGCGGGCTCAGGCAGCCTCCTGGGGCTGCAT
     CTGGGTGAGGCCCCCAGCCCTCTGCGACTGCGCCCGAGTCCGTCGGAGGACCCCGACGTA

481  CAAAGCTGCATCCTGGCCGGACTGTGTAAAGTCTGGATTCTGGCGCACCCCCTCTCCGG
     GTTTCGACGTAGGACCGGCCTGACACATTTCAGACCCTAAGACCGCGTGGGGGAGAGGCC

541  GTTCGCCCCCGGCCGGGGGCGAACACCCCGAACAGACGGCCACCGCGTCCCGCCCGGCGA
     CAAGCGGGGGCCGGCCCCCGCTTGTGGGGCTTGTCTGCCGGTGGCGCAGGGCGGGCCGCT

601  CCCCCGGCCGGGGCGCCGTCCTCCCGGCCGCCGCCCCCGGGGCGCCACCACACGCCCCAC
     GGGGGCCGGCCCCGCGGCAGGAGGGCCGGCGGCGGGGGCCCCGCGGTGGTGTGCGGGGTG

661  GCCCAGGAGGACGGGCACCCCCCACCCCAAACAGCCCCTGACCAGCACAGACACCCACCG
     CGGGTCCTCCTGCCCGTGGGGGGTGGGGTTTGTCGGGGACTGGTCGTGTCTGTGGGTGGC

721  CAGCCACCCGCTCGCCCCCGGCCCGACCCCACCCCAACACCCCCGAACACCCCGGCTCGA
     GTCGGTGGGCGAGCGGGGGCCGGGCTGGGGTGGGGTTGTGGGGGCTTGTGGGGCCGAGCT

781  CCCACCCGAACACCACCCCGGCGGCACCCCGGCGGCACCCTCCGGAACACCCCAAAACAG
     GGGTGGGCTTGTGGTGGGGCCGCCGTGGGGCCGCCGTGGGAGGCCTTGTGGGGTTTTGTC

841  TTGAACATTCAATTTCCCCATCGGCGAAGCACCGGCCACCACACCCAACCTGACGGCCCG
     AACTTGTAAGTTAAAGGGGTAGCCGCTTCGTGGCCGGTGGTGTGGGTTGGACTGCCGGGC

901  TCACACACCCATCGGCGAACACAAACGGTCGACCACAACACCCCCAAACGGCAAAAGCGAT
     AGTGTGTGGGTAGCCGCTTGTGTTTGCCAGCTGGTGTTGTGGGGTTTGCCGTTTTCGCTA

961  GACCCCGAAATACCCCATTAGCCGGAACCCGGAAAACCGCCCGGAAAAGCCTTTCGGCTC
     CTGGGGCTTTATGGGGTAATCGGCCTTGGGCCTTTTGGCGGGCCTTTTCGGAAAGCCGAG
```

1021 CCACCCCCGAAACGGGGCGGGAGCCGAAATGAAAAGAGACGGAGAAACGGCCGGGGAGCC
GGTGGGGGCTTTGCCCCGCCCTCGGCTTTACTTTTCTCTGCCTCTTTGCCGGCCCCTCGG

1081 GGTTACCGGCCCCCCGGATCAGGTGGTGCGGCGGTCGAGGGTGCGGGCCAGGGCTTCGGT
CCAATGGCCGGGGGGCCTAGTCCACCACGCCGCCAGCTCCCACGCCCGGTCCCGAAGCCA

1141 GACGCGGTCGTCCTGGTCTGCGATGCGGATCAGGGTGCGCAGTATCTCGGACTTGTCGAC
CTGCGCCAGCAGGACCAGACGCTACGCCTAGTCCCACGCGTCATAGAGCCTGAACAGCTG

1201 CGGGCGGCCGGTGCGGCGGCGCAGGGTGAGGGCGAGCTGGTCGAGGGTGAGGGCGTCGTC
GCCCGCCGGCCACGCCGCCGCGTCCCACTCCCGCTCGACCAGCTCCCACTCCCGCAGCAG

1261 CTGGTCGAGGAGCAGCGTGTACTTGCTGGGCTTGGGCGCGGGTGCCGTCGTGCCGTTCTG
GACCAGCTCCTCGTCGCACATGAACGACCCGAACCCGCGCCCACGGCAGCACGGCAAGAC

1321 CCGTTCTGCCGTTCTGCCGTCATGGGGCGTTTGAACTGGGGTTTCTTCCTTGAAACGGCT
GGCAAGACGGCAAGACGGCAGTACCCCGCAAACTTGACCCCAAAGAAGGAACTTTGCCGA

1381 GGACAGGCCGGTGAGAAGGGCGTTCACGCCGCCGGTGTCCACGGTGAAGTCGTTCTTCTT
CCTGTCCGGCCACTCTTCCCGCAAGTGCGGCGGCCACAGGTGCCACTTCAGCAAGAAGAA

1441 GCTGCTCATGCTGCGGGGGCCTCCGGGAGGTACCCGGCCTTCACGGCCAGGTCGCGGAAC
CGACGAGTACGACGCCCCCGGAGGCCCTCCATGGGCCGGAAGTGCCGGTCCAGCGCCTTG

1501 ATGCGCGCGACGTCGGCCGTGCGCGCGTACTGGGCGAGGGTCTGGCCCTGGTTCTTGGCT
TACGCGCGCTGCAGCCGGCACGCGCGCATGACCCGCTCCCAGACCGGGACCAAGAACCGA

1561 TCGGCCACGGCGGTGCGCAGCGGGATCTCCCCGACGACGGGGAGGGGAAGTCGGTGCAGC
AGCCGGTGCCGCCACGCGTCGCCCTAGAGGGGCTGCTGCCCCTCCCCTTCAGCCACGTCG

1621 GCGTCGTAGACCGCCTGGTGGAAGCCGGAGAAGGGGCGGGCGACGCGGTTGATGACCAGG
CGCAGCATCTGGCGGACCACCTTCGGCCTCTTCCCCGCCCGCTGCGCCAACTACTGGTCC

1681 CCCCGGTAGCGGGGCGGTGCCATGCGGGTCTCGCGGACGAGGGTCTGCACCTGACCGAGG
GGGGCCATCGCCCCGCCACGGTACGCCCAGAGCGCCTGCTCCCAGACGTGGACTGGCTCC

1741 AGGAGTTCCAGGGCTTCGATGGAGAACTCGTCGACGTCCACGGGGATCAGGATGTCGTGA
TCCTCAAGGTCCCGAAGCTACCTCTTGAGCAGCTGCAGGTGCCCCTAGTCCTACAGCACT

1801 GCCCAGCGCAGGACGTTGTCCGTATCGACGTCGAGCGCGGGGCGGCAGTCGACAAGGCAG
CGGGTCGCGTCCTGCAACAGGCATAGCTGCAGCTCGCGCCCCGCCGTCAGCTGTTCCGTC

1861 TGGTCGAAGCCGTCCTCCAGGTGCTCCAGGACCCAGGACAACCGCTCTTCCTTCGAGCGG
ACCAGCTTCGGCAGGAGGTCCACGAGGTCCTGGGTCCTGTTGGCGAGAAGGAAGCTCGCC

1921 GCGCTGTGGAGCTGCCGCGGCAGGGTGAACAGGTCCAGCGCGGACGGGATCACCCACAGG
CGCGACACCTCGACGGCGCCGTCCCACTTGTCCAGGTCGCGCCTGCCCTAGTGGGTGTCC

1981 TTCTCGGAGTGCCGACGGACCAGGGACTGGGCGCCCGCCCGGTCGCAGCCCTCCAGCACC
AAGAGCCTCACGGCTGCCTGGTCCCTGACCCGCGGGCGGGCCAGCGTCGGGAGGTCGTGG

```
2041 GCGCGGGCCAGCGTCAGCTCACCCGCAGGGTTGAGCCGGGGCAGTTTCAGGGCGGTGGTC
     CGCGCCCGGTCGCAGTCGAGTGGGCGTCCCAACTCGGCCCCGTCAAAGTCCCGCCACCAG

2101 AGGTTCCCCTGCCCGTCGAGGTCCACCAGGAGTGTCCGGCGGCCCGCAGCCGCCAGCGTG
     TCCAAGGGGACGGGCAGCTCCAGGTGGTCCTCACAGGCCGCCGGGCGTCGGCGGTCGCAC

2161 CCGCCCAGATGGAGCGTGGTCGTCGTCTTGGCGACGCCGCCCTTCTGGTTCAGCAGCGCG
     GGCGGGTCTACCTCGCACCAGCAGCAGAACCGCTGCGGCGGGAAGACCAAGTCGTCGCGC

2221 TCGCGTGTCATTGCCATGCGGCAGACCGTACCGCCGTCACGACGGCACGCCGCAGTTACG
     AGCGCACAGTAACGGTACGCCGTCTGGCATGGCGGCAGTGCTGCCGTGCGGCGTCAATGC
         <-- ORF 1
2281 GCGTGCCGTCGTGACGGTGTGACGGCAGCAAGCCTGTGAACTGGGGCCGTAGTCCGCTGC
     CGCACGGCAGCACTGCCACACTGCCGTCGTTCGGACACTTGACCCCGGCATCAGGCGACG

2341 GGCCCCGGGTTCACCGGACCGGCGCCGCGCCCTCGGCCCCGGGGTCGGCTGTATGGGTTG
     CCGGGGCCCAAGTGGCCTGGCCGCGGCGCGGGAGCCGGGGCCCCAGCCGACATACCCAAC

2401 TCTGGGGTGTCTGGGGTGGTGGCCCTGTATGGGGTGTCTGGGGTCCAGCAGACCCCATCG
     AGACCCCACAGACCCCACCACCGGGACATACCCCACAGACCCCAGGTCGTCTGGGGTAGC

2461 AGGGGGGTCTTCCTGATCCCAGACACCCCATACACCGCAGGTCACGACCCCCTACAGGGG
     TCCCCCCAGAAGGACTAGGGTCTGTGGGGTATGTGGCGTCCAGTGCTGGGGGATGTCCCC

2521 CATCACCCCAGACACGCCGTCTGGCACCTGCTGTCTGGGGCAGGATGGCCGCGTGACCGC
     GTAGTGGGGTCTGTGCGGCAGACCGTGGACGACAGACCCCGTCCTACCGGCGCACTGGCG

2581 CCACCGCCGCACCCTCGGGACCGGGCCCGCGCCCGCGCCCGCCGTTCCCGCGCCGCCCCG
     GGTGGCGGCGTGGGAGCCCTGGCCCGGGCGCGGGCGCGGGCGGCAAGGGCGCGGCGGGGC

2641 GCCCCGGCCGCCCCTCGCCGCCGAACCCCCGGCCGGGCACGAGCGCCAGGACGCGGCCGG
     CGGGGCCGGCGGGGAGCGGCGGCTTGGGGGCCGGCCCGTGCTCGCGGTCCTGCGCCGGCC

2701 ACCACCCGCTCGCCCCGGGGCCGGACCCGCCGGACAACGGCGCGCCCTCGGCCCCGGCCG
     TGGTGGGCGAGCGGGGCCCCGGCCTGGGCGGCCTGTTGCCGCGCGGGAGCCGGGGCCGGC

2761 GGGCAGGAAGACCGCCGACTGACCCGGCCGCCCTTACGCGTGCGTGCGCGGTACGCACGG
     CCCGTCCTTCTGGCGGCTGACTGGGCCGGCGGGAATGCGCACGCACGCGCCATGCGTGCC

2821 CACGGGTCACGCGCGGCACGCGTTGGACCCGGCCCGGGTGGTCAGACCGGAGCGTTCTCC
     GTGCCCAGTGCGCGCCGTGCGCAACCTGGGCCGGGCCCACCAGTCTGGCCTCGCAAGAGG

2881 GCCGGGTAGGGCACGCCCGGCATCCACGTGCTCACCGTGTCCCGCAGGACCCACTCCACC
     CGGCCCATCCCGTGCGGGCCGTAGGTGCACGAGTGGCACAGGGCGTCCTGGGTGAGGTGG

2941 CTCGTGCCACCGGCCGCCCGGTCCTCCGCCCGCCACGCGCGCCCGGCATGGGGCCCGGCG
     GAGCACGGTGGCCGGCGGGCCAGGAGGCGGGCGGTGCGCGCGGGCCGTACCCGGGCCGC

3001 TGCCCGGCCGCCAGCAGGCACGGCACCGCCATGTCCGGCGCCATCGTGTAGCACTCCACC
     ACGGGCCGGCGGTCGTCCGTGCCGTGGCGGTACAGGCCGCGGTAGCACATCGTGAGGTGG
```

3061 CGCGCCCACTCCATCCACTCCCGAGCGCCCCCCGCCCGGTGAGGGCCCCAGTGCCCCGCC

GCGCGGGTGAGGTAGGTGAGGGCTCGCGGGGGGCGGGCCACTCCCGGGGTCACGGGGCGG

3121 GGATCGTCACACACCAGTCCGTCAAGCCCGGCTCCGCACAACGCCGTCGTCGTCATCGTC

CCTAGCAGTGTGTGGTCAGGCAGTTCGGGCCGAGGCGTGTTGCGGCAGCAGCAGTAGCAG

3181 ATCGTCGTCTCCCCTTCCGTTGGTGAGGGGCGAGCGTATCCACCTATACCGGCCCATACC

TAGCAGCAGAGGGGAAGGCAACCACTCCCCGCTCGCATAGGTGGATATGGCCGGGTATGG

3241 AGCAAATACCAGCCAATCCCAGGGCACATGGACCCGGCCCCGGCCCTTCCTCGCGTGTGG

TCGTTTATGGTCGGTTAGGGTCCCGTGTACCTGGGCCGGGGCCGGGAAGGAGCGCACACC

3301 GCGTGTGGGCGTGTGGGCGCCTGGGCGTTGGACCCCCGCCAGGACCCCTCAGAGCCCGAC

CGCACACCCGCACACCCGCGGACCCGCAACCTGGGGGCGGTCCTGGGGAGTCTCGGGCTG

3361 ATGGCCCCGACATACGGCCCGACATCGGCCCGACATAGCGGCCCCCGGTATGTCGGGCCG

TACCGGGGCTGTATGCCGGGCTGTAGCCGGGCTGTATCGCCGGGGGCCATACAGCCCGGC

3421 GGGGCCGCCCATGTCGGGGCCATGTCGGGGCCTATGTCGGGTCGTATGTCGGGCCTGACC

CCCCGGCGGGTACAGCCCCGGTACAGCCCCGGATACAGCCCAGCATACAGCCCGGACTGG

3481 TGCGGAAACCGCCTCCGCCGGCCGCGCGTGCGCATCCGGCCCCGGCGGGCCGCCGGGGCC

ACGCCTTTGGCGGAGGCGGCCGGCGCGCACGCGTAGGCCGGGGCCGCCCGGCGGCCCCGG

3541 GAGGACGCCGACGGCCACCGGGGCCAGCCCCCGCGCGTGCGCATCACGCGTGCGCCTTGG

CTCCTGCGGCTGCCGGTGGCCCCGGTCGGGGGCGCGCACGCGTAGTGCGCACGCGGAACC

3601 ACCCCCGGCCGTG[illegible]CCCTCCGGCCGGTGATGAAGATGAAGAAGAAGCGGGCGCCCGCCCC

TGGGGGCCGGCAC[illegible]GGAGGCCGGCCACTACTTCTACTTCTTCTTCGCCCGCGGGCGGGG

3661 GGGCCGCCCCGGAATGACCGTCATCCGGACCGTCATGACCACCGTCACCGCACCGTCAtG

CCCGGCGGGGCCTTACTGGCAGTAGGCCTGGCAGTACTGGTGGCAGTGGCGTGGCAGTAC

3721 GACGCGGTGGCCATGACGGTCCGGCCCATCGCCATGACGGTCCCATGACGGTCCGGATGA

CTGCGCCACCGGTACTGCCAGGCCGGGTAGCGGTACTGCCAGGGTACTGCCAGGCCTACT

3781 CGGTGCCCATGACGGTGCTGCCACGGGCGACAACCCGGCACCTTCTCGTGCGCGCACGAC

GCCACGGGTACTGCCACGACGGTGCCCGCTGTTGGGCCGTGGAAGAGCACGCGCGTGCTG

3841 ACGCGCGCACCTTGGGCCCGGCCCCGGAGGCCCGCCCGGCGGTGAAGAAGAAGAGGAAGT

TGCGCGCGTGGAACCCGGGCCGGGGCCTCCGGGCGGGCCGCCACTTCTTCTTCTCCTTCA

3901 TCATTGATCCTGGGTGCCGTGTGGTCCGCCGGTTCGGCGGGCCCACGACCCCGGACGTCC

AGTAACTAGGACCCACGGCACACCAGGCGGCCAAGCCGCCCGGGTGCTGGGGCCTGCAGG

3961 TGTGCGGCGATCTGACGGCCGATCTCCGATCCCCCGGCCCCGGAGGGCGCCTGACGGATC

ACACGCCGCTAGACTGCCGGCTAGAGGCTAGGGGGCCGGGGCCTCCCGCGGACTGCCTAG

4021 ATCGGGCCGGAGATCGCCGCACAGGCGGCTTCCCGGGCGGGCTAGCCGATCTGCTGATCG

TAGCCCGGCCTCTAGCGGCGTGTCCGCCGAAGGGCCCGCCCGATCGGCTAGACGACTAGC

4081 TCCCAGAACTCCGGCTGCTCCGGGATCAGGCCCGCCGCCACCGCCTGCTCCTCGGTCAGC
AGGGTCTTGAGGCCGACGAGGCCCTAGTCCGGGCGGCGGTGGCGGACGAGGAGCCAGTCG

4141 GTGCCCGCCGGGATCGGGACGTTGCCCGCGTACTGGACCGCCGTCCACGGCCACGTCCCG
CACGGGCGGCCCTAGCCCTGCAACGGGCGCATGACCTGGCGGCAGGTGCCGGTGCAGGGC

4201 CCCTCGCAGAACTGGGCTGGCCACTCACGCTCGGACCGGGGCCCGCGCCAGGGCAGCCAG
GGGAGCGTCTTGACCCGACCGGTGAGTGCGAGCCTGGCCCCGGGCGCGGTCCCGTCGGTC

4261 TCCATCAGCCGCAGCTCCTCCGCCCCCATCCCGGTCGCCGGGGACAGCCCGAACCCGAAC
AGGTAGTCGGCGTCGAGGAGGCGGGGGTAGGGCCAGCGGCCCCTGTCGGGCTTGGGCTTG

4321 TCCGGCCACCGCAGCCACAGGCTGCTCCCGATCGGGCGCAGGTCCCGGCGCCTGACCCCG
AGGCCGGTGGCGTCGGTGTCCGACGAGGGCTAGCCCGCGTCCAGGGCCGCGGACTGGGGC

4381 GGCGCCGCGTGCGGGGAGTGGGCCTCCAGCAGCATCGCCGCGCCGTCCGACGCCTGACGG
CCGCGGCGCACGCCCCTCACCCGGAGGTCGTCGTAGCGGCGCGGCAGGCTGCGGACTGCC

4441 ACCATCTCCAGCGCCGACATGAGCGGCCGGGCCGCCTCCTCGTCGCTCGCCCCGCTGGTG
TGGTAGAGGTCGCGGCTGTACTCGCCGGCCCGGCGGAGGAGCAGCGAGCGGGGCGACCAC

4501 GTCAACTTGTACAGCGGGCCGATGGACATCAGGTCCGGCCGCACCCGCTCGACCGTCCGG
CAGTTGAACATGTCGCCCGGCTACCTGTAGTCCAGGCCGGCGTGGGCGAGCTGGCAGGCC

4561 GCCAGCCACGCCCGGTCGGCCGCCGTGGTCAGGTCCAGGCCCTCCGGCCGGACCTCGATG
CGGTCGGTGCGGGCCAGCCGGCGGCACCAGTCCAGGTCCGGGAGGCCGGCCTGGAGCTAC

4621 TGCATCCGGCCCCGGGCCACCTCCGCGCCCTCCCCGGCCGCCGCCGCCGCCATCCGCCGC
ACGTAGGCCGGGGCCCGGTGGAGGCGCGGGAGGGGCCGGCGGCGGCGGCGGTAGGCGGCG

4681 AGCCACGGCCGCACCTGCGAGGCCGCGTTCTCCGCGTCGACCAGGAGCACGCGCACCGGC
TCGGTGCCGGCGTGGACGCTCCGGCGCAAGAGGCGCAGCTGGTCCTCGTGCGCGTGGCCG

4741 TCGATCCGGGCCCGCCGCCATGGGTGCAGACCGGCCGCCGCCCGCACCAGGATCTGACGC
AGCTAGGCCCGGGCGGCGGTACCCACGTCTGGCCGGCGGCGGGCGTGGTCCTAGACTGCG

4801 AGCAACAGCGACTTGCCGCCGCCCTCCCCGGCCGTGATCATCAGCCGGTCCCAGCGCGCC
TCGTTGTCGCTGAACGGCGGCGGGAGGGGCCGGCACTAGTAGTCGGCCAGGGTCGCGCGG

4861 AGCACCCCCGGGATCACCCACTCCGGCTCGTCGTCCGCCTGGTCCAGGAAATCGATCAGA
TCGTGGGGGCCCTAGTGGGTGAGGCCGAGCAGCAGGCGGACCAGGTCCTTTAGCTAGTCT

4921 TCGGCTGGCGGGGCATCCGCCGCCGCGAGGCCACGATCGCGGACCTCCCGCACCGCCCCC
AGCCGACCGCCCCGTAGGCGGCGGCGCTCCGGTGCTAGCGCCTGGAGGGCGTGGCGGGGG

4981 ACGGCGGTCTCCGCCGCCTCGGCTGCGGTGCCCTCCCGGCGCTGAGCCGCCGTCACCGTC
TGCCGCCAGAGGCGGCGGAGCCGACGCCACGGGAGGGCCGCGACTCGGCGGCAGTGGCAG

5041 CGCACCCCCGCCGCCAGCAGCGCCCGCCACCGCGACTCCTCGGCCACGATCTCCGCGTAC
GCGTGGGGGCGGCGGTCGTCGCGGGCGGTGGCGCTGAGGAGCCGGTGCTAGAGGCGCATG

5101 CGCAGCGCGTGCGCCGTCGTCGGCACCGCGTCGTGCAGCGCGTGCAGCCACACCGCCCCG
GCGTCGCGCACGCGGCAGCAGCCGTGGCGCAGCACGTCGCGCACGTCGGTGTGGCGGGGC

5161 CCGATGCGCTCCAGGTCGCCCCGCTTGCGCAGCACCGCCGACACCGTCACCGGGTCGCAC
GGCTACGCGAGGTCCAGCGGGGCGAACGCGTCGTGGCGGCTGTGGCAGTGGCCCAGCGTG

5221 GGCGTCCCCCGCGCCCACAGCGACGCCAGCAGATCCCACAACTCCCCGTACCGGGGCTCC
CCGCAGGGGGCGCGGGTGTCGCTGCGGTCGTCTAGGGTGTTGAGGGGCATGGCCCCGAGG

5281 CGGAAATCCGTCGGGGCCAGGGCGGAGACCAGATCCGGCACCAGATCCCGAGCCTCCGGC
GCCTTTAGGCAGCCCCGGTCCCGCCTCTGGTCTAGGCCGTGGTCTAGGGCTCGGAGGCCG

5341 CGCAGCAGCGTCCCCAGCACCGCGTCCTCGGCCTCCCGCACCTCAGCCGCCCGCAGCTCC
GCGTCGTCGCAGGGGTCGTGGCGCAGGAGCCGGAGGGCGTGGAGTCGGCGGGCGTCGAGG

5401 TCCGACTCCATGCTCACAGCCCGCCCTCCCCCGCCAGCCGCGCCCGCAGCACCGCCAGCT
AGGCTGAGGTACGAGTGTCGGGCGGGAGGGGGCGGTCGGCGCGGGCGTCGTGGCGGTCGA

←ORF 22

5461 CCTCCGGCGACACCCTCGGCCGCCCACCCGGACCACCCGGGCCACCCGCAGCCACAGGCG
GGAGGCCGCTGTGGGAGCCGGCGGGTGGGCCTGGTGGGCCCGGTGGGCGTCGGTGTCCGC

5521 CCTGAGACGGGCTGGACTCCATGTGCTGCACCAGGCTCATCCACGTTGGGTGGTCGGCCA
GGACTCTGCCCGACCTGAGGTACACGACGTGGTCCGAGTAGGTGCAACCCACCAGCCGGT

5581 TCACCGCCGCAGCAGCCTCCAGCCGCTGAGCAGGCAGCCCCTGGGACAGCAGCCACCCGG
AGTGGCGGCGTCGTCGGAGGTCGGCGACTCGTCCGTCGGGGACCCTGTCGTCGGTGGGCC

5641 CCTCCCGGGCGAACGCCGCCACCCGGCCCGGAGACGTCACCACCCCCGCAGCCGCCCGCC
GGAGGGCCCGCTTGCGGCGGTGGGCCGGGCCTCTGCAGTGGTGGGGGCGTCGGCGGGCGG

5701 GCTCCGCCCACCGCGCAGCGATCCGCTCAGCGGACTGCTCAGCCACCCCCGCCCCCGGCG
CGAGGCGGGTGGCGCGTCGCTAGGCGAGTCGCCTGACGAGTCGGTGGGGGCGGGGGCCGC

5761 CAGCCGCCTCTCTCTCTCCCCCCGCGTCCGGCCCGACGCCCGAGGGACGACCGGGACCAG
GTCGGCGGAGAGAGAGAGGGGGGCGCAGGCCGGGCTGCGGGCTCCCTGCTGGCCCTGGTC

5821 AGAGAGAGATCCCTTCGGGATCACTAAGGGATGGTGAGTCATTTTGACTCTCACTTACAT
TCTCTCTCTAGGGAAGCCCTAGTGATTCCCTACCACTCAGTAAAACTGAGAGTGAATGTA

5881 GAGTCATTTTGACTCCGGGTAGGAGTCGATTTGACTCCGGGTAGGAGTCAGAATGACGTG
CTCAGTAAAACTGAGGCCCATCCTCAGCTAAACTGAGGCCCATCCTCAGTCTTACTGCAC

5941 TCAGCAGGTCAGAGGGGGGTTCCGGGTGGTCCGGGGTCCCGCGCAGCGTCGCCTCGGGGT
AGTCGTCCAGTCTCCCCCCAAGGCCCACCAGGCCCCAGGGCGCGTCGCAGCGGAGCCCCA

6001 CGCCCTGGGTGTCGTCGTCGAACGTGATCTGGTCGATGATGTTGTCGTCGTAGGTCTTCT
GCGGGACCCACAGCAGCAGCTTGCACTAGACCAGCTACTACAACAGCAGCATCCAGAAGA

6061 CCTCGCGCGCCATGGACGCGAGCAGCGGCAGGTTGACCCGTGACAGGTTGGAGATCGGCT
GGAGCGCGCGGTACCTGCGCTCGTCGCCGTCCAACTGGGCACTGTCCAACCTCTAGCCGA

```
6121 CGCCCTTGCTGTTCTTCCGGCGCACGGACCTGAGCAGGTTCTTCTTGACGAGCTTGCGCT
     GCGGGAACGACAAGAAGGCCGCGTGCCTGGACTCGTCCAAGAAGAACTGCTCGAACGCGA

6181 TGGTCCTCTGGACCGTACTGCGGGAGGTCCCACAGTCGTCCACGAGCCGGTCCTCAGACG
     ACCAGGAGACCTGGCATGACGCCCTCCAGGGTGTCAGCAGGTGCTCGGCCAGGAGTCTGC

6241 GCCAGCAGTAGCCGTAGGGGTCGGTCCAGTTCGTGTAACCCAGGAGCAGCAGCTTCTCCG
     CGGTCGTCATCGGCATCCCCAGCCAGGTCAAGCACATTGGGTCCTCGTCGTCGAAGAGGC

6301 AGCCGTCCAGCCCCTCGGCCGCGAACACCATGCCCATGTGCTCCGCACTCACCGGGCCAC
     TCGGCAGGTCGGGGAGCCGGCGCTTGTGGTACGGGTACACGAGGCGTGAGTGGCCCGGTG

6361 CCCCACCCCGGACGCGCTGAAGTGGTAGCAGCCGCAGGCGCCCTTCCCGGAGCACGGGCC
     GGGGTGGGGCCTGCGCGACTTCACCATCGTCGGCGTCCGCGGGAAGGGCCTCGTGCCCGG

6421 CTCGATGCCCAGGTGGTCCTGTGCGCTGTGGCGGCAGTGACGGCACGCGGGCATGCCGAC
     GAGCTACGGGTCCACCAGGACACGCGACACCGCCGTCACTGCCGTGCGCCCGTACGGCTG

6481 GTGCCAATCAGCTTCGGCGGCCCGGAACCTTTCGCATGAGCAGTCGAGCCACTCATCCCA
     CACGGTTAGTCGAAGCCGCCGGGCCTTGGAAAGCGTACTCGTCAGCTCGGTGAGTAGGGT

6541 GCCCGCCACCGTGCAGCCCGCCGTACCCCGGTGGAACGAAAGCTCGTGCCCGCAGCGGCA
     CGGGCGGTGGCACGTCGGGCGGCATGGGGCCACCTTGCTTTCGAGCACGGGCGTCGCCGT

6601 CCGCGCGTCCTTCGAGACGGAGGGTCGCGGGGGGATCACCCGGTGTGAAGGCTGCTGCGC
     GGCGCGCAGGAAGCTCTGCCTCCCAGCGCCCCCCTAGTGGGCCACACTTCCGACGACGCG

6661 CGGGATCTCCGGGAGACCGGCCGTGATCGCCTCCTCCACCTTGAGTACCGGGTTGTCTGA
     GCCCTAGAGGCCCTCTGGCCGGCACTAGCGGAGGAGGTGGAACTCATGGCCCAACAGACT

6721 GCCGAAGTCGAACCACTCCCCGTGTTGCCTGAGCTGGGCGAACCGCAGGTGAAGCGCGGC
     CGGCTTCAGCTTGGTGAGGGGCACAACGGACTCGACCCGCTTGGCGTCCACTTCGCGCCG

6781 CTCCATCGACCGGCCGCCGGGTGCCGACCAAAGGATCGCCAGAGGGACCGGCGACGAGAG
     GAGGTAGCTGGCCGGCGGCCCACGGCTGGTTTCCTAGCGGTCTCCCTGGCCGCTGCTCTC
      ORF 5

6841 CTGAAGCGCCCGCAAGCGCCGCGCGAGCTGAGTGGTCGTCCCAACCTTGACCAGGCCGCC
     GACTTCGCGGGCGTTCGCGGCGCGCTCGACTCACCAGCAGGGTTGGAACTGGTCCGGCGG

6901 CCCCGGCGCCCCGATCAGATAGACGCAGCTCTCACCTGCGGTGTCCTGGCCGTCTGGCCC
     GGGGCCGCGGGGCTAGTCTATCTGCGTCGAGAGTGGACGCCACAGGACCGGCAGACCGGG

6961 GCAACCCGGGTTTTCTCGTACGCTCAAGGGAGCGCACCTCCTTGCTTGCCTGCTGGGGGT
     CGTTGGGCCCAAAAGAGCATGCGAGTTCCCTCGCGTGGAGGAACGAACGGACGACCCCCA

7021 TGAAGCGCTTCTCCACCCGGGCCGTGAACCCGGGTGGACGTCAGGCCCCGCCGATGGCGG
     ACTTCGCGAAGAGGTGGGCCCGGCACTTGGGCCCACCTGCAGTCCGGGGCGGCTACCGCC

7081 GGCCTGCTGCGTGTGCGGGGCCATGCTCCCGCACACGCCGATGCGAGTCCGTGAGCGTTC
     CCGGACGACGCACACGCCCCGGTACGAGGGCGTGTGCGGCTACGCTCAGGCACTCGCAAG
```

```
7141 AGCTCACGTCGTAGTCGTAAATCGACTCTCGACCATTTCCGTTTATGCTGACGCCGTATT
     TCGAGTGCAGCATCAGCATTTAGCTGAGAGCTGGTAAAGGCAAATACGACTGCGGCATAA

7201 CAACGGCGACACTGTTTGCGTCCCGGAGAATGGTGCGTGCCTCCTTGATTGCCGCAGGCA
     GTTGCCGCTGTGACAAACGCAGGGCCTCTTACCACGCACGGAGGAACTAACGGCGTCCGT

7261 ATTCGATCCCAAGGAGCTGAGCGTCCTGCTCGGTAGCCATTCGGGCGGAAATTCGATCCT
     TAAGCTAGGGTTCCTCGACTCGCAGGACGAGCCATCGGTAAGCCCGCCTTTAAGCTAGGA

7321 GCCCGTAAACGGCTGTCATTCCGGTCTGCTCTTCGAGATATGCCCAACTACCCTCCTTGA
     CGGGCATTTGCCGACAGTAAGGCCAGACGAGAAGCTCTATACGGGTTGATGGGAGGAACT

7381 TTCGCTCTGTTTTGAGGAGCTTGGGGGCTGCTGCTGCGAGGTTGCCGTCATACCAGCTTG
     AAGCGAGACAAAACTCCTCGAACCCCCGACGACGACGCTCCAACGGCAGTATGGTCGAAC

7441 TGGACGCGCTGATGGGCTGGTTTTCGGCGTTGTAGGTGATGCGAATGCGCCGAATTACGG
     ACCTGCGCGACTACCCGACCAAAAGCCGCAACATCCACTACGCTTACGCGGCTTAATGCC

7501 CGGCGCCTTCCTGAACGGAAAGAACCTCGGCCACATCGGCGGGAGCCGGTGCCAGTTCGG
     GCCGCGGAAGGACTTGCCTTTCTTGGAGCCGGTGTAGCCGCCCTCGGCCACGGTCAAGCC

7561 CGCTGACGATTCGGGCGTACTCGCCCGCTGTGTAGATCCGGCCCGTCCTGCGTACCGAGG
     GCGACTGCTAAGCCCGCATGAGCGGGCGACACATCTAGGCCGGGCAGGACGCATGGCTCC

7621 CGGCGCGGTCCCGGCCGCTGCGGTGGAGGTGGTCGGCCGCAGCCCTGACCCGTGTTCCTG
     GCCGCGCCAGGGCCGGCGACGCCACCTCCACCAGCCGGCGTCGGGACTGGGCACAAGGAC

7681 ACCCTTGGACTGCCTCTACCAGGCCCTGTGCCTTGAGCGCGGTGATCACTTTCTGGGCGG
     TGGGAACCTGACGGAGATGGTCCGGGACACGGAACTCGCGCCACTAGTGAAAGACCCGCC

7741 TCGCGCGGGAGATGCCCCACTCCTGGGCGATGTTTCGGTCGGACGGATGAGATCCCCCG
     AGCGCGCCCTCTACGGGGTGAGGACCCGCTACAAAGCCAGCCTGCCCTACTCTAGGGGGC

7801 CCCGGAGTTCGCCGGAGTCGATCTGATCCCGGATATGGGCGATGACCTGCATGTATGGCG
     GGGCCTCAAGCGGCCTCAGCTAGACTAGGGCCTATACCCGCTACTGGACGTACATACCGC

7861 GCAATGGCCGCTCGATCTTCGGCATTTCTACCTCCCGGTGTGGCCTGTGCTGACGACAGC
     CGTTACCGGCGAGCTAGAAGCCGTAAAGATGGAGGGCCACACCGGACACGACTGCTGTCG
                                                  <-- Large RNA
7921 AGGCTAGCCCATGGTCTTGACATGGGCTCGCGGCTAGGCCACTTTGTTCCCACGGTCAGT
     TCCGATCGGGTACCAGAACTGTACCCGAGCGCCGATCCGGTGAAACAAGGGTGCCAGTCA

7981 TTACCGGCCGTGACGGGGCAGCTCCGGTTGGAAGCCGGAGCCGCCCCGTGCCGGGCCGGT
     AATGGCCGGCACTGCCCCGTCGAGGCCAACCTTCGGCCTCGGCGGGGCACGGCCCGGCCA

8041 GTTGACGCACCGGCCGGGCGGGTGGCACCGGACCTGTTGACGCAGGTCAGGGGCCATTTG
     CAACTGCGTGGCCGGCCCGCCCACCGTGGCCTGGACAACTGCGTCCAGTCCCCGGTAAAC
              ORF 3 -->
8101 GGGCGGTGGTCGGGTGTTGGAAGCACCTCGACCGCTGTCGGACCCGACGTGTCTAGCGAG
     CCCGCCACCAGCCCACAACCTTCGTGGAGCTGGCGACAGCCTGGGCTGCACAGATCGCTC
```

8161 GAGATCCGACGTGTCGATTGTGGCGGAGCGTCGTGTGGTGGCGCAGGTGGGTGTCCGGCC
CTCTAGGCTGCACAGCTAACACCGCCTCGCAGCACACCACCGCGTCCACCCACAGGCCGG

8221 GGTCGTGGCTGGGCTGGCTGAGCAGGTCGAGGAGTGGGAGCGGGAGGACGACGCGCAGGG
CCAGCACCGACCCGACCGACTCGTCCAGCTCCTCACCCTCGCCCTCCTGCTGCGCGTCCC

8281 GCCGACGGACGGCCCGGAGGGTGGGGTGCCTCGGTGGTCGGAGCGTCAGCGGGAGCTGGG
CGGCTGCCTGCCGGGCCTCCCACCCCACGCAGCCACCAGCCTCGCAGTCGCCCTCGACCC

8341 GGCGCGGCGGGCGGTCGCGGTGCCGGGACAGCGTCGGGTGACGCCGTCGTACACCGCGCC
CCGCGCCGCCCGCCAGCGCCACGGCCCTGTCGCAGCCCACTGCGGCAGCATGTGGCGCGG

←Small RNA

8401 GGATCGGGTGGATCTGCGGACGGCGGAGGACGTGCCGCTGACGCCGGTCCTGGCGGCGCG
CCTAGCCCACCTAGACGCCTGCCGCCTCCTGCACGGCGACTGCGGCCAGGACCGCCGCGC

8461 GGTGGCGGCGCTGGTGGAGGAGCACGGCGATCATCTGGTGCGGTACGTCGCTGCGCGGCT
CCACCGCCGCGACCACCTCCTCGTGCCGCTAGTAGACCACGCCATGCAGCGACGCGCCGA

8521 GCGTGATGCGGAGTGGGCGTACTGGGCGCGGCGGAGGACGTGGCCCAGGACGTGTGGCT
CGCACTACGCCTCACCCGCATGACCCGCGCCCGCCTCCTGCACCGGGTCCTGCACACCGA

8581 GGACGTGGCGCGGGGGCGGGTGCCGGAGCTGCTGGAGGAGGTGCCGGACCTGCCGTGGCC
CCTGCACCGCGCCCCCGCCCACGGCCTCGACGACCTCCTCCACGGCCTGGACGGCACCGG

8641 GCGTCTGGCGGCTGCGGCGAAGTGGTCGCTGCTGGACAACACCAGGACGAGGCGGCGCCG
CGCAGACCGCCGACGCCGCTTCACCAGCGACGACCTGTTGTGGTCCTGCTCCGCCGCGGC

8701 TGAGTGGCTGGTGCGGGTGCCGGAGGACCGGTCGGCCGATGAGGTGCTGGAAGCGCTCGC
ACTCACCGACCACGCCCACGGCCTCCTGGCCAGCCGGCTACTCCACGACCTTCGCGAGCG

8761 GGGCCCGGGGCCGGACTCCACGGTGTGCGCGGTGGATGAACTCATGGAGCCGGAGCCGGA
CCCGGGCCCCGGCCTGAGGTGCCACACGCGCCACCTACTTGAGTACCTCGGCCTCGGCCT

8821 GCCGGGGGGGTGGGCACCGGCCTGCTACGCGGAGCGGATCGCGGCTCTGCCGCCGCGTCA
CGGCCCCCCCACCCGTGGCCGGACGATGCGCCTCGCCTAGCGCCGAGACGGCGGCGCAGT

8881 GCGGGAGGTGCTGGAGCTGCGCTGCACCGAGGGCATGACCACCCCGGCGATCGCGGCCCG
CGCCCTCCACGACCTCGACGCGACGTGGCTCCCGTACTGGTGGGGCCGCTAGCGCCGGGC

8941 GCTCGGCATCTCCCGGCAGAGTGTCGACCGCGCCCTGCGCTTTGCGGTCACCGCTCTGGG
CGAGCCGTAGAGGGCCGTCTCACAGCTGGCGCGGGACGCGAAACGCCAGTGGCGAGACCC

9001 CAGCCCCGGCACGGTCGTCCGGCAGCGCTCGCGCGGCGCCGGGCAGCCCCTCCCCGCCGG
GTCGGGGCCGTGCCAGCAGGCCGTCGCGAGCGCGCCGCGGCCCGTCGGGGAGGGGCGGCC

9061 GTGGGAGCGCGTCCTGGACAGGCTGCCGAACCGGACACAGCGCGACGTGGTCCGGCTGCG
CACCCTCGCGCAGGACCTGTCCGACGGCTTGGCCTGTGTCGCGCTGCACCAGGCCGACGC

9121 CGCGGGCGGCGCGTCGTTCGGGGAGCTGGGCGAGCAGCTCGGTCTCCACCGGGGCTATGC
GCGCCCGCCGCGCAGCAAGCCCCTCGACCCGCTCGTCGAGCCAGAGGTGGCCCCGATACG

9181 GCACGAGCTGTACACGCGGGCGCTGCGCTCGCTGCGGGAGATGGTGCAGGACCAGCGGCT
CGTGCTCGACATGTGCGCCCGCGACGCGAGCGACGCCCTCTACCACGTCCTGGTCGCCGA

9241 GGACCCGGTCCCCGCCGCGCCCGCCAGGAGCAAGCCGAAGCCGCACCCGGCCCCGGCCGG
CCTGGGCCAGGGGCGGCGCGGGCGGTCCTCGTTCGGCTTCGGCGTGGGCCGGGGCCGGCC

9301 GTGCGCCCGGACGTCGTGTGTGTCGGGGTGCTACCTGCGCACCGCCCGGACGGGGGCGGC
CACGCGGGCCTGCAGCACACACAGCCCCACGATGGACGCGTGGCGGGCCTGCCCCCGCCG

9361 GTCGTGAGCGGCCGTCGGCGCGTGGTGGCCGTGGGTCGGTATCTGTGGCAGTGCGTCCGG
CAGCACTCGCCGGCAGCCGCGCACCACCGGCACCCAGCCATAGACACCGTCACGCAGGCC

9421 TGCGGGACGTGGACGACGTCGCCGGGCGGCCGGTGCGCGCTGTGCCGGTGAGGTGATGGC
ACGCCCTGCACCTGCTGCAGCGGCCCGCCGGCCACGCGCGACACGGCCACTCCACTACCG

9481 TGCCTGTCCATCCCCGCCGGTGGCTACTTTGGGTGATTACTCCTTCCTTCCTCCTTCCCG
ACGGACAGGTAGGGGCGGCCACCGATGAAACCCACTAATGAGGAAGGAAGGAGGAAGGGC

9541 GAAAAGCCCTGGATCGGGCCCGCATGGGGCCGGGTCGGCGCCTCTCGGGAGCGAA[illegible]G
CTTTTCGGGACCTAGCCCGGGCGTACCCCGGCCCAGCCGCGGAGAGCCCTCGCTT[illegible]C

9601 TAAGGAGTAATCACCCAAAGTAGCCACCGGCGGGGGCGGGCGGGGAGCCGGGCACCCGCC
ATTCCTCATTAGTGGGTTTCATCGGTGGCCGCCCCCGCCCGCCCCTCGGCCCGTGGGCGG

ORF 45 →

9661 CGGCCGGTTCGGCGCGACGGAGAGGGAACGAGGAGATGGGCAGGGAAGAGACGGGGCCGG
GCCGGCCAAGCCGCGCTGCCTCTCCCTTGCTCCTCTACCCGTCCCTTCTCTGCCCCGGCC

9721 GGGTGCGGCCTCAGTGGGTGTTGCGGACGCTGCTGGGCGGTGCGGTGGCGGTCGGTGCGG
CCCACGCCGGAGTCACCCACAACGCCTGCGACGACCCGCCACGCCACCGCCAGCCACGCC

9781 TGGCGATGGCGGCGTCCGCGGTGACGCTGGCGGAGCTGGGGCACGCGGTCGGCTGGGTGC
ACCGCTACCGCCGCAGGCGCCACTGCGACCGCCTCGACCCCGTGCGCCAGCCGACCCACG

9841 AGTGGGACGGGCGGCTGTCGTGGTCCCTGCCTGTGGCGGTGGATCTGCTGGCGCTGGTGG
TCACCCTGCCCGCCGACAGCACCAGGGACGGACACCGCCACCTAGACGACCGCGACCACC

9901 CGGGCGCGGTGTGGCTGTCCCCGGGCATGACCGCGCGGTCCCGGAGCCTGGCCCGGGGGA
GCCCGCGCCACACCGACAGGGGCCCGTACTGGCGCGCCAGGGCCTCGGACCGGGCCCCCT

9961 TCACGCTGGGCGCGGTCGCCGGGTCGGTCGCCCTCAACGCCATCGGACACCTGGTGGAGT
AGTGCGACCCGCGCCAGCGGCCCAGCCAGCGGGAGTTGCGGTAGCCTGTGGACCACCTCA

10021 CCGGGGACATGAAGGTGACCCCGCTCCTGCGGATCGCGGTGTCCGCCGTGCCGCCGCTGG
GGCCCCTGTACTTCCACTGGGGCGAGGACGCCTAGCGCCACAGGCGGCACGGCGGCGACC

10081 TGGCCGCGCTGGTGGTCCATCTCGTCGGGGTGGTCCTGGCCCCCGTCCCCGGGGCGGACC
ACCGGCGCGACCACCAGGTAGAGCAGCCCCACCAGGACCGGGGGCAGGGGCCCCGCCTGG

10141 ACGGCCAGGACCAGGCCGTCGAGCTGGAGGAGCAGCGGGAGGCCGCCACCATCTCCGCAG
TGCCGGTCCTGGTCCGGCAGCTCGACCTCCTCGTCGCCCTCCGGCGGTGGTAGAGGCGTC

10201 CTCATGACCCTGCTGACGGCACGACGGC[illegible]TCACGACGGCACGACGGCACGGAACG
GAGTACTGGGACGAC[illegible]CGTCTGG[illegible]TGCTGCCGTGCTGCCGTGCCTTGC

10261 CCCCGGCCCCGCAGCCCCGTCCGGCACCCGCCCCGGACGCGCCGGTACGTGTACCTGCCG
GGGGCCGGGGCGTCGGGGCAGGCCG[illegible]GGCGGGGCCTGCGCGGCCATGCACATGGACGGC

10321 CGCCCGCAGCCGACGCCCGCACGGCAG[illegible]GACGACACCCCTGCACCGCCCCCCGTGCCCG
GCGGGCGTCGGCTGCGGGCGTGCCGTCGCCTGCTGTGGGGACGTGGCGGGGGCACGGGC

10381 CAGCCGCGCAGACCGGCCCGCCCCGGGTCACCGCCCGACTGGGCGACCGCCCGGACGATT
GTCGGCGCGTCTGGCCGGGCGGGGCCCAGTGGCGGGCTGACCCGCTGGCGGGCCTGCTAA

10441 CCCAAATTTGGGAACGGCCCGCCCCCGTGCGGGACTGGGACCGACCCGACCTCCCCGCCG
GGGTTTAAACCCTTGCCGGGCGGGGGCACGCCCTGACCCTGGCTGGGCTGGAGGGGCGGC

10501 ACTGCGCACCCGGCACCGAGCCGCTGCTCCTCACCCCGGACCAGTCCCGCGCCCGCATCC
TGACGCGTGGGCCGTGGCTCGGCGACGAGGAGTGGGGCCTGGTCAGGGCGCGGGCGTAGG

10561 TCTGGGCGCTCCAGAACACCGACTGGACCCAAAAGCGAATCGGCGAATTTGCCGGGCGCA
AGACCCGCGAGGTCTTGTGGCTGACCTGGGTTTTCGCTTAGCCGCTTAAACGGCCCGCGT

10621 GCGCCAGCGTGGTCAACACCCGGAAAGCAGAGGCCATTCGGGACGGCATTCTGATCTGAT
CGCGGTCGCACCAGTTGTGGGCCTTTCGTCTCCGGTAAGCCCTGCCGTAAGACTAGACTA

10681 TTCAGAGAATGCCCGAGTGTGAATACAGCGACCCCGAACGAAAATCCGTTCGGGGTCGCT
AAGTCTCTTACGGGCTCACACTTATGTCGCTGGGGCTTGCTTTTAGGCAAGCCCCAGCGA

10741 GTATTTCTGTTGGCGGCGTTCGGGCGACCTCCTGACTGCGCCGATGGGTGTCTGATGGT
CATAAAGACAACCCGCCGCAAGCCCGCTGGAGGACTGACGCGGCTACCCACAGACTACCA

10801 GTGTCAGGTTGGGTGTGGTGGCCGGTGCTTCGCCGATGGGGAAATTGAATGTTCAACTGT
CACAGTCCAACCCACACCACCGGCCACGAAGCGGCTACCCCTTTAACTTACAAGTTGACA

10861 TTTGGGGTGTTCCGGAGGGTGCCGCCGGGGTGCCGCCGGGGTGGTGTTCGGGTGGGTCGA
AAACCCCACAAGGCCTCCCACGGCGGCCCCACGGCGGCCCCACCACAAGCCCACCCAGCT

10921 GCCGGGGTGTTCGGGGGTGTTGGGGTGGGGTCGGGCCGGGGGCGAGCGGGTGGCTGCGGT
CGGCCCCACAAGCCCCCACAACCCCACCCCAGCCCGGCCCCCGCTCGCCCACCGACGCCA

10981 GGGTGTCTGTGCTGGTCAGGGGCGTGTTGGGGTGGGGGTGCCCGTCCTCCTGGGCGTGG
CCCACAGACACGACCAGTCCCCGCACAACCCCACCCCCACGGGCAGGAGGACCCGCACC

11041 GGCGTGTGGTGGCGCCCCGGGGGCGGCGGCCGGGAGGACGGCGCCCCGGCCGGGGGTCGC
CCGCACACCACCGCGGGGCCCCCGCCGCCGGCCCTCCTGCCGCGGGGCCGGCCCCCAGCG

11101 CGGGCGGGACGCGGTGGCCGTCTGTTCGGGGTGTTCGCCCCCGGCCGGGGGCGAACCCGG
GCCCGCCCTGCGCCACCGGCAGACAAGCCCCACAAGCGGGGGCCGGCCCCCGCTTGGGCC

11161 AGAGGGGGTGCGCCAGAATCCCAGACTTTACACAGTCCGGCCAGGATGCAGCTTTGATGC
TCTCCCCCACGCGGTCTTAGGGTCTGAAATGTGTCAGGCCGGTCCTACGTCGAAACTACG

```
11221 AGCCCCAGGAGGCTGCCTGAGCCCGCGTCAGCGTCTCCCGACCCCCGGAGTGGGTCTCTG
      TCGGGGTCCTCCGACGGACTCGGGCGCAGTCGCAGAGGGCTGGGGGCCTCACCCAGAGAC

11281 TGTCCATCCTGGGGTGAACGGCTGTCAGATCGGCGCACAGCGCCTCGTCGGGGGGTGTGG
      ACAGGTAGGACCCCACTTGCCGACAGTCTAGCCGCGTGTCGCGGAGCAGCCCCCCACACC

11341 CCCGGATCTGCCAACCGTACATGCCGCATCTTTTGATGCAGGCTCAGAACCCGCACCCAA
      GGGCCTAGACGGTTGGCATGTACGGCGTAGAAAACTACGTCCGAGTCTTGGGCGTGGGTT

11401 CTCCACCACCCCCGCCCCCACTCCCCGTAGCCAGCCGCCGCGAAGCGGCCAGGGCCCGGA
      GAGGTGGTGGGGGCGGGGGTGAGGGGCATCGGTCGGCGGCGCTTCGCCGGTCCCGGGCCT

11461 GCGCAGCGGAGCGGCCTCGGCCCCGGGCGCAGCCCGGGGCCCATTCCAGAGTTCATCCAG
      CGCGTCGCCTCGCCGGAGCCGGGGCCCGCGTCGGGCCCCGGGTAAGGTCTCAAGTAGGTC

11521 GCCAGCACTCTGAAGCACAGCGGAAGAGTGCCCCTCGCAGCGCAGCGGAGAGGACAGGCC
      CGGTCGTGAGACTTCGTGTCGCCTTCTCACGGGGAGCGTCGCGTCGCCTCTCCTGTCCGG

11581 CGCGCAGCGGGCCCAACGGAGCCCGCTTGCGGGCTCCGTGACACTCCGTCACTGCCCCCG
      GCGCGTCGCCCGGGTTGCCTCGGGCGAACGCCCGAGGCACTGTGAGGCAGTGACGGGGGC

11641 CGCAGCGGGGGCATGCACTCGCTAGGCGCGCAGCGCCTAGGTACCCGCTCCGCGGG
      GCGTCGCCCCCGTACGTGAGCGATCCGCGCGTCGCGGATCCATGGGCGAGGCGCCC
```

Fig. 4. Alignment of the 969 bp left terminal sequence of pSCL (PSCLLEFT) with the 969 bp right terminal sequence of pSCL (PSCLRIGHT). The symbol (:) indicates identical nucleotides in the two sequences. Numbering is initiated from the 5' nucleotide sequence of each strand.

```
PSCLLEFT  - CCCGCGGAGCGGGTACCTAGGCGCTGCGCGCCTAGCGAGTGCATGCCCCC -50
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - CCCGCGGAGCGGGTACCTAGGCGCTGCGCGCCTAGCGAGTGCATGCCCCC -50

PSCLLEFT  - GCTGCGCGGGGGCAGTGACGGAGTGTCACGGAGCCCGCAAGCGGGCTCCG -100
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - GCTGCGCGGGGGCAGTGACGGAGTGTCACGGAGCCCGCAAGCGGGCTCCG -100

PSCLLEFT  - TTGGGCCCGCTGCGCGGGCCTGTCCTCTCCGCTGCGCTGCGAGGGGCACT -150
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - TTGGGCCCGCTGCGCGGGCCTGTCCTCTCCGCTGCGCTGCGAGGGGCACT -150

PSCLLEFT  - CTTCCGCTGTGCTTCAGAGTGCTGGCCTGGATGAACTCTGGAATGGGCCC -200
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - CTTCCGCTGTGCTTCAGAGTGCTGGCCTGGATGAACTCTGGAATGGGCCC -200

PSCLLEFT  - CGGGCTGCGCCCGGGCCGAGGCCGCTCCGCTGCGCTCCGGGCCCTGGCC -250
            :::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - CGGGCTGCGCCCGGGCCGAGGCCGCTCCGCTGCGCTCCGGGCCCTGGCC -250

PSCLLEFT  - GCTTCGCGGCGGCTGGCTACGGGGAGTGGGGGCGGGGTGGTGGAGTTGG -300
            :::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - GCTTCGCGGCGGCTGGCTACGGGGAGTGGGGGCGGGGTGGTGGAGTTGG -300

PSCLLEFT  - GTGCGGGTTCTGAGCCTGCATCAAAAGATGCGGCATGTACGGTTGGCAGA -350
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - GTGCGGGTTCTGAGCCTGCATCAAAAGATGCGGCATGTACGGTTGGCAGA -350

PSCLLEFT  - TGCGGGCCACACCCCCCGACGAGGCGCTGTGCGCCGATCTGACAGCCGTT -400
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - TCCGGCCCACACCCCCCGACGAGGCGCTGTGCGCCGATCTGACAGCCGTT -400

PSCLLEFT  - CACCGCAGGATGGACACAGAGACCCACTCCGGGGGTCGGGACACGCTGAC -450
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - CACCGCAGGATGGACACAGAGACCCACTCCGGGGGTCGGGAGACGCTGAC -450

PSCLLEFT  - GCGGGCTCAGGCAGCCTCCTGGGGCTGCATCAAAGCTGCATCCTGGCCGG -500
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - GCGGGCTCAGGCAGCCTCCTGGGGCTGCATCAAAGCTGCATCCTGGCCGG -500

PSCLLEFT  - ACTGTGTAAAGTCTGGGATTCTGGCGCACCCCCTCTCCGGGTTCGCCCCC -550
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - ACTGTGTAAAGTCTGGGATTCTGGCGCACCCCCTCTCCGGGTTCGCCCCC -550

PSCLLEFT  - GGCCGGGGGCGAACACCCCGAACAGACGGCCACCGCGTCCCGCCCGGCGA -600
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - GGCCGGGGGCGAACACCCCGAACAGACGGCCACCGCGTCCCGCCCGGCGA -600

PSCLLEFT  - CCCCCGGCCGGGGCGCCGTCCTCCCGGCCGCCGCCCCCGGGGCGCCACCA -650
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - CCCCCGGCCGGGGCGCCGTCCTCCCGGCCGCCGCCCCCGGGGCGCCACCA -650

PSCLLEFT  - CACGCCCCACGCCCAGGAGGACGGGCACCCCCCACCCCAAACAGCCCCTG -700
            ::::::::::::::::::::::::::::::::::::::::   :::::::
PSCLRIGHT - CACGCCCCACGCCCAGGAGGACGGGCACCCCCCACCCCAACACGCCCCTG -700

PSCLLEFT  - ACCAGCACAGACACCCACCGCAGCCACCCGCTCGCCCCCGGCCCGACCCC -750
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - ACCAGCACAGACACCCACCGCAGCCACCCGCTCGCCCCCGGCCCGACCCC -750

PSCLLEFT  - ACCCCAACACCCCCGAACACCCCGGCTCGACCCACCCGAACACCACCCCG -800
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - ACCCCAACACCCCCGAACACCCCGGCTCGACCCACCCGAACACCACCCCG -800

PSCLLEFT  - GCGGCACCCCGGCGGCACCCTCCGGAACACCCCAAAACAGTTGAACATTC -850
            ::::::::::::::::::::::::::::::::::::::::::::::::::
PSCLRIGHT - GCGGCACCCCGGCGGCACCCTCCGGAACACCCCAAAACAGTTGAACATTC -850

PSCLLEFT  - AATTTCCCCATCGGCGAAGCACCGGCCACCACACCCAACCTGACGGCCCG -900
            ::::::::::::::::::::::::::::::::::::::::::::   :: 
PSCLRIGHT - AATTTCCCCATCGGCGAAGCACCGGCCACCACACCCAACCTGACACACCA -900

PSCLLEFT  - TCACACACCCATCGGCGAACACAAACGGTCGACCACAACACC-CCAAACG -949
            ::: ::::::::::::: :  ::   ::::: ::  ::: :: :: ::: 
PSCLRIGHT - TCAGACACCCATCGGCGCAGTCAGGAGGTCGCCCG-AACGCCGCCCAACA -949

PSCLLEFT  - GCAAAAGCGATGACCCCGAA  -969
            : :::  :   :::::::::
PSCLRIGHT - G-AAATACAGCGACCCCGAAC -969

Identity  : 943 ( 97.3%)
```

50% longer than those of pSLA2. The significance of the length of these long inverted terminal repeats remains unclear.

3.3.2 Internal Inverted Repeats

From the sequence of pSCL presented in Fig. 3, a number of small internal inverted repeats (indicated by arrows between the double-stranded DNA sequences) have been detected. Most of them are clustered within the long inverted terminal repeats. Among them, only four perfect palindromic sequences were found at the positions 541-564, 601-613, 11084-11096 and 11133-11156. Interestingly, the small inverted repeats clustered at the termini of pSCL are quite similar. Inverted repeat sequences are thought possibly to be protein-binding sites. The functions of the small inverted repeats found within the long inverted terminal repeats might be DNA-specific protein binding sites as suggested by Hirochika *et al.* (1984). Some of the small inverted repeats located outside of the long inverted terminal repeats are associated with some ORFs and the RNA transcripts, which will be discussed later.

3.3.3 Comparison of the Long Inverted Terminal Repeats of pSCL with Those of pSLA2

Comparison of the inverted terminal repeats of pSCL with those of *S. rochei* revealed some interesting similarities. The comparison is shown in Fig. 5, wherein the 969 bp of the left terminal sequence of pSCL (PSCLLEFT) was aligned with the 838 bp of the left terminal sequence of pSLA2 (SRPLL). The results clearly show that the two sequences have a quite high degree of similarity (65.4%). The right terminal sequences of these two plasmids also contain 59.4% similarity (the data are not shown). The similarity between the first 366 bp of the pSCL left terminal sequence and the first 357 bp of the pSLA2 left terminal sequence is even higher (71.7%). Furthermore, the first 12 bp at the termini of both plasmids are completely identical, which indicates that the two plasmids may have a common replication mechanism or a common origin, or may have the same (or

Fig. 5. Comparison of [illegible] bp left terminal sequence of pSCL (PSCLLEFT) with the 838 bp left terminal sequence of pSLA2 (SRPSLL). The symbol (:) indicates identical nucleotides in the two sequences. Dashed lines indicate gaps. Numbering begins with the 5' nucleotide in each strand.

```
PSCLLEFT  - CCCGCGGAGCGGGTACCTAGGCGCTGCGCGC[illegible]AGCGAGTG-CATGCCCC -49
            :::::::::::: : :::   : ::: ::: ::   :: :: : :: :::::
SRPSLL    - CCCGCGGAGCGG-TCCCTGCGGGCTTCGC-CCGCGC-AGAGACACGCCCC -47

PSCLLEFT  - CGCTGCGCGGGGGCA-GTGA-CGGAGTGTCACGGAGCCCGCAA-GCGGGC -96
             ::::::::::::: :::: ::::: :::::   ::::::   ::::::
SRPSLL    - -GCTGCGCGGGGGCGTGTGAACGGAGGGTCACC--GCCCGCCCTGCGGGC -94

PSCLLEFT  - TC-CGTTGGGCCCGCTGCGCGG---GCCTGTCCTCTCCGC-TGCGCTGCG -141
                :::: :::: :::::::::   ::: :  : ::: :: ::::    ::
SRPSLL    - GGGCGTTCGGCC-GCTGCGCGGCCTGCCGGCTCGCTCGGCCTGCGGCCCG -143

PSCLLEFT  - AGGGGCACTCTTCCGCTGTGCTTCAGAGTGCTGGCCTGGATGAACTCTGG -191
            :: ::: :     :: ::: ::: :    ::::      :: :      ::: :
SRPSLL    - AGCGGCGCGGG-CCACTGCGCTCCCC-GTGCGAATCTCGC---GCTCCGC -188

PSCLLEFT  - AATGGGCCCCGGGCTGCGCCCGGGGCCGAGG--CCGCTCCGCTGCGCTCC -239
                :  : :::::  :::  :: :::::::  ::   ::::: :::: ::
SRPSLL    - G--CGTACACGGGC--CGCAGGGCGCCGAGGGGCCCGTCCG[illegible]GCGC-CC -232

PSCLLEFT  - GGGCCCT-GGCCGCTTCGCGGC--GGCTGGCTACGGGGAGTGGGGGCGGG -286
            :: :::  :::: :: :: :::  ::::::::::: :: : :::: ::::
SRPSLL    - GGACCCGAGGCCCCTGCGGGGCCTGGCTGGCTACGAGGGGAGGGGTCGGG -282

PSCLLEFT  - GGTGGTGGAGTTGGGTGCGGGTTCTGAGCCTGCATCAAAAGATGCGGCAT -336
            :  ::::: :      : ::::   : : :  :  : :   ::::: ::::
SRPSLL    - GCGGGTGG-GACAACT-CGGGGGATCATCACGG-TGAGTTGATGCAGCAT -329

PSCLLEFT  - GTACGGTTGGCAGATCCGGGCCACACCCCCCGACGAGGCGCTGTGCGCCG -386
            :: : :::::::::::   ::  :::::::       : :::      :
SRPSLL    - GTGCTGTTGGCAGATCT--GCATCACCCCC-------GTGCTT----CAA -366

PSCLLEFT  - ATCTGACAGCC-GTTCACCCCAGGATGGACACAGAG[illegible]CCACTCCGGGG -434
            :: :::: ::  :  :  : : :::    : :    : : :: :: :::::
SRPSLL    - ATGTGACGGCTAGACCGACTCTGGAGCCGCGCC[illegible]TGCGGGG -416

PSCLLEFT  - GTCGGGAGACGCTG---ACGCGGGCTCAGGCAGC[illegible]GGCTGCATC -481
             : :  :: : :::   :: : ::  :: : ::::      . ::: ::  :
SRPSLL    - -TGGACAGGCCCTGTCAACACCGGAACAAG-AGCCAGGCGCGGC-GCGGC -463

PSCLLEFT  - AAAGCTGCATCCTGGCCGGACTGTGTAAAGTCTGGGATTCTGGCGCACCC -531
            :  ::::  :::   :::::    : :   :: ::      :::: : :
SRPSLL    - ACCGCTG--TCC---CCGGA----GCA--CTCGGGCGCGAGGGCGAAGCA -502

PSCLLEFT  - CCTCTCCGGGTTCGCCCCCG-GCCGGGGGCGAACACCCCGAACAGACGGC -580
                :  :::  :  : : : : :   ::  :::::: :    ::::    :
SRPSLL    - ----TGGGGGAGCAGCGCTGTGACCACGGTAAACACCGC---CAGAGACC -545

PSCLLEFT  - CACCGCGTCCCGCCCGGCGACCCCCGGCCGGGGCGCCGTCCTCCCGGCCG -630
            :::     ::: :: ::              :  ::::   :::: ::::
SRPSLL    - CAC----TCCGGCACGT------------GTTGCGCAC-CCTCACGGC-- -576

PSCLLEFT  - CCGCCCCCGGGGCGCCACCACACGCCCCACGCCCAGGAGGACGGGCACCC -680
            ::::: ::: : ::   :: ::::  ::   :::    :    ::  :::: :
SRPSLL    - CCGCCGCCGAG-CGGACCCGCACGGACC--GCC---GGCCACCCGCACTC -620

PSCLLEFT  - CCCACCCCAAACAGCCCCTGACCAGCACAGACACCCACCGCAGCCACCCG -730
            :   ::::  ::         :: : :::: :    :  : :  ::: :
SRPSLL    - CGAGCCCCGGAC-------GAGCTGCACGGACACTTCTGACGCCTCATC -663

PSCLLEFT  - CTCGCCCCCGGCCCGACCCCACCCCAACACCCCCGAACACCCCGGCTCGA -780
              :::: : :::  ::    :    :   : ::::  ::::      :: : ::
SRPSLL    - AGCGCCACAGGC--GAATACTATCGGTCGCCCCT-AACAA---GGTT-GA -706

PSCLLEFT  - [illegible]CCACCCGAACACCACCCCGGCGGCACCCCGGCGGCACCCTCCGGAACAC -830
            : : :::::: :  : :: : :   : : ::     : ::: ::::::::
SRPSLL    - CAC-CCCGAAAAAAAACCAGCCCTAAACACGAAA-CGCCC-CCGGAACA- -752

PSCLLEFT  - CCCAAAACAGTTGAACATTCAATTTCCCCATCGGCGAAGCACCGGCCACC -880
                 :: ::     ::        :::: :  ::::      :::  ::
SRPSLL    - -----AAAAG-----CA--------TCCCGA--GGCG---TTCCG--CAG- -777

PSCLLEFT  - ACACCCAACCTGACGGCCCGTCACACACCCATCGGCGAACACAAACGGTC -930
            :: ::: :  : :::   :: :      :   :: :    :: : :::::
SRPSLL    - ACTCCCGAA-TTACGCGACGGCTTTGGCT--TCAG----CAGATACGGT- -819

PSCLLEFT  - GACCACAACACCCCAAACGGCAAAAGCGATGACCCCGAA -969
                        :: :  : :  ::  :
SRPSLL    - --------GTTCCGATTCCGGGAATTC             -838
```

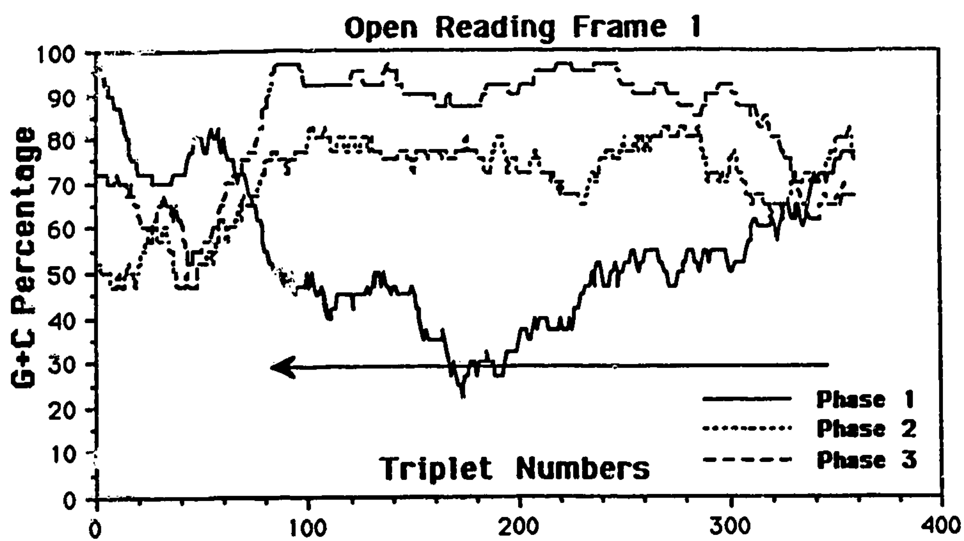
Identity : 548 (65.4%)

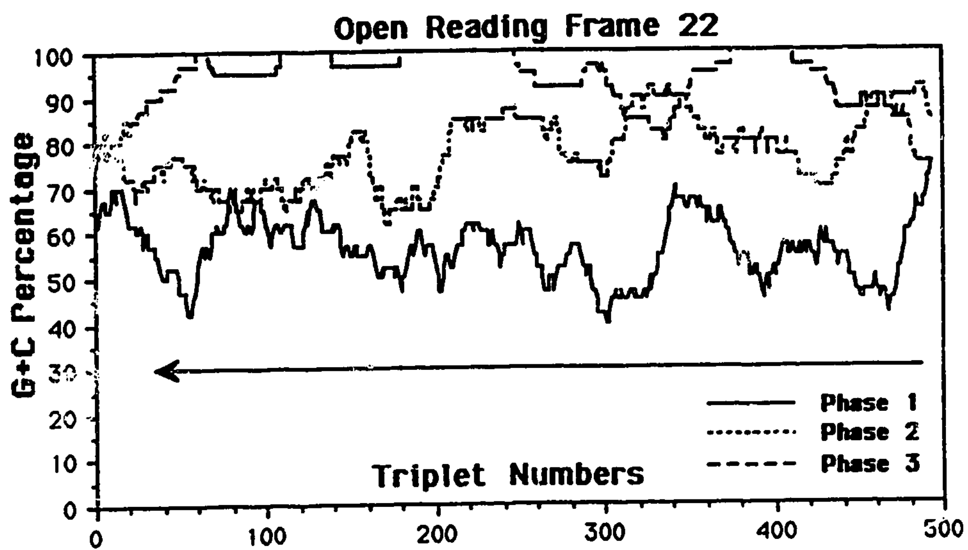
similar) terminal proteins. More interestingly, the small inverted repeats (indicated by arrows) at the termini found in both plasmids are highly conserved; not only do they contain quite similar sequences, but also they are located at almost the same distances from the ends. This suggests that the same (or similar) DNA binding protein(s) may be associated with these sites if these inverted repeats are truly protein binding sites. On the other hand, the terminal sequence of pSLA2 lacks the two perfect palindromic sequences found in pSCL. The function of these palindromes remains in question.

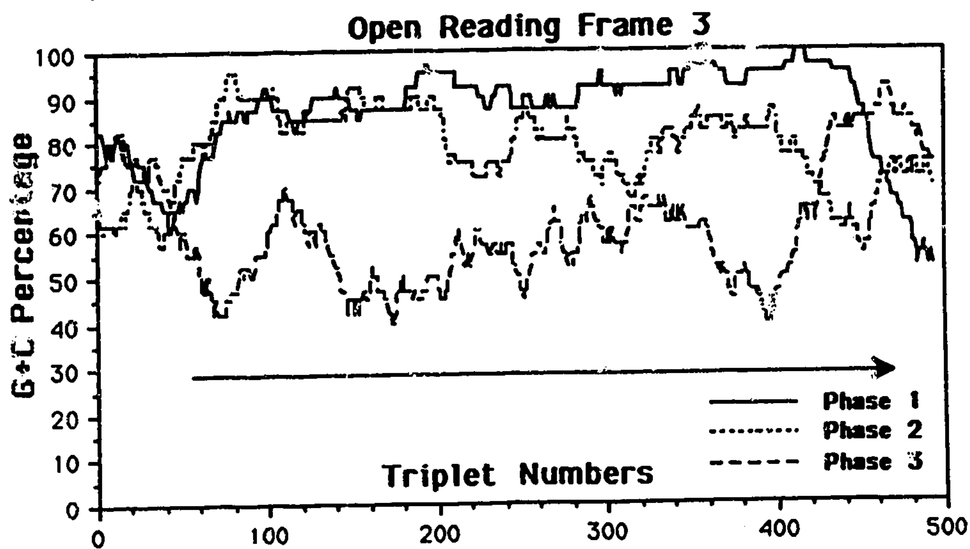
3.4 Open Reading Frame (ORF) Analysis

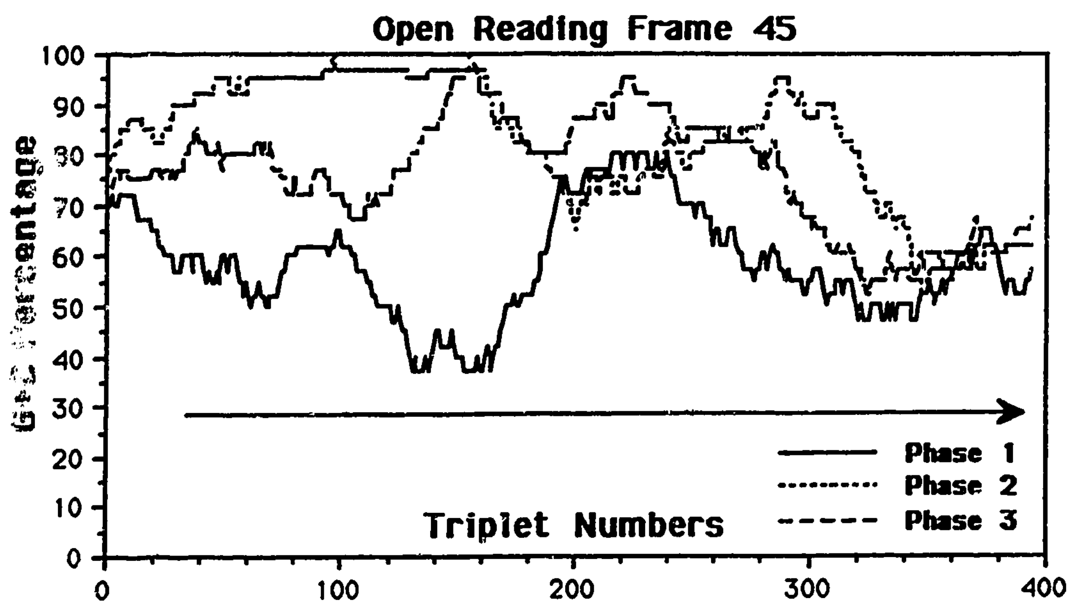
After all sequencing data was collected and entered into a computer, analysis of potential protein-coding ORFs was conducted. This analysis was performed according to Fickett (1982), using ATG and GTG as initiation codons and TAA, TGA and TAG as stop codons. Not surprisingly, many ORFs, in all possible phases, were predicted throughout the entire pSCL. Bibb *et al.* (1984) reported the self-evident fact that because of the high G+C content of *Streptomyces* DNA, the protein-coding ORFs of this organism appear to have a strong, biased codon usage. For a given protein-coding ORF, the G+C percentage at each position within the codons is uneven; about 70% at the first position, about 50% at the second position and more than 90% at the third position. Based on this observation, analysis of G+C percentage at each position within triplets becomes a useful diagnostic test for protein-coding genes in DNA isolated from *Streptomyces*. The FRAME computer program, originally written by Bibb *et al.* (1984) and modified by C. Jensen for use on the Macintosh computer, was used to analyze the entire sequence of pSCL. In this computer program, the G+C compositions at each of the 3 positions in triplets are displayed graphically. Any region which shows the characteristic distribution of G+C percentage at each of the 3 positions was considered to be a potential protein-coding region. As a result of the %G+C analysis shown in Fig. 6, surprisingly, only five regions were found to have the typical distribution of G+C percentage for *Streptomyces* structural genes. Based on

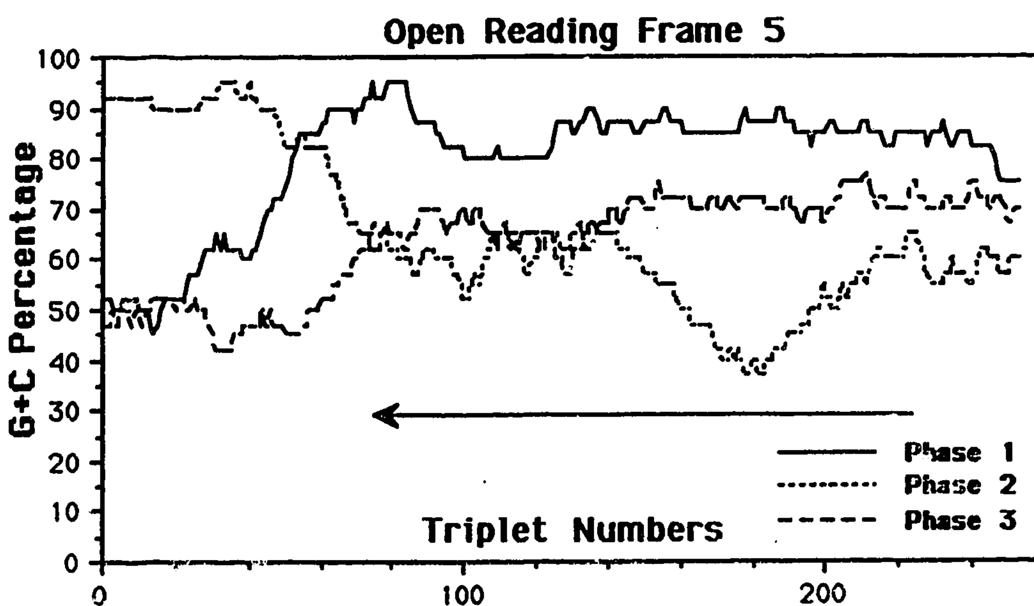
Fig. 6. Graphs illustrating the %G+C in the three codon positions for the five most probable ORFs. The percentage of G+C at each position was calculated using a window of 40 triplets. Arrows indicate the directions and sizes of the ORFs. Phase 1 begins with the first nucleotide reading in the 5' to 3' direction from the extreme left end; phase 2 begins with the second nucleotide; phase 3 begins with the third nucleotide. The sequence ranges chosen for the %G+C analysis are: ORF 1 from 1201 to 2400, ORF 22 from 3901 to 5500, ORF 3 from 7951 to 9550, ORF 45 from 9601 to 10900 and ORF 5 from 6121 to 7000.


Open Reading Frame 1
G+C Percentage
Triplet Numbers
Phase 1
Phase 2
Phase 3
0
10
20
30
40
50
60
70
80
90
100
0
100
200
300
400


Open Reading Frame 22
G+C Percentage
Triplet Numbers
Phase 1
Phase 2
Phase 3
0
10
20
30
40
50
60
70
80
90
100
0
100
200
300
400
500


Open Reading Frame 3
G+C Percentage
Triplet Numbers
Phase 1
Phase 2
Phase 3
0
10
20
30
40
50
60
70
80
90
100
0
100
200
300
400
500


Open Reading Frame 45
G+C Percentage
100
90
80
70
60
50
40
30
20
10
0
Phase 1
Phase 2
Phase 3
Triplet Numbers
0
100
200
300
400


Open Reading Frame 5
G+C Percentage
100
90
80
70
60
50
40
30
20
10
0
Phase 1
Phase 2
Phase 3
Triplet Numbers
0
100
200

the results of the ORF analysis, the ORFs corresponding to these 5 regions were examined. They were compared with a table of codon usage constructed from the sequences of 31 previously published *Streptomyces* genes (see Appendix). Those ORFs which closely matched the codon usage were selected for further examination. Finally, 5 potential protein-coding ORFs were chosen. The results of the %G+C analysis for these 5 ORFs are presented in Fig 6 (the sizes and directions of these ORFs are indicated by arrows). The major characteristics of these ORFs are summarized in Table 4. From these results, two of these ORFs (ORF 3 and ORF 45) were located on one strand of the plasmid, while the remainders (ORF 1, ORF 22 and ORF 5) were on the other strand. They are all preceded by ribosome binding sites. It can also be seen that ORF 45 and ORF 5 did not show an ideal response to the FRAME analysis. ORF 5 is smaller than the region predicted by the %G+C analysis. Besides, this ORF does not appear to be preceded by a good ribosome binding site. But, ORF 5 is the only ORF in this region which has the typical G+C percentage at each position of its codons expected of *Streptomyces* genes. For ORF 45, the first half of the ORF shows the typical *Streptomyces* %G+C distribution in the codons in the expected manner, but the second half of the ORF displays a higher G+C percentage than expected at all 3 positions of its codons. The reasons for choosing this ORF can be explained by the following arguments. First, the overall G+C percentage at each position of its codons is in good agreement with the known *Streptomyces* genes. This is 79.1% at the first position, 58% at the second and 86% at the third. Other predicted ORFs do not demonstrate this distribution of %G+C in their codons. In other words, they have higher G+C percentages in the first or the second position, and lower G+C percentages in the third position, than expected. Therefore, they were not normally considered to be potential protein-coding ORFs. Second, ORF 45 does have a good ribosome binding site preceding it and a good potential transcription terminator

Table 4. Summary of positions of the most probable ORFs in pSCL. Lines under the sequences indicate potential ribosome binding sites predicted by comparison with the 3' terminal sequence of the 16S rRNA of *S. lividans*.

Summary of the Most Probable ORFs in pSCL

RF	Ribosome binding site	Start codon	Stop codon	Predicted size (amino acid)	Predicted molecular weight	Location	%G+C at codon position		
							1st	2nd	3rd
F 1	ACGGCGGTACGGTCTGCCGC	ATG	TGA	263	28,896	1446-2237	75.2%	45.1%	92.6%
F22	CTGGCGGGGGAGGGCGGGCT	GTG	TGA	451	49,247	4062-5417	77.8%	54.4%	95.3%
F 3	CCATTTGGGGCGGTGGTCGG	GTG	TGA	417	45,825	8114-9367	90.3%	56.4%	89.9%
F45	GACGGAGAGGGAACGAGGAG	ATG	TGA	327	33,874	9696-10679	79.1%	58.0%	86.0%
F 5	CGGCACCCGGCGGCCGGTCG	ATG	TGA	145	15,848	6349-6786	68.2%	55.1%	87.4%
F L	GGCCACACCGGGAGGTAGAA	ATG	TGA	248	26,956	7885-7139	65.6%	48.7%	72.1%

(ΔG = -45.2 kcal) at the end. Moreover, the N-terminus (about 150 amino acids) of the polypeptide predicted by ORF 45 contains transmembrane domains (to be discussed later). Taking all these considerations together, ORF 45 was chosen as a potential protein-coding ORF.

3.5 Properties of the 5 Predicted Polypeptides Deduced from the 5 Most Probable ORFs

After selection of the 5 most probable ORFs, their corresponding polypeptide sequences were deduced. Each sequence was analyzed for sequence similarity (and possible homology) to any known proteins, DNA or RNA binding domains, ATP/GTP-binding sites, hydrophobicity and secondary structure of these proteins.

These 5 predicted polypeptides did not show any significant sequence similarity to any known proteins when they were compared with all known proteins in the SWISS-PROT protein sequence data base. Moreover, these polypeptides showed no unusual structural features. Analysis of DNA-binding domains on these polypeptides did not detect any obvious signature of positive, or negative and sigma-type regulatory proteins (Pabo and Sauer, 1984; Wharton and Ptashne, 1986). However, an ATP or GTP-binding site motif A (Walker *et al.*, 1982) was found in the polypeptide predicted by ORF 22 at the amino acid positions between 196 and 203. The significance of this site in the polypeptide remains unknown. From analysis of hydrophobicity of these polypeptides, the N-terminal sequence of the polypeptide predicted by ORF 45 displayed an unusually hydrophobic pattern. Using the methods of Eisenberg *et al.* (1984) and Rao and Argos (1986) to predict membrane associated helices, 4 putative transmembrane helices were detected. Interestingly, these 4 transmembrane helices are all clustered at the N-terminus of the protein within the first 150 amino acids. This suggests that the polypeptide is likely to be a membrane associated protein.

It is also interesting to note that a "leucine zipper"-like structure (Landschulz *et al.*, 1988) has been noticed in the polypeptide predicted by ORF 1 between the amino acid positions 113 and 173. This segment of the polypeptide displays a periodic array of leucine residues of the type like 'leucine-6 amino acids-leucine-7 amino acids-leucine'. This pattern has been found to repeat 4 times except for the presence of a phenylalanine residue instead of a leucine at the position of 158. At present, it is not clear whether this leucine zipper-like structure might play a role in the biological function of the polypeptide, although the leucine zipper structure has been found in some eukaryotic gene regulatory proteins, such as the octamer-binding transcription-factor 2 (OTF-2, Muller *et al.*, 1988).

Jung *et al.* (1987a) reported that three domains are conserved in a number of DNA polymerases (designated as the B family), including DNA polymerases from human, yeast, eukaryotic viruses, bacteriophages and all known linear plasmids with available sequences. As mentioned in the Introduction, some of these viruses or bacteriophages share similar structural characteristics with linear plasmids. Surprisingly, these conserved domains were not found in any of the polypeptides predicted by the 5 most probable ORFs, or even the polypeptides encoded by other ORFs which were not identified as probable by %G+C analysis. The polypeptides predicted by the 5 most probable ORFs were compared with other DNA polymerases than the members of the DNA polymerase B family, such as T7 DNA polymerase (Dunn and Studier, 1984), *E. coli* DNA polymerase I (Joyce *et al.*, 1982) and the DNA polymerase from *Bacillus* bacteriophage SPO2 (Raden *et al.*, 1984). No significant similarity could be found.

3.6 RNA Transcripts from pSCL

After determination of the most probable protein-coding ORFs on pSCL, the question arose whether these ORFs are transcribed *in vivo*. An obvious approach would be to detect RNA transcripts using pSCL specific probes. Northern transfer and hybridization were performed and the results are shown in Fig. 7. Only two RNA transcripts from pSCL were detected, with sizes of about 980 nucleotides and 450 nucleotides (Lane B). As shown in Fig. 7, the DNA probe did not hybridize to RNA isolated from *S. lividans* (Lane A), indicating that the RNA bands detected were pSCL-specific. To avoid missing weak hybridization signals, the filter was exposed to X-ray film for a longer period of time (more than 3 days). No further bands were detected. Using the M13 hybridization method, the two RNA species were shown to be transcribed from one strand of pSCL. Moreover, the two RNA molecules were transcribed from sequences near to each other which were located in a region between 6930 and 8340 on pSCL. This was demonstrated using various DNA fragments generated for sequencing as probes. Unexpectedly, neither RNA transcript corresponded to any of the five most probable ORFs.

To analyze the RNA transcripts further, the 5' termini of the two RNAs were determined by the 5' primer extension approach. Basically, in the 5' primer extension experiments, oligonucleotides complementary to sequences downstream from the 5' ends of the two RNA transcripts were hybridized to total RNA preparations from *S. clavuligerus*. Then, cDNAs were synthesized and extended to the 5' termini of the two RNA transcripts by using AMV reverse transcriptase. The final cDNA products were denatured and resolved on sequencing gels. The results are shown in Fig. 8 and Fig. 9. By comparing the mobilities of these cDNA with mobilities of sequencing fragments obtained using the same oligonucleotides as primers, it is clear that transcription of the large RNA started at a unique site with a G as the first nucleotide (Fig. 8, Lane 2). On the other hand, transcription of the small RNA started from two sites, and both began with G as the

Fig. 7. Northern transfer and hybridization analysis of total RNA isolated from *S. lividans* and *S. clavuligerus*, using pSCL DNA as a probe. Lane A: total RNA isolated from *S. lividans*. Lane B: total RNA isolated from *S. clavuligerus*. Hybridization bands are indicated by arrows. L indicates the large RNA transcript. S indicates the small RNA transcript. Ribosomal RNA species are indicated by 16S and 23S. Markers are fragments of λ DNA (size in kilobase) generated by digestion with *Cla* I. The DNA markers were denatured by glyoxal and electrophoresed along with the RNAs. The marker lane was cut and treated with 50 mM sodium hydroxide, followed by neutralization with 50 mM sodium phosphate (pH 6.5). Then, it was stained with ethidium bromide (0.5 μg/ml) for 60 minutes and photographed under ultraviolet (302 nm) illumination.

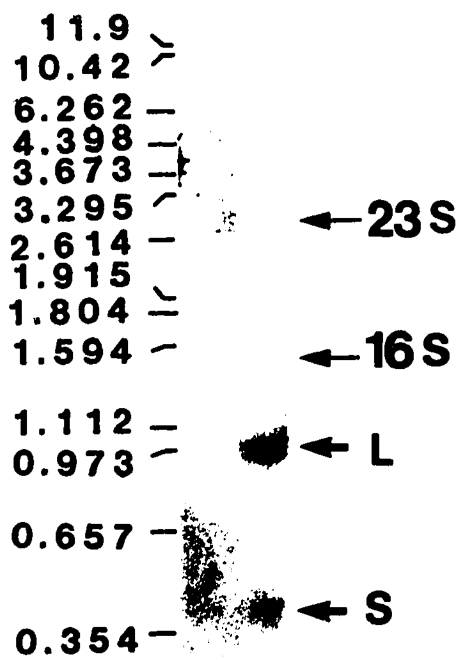

A B
11.9
10.42
6.262
4.398
3.673
3.295
2.614
1.915
1.804
1.594
1.112
0.973
0.657
0.354
23S
16S
L
S

Fig. 8. Autoradiograph of a 6% polyacrylamide denaturing gel used for mapping the 5' terminus of the large RNA transcript of pSCL. The primer extension reaction was carried out using the 5' labeled oligonucleotide (GATGTTTCGGTCGGACG) complementary to a sequence downstream from the 5' end of the pSCL large RNA transcript. The oligonucleotide was hybridized to total RNA preparations from *S. lividans* and *S. clavuligerus* (Lane 1 and Lane 2) respectively and cDNA was made by using AMV reverse transcriptase (see Materials and Methods). The cDNA products were then denatured and electrophoresed on a 6% polyacrylamide denaturing gel, along with a DNA sequence ladder. The DNA sequence ladder was prepared using Sanger's chain termination method. In the sequencing reactions, the same oligonucleotide was used as a primer and the DNA template was the single-stranded, cloned pSCL 4.3 kb DNA in M13, which is complementary to the RNA transcripts. The arrow indicates the cDNA produced by the 5' primer extension reaction. The asterisk indicates the start position of transcription on the complementary DNA strand.

1 2 G A T C

GGACGACAGC*AGTCGTGTCC

Fig.9. Autoradiograph of a 6% polyacrylamide denaturing gel used for mapping the 5' terminus of the small RNA transcript of pSCL. The primer extension reaction was carried out using the 5' labeled oligonucleotide (AGCAGGTCGAGGAGTG) complementary to a sequence downstream from the 5' end of the pSCL small RNA transcript. The oligonucleotide was hybridized to total RNA preparations from *S. lividans* and *S. clavuligerus* (Lane 1 and Lane 2) respectively and cDNAs were made by using AMV reverse transcriptase (see Materials and Methods). The cDNA products were then denatured and electrophoresed on a 6% polyacrylamide denaturing gel, along with a DNA sequence ladder. The DNA sequence ladder was prepared using Sanger's chain termination method. In the sequencing reactions, the same oligonucleotide was used as a primer and the DNA template was the single-stranded, cloned pSCL 4.3 kb DNA in M13, which is complementary to the RNA transcripts. Arrows indicate the cDNAs produced by the 5' primer extension reaction. Asterisks indicate the start positions of transcription on the complementary DNA strand.

GATC 1 2
G
T
G
G
G
C
T
A
G
G
C
C
G
C
G
C
C
A
C
A
T
G

first nucleotide (Fig. 9, Lane 2). Because the two cDNA bands have the same intensity, it is unlikely that they were artifacts caused by degradation of the RNA. The more reasonable explanation for this would be heterogeneity in initiation at the 5' terminus of the small RNA transcript. RNA polymerase may start the transcription at a site involving two G residues and might not distinguish between these G residues. Additionally, a weak cDNA band could be seen just 9 nucleotides upstream of the two major transcription start sites of the small RNA transcript, indicating that some of the small RNA transcript may be transcribed from a different start point.

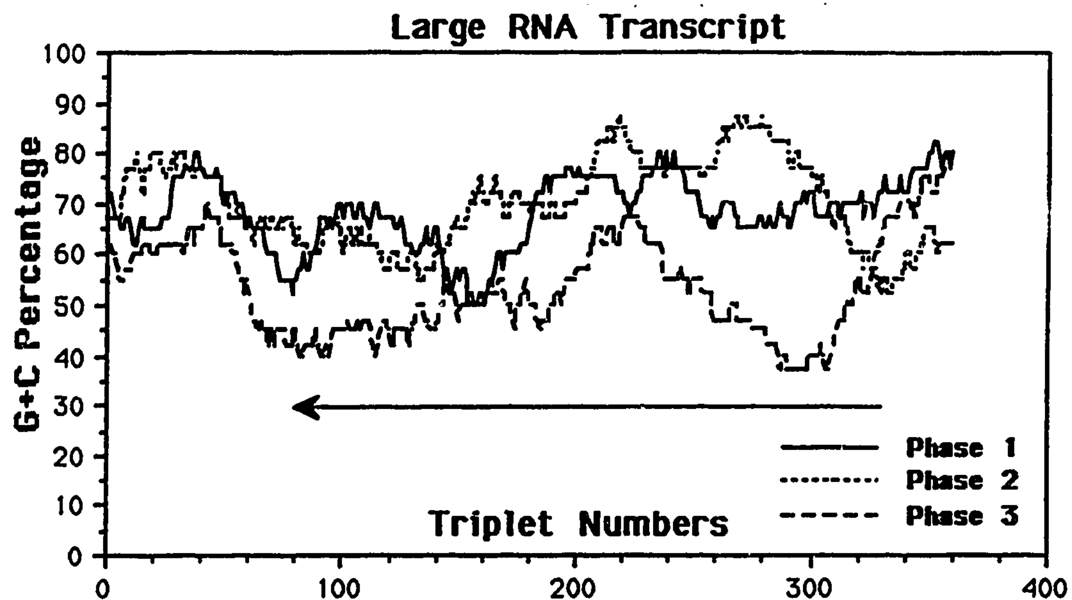
3.7 Characterization of Two RNA Transcripts

Once the 5' termini of the RNA transcripts were determined, the next question was whether the two RNAs encode any proteins. ORF analysis has revealed that the large RNA transcript indeed encompasses an ORF (designated as ORF L). As summarized in Table 4, ORF L is preceded by a good ribosome binding site and encodes a polypeptide chain of 248 amino acids. In contrast, the small RNA does not encompass any potential ORF. The sequences of the two RNAs were subjected to %G+C analysis (shown in Fig. 10). Neither of them showed the typical %G+C distribution of *Streptomyces* genes. The codon usage of ORF L was a reasonable match of the typical codon usage of *Streptomyces* genes, despite the fact that the overall G+C percentage at each position in the codons is lower than expected. In particular, if the large RNA is a mRNA for a protein, its %G+C distribution would be exceptional.

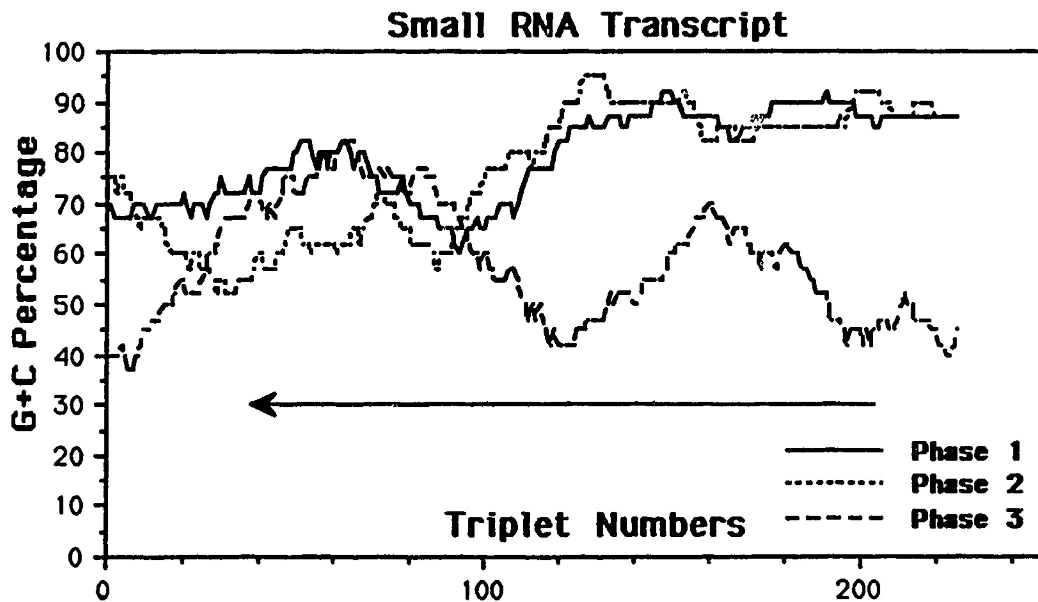
The polypeptide predicted by ORF L was subjected to further analysis as were the other hypothetical polypeptides described previously. It neither showed homology to any known proteins nor displayed the specific structural features in the known regulatory proteins, e.g. helix-turn-helix motif.

The determination of the 5' termini of the two RNA transcripts also allowed us to look for promoter regions for the two RNAs. The sequences of the promoter regions for

Fig. 10. Graphs illustrating the %G+C in the three codon positions for the two RNA molecules shown to be transcripts of pSCL. The percentage of G+C at each position was calculated using a window of 40 triplets. Arrows indicate the directions and sizes of the RNA transcripts. Phase 1 begins with the first nucleotide reading in the 5' to 3' direction from the extreme left end; phase 2 begins with the second nucleotide; phase 3 begins with the third nucleotide. The sequence ranges chosen for the %G+C analysis are: for the large RNA, from 6901 to 8100, and for the small RNA, from 7801 to 8600.



Large RNA Transcript
G+C Percentage
100
90
80
70
60
50
40
30
20
10
0
Triplet Numbers
0
100
200
300
400
Phase 1
Phase 2
Phase 3



Small RNA Transcript
G+C Percentage
100
90
80
70
60
50
40
30
20
10
0
Triplet Numbers
0
100
200
Phase 1
Phase 2
Phase 3

Table 5. Summary of the anticipated promoter regions of the two RNA transcripts from pSCL. The P1 promoter of the thiostrepton resistance gene from *S. azureus* (Janssen *et al.*, 1985) is indicated by tsr/P1.

Summary of the Anticipated Promoter Regions of the Two RNA Transcripts from pSCL

Gene/Promoter	-35 region	Space (bp)	-10 region	Space (bp)	+1
E. coli consensus	TTGACa	17	TAtAat		
Streptomyces consensus	TTGaca	18	tAGgaT		
tsr/P1	GGCAGCCAT	12	TAGGGT	7	U̇
Large RNA/P	GCGAGCCCAT	15	TAGCCT	6	G
Small RNA/P	GTCAGCGGCA	16	CAGATC	10	G/G

the two RNAs are summarized in Table 5. Interestingly, the promoter region for the large RNA has some similarity to the P1 promoter of the thiostrepton resistance gene (tsrP1, Janssen *et al.*, 1985). Both promoters also show some homology to the *E. coli* consensus sequence in the -10 region, but little in the -35 region.

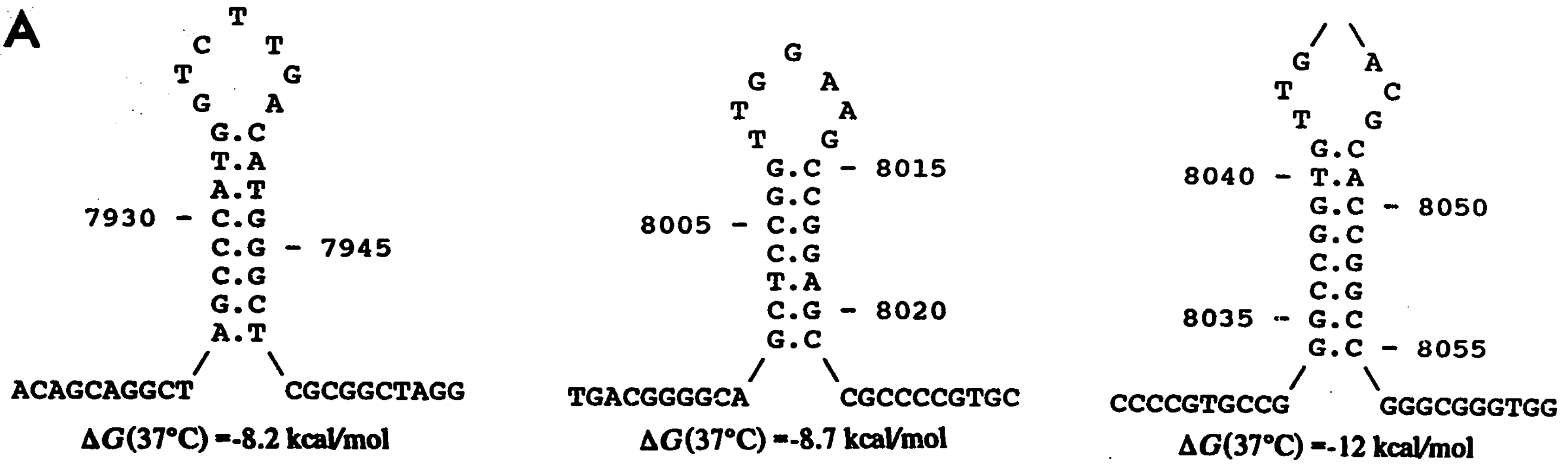
Several inverted repeats were found close to the 3' termini of the two transcripts. These inverted repeats have the potential to form 3 stem-loop structures at the 3' end of each RNA transcript. They are shown in Fig. 11. These potential stem-loop structures are likely to serve as terminators of transcription of the two RNAs. Since the exact sequences at the 3' termini of the two RNA transcripts have not been mapped, which of the stem-loop structures are directly involved in the termination process remains to be determined.

A second question to be answered is what functions the two RNA transcripts have. From the ORF analysis it seems probable that the large RNA transcript is a mRNA for a particular protein. The small RNA transcript, however, does not contain any convincing ORF, but its 5' terminus is complementary to the 5' terminus of ORF 3 (as shown in Fig. 3). The overlapped sequence is 237 bp in length. This suggests that the small RNA transcript may function as an antisense RNA to control the expression of ORF 3. It should be emphasized that this speculation is only based on the sequence. Whether the small RNA plays a role in control of the expression of ORF 3 *in vivo* requires further investigation.

To study the transcription of the two RNAs at different stages of cell growth, total RNA from different stages (from 12 hours to 48 hours after spore inoculation) was prepared. RNAs containing the two transcripts were subjected to analysis by electrophoresis, Northern transfer and hybridization. The results are shown in Fig. 12. The hybridization clearly indicates that the level of transcription of the large RNA varies with the stage of growth. In the early stage, the large RNA was transcribed at a high level (Lane B), then decreased and remained at a relatively constant level through later stages (Lanes C, D and E). The transcription level at 12 hours was at least 5 times higher than that at 48 hours. In contrast to the large RNA, the transcription of the small RNA remained

Fig. 11. Potential stem-loop structures close to the 3' termini of the two RNA transcripts from pSCL. A: 3 potential stem-loop structures close to the 3' terminus of the large RNA transcript. B: 3 potential stem-loop structures close the 3' terminus of the small RNA transcript. The free energy values (ΔG : the change in free energy) were calculated according to Freier *et al.* (1986).

A



B

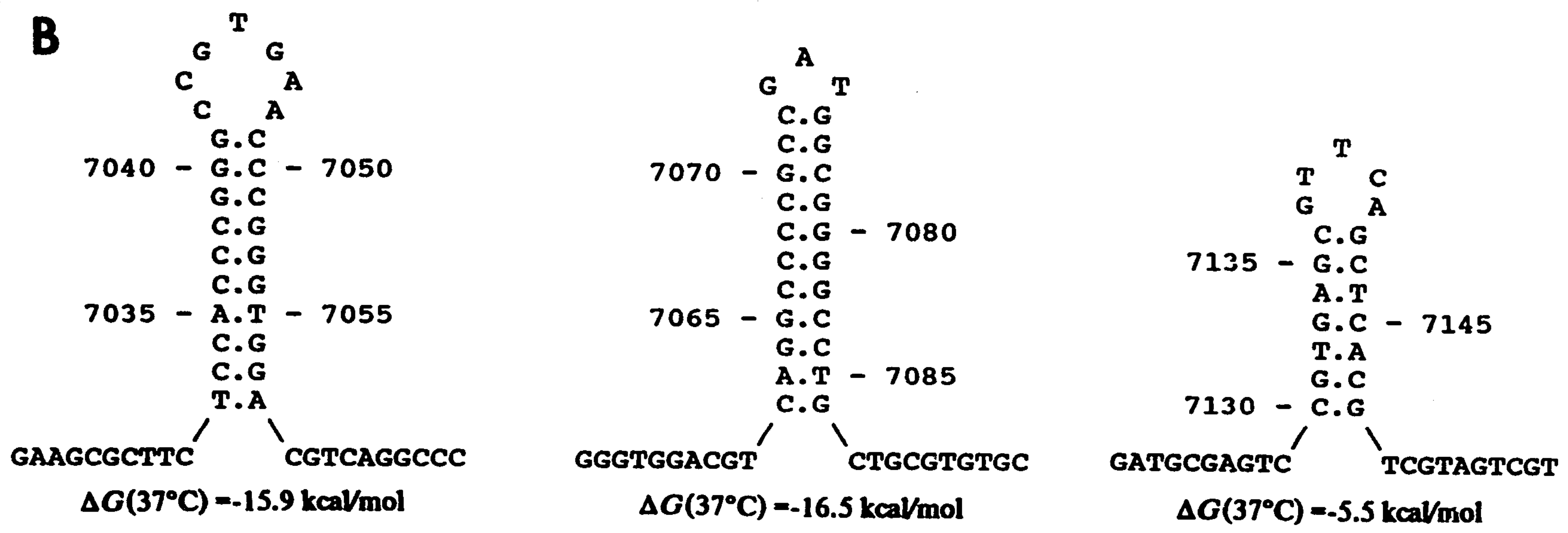
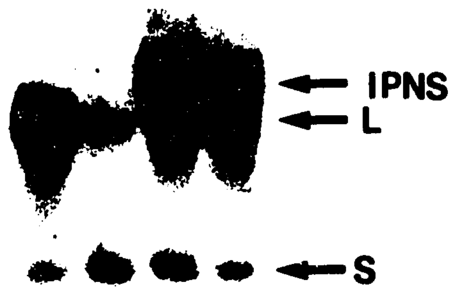


Fig. 12. Northern transfer and hybridization analysis of total RNA preparations from different stages of cell growth. Lane A: total RNA isolated from *S. lividans*. Lanes B to E: total RNA isolated from *S. clavuligerus* at 12 hours, 24 hours, 36 hours and 48 hours of culture after spore inoculation. Hybridization bands are indicated by arrows. L indicates the large RNA transcript. S indicates the small RNA transcript. IPNS indicates hybridization of the putative mRNA of the isopenicillin N synthase gene from *S. clavuligerus* (Leskiw *et al.*, 1988).

A B C D E

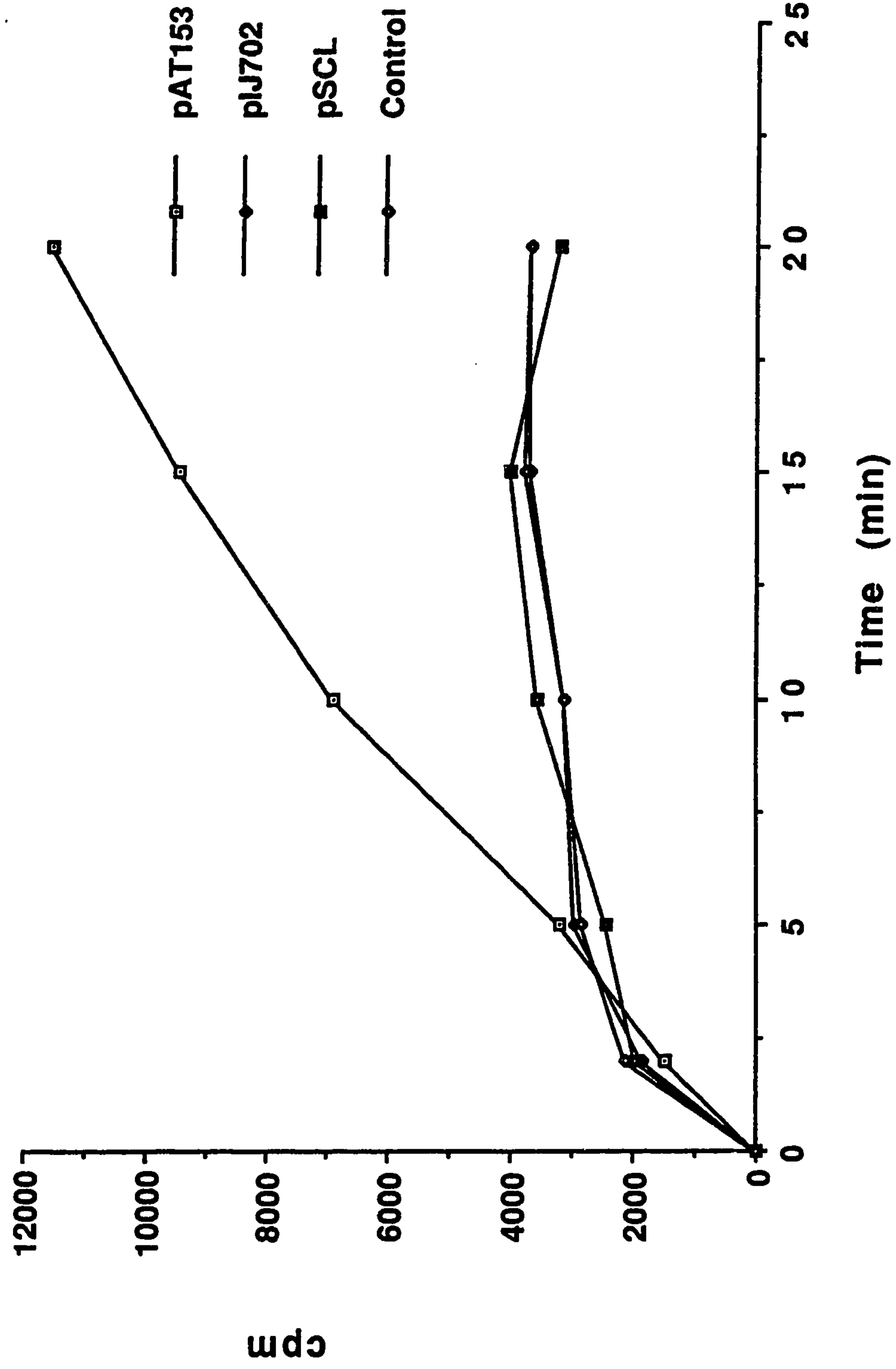


at a relatively constant level through all stages of cell growth. Besides, the level of the large RNA on average was approximately two fold higher than that of the small RNA. In this experiment, the isopenicillin N synthase gene (IPNS, Leskiw *et al.*, 1988) was used as a control. As shown in Fig. 12, the mRNA of the IPNS gene was detected only after 36 hours of cell growth (Lanes D and E). It has been demonstrated that the enzymatic activity of isopenicillin N synthase appears only after 36 hours of cell growth (Jensen, S.E., personal communication).

3.8 *In Vitro* Coupled Transcription and Translation Using *S. lividans* Cell Extracts

Thompson *et al.* (1984) first reported *in vitro* coupled transcription and translation using *S. lividans* cell extracts. We have attempted to use this system to detect the proteins encoded by pSCL. *S. lividans* cell extracts have been prepared and assays of transcription and translation directed by DNAs from various sources have been performed in the cell extracts as suggested by Thompson *et al.* (1984). The results are presented in Fig. 13. As expected, the *E. coli* plasmid pAT153 was actively transcribed and the transcripts were translated into protein products as described by Thompson *et al.* (1984), since *E. coli* promoters can be recognized by *Streptomyces* RNA polymerase (Hopwood *et al.*, 1986a). This indicated that the cell extract was active in terms of transcription and translation. On the other hand, when the *Streptomyces* DNAs, including pSCL, were added to the system, very little ^{35}S-methionine was incorporated. Numerous variations of the assay conditions, including changes in the concentration of Mg^{++}, the amount of S30 cell extracts, amount of input DNA, and addition of *Streptomyces* tRNA, were all performed, as suggested by Chen and Zubay (1982). But, little improvement was seen. This system has been reported to respond well to linear DNA templates. Further efforts should be made to find the optimal conditions in the *S. lividans* cell extracts we prepared for expressing genes on *Streptomyces* DNA.

Fig. 13. Incorporation of ^{35}S-methionine into hot TCA-precipitable material, using various DNA templates with *S. lividans* cell extracts. All assays were conducted under the same conditions except using different DNA templates. The DNA templates used in the assays were as follows. pAT153: the *E. coli* plasmid pAT153; pIJ702: the *Streptomyces* plasmid pIJ702; pSCL: the linear plasmid pSCL. Control: no DNA was added. The plasmids pAT153 and pIJ702 were used as covalently closed circular forms. In this experiment, ^{35}S-methionine was diluted with unlabeled methionine to a final specific activity of 120 Ci/mmol (4.5 TBq/mmol).



cpm
12000
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
Time (min)
pAT153
pIJ702
pSCL
Control

4. DISCUSSION

The complete nucleotide sequence of a linear plasmid (pSCL) from *S. clavuligerus* has been determined. This is the first prokaryotic linear plasmid that has been completely sequenced. The linear plasmid is 11,696 base pairs in length and has a G+C content of 71.9%, which is typical of *Streptomyces* DNA.

To accomplish this sequencing project, the DNA fragments of pSCL, including the two terminal fragments, generated by digestion with *Sal* I, were first cloned into *E. coli* cloning vectors. The internal DNA fragments were directly cloned into M13 single-stranded DNA phage vectors (mp18 and mp19) or plasmid vectors (pUC118 and pUC119). Two approaches were used to clone the terminal fragments of pSCL. The first approach using a tailing method took the advantage of the *E. coli in vivo* repair system to remove the amino acids covalently linked to one end of each terminal fragment. Although the two terminal fragments of pSCL were successfully cloned into the *E. coli* vector pUC19 using this approach, several problems arose from the dG:dC tails at one end of each terminal fragment. First, the tail of dG and dC at one end of the terminal fragments led to difficulty in identifying the last few nucleotides at the ends because the native termini of pSCL consist of only dG and dC. Second, the formation of a long GC homopolymer resulting from the tailing at one end of the terminal fragments made it almost impossible to sequence this region since the strong secondary structures due to the homopolymer prevent DNA polymerase from synthesizing a copy through it. It was estimated that the tailing method generated a dG:dC homopolymer of about 28 bp at the left terminus and of about 34 bp at the right. Also, when recombinant M13 phages containing such tailed DNA fragments were transformed into *E. coli* hosts with either a $RecA^+$ or $RecA^-$ phenotype, some evidence of recombination between the long GC homopolymer and internal GC rich regions within the terminal fragments was observed (data not shown). Thus, the presence of this long GC homopolymer is likely to be one of the reasons for the instability of the

tailed right end fragment in M13 vectors. Because of the problems mentioned above, it became necessary to reclone the terminal fragments of pSCL.

The second approach taken was based on the fact that the amino acids linked to the 5' ends of the linear DNA plasmid pSLA2 can be removed *in vitro* by alkali treatment (Hirochika *et al.*, 1984). Thus, by treatment with NaOH to remove amino acids at the 5' ends of the terminal fragments, and renaturation of the denatured DNA, the two terminal fragments of pSCL were recloned into *E. coli* vectors. Subsequent sequence analysis including the cloned terminal fragments (Fig. 1) and the terminal sequences of the native pSCL DNA (Fig. 2) revealed that the complete terminal sequences had been obtained. Although nothing is known about the terminal proteins of linear plasmids in *Streptomyces,* three amino acids (serine, threonine and tyrosine) in terminal proteins have previously been shown to be covalently linked to 5' phosphate groups of DNA, based on structural analysis of several viruses and phages. In the cases of adenoviruses and phage ϕ29, a serine residue in the terminal protein is responsible for this linkage, while the Cp-1 phage of *Streptococcus pneumoniae* has a threonine residue in its terminal protein linked to the DNA genome. It has been shown that chemical linkages between the 5' phosphate group of DNA and a OH group of a serine or threonine residue in proteins are alkali-labile (Rekosh *et al.*, 1977; Salas *et al.*, 1978). On the other hand, if the linkage is between the 5' phosphate group of DNA and a OH group of a tyrosine residue, it cannot be cleaved by alkaline hydrolysis. Such a linkage was proven to exist in the case of the PRD1 phage (Bamford and Mindich, 1984). Therefore, the alkali lability of the chemical bond between the terminal proteins and the pSLA2 and pSCL linear plasmid DNA suggests that the amino acid residue involved in the linkage is either a serine or threonine.

An important decision for the cloning of pSCL DNA was to use the *E. coli* vectors pUC118 and pUC119, which are single-stranded DNA producing plasmids (phagemids). As pointed out previously, spontaneous deletions occurred when the terminal fragments of pSCL were cloned in M13 vectors. It is believed that these deletions were caused by

homologous recombination between GC rich regions. On the other hand, these DNA fragments were quite stable in pUC118 or pUC119. Even when single-stranded DNA was produced from these recombinant plasmids in the presence of an M13 helper phage, no deletions or other signs of recombination were observed. This suggests that the deletion events observed in the M13 system may have been somehow related to the phage replication processes. Plasmid replication is independent of phage replication and it is only in the final step that the double-stranded plasmid is converted into single-stranded plasmid DNA by viral gene products carried on the M13 helper phage. Generally speaking, the *E. coli* plasmid vectors pUC118 and pUC119 have several advantages over M13 vectors, particularly for these GC rich DNA molecules. Besides their stability, they can carry relatively large inserts compared to the M13 vectors. In addition, because of their small size, it is easier to screen recombinant clones with various deletions when unidirectional deletion procedures are used.

After digestion of pSCL with the *Sal* I endonuclease, three small DNA fragments were released, in addition to the two large fragments and two terminal fragments. These small *Sal* I fragments were cloned directly into M13 vectors and subjected to sequencing. However, because these fragments are small and there are few suitable restriction endonuclease cleavage sites within them, the determination of their correct orientations was difficult. To overcome this problem, the PCR technique was employed to obtain a DNA fragment encompassing the entire region of these small *Sal* I fragments. With such a DNA fragment, not only the correct orientations of the three small fragments can be obtained, but also the number of the *Sal* I sites between these fragments can be confirmed. As a result, errors due to multiple, very closely spaced *Sal* I sites (leading to additional tiny fragments) would be avoided. Similarly, all other *Sal* I cleavage sites used for cloning have also been confirmed by using PCR to generate DNA fragments extending through these sites. It should be noted that because of the high G+C content in *Streptomyces* DNA, exposure of the DNA to a high temperature for a short period of time during PCR might not have been

sufficient to denature the double-stranded DNA. Consequently, the subsequent amplification of the DNA fragment might not occur. This was exactly what happened in the first few unsuccessful trials, in which the specific DNA fragment desired was not amplified. For this reason, the plasmid DNA template was first denatured by treatment with NaOH. Then, in the subsequent PCR reactions, DMSO was added to a final concentration of 10% to help disrupt bonding between the two DNA strands.

As mentioned in previous chapters, *Streptomyces* DNA has an unusually high G+C content. Such G+C rich DNA is very difficult to sequence. Band compressions, i.e. abnormal migration of certain DNA fragments, were often observed while sequencing. These compressions are believed to be caused by G+C rich sequences which can form very stable "hairpin" structures. These intramolecular secondary structures are thought to be at least partly a consequence of Hoogsteen base-pairing involving both N-7 and the exocyclic NH_2 group at the second position of the guanine ring. Therefore, 7-deaza dGTP was used to replace normal dGTP in sequencing mixes (Mizusawa *et al.*, 1986). Since the nitrogen atom at the seventh position is replaced by a carbon atom, Hoogsteen base-pairing is impossible and the formation of secondary structures is reduced. Although most compressions in the sequences of pSCL were resolved efficiently by using 7-deaza dGTP in the sequencing mixes, some compressions in regions with continuous dG and dC sequences still persisted. This indicates that these particular compressions may not have been caused by secondary structures involving Hoogsteen base-pairing. The reason for the formation of these compressions is more likely to be Watson-Crick base pairing. During the early stages of the sequencing project, sequencing gels containing 40% formamide were used in attempts to resolve these compressions. However, the results obtained from such sequencing gels seemed inconsistent and sometimes unreadable. Accurate sequence information for these regions with stable compressions had to be obtained by sequencing of the other strand of pSCL DNA since compressions rarely happen on both strands in the

same region. Fortunately, the sequence data collected from the other strand of pSCL DNA in the regions of stable compressions were unambiguous and reliable.

Many strategies have been developed to sequence large DNA fragments cloned in *E. coli* vectors. In this study, the unidirectional deletion method developed by Henikoff (1984, 1987) has been used successfully. The major advantage of this method, which uses exonuclease III, is that the deletions are independent of restriction endonuclease cleavage sites in the DNA fragment to be sequenced. This method is particularly useful for the sequencing of GC rich *Streptomyces* DNA since many of the commonly used restriction enzymes either do not cut *Streptomyces* DNA at all, or do so at a low frequency. Consequently, the division of large *Streptomyces* DNA fragments into smaller segments for sequencing becomes difficult. The procedures involved in the Henikoff deletion method are simple. The reactions can be carried out directly in one tube using DNA prepared on a small scale. All reagents used are inexpensive and easily obtained. The entire procedure, for the cloned 4.3 kb pSCL DNA fragment, from the preparation of the DNA template to recovery of large numbers of overlapping clones, can be finished in one week.

Although this method allowed collection of many overlapping clones in a short period of time, a problem associated with this approach is that about 10% of the recovered clones could not be sequenced due to the loss of the sequencing primer binding site. Exonuclease III is known to digest double-stranded DNA at a 3' recessed or blunt end, but it does not attack a 3' protruding end. However, it will initiate digestion from nicks. Therefore, it is assumed that the loss of the primer binding site was the consequence of exonuclease III attacking nicks around a primer site, particularly when DNA from small scale, rapid preparations was used. Alternatively, the exonuclease used may have been contaminated with enzymes which were able to digest DNA from both 3' and 5' protruding ends. One way to diagnose such a problem before sequencing is to digest the selected DNA clones with restriction enzymes close to the sequencing primer site, such as *Hin*dIII

in pUC118 and *Eco*RI in pUC119, to ensure the presence of the primer binding sites. Alternatively, hybridization of labeled sequencing primer to Southern transfers of the deleted clones would reveal which were suitable.

Nevertheless, by combining all of the techniques available in this laboratory, such as unidirectional deletion by exonuclease III digestion, polymerase chain reactions, use of the single-stranded plasmid DNA system, etc., completion of the sequencing on both strands of the 12 kb linear plasmid within one year would now seem to be reasonable. Also, it should be noted that extensive use of a DNA synthesizer undoubtedly helped in speeding up the whole process by providing oligonucleotide primers for sequencing very difficult regions and for priming in polymerase chain reactions.

In previous studies, the evidence obtained by cross hybridization strongly suggested the presence of 900 bp inverted terminal repeats in pSCL (Keen *et al.*, 1988). Sequence analysis has confirmed this observation and indicates that pSCL indeed contains 969 bp inverted terminal repeats. When comparing the left and right terminal sequences, a few slight differences were observed (Fig. 4). Several clones containing the terminal fragments have been sequenced. The results showed that these differences exist in all of these clones, excluding the possibility that they were caused by the cloning procedures.

A comparison of the terminal sequences of the linear plasmid pSLA2 from *S. rochei* with pSCL revealed that the long inverted terminal repeats in both plasmids have a fairly high degree of similarity (Fig. 5). In particular, the first 12 bp of the two terminal sequences are exactly the same. This result is especially interesting with regard to the replication of linear plasmids. As discussed before, it has been demonstrated that the terminal sequences of adenoviruses and phage ϕ29 are very important for initiation of their replication. The terminal sequences of all adenoviruses are conserved and they all use a similar viral DNA polymerase. This is also true for each of the bacteriophage families represented by ϕ29, Cp-1 and PRD1. The eukaryotic linear plasmids, S1 and S2, have identical terminal sequences of 208 bp. Apparently, both plasmids use a DNA polymerase

encoded by S1 plasmid, while the S2 plasmid encodes only a putative RNA polymerase. The two linear plasmids pGKl1 and pGKl2 from *K. lactis* have quite different terminal repeats and terminal proteins (Hishinuma *et al.*, 1984; Stam *et al.*, 1986), and they obviously encode their own DNA polymerases with only a 46% conservation of amino acids (Tommasino *et al.*, 1988). Thus, it is reasonable to assume that there is a correlation between terminal sequences and use of a particular DNA polymerase. Based on this assumption, it is possible that the two plasmids (pSLA2 and pSCL) may have a common replication mechanism which is likely to be initiated by the same (or a similar) DNA polymerase. Besides, the fact that the terminal sequences in the two plasmids are identical leads to the suggestion that they may have the same or similar terminal proteins. Further evidence to support the speculation that the two linear plasmids share a common replication mechanism comes from the highly conserved small inverted repeats found at the termini of the two plasmids. These small inverted repeats are located at almost the same distance from the termini in both plasmids (Fig. 5). The small inverted repeats in pSLA2 have been suggested to be the sites for specific DNA-binding proteins. These hypothetical DNA-binding proteins have been postulated to hold the two double-stranded DNA ends together after their binding to these small inverted repeats and to form a so-called quadruple-stranded DNA structure (Hirochika *et al.*, 1984). This proposed racket frame-like structure is assumed to play a key role in linear plasmid replication, or integration into or excision from their host chromosomes (Sakaguchi, 1990). The *in vitro* replication systems of adenoviruses and phage ϕ29 have demonstrated that specific DNA-binding proteins encoded by the virus or phage are important for initiation of viral replication. In the case of adenoviruses, the viral DNA-binding protein is specifically associated with the termini of the viral genome. Therefore, by analogy with the viral replication systems, it seems likely that the small inverted repeats found in both plasmids are DNA-specific protein binding sites and binding of specific proteins to these sites may be involved in the initiation of plasmid replication. In addition, the conservation of these small inverted repeats in both

plasmids suggests that the same or similar DNA binding protein(s) may be associated with these sites. At present, there is no evidence to indicate whether the putative DNA binding proteins come from cellular sources or are encoded by the plasmids. Analysis of the polypeptides predicted by the 5 most probable ORFs revealed no structural similarity between these polypeptides and any known DNA-binding proteins (The search was performed using programs resident in the PCGENE computer software).

The presence of the long terminally repeated sequences in pSCL gives rise to speculations on the ability of the linear plasmid to integrate into the host chromosome, since many transposable elements in prokaryotes and eukaryotes contain inverted terminal repeats. As mentioned earlier, all linear plasmids so far investigated have inverted terminal repeats. In some cases, linear plasmids have been found to integrate into mitochondrial or chromosomal genomes. For instance, the linear plasmids S1 and S2 have been detected in an integrated form in the maize mitochondrial genome (Kemble and Mans, 1983). Genetic evidence also has strongly suggested that the linear plasmid SCP1 is able to integrate into various locations in the *S. coelicolor* chromosome (Hopwood and Wright, 1976) although hybridization experiments, using parts of SCP1 DNA as a probe, only gave sketchy evidence for this integration (Kinashi *et al.*, 1987). In the case of pSCL, attempts have been made by several laboratories to detect whether pSCL integrates into its host chromosome (Roy, K.L.; Jensen, S.E.; Aharonowitz, Y., personal communications). But, the results obtained from these experiments suggested that pSCL does not integrate into the *S. clavuligerus* chromosome DNA. There is no available evidence to indicate that pSLA2 from *S. rochei* integrates into its host chromosome. On the other hand, the information obtained so far cannot exclude the possibility that the small linear plasmids in *Streptomyces* may have lost the ability to integrate during evolution, while the large ones have retained this capacity.

The complete nucleotide sequence of pSCL permits us to examine the gene organization of the linear plasmid. Combining ORF predication with %G+C analysis, 5

highly probable ORFs were predicted. Although two of the ORFs (ORF 45 and ORF 5) did not match perfectly the patterns predicted by %G+C analysis, the characteristic %G+C distribution in these regions are unlikely to have occurred accidentally, since only 5 such regions have been found in a 12 kb plasmid! It should also be pointed out that most meaningful ORFs in known *Streptomyces* genes are detected by %G+C analysis, but there are exceptions. In the case of an ORF for the carbomycin resistance gene isolated from *S. thermotolerans* (Epp *et al.*, 1987), the first 50 codons did not match the %G+C analysis. In the case of an ORF overlapping the large RNA transcript from pSCL, the typical %G+C distribution normally found in known *Streptomyces* genes was not seen, although it is likely to be a mRNA for a protein (to be discussed later). Furthermore, some *Streptomyces* genes do not appear to have typical ribosome binding sites (reviewed by Seno and Baltz, 1989). Thus, the %G+C distribution and ribosome binding sites cannot be used as absolute criteria for selection of meaningful protein-coding ORFs. The same argument has been made by Kendall and Cohen (1988).

Since most linear plasmids so far reported contain genes putatively encoding their own DNA polymerases, a search for a DNA polymerase gene in pSCL has been conducted, based on the three main conserved domains of the B family DNA polymerases. Surprisingly, the three conserved domains found in almost all known linear plasmids were not detected in any of the polypeptides predicted by the 5 ORFs or by other ORFs which were not identified as probable by %G+C analysis. Because of the unique structural features of the linear plasmid, it is anticipated that a specific DNA polymerase should be encoded for its replication. There are several plausible explanations for not finding such a gene. First, the linear plasmid may encode an entirely different DNA polymerase from those of the B family DNA polymerases. However, comparison of the polypeptides predicted from the 5 ORFs with other known DNA polymerases did not reveal any significant homology either. This possibility requires further investigation. Second, the DNA polymerase required for pSCL replication may be provided by other sources, such as

other undiscovered linear plasmids in *S. clavuligerus*. This type of situation was observed in the S plasmids in *Zea mays*. The S1 and S2 plasmids coexist in the same host, but the replication of both plasmids depends on the DNA polymerase encoded by the S1 plasmid. In *Streptomyces*, many strains also harbor more than one linear plasmid. In particular, *S. coelicolor* has a series of large linear plasmids closely related to each other (Kinashi *et al.*, 1987). If this was the case with pSCL, the ^{32}P labeled intact pSCL DNA should hybridize to high molecular weight DNA since an unknown linear plasmid(s) providing the DNA polymerase would be expected to have terminal sequences similar to those of pSCL. But, as pointed out earlier, such a search did not reveal any positive signal other than pSCL. Nevertheless, the search for other linear plasmids in *S. clavuligerus* should be conducted further by using pulse-field electrophoretic techniques, since conventional methods may lead to degradation or even loss of large linear plasmids as observed in the case of SCP1. Finally, because of the unique structure of the linear plasmid, it is very unlikely that any of cellular DNA polymerases is directly used for the plasmid replication. However, it is possible that DNA polymerase activity might require components contributed by both the host cell and the plasmid, as in the case of the bacteriophage T7. The DNA polymerase of phage T7 consists of two polypeptides: one encoded by gene 5 of the phage and the other (the thioredoxin protein) encoded by the host cell, i.e. *E. coli* (Modrich and Richardson, 1975; Mark and Richardson, 1976). High polymerase activity and double-stranded exonuclease activity are present only when thioredoxin is associated with the gene 5 protein, although all of the catalytic activities of the enzyme are contributed by the gene 5 protein (Tabor *et al.*, 1987). Thus, it is possible that the linear plasmid encodes a DNA polymerase protein which requires a cellular component for activity. So far, there is no evidence to support this speculation.

Despite the fact that the 5 polypeptides predicted by the 5 presumptive ORFs of pSCL show no homology to any known proteins, a polypeptide deduced from ORF 45 shows an unusual hydrophobic pattern at its N-terminal sequence. Interestingly enough,

the putative *spd B* protein in the *Streptomyces* plasmid pIJ101, which is responsible for the "spread" phenotype, also contains three strongly hydrophobic domains in its N-terminal half (Kendall and Cohen, 1988). Since nothing is known about pock formation in *S. clavuligerus*, it is impossible to predict whether the putative ORF 45 is involved in the "spread" function or not. Similarly, no functions can be assigned to the other polypeptides deduced from the remaining probable ORFs at present.

To examine gene expression due to pSCL, transcript analysis was conducted. Northern transfer and hybridization experiments indicated that at least two RNA transcripts from pSCL were present during *S. clavuligerus* growth in liquid culture (Fig. 7). It was surprising that both RNA species were transcribed from one strand of pSCL and located in a region which does not correspond to any of the five most probable ORFs. Further characterization of the two RNA transcripts revealed that the larger transcript does contain an ORF, while the small RNA contains no ORF at all. The %G+C analysis showed that neither RNA gave the typical %G+C distribution common in *Streptomyces* structural genes. If the large RNA transcript is indeed an mRNA for a particular protein, its %G+C distribution is an exceptional case which cannot be predicted by conventional %G+C analysis. This raises the question as to the existence of other protein-coding ORFs which were not predicted by the %G+C analysis. Because of the lack of genetic information on pSCL, it is very difficult to draw any conclusions. It is possible that such an unusual example may be plasmid-specific. To our knowledge, this is the first *Streptomyces* structural gene which gives a completely unexpected %G+C distribution, if the large RNA transcript is in fact translated into a protein *in vivo*. Examination of the ORF encoded by the large RNA transcript reveals that it is preceded by a good ribosome binding site and in spite of the unexpected G+C distribution, it has a codon usage matching the overall usage of known *Streptomyces* genes. The protein (248 amino acid) deduced from this ORF may be too small to be a DNA polymerase since the smallest prokaryotic DNA polymerase (from PRD1 phage) is 553 amino acids in length (Jung *et al.*, 1987a). Although no DNA-

binding domains were found in this putative protein, it is still possible that it may function as a regulatory factor or a terminal protein.

A re-examination of the sequence did not reveal any meaningful ORF in the small RNA transcript. It was observed that the 5' terminal sequence of this transcript is complementary to the 5' terminal sequence of the putative ORF 3, suggesting that the small transcript may function as an antisense RNA for control of ORF 3 expression. The regulation of gene expression by possible antisense RNAs in *Streptomyces* has been proposed on several occasions. In the case of methylenomycin production and resistance genes in the linear plasmid SCP1, the mRNAs for the genes *mmy* and *mmr* overlapped each other at their 3' termini. A possible regulation of gene expression by such overlapping of the two mRNAs was postulated (Chater and Hopwood, 1989). Several RNA transcripts have been found to overlap the 5' terminal sequence of the neomycin resistance gene (*aph*) in *S. fradiae*, but in the opposite direction from the *aph* mRNA (Bibb and Janssen, 1987). It is possible that they may act as antisense RNAs to control *aph* gene expression, although it is not yet known whether those RNA species are involved in expression of a gene upstream of *aph* or not. In the case of pSCL, the role of the small RNA as an antisense RNA is proposed solely on the sequence, but its possible functions *in vivo* require further study. If it does act as an antisense RNA *in vivo*, it would be the first example of antisense RNA control of gene expression in a streptomycete.

The determination of the 5' ends of the two RNA transcripts allowed us to examine their probable promoter regions. Interestingly, the promoter region of the large RNA has some similarity to the P1 promoter of the thiostrepton resistance gene (Janssen *et al.*, 1985) in both the -10 and -35 regions. The two promoters have some similarity to the *E. coli* consensus sequence in the -10 region, but no similarity in the -35 region (Table 4). The sequence similarity between the two promoters suggests that transcription of the two genes may be controlled by a similar sigma factor. Control of transcription of the large RNA is further complicated by the fact that it is very actively transcribed soon after spore

germination in liquid culture. The level of transcription then decreases and remains relatively constant during the rest of cell growth (Fig. 12). Such differential expression of the large RNA may reflect the expression level of a specific sigma factor in *S. clavuligerus*. But, this does not exclude the expression of other factors, such as a repressor, at later stages of cell growth. It is possible that expression of the large RNA is required for replication of the plasmid since in the early stage of cell growth, cellular DNA and protein synthesis are always active in order to increase the total cell mass. Therefore, an increase in the copy number of the linear plasmid to adapt to rapid cell division is conceivable.

Unlike the promoter of the large RNA transcript, the presumed promoter region of the small RNA transcript shows little similarity to any known *Streptomyces* promoters or to the *E. coli* promoter consensus sequence (Table 4). This suggests that transcription of the small RNA is controlled by a different sigma factor from the one responsible for the large RNA. This conclusion is consistent with and reinforced by the observation that the level of transcription of the small RNA remains relatively constant through all stages of cell growth.

Taken all together, the data suggest that the different levels of transcription for the two transcripts from pSCL detected by the hybridization are likely to reflect the situation occurring *in vivo*. Why the large RNA was actively transcribed in the early stage, but much less in the later stage of cell growth, and what control mechanism was involved, require further study.

It is interesting to note that a weak cDNA band can be detected 9 nucleotides upstream of the two major cDNA bands in the 5' primer extension experiments (Fig. 9). This might indicate that an additional promoter for transcription of the small RNA may exist. It has been observed that many *Streptomyces* genes have multiple promoters (Seno and Baltz, 1989). The upstream sequences (-10 and -35 regions) of the minor RNA transcript show no homology to the promoters of the two major RNA transcripts or to any known *Streptomyces* promoters. Therefore, it is possible that the small RNA is

transcribed from two different promoters, a strong promoter and a weak one. The minor cDNA band may also be caused by artifacts. This possibility requires further investigation.

Both promoters for the large and small RNAs are quite different from the *E. coli* promoter consensus sequence. This suggests that the two promoters do not belong to that group of *Streptomyces* promoters which has shown the closest resemblance to the *E. coli* consensus sequence (Hopwood *et al.*, 1986a). It is therefore not surprising that the two promoters are not recognized by the *E. coli* RNA polymerase. In fact, the linear plasmid pSCL was used as a template in *E. coli* cell extracts for *in vitro* coupled transcription and translation and no positive results were obtained (data not shown).

Three inverted repeats were detected near the sequences where each of the two RNA transcripts terminates. These inverted repeats are capable of forming stem-loop structures (Fig. 12). It has been shown that many transcription terminators in *Streptomyces* are quite similar to rho-independent terminators in *E. coli*, except that poly dT sequences at the 3' end of stem-loop structures are not found in the DNA strand from which the mRNA is to be transcribed (Hopwood *et al.*, 1986a). Besides, some mRNAs in *Streptomyces* are found to terminate at the 5' end of stem-loop structures, rather than at the 3' end, as is normal in *E. coli*. Deng *et al.* (1987) reported that the terminator of the pIJ101C transcript effectively terminated transcription at the same nucleotides in both *E. coli* and *S. lividans*. The evidence indicates that it is very likely that the presumptive stem-loop structures near the 3' ends of the two RNA transcripts function as transcription terminators.

The detection of only two RNA transcripts from pSCL through all stages of cell growth in liquid culture is surprising. Since an independent research group at the University of Tel Aviv in Isreal also found only the same two RNA transcripts from pSCL (Aharonowitz, Y., personal communication), it is hard to imagine that some RNA species from pSCL were missed. However, it is conceivable that some RNA species are expressed at such low levels that they cannot be detected. Alternatively, they may be extremely labile

or degraded very rapidly. It seems highly unlikely that the single gene product from the large RNA transcript would be able to support plasmid replication unless additional necessary components are supplied by the host cell or by other linear plasmids, if any. There is a possibility that expression of some genes in pSCL may require specific sigma factors or other components which are only present at other stages of differentiation of *S. clavuligerus*. Since differentiation of *Streptomyces* in liquid culture is limited, the expression of these genes on pSCL in liquid culture may be impossible. On the other hand, the fact that the linear plasmid persists, and is probably replicated, in liquid culture seems to be paradoxical to this possibility unless expression of these genes in pSCL is not required for replication of pSCL in liquid culture.

To study gene expression in pSCL further, detection of gene products encoded by pSCL was attempted using *S. lividans* cell extracts as an *in vitro* coupled transcription and translation system (Thompson *et al.*, 1984). As seen in the Results, the system failed to reveal any convincing coupled transcription and translation activity using pSCL as a template (Fig. 13). The active transcription and translation of the β-lactamase gene of pAT153 in these cell extracts indicated that most components required for both RNA transcription and protein synthesis were present in the system. It is likely that the failure by the system to use the pSCL template was due to the lack of active transcription of pSCL. Probably a specific sigma factor or other components required for expression of the genes on pSCL are completely absent from the *S. lividans* cell extracts. If this is the case, cell extracts prepared from the original pSCL host, *S. clavuligerus*, might support the pSCL transcription *in vitro*. Unfortunately, cell extracts prepared from different stages of cell growth in *S. clavuligerus* were not active at all. Another explanation would be that some factors required for the expression of genes on pSCL may be present only in the later stages of the *S. lividans* cell growth. Thus, *S. lividans* cell extracts prepared from the early stage (16 hours after spore inoculation; this was the only stage which gave active cell extracts) may result in the failure of pSCL transcription. Nevertheless, all of these

explanations are hard to apply in the case of pIJ702 since this multicopy plasmid has been used in both *S. lividans* and *S. clavuligerus* as a cloning vector. It is also possible that the failure to express *Streptomyces* DNA in the *S. lividans* cell extracts may be due to technical problems. Further investigations should definitely be conducted to verify the possibilities or problems mentioned above.

As discussed in the Introduction, genetic and biochemical evidence in several cases strongly suggest that some linear plasmids carry antibiotic biosynthesis and resistance genes. *S. clavuligerus* is known to produce clavulanic acid and a number of β-lactam antibiotics, such as penicillin N, cephamycin C, etc. (Jensen, 1986). It also produces non-β-lactam antibiotics like tunicamycin and holomycin (Kenig and Reading, 1979). The structural genes for biosynthesis of β-lactam antibiotics have been demonstrated to be either chromosomal or on a very large plasmid (Leskiw *et al.*, 1988). Based on genetic evidence, Kirby (1978) located the genes responsible for holomycin production on the *S. clavuligerus* chromosome. It seems probable that all biosynthetic genes for antibiotics in *S. clavuligerus* reside on its chromosome. Kirby (1978), however, showed that an extrachromosomal element in *S. clavuligerus* controlled the production of holomycin on solid media. This element, designated as SUE1, was lost after UV irradiation of SUE1^{+} strains, as was the phenotype of holomycin production. Moreover, genetic analysis indicated that the element was transmissible in crosses from Hol^{+} to Hol^{-} strains, in which the phenotype of holomycin production was acquired by the Hol^{-} strains. The transfer frequency of the Hol^{+} phenotype was about 13% nonselectively and about 80% if chromosomal recombinants were selected, indicating that transfer of the Hol^{+} phenotype was not due to chromosomal recombination. Interestingly, the control of holomycin production by this element seems to be related to differentiation, as the strain gave the Hol^{-} phenotype only on solid media, but produced holomycin in submerged liquid culture where differentiation does not occur. Thus, it was assumed that the element may encode regulatory genes that control holomycin biosynthesis and that such regulatory functions

may be expressed only when *S. clavuligerus* undergoes differentiation. At present, it is not clear if the extrachromosomal element postulated by Kirby (1978) plays a positive or negative role in the control of holomycin production.

This extrachromosomal element has never been isolated. On the basis of these observations and the fact that no covalently closed circular plasmids have ever been detected in *S. clavuligerus*, it is reasonable to speculate that the linear plasmid pSCL is the presumptive extrachromosomal element involved in control of holomycin production in *S. clavuligerus*. If so, such a regulatory control of holomycin production by pSCL would be at some stage of differentiation since holomycin was produced in liquid culture regardless of the presence of the extrachromosomal element. It is possible that pSCL may encode a gene for a sigma-like factor, which is required for holomycin production on solid media. When *S. clavuligerus* is growing in liquid media, in which the expression of this factor might become impossible, a cellular sigma-like factor may be expressed and replace the one encoded by pSCL. Thus, holomycin production would be no longer controlled by pSCL. In addition, the fact that most genes on pSCL seem to remain silent in liquid culture might support this speculation, if expression of these genes on pSCL were only possible during differentiation.

A fundamental question regarding linear plasmids in *Streptomyces* is their origin. From the accumulated knowledge of linear plasmids it seems that they are likely to have originated from viruses. In particular, the finding of the virus-like DNA and RNA polymerase genes in eukaryotic linear plasmids supports this speculation. Despite the fact that pSCL is the first *Streptomyces* linear plasmid for which the complete sequence is available, it is reasonable to postulate that *Streptomyces* linear plasmids may be derived from viruses simply because they share great structural similarities with many viruses and bacteriophages, such as adenoviruses and ϕ29. Thus, based on this assumption, it is likely that *Streptomyces* linear plasmids use a replication mechanism quite similar to the one proposed for adenoviruses and bacteriophage ϕ29. Although a virus-like DNA polymerase

gene could not be detected in the pSCL sequence, it is possible that linear plasmids in *Streptomyces* may be derived from another group of bacteriophages, which use a different class of DNA polymerases for their replication. At present we do not know whether there are other linear plasmids in *S. clavuligerus*. Therefore, it is possible that a virus-like DNA polymerase gene similar to those found in the known eukaryotic linear plasmids may exist in undiscovered linear plasmids in *S. clavuligerus*, and thus support the replication of pSCL.

The availability of the complete nucleotide sequence of pSCL opens many new doors to research in the replication and gene regulation of *Streptomyces* linear plasmids.

To verify whether pSCL confers any phenotypes on the host, it would be worthwhile to cure *S. clavuligerus* of pSCL. Although preliminary trials of protoplast regeneration failed to eliminate pSCL from *S. clavuligerus*, several other methods can be used, such as UV irradiation and ethidium bromide or acridine dye treatments. If these methods still fail to cure pSCL, introduction of pSCL into another host, such as *S. lividans* may be possible. Protoplast fusion may be a way to achieve this. Another approach would be to isolate the intact DNA-protein complex of pSCL by sucrose gradient sedimentation. Then, the purified DNA-protein complex of pSCL could be introduced into another host by electroporation. It is possible that pSCL may not confer any observable phenotype on the host, but the successful introduction of pSCL into another host may enable us to study the relationship between pSCL and the original host, for instance, the potential involvement of cellular components in pSCL replication and gene regulation. Alternatively, it might be worthwhile to search for naturally occurring transfer of pSCL, by looking for pock formation, since one *Streptomyces* linear plasmid, pSRM from *S. rimosus*, has been shown to give the pock formation phenotype (Chardon-Loriaux *et al.*, 1986). If pSCL could be transferred this way, the introduction of pSCL into other hosts would be much easier.

To study gene expression in pSCL, an *in vitro* transcription system could be utilized in which bacteriophage T3 or T7 promoters and RNA polymerases are used. Each of the 5 predicted, most probable protein-coding ORFs, with their flanking sequences, would be cloned into a vector containing a T3 or T7 phage promoter. Thus, RNAs corresponding to each ORF could be transcribed *in vitro*, using T3 or T7 phage RNA polymerase (commercially available). Then, the RNAs obtained *in vitro* will be introduced into *S. lividans* cell extracts to examine their capacity to be translated into protein products. In this way, the absence of specific sigma factors or other components in the *S. lividans* cell extracts, which may have led to the failure to identify gene products from pSCL, can be avoided. The capability of these mRNAs to be translated in the cell extracts may prove the protein-coding potential of these ORFs *in vivo*. Similarly, this approach could also be applied to the large RNA transcript of pSCL to clarify whether the transcript is an mRNA for a protein.

Although only two RNA transcripts from pSCL could be detected in *S. clavuligerus* grown in liquid culture, it is not known whether the expression of other genes in pSCL occur at other stages of cell differentiation. Thus, it would be very interesting to examine RNA transcripts isolated from cultures grown on solid media. It might be difficult to obtain large amounts of mycelia, but it may still be possible to obtain sufficient mRNA for hybridization if some genes in pSCL are expressed in other stages of differentiation.

In light of the finding that giant linear plasmids are present in many species of *Streptomyces*, it is logical to examine *S. clavuligerus* for the presence of giant linear plasmids by using pulse-field gel electrophoresis techniques. In particular, such a search for other linear plasmids may help us to resolve the puzzle that a virus-like DNA polymerase could not be found in pSCL.

In some studies of adenovirus and bacteriophage ϕ29 replication, it has been demonstrated that the terminal proteins on their parental DNA genomes are not absolutely required for initiation of replication of adenovirus and ϕ29 DNAs, at least *in vitro*. These

observations may suggest the possibility of construction of new linear plasmids *in vitro*, without terminal proteins but with terminal sequences and an antibiotic resistance gene. These constructs could be introduced by electroporation into a host harboring a natural linear plasmid. The new linear plasmid may replicate by support of the existing linear plasmid in the host if the terminal proteins on the parental DNA are not needed for the initiation of plasmid replication *in vivo*. If such constructs cannot replicate, homologous recombination between the new constructs and the existing linear plasmid may occur under selective conditions, e.g. the presence of an antibiotic. Thus, by selection of different flanking sequences around the antibiotic resistance gene, new linear plasmids with the antibiotic resistance gene in different locations may be constructed *in vivo* by recombination. Based on observations that many eukaryotic organisms and *Streptomyces* species are found to harbor multiple linear plasmids, these newly constructed linear plasmids might coexist with the natural linear plasmid in the same host. This approach may enable us to identify the genes or regions of the linear plasmid essential for its replication. In addition, insertion into non-essential regions might permit us to introduce foreign genes into the host. The knowledge of the complete sequence of pSCL certainly has advantages for the construction of such "new linear plasmids".

In order to investigate the possible relationship between an extrachromosomal element and holomycin production in *S. clavuligerus*, it would be interesting to see whether pSCL is involved in the regulation of holomycin biosynthesis. A simple way to test such a possibility might be to use pSCL DNA as a probe to hybridize the total DNA from Hol$^-$ strains obtained by UV irradiation. If pSCL is lost in all Hol$^-$ but not in Hol$^+$ strains, a relationship between pSCL and the regulation of holomycin production could be established. Alternatively, isolation of pSCL from Hol$^-$ strains might yield a truncated linear plasmid, which may result from deletions or recombination *in vivo*. Then, further restriction endonuclease analysis may be required. Combined with examination of pSCL DNA from Hol$^-$ strains and RNA transcript analysis from cultures grown on solid media, it

may be possible to reveal the regulatory functions of pSCL which control holomycin production on solid media.

In summary, the determination of the first complete sequence of a prokaryotic linear plasmid from a streptomycete is undoubtedly an important step towards a better understanding of the biological functions of *Streptomyces* linear plasmids. Information on gene organization and regulation in pSCL will certainly help to extend our knowledge of gene expression and regulation in *Streptomyces*. Further understanding of the mechanism of linear plasmid replication may make it possible for us to develop new cloning vectors for genetic engineering of this organism. Genetic and biochemical studies on pSCL will provide us with a model to further explore the genetic behavior of other *Streptomyces* linear plasmids, particularly the giant linear plasmids, as well as their possible roles in antibiotic biosynthesis.

5. BIBLIOGRAPHY

Aguilar, A., and D.A. Hopwood. 1982. Determination of methylenomycin A synthesis by the pSV1 plasmid from *Streptomyces violaceus-ruber* SANK-95570. J. Gen. Microbiol. **128**: 1893-1902.

Akagawa, H. 1987. Characterization and localization of plasmid functions involved in pock formation and pock resistance of plasmid pSK3* of *Streptomyces kasugaensis* MB273. J. Gen. Microbiol. **133**: 1951-1958.

Antonov, P.P., I.G. Ivanov, and G.G. Markov. 1977. Heterogeneity of *Streptomyces* DNA. FEBS Letters **79**: 151-154.

Bamford, D.H., T. McGraw, G. Mackenzie, and L. Mindich. 1983. Identification of a protein bound to the termini of bacteriophage PRD1 DNA. J. Virol. **47**: 311-316.

Bamford, D.H., and L. Mindich. 1984. Characterization of the DNA-protein complex at the termini of the bacteriophage PRD1 genome. J. Virol. **50**: 309-315.

Barbour, A.G., and C.F. Garon. 1987. Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. Science **237**: 409-411.

Baroudy, B.M., S. Venkatesan, and B. Moss. 1983. Structure and replication of vaccinia virus telomeres. Cold Spring Harbor Symp. Quant. Biol. **47**: 723-729.

Benigni, R., P.P. Antonov, and A. Crere. 1975. Estimate of the genome size by renaturation studies in *Streptomyces*. Appl. Microbiol. **30**: 324-326.

Benton, W.D., and R.W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**: 180-182.

Bertrand, H., and A.J.F. Griffiths. 1989. Linear plasmids that integrate into mitochondrial DNA in *Neurospora*. Genome **31**: 155-159.

Bibb, M.J., and D.A. Hopwood. 1981. Genetic studies of the fertility of plasmid SCP2 and its SCP2* variants in *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. **126**: 427-442.

Bibb, M.J., and G.R. Janssen. 1987. Unusual features of transcription and translation of antibiotic resistance genes in antibiotic-producing *Streptomyces*. *In* Genetics of industrial microorganisms. Part B. *Edited by* M. Alacevic, D. Hranueli, and Z. Toman. Pliva, Zagreb, pp. 309-318.

Bibb, M.J., R.F. Freeman, and D.A. Hopwood. 1977. Physical and genetical characterization of a second sex factor, SCP2, for *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. **154**: 155-166.

Bibb, M.J., J.M. Ward, T. Kieser, S.N. Cohen, and D.A. Hopwood. 1981. Excision of chromosomal DNA sequences from *Streptomyces coelicolor* forms

a novel family of plasmids detectable in *Streptomyces lividans*. Mol. Gen. Genet. **184**: 230-240.

Bibb, M.J., P.R. Findlay, and M.W. Johnson. 1984. The relationship between base composition and codon usage in bacterial genes and its use in the simple and reliable identification of protein coding sequences. Gene **30**: 157-166.

Birnboim, H.C. 1983. A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol. **100**: 143-155.

Blanco, L., and M. Salas. 1984. Characterization and purification of a phage ϕ29-encoded DNA polymerase required for the initiation of replication. Proc. Natl. Acad. Sci. USA **81**: 5325-5329.

Blanco, L., and M. Salas. 1985. Replication of phage ϕ29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length ϕ29 DNA. Proc. Natl. Acad. Sci. USA **82**: 6404-6408.

Bruton, C.J., and K.F. Chater. 1987. Nucleotide sequence of IS*110*, an insertion sequence of *Streptomyces coelicolor* A3(2). Nucl. Acids Res. **15**: 7053-7065.

Calzon, F.J., R.J. Britten, and E.H. Davidson. 1987. Mapping of gene transcripts by nuclease protection assays and cDNA primer extension. Methods Enzymol. **152**: 611-632.

Challberg, M.D., and T.J. Kelly, Jr. 1979. Adenovirus DNA replication *in vitro*. Proc. Natl. Acad. Sci. USA **76**: 655-659.

Challberg, M.D., S.V. Desiderio, and T.J. Kelly, Jr. 1980. Adenovirus DNA replication *in vitro*: characterization of a protein covalently linked to nascent DNA strands. Proc. Natl. Acad. Sci. USA **77**: 5105-5109.

Chardon-Loriaux, I., M. Charpentier, and F. Percheron. 1986. Isolation and characterization of a linear plasmid from *Streptomyces rimosus*. FEMS Microbiol. Letters **35**: 151-155.

Chater, K.F. 1986. *Streptomyces* phage and their applications to *Streptomyces* genetics. *In* Antibiotic-producing *Streptomyces*, The bacteria, Vol. IX. *Edited by* S.W. Queener, and E. Day. Academic Press, New York, pp. 119-158.

Chater, K.F., and D.A. Hopwood. 1983. *Streptomyces* genetics. *In* The biology of actinomycetes. *Edited by* M. Goodfellow, M. Mordarski, and S.T. Williams. Academic Press, London, pp. 229-286.

Chater, K.F., and D.A. Hopwood. 1989. Antibiotic biosynthesis in *Streptomyces*. *In* Genetics of bacterial diversity. *Edited by* D.A. Hopwood and K.F. Chater. Academic Press, San Diego, pp. 129-150.

Chater, K.F., D.J. Henderson, M.J. Bibb, and D.A. Hopwood. 1988. Genome flux in *Streptomyces coelicolor* and other streptomycetes and its possible relevance to the evolution of mobile antibiotic resistance determinants. *In* Transposition. *Edited by* A.J. Kingsman, K.F. Chater, and S.M. Kingsman. Cambridge University Press, Cambridge, pp. 1-42.

Chen, H-Z., and G. Zubay. 1983. Prokaryotic coupled transcription-translation. Methods Enzymol. **101**: 647-690.

Chung, S.T. 1982. Isolation and characterization of *Streptomyces fradiae* plasmids which are prophage of the actinophage ϕSF1. Gene **17**: 239-246.

Chung, S.T. 1987. Tn*4556*, a 6.8 kb transposable element of *Streptomyces fradiae*. J. Bacteriol. **169**: 4436-4441.

Chung, S.T., and M.B. Molnar. 1983. Characterization and development of *Streptomyces* gene cloning vectors I. *Streptomyces fradiae*. *In* Trends in antibiotic research. *Edited by* H. Umezawa, A.L. Demain, T. Hata, and C.R. Hutchinson. Japan Antibiotic Research Association, Tokyo, pp. 42-47.

Cullum, J., J. Altenbuchner, F. Flett, and W. Piendl. 1986. DNA amplification and genetic instability in *Streptomyces*. Biotechnol. Genet. Eng. Rev. **4**: 59-78.

Curtiss III, R., and D.R. Stallions. 1969. Probability of F integration and frequency of stable Hfr donors in F^{+} populations of *E. coli* K-12. Genetics **63**: 27-38.

Deng, Z., T. Kieser, and D.A. Hopwood. 1987. Activity of a *Streptomyces* transcriptional terminator in *Escherichia coli*. Nucleic Acid Res. **15**: 2665-2675.

Denhardt, D.T. 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Comm. **23**: 641-646.

Dunn, J.J., and F.W. Studier. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and locations of T7 genetic elements. J. Mol. Biol. **166**: 477-535.

Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179**: 125-142.

Epp, K.J., S.G. Burgett, and B.E. Schoner. 1987. Cloning and nucleotide sequence of a carbomycin-resistance gene from *Streptomyces thermotolerans*. Gene **53**: 73-83.

Esser, K., and F. Kempken. 1986. Structure and function of linear extrachromosomal DNA in eukaryotes. Process Biochem. **34**: 69-76.

Feinberg, A., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**: 6-13.

Feinberg, A., and B. Vogelstein. 1984. Addendum. Anal. Biochem. **137**: 266-267.

Fickett, J.W. 1982. Recognition of protein coding regions in DNA sequences. Nucl. Acids Res. **10**: 5303-5318.

Freier, S.M., R. Kierzek, J.A. Jaeger, N. Sugimoto, M.H. Caruthers, T. Neilson, and D.H. Turner. 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83**: 9373-9377.

Garcia, E., A. Gomez, C. Ronda, C. Escarmis, and R. Lopez. 1983a. Pneumococcal bacteriophage Cp-1 contains a protein tightly bound to the 5' termini of its DNA. Virology **128**: 92-104.

Garcia, J.A., R. Pastrana, I. Prieto, and M. Salas. 1983b. Cloning and expression in *Escherichia coli* of the gene coding for the protein linked to the ends of *Bacillus subtilis* phage ϕ29 DNA. Gene **21**: 65-76.

Garcia, J.A., M.A. Penalva, L. Blanco, and M. Salas. 1984. Template requirements for the initiation of phage ϕ29 DNA replication *in vitro*. Proc Natl. Acad. Sci. USA **81**: 80-84.

Garcia, P., J.M. Hermoso, J.A. Garcia, E. Garcia, E. Lopez, and M. Salas. 1986. Formation of a covalent complex between the terminal protein of pneumococcal bacteriophage Cp-1 and 5'-dAMP. J. Virol. **58**: 31-35.

Garvey, K.J., H. Yoshikawa, and J. Ito. 1985. The complete sequence of the *Bacillus* phage ϕ29 right early region. Gene **40**: 301-309.

Geshelin, P., and K.I. Berns. 1974. Characterization and localization of the naturally occurring cross-links in vaccinia virus DNA. J. Mol. Biol. **88**: 785-796.

Graham, F.L. 1984. Covalently closed circles of human adenovirus DNA are infectious. EMBO J. **3**: 2917-2922.

Graham, F.L., J. Rudy , and P. Brinkley. 1989. Infectious circular DNA of human adenovirus type 5: regeneration of viral DNA termini from molecules lacking terminal sequences. EMBO J. **8**: 2077-2085.

Grunstein, M., and D.S. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**: 3961-3965.

Gunge, N., A. Tamaru, F. Ozawa, and K. Sakaguchi. 1981. Isolation and characterization of linear deoxyribonucleic acid plasmids from *Kluyveromyces lactis* and the plasmid associated killer character. J. Bacteriol. **145**: 382-390.

Gunge, N., K. Murata, and K. Sakaguchi. 1982. Transformation of *Saccharomyces cerevisiae* with linear DNA killer plasmids from *Kluyveromyces lactis*. J. Bacteriol. **151**: 462-464.

Gutierrez, J., J. Vinos, I. Prieto, E. Mendez, J.M. Hermoso, and M. Salas. 1986. Signals in the ϕ29 DNA-terminal protein template for the phage ϕ29 DNA replication. Virology **155**: 474-483.

Hanahan, D. 1985. Techniques for transformation of *E. coli*. *In* DNA cloning - A Practical Approach. Vol. 1. *Edited by* D.M. Glover. IRL Press, Oxford, Washington DC, pp. 109-136.

Hay, R.T., N.D. Stow, and I.M. McDougall. 1984. Replicauon of adenovirus minichromosomes. J. Mol. Biol. **175**: 493-510.

Hayakawa, T., T. Tanaka, K. Sakaguchi, N. Otake, and H. Yonehara. 1979. A linear plasmid-like DNA in *Streptomyces* spp. producing lankacidin group antibiotics. J. Gen. Appl. Microbiol. **25**: 255-260.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351-359.

Henikoff, S. 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. **155**: 156-165.

Hirochika, H., and K. Sakaguchi. 1982. Analysis of linear plasmids isolated from *Streptomyces*: association of protein with the ends of the plasmid DNA. Plasmid **7**: 59-65.

Hirochika, H., K. Nakamura, and K. Sakaguchi. 1984. A linear DNA plasmid from *Streptomyces rochei* with an inverted terminal repetition of 614 base pairs. EMBO J. **3**: 761-766.

Hishinuma, F., K. Nakamura, K. Hirai, R. Nishizawa, N. Gunge, and T. Maeda. 1984. Cloning and nucleotide sequences of the linear DNA killer plasmids from yeast. Nucl. Acids Res. **12**: 7581-7597.

Hodgson, D.A., and K.F. Chater. 1981. A chromosomal locus controlling extracellular agarase production by *Streptomyces coelicolor* A3(2), and its inactivation by chromosomal integration of plasmid SCP1. J. Gen. Microbiol. **124**: 339-348.

Hopwood, D.A. 1981. Genetic studies with bacterial protoplasts. Ann. Rev. Microbiol. **35**: 237-272.

Hopwood, D.A., and H.M. Wright. 1973. Transfer of a plasmid between *Streptomyces* species. J. Gen. Microbiol. **77**: 187-195.

Hopwood, D.A., and H.M. Wright. 1976. Genetic studies on SCP1-prime strains of *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. **95**: 107-120.

Hopwood, D.A., G. Hintermann, T. Kieser, and H.M. Wright. 1984a. Integrated DNA sequences in three streptomycetes form related autonomous plasmids after transfer to *Streptomyces lividans*. Plasmid **11**: 1-16.

Hopwood, D.A., D.J. Lydiate, F. Malpartida, and H.M. Wright. 1984b. Conjugative sex plasmids in *Streptomyces*. *In* Plasmids in bacteria. *Edited by* D.R. Helinski, S.N. Cohen, D.B. Clewell, D.A. Jackson, and A. Hollaender. Plenum Press, New York, pp. 615-634.

Hopwood, D.A., F. Malpartida, H.M. Kieser, H. Ikeda, J. Duncan, I. Fujii, B.A.M. Rudd, H.G. Floss, and S. Omura. 1985a. Production of 'hybrid' antibiotics by genetic engineering. Nature **314**: 642-644.

Hopwood, D.A., M.J. Bibb, K.F. Chater, T. Kieser, C.J. Bruton, H.M. Kieser, D.J. Lydiate, C.P. Smith, J.M. Ward, and H. Schrempf. 1985b. Genetic manipulation of *Streptomyces* - A Laboratory Manual. The John Innes Foundation, Norwich, U.K.

Hopwood, D.A., M.J. Bibb, K.F. Chater, G.R. Janssen, F. Malpartida, and C.P. Smith. 1986a. Regulation of gene expression in antibiotic-producing *Streptomyces*. *In* Regulation of gene expression 25 years on. *Edited by* I.R. Booth, and C.F. Higgens. Cambridge University Press, London, pp. 251-276.

Hopwood, D.A., T. Kieser, D.J. Lydiate, and M.J. Bibb. 1986b. *Streptomyces* plasmids: their biology and use as cloning vector. *In* Antibiotic-producing *Streptomyces*, The bacteria. Vol. IX. *Edited by* S.W. Queener, and E. Day. Academic Press, New York, pp. 159-229.

Hopwood, D.A., M.J. Bibb, K.F. Chater, and T. Kieser. 1987. Plasmid and phage vectors for gene cloning and analysis in *Streptomyces*. Methods Enzymol. **153**: 116-165.

Hu, N-T., and J. Messing. 1982. The making of strand-specific M13 probes. Gene **17**: 271-277.

Hutter, R., and G. Hintermann. 1985. Genetic instability in streptomyces. *In* Industrial aspects of biochemistry and genetics. *Edited by* N.G. Alaeddinoglu, A.L. Demain, and G. Lancini. Plenum Press, New York, pp. 24-34.

Ito, J. 1978. Bacteriophage ϕ29 terminal protein: its association with the 5' termini of the ϕ29 genome. J. Virol. **28**: 895-904.

Janssen, G.R., M.J. Bibb, C.P. Smith, J.M. Ward, T. Kieser, and M.J. Bibb. 1985. Isolation and analysis of *Streptomyces* promoters. *In* Microbiology - 1985. *Edited by* L. Leive. American Society for Microbiology, Washington, DC, pp. 392-396.

Jaurin, B., and S.N. Cohen. 1984. *Streptomyces lividans* RNA polymerase recognizes and uses *Escherichia coli* transcription signals. Gene **28**: 83-91.

Jensen, S.E. 1986. Biosynthesis of cephalosporins. CRC Crit. Rev. Biotechnol. **3**: 277-301.

Joyce, C.M., W.S. Kelley, and N.D.F. Grindley. 1982. Nucleotide sequence of the *Escherichia coli polA* gene and primary structure of DNA polymerase I. J. Biol. Chem. **257**: 1958-1964.

Jung, G., M.C. Leavitt, J.C. Hsieh, and J. Ito. 1987a. Bacteriophage PRD1 DNA polymerase: evolution of DNA polymerases. Proc. Natl. Acad. Sci. USA **84**: 8287-8291.

Jung, G., M.C. Leavitt, and J. Ito. 1987b. Yeast killer plasmid pGKl1 encodes a DNA polymerase belonging to the family B DNA polymerase. Nucl. Acids Res. **15**: 9088.

Katz, E., C.J. Thompson, and D.A. Hopwood. 1983. Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. J. Gen. Microbiol. **129**: 2703-2714.

Keen, C.L., S. Mendelovitz, G. Cohen, Y. Aharonowitz, and K.L. Roy. 1988. Isolation and characterization of a linear DNA plasmid from *Streptomyces clavuligerus*. Mol. Gen. Genet. **212**: 172-176.

Kemble, R.J., and R.J. Mans. 1983. Examination of the mitochondrial genome revertant progeny from S *cms* maize with cloned S1 and S2 hybridization probes. J. Mol. Appl. Genet. **2**: 161-171.

Kendall, K., and J. Cullum. 1986. Identification of a DNA sequence associated with plasmid integration in *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. **202**: 240-245.

Kendall, K.J., and S.N. Cohen. 1987. Plasmid transfer in *Streptomyces lividans*: identification of a *kil-kor* system associated with the transfer region of pIJ101. J. Bacteriol. **169**: 4177-4183.

Kendall, K.J., and S.N. Cohen. 1988. Complete nucleotide sequence of the *Streptomyces lividans* plasmid pIJ101 and correlation of the sequence with genetic properties. J. Bacteriol. **170**: 4634-4651.

Kenig, M., and C. Reading. 1979. Holomycin and an antibiotic (MM19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. J. Antibiotics **32**: 549-554.

Kieser, T. 1984. Factors affecting the isolation of cccDNA from *Streptomyces lividans* and *Escherichia coli*. Plasmid **12**: 19-36.

Kieser, T., D.A. Hopwood, H.M. Wright, and C.J. Thompson. 1982. pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors. Mol. Gen. Genet. **185**: 223-238.

Kinashi, H., and M. Shimaji. 1987. Detection of giant linear plasmids in antibiotic producing strains of *Streptomyces* by the OFAGE technique. J. Antibiotics **40**: 913-916.

Kinashi, H., M. Shimaji, and A. Sakai. 1987. Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. Nature **328**: 454-456.

Kirby, R. 1978. An unstable genetic element affecting the production of the antibiotic holomycin by *Streptomyces clavuligerus*. FEMS Microbiol. Letters **3**: 283-286.

Kirby, R., and D.A. Hopwood. 1977. Genetic determination of methylenomycin biosynthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. **98**: 239-252.

Kitada, K., and F. Hishinuma. 1987. A new linear DNA plasmid isolated from the yeast *Saccharomyces kluyveri*. Mol. Gen. Genet. **206**: 377-381.

Kleppe, K., E. Ohtsuka, R. Kleppe, I. Molinneux, and H.G. Khorana. 1971. Studies on polynucleotides XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. J. Mol. Biol. **56**: 341-361.

Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**: 495-508.

Kutzner, H.G. 1981. The family Streptomycetaceae. *In* The prokaryotes. *Edited by* M.P. Starr, H. Stolp, H.G. Truper, A. Balows, and H.G. Schlegel. Springer, New York, pp. 2028-2082.

Kuzmin, E.V., and J.V. Levchenko. 1987. S1 plasmid from cms-S-maize mitochondria encodes a viral type DNA polymerase. Nucl. Acids Res. **15**: 6758.

Kuzmin, E.V., J. V. Levchenko, and G.N. Zaitseva. 1988. S2 plasmid from cms-S-maize mitochondria potentially encodes a specific RNA polymerase. Nucl. Acids Res. **16**: 8729.

Landschuiz, W.H., P.F. Johnson, and S.L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**: 1759-1764.

Lechner, R.L., and T.J. Kelly, Jr. 1977. The structure of replicating adenovirus 2 DNA molecules. Cell **12**: 1007-1020.

Leskiw, B.K., Y. Aharonowitz, M. Mevarech, S. Wolfe, L.C. Vining, D.W.S. Westlake, and S.E. Jensen. 1988. Cloning and nucleotide sequence determination of the IPNS gene from *Streptomyces clavuligerus*. Gene **62**: 187-196.

Levings III, C.S., and R.R. Sederoff. 1983. Nucleotide sequence of the S2 mitochondrial DNA from the S cytoplasm of maize. Proc. Natl. Acad. Sci. USA **80**: 4055-4059.

Levings III, C.S., B.D. Kim, D.R. Pring, M.F. Conde, R.J. Mans, J.R. Laughnan, and S.J. Gabay-Laughnan. 1980. Cytoplasmic reversion of cms-S in maize: association with a transpositional event. Science **209**: 1021-1024.

Lis, J. 1980. Fractionation of DNA fragments by polyethylene glycol induced precipitation. Methods Enzymol. **65**: 347-353.

Lydiate, D.J., H. Ikeda. and D.A. Hopwood. 1986. A 2.6 kb DNA sequence of *Streptomyces coelicolor* A3(2) which functions as a transposable element. Mol. Gen. Genet. **203**: 79-88.

Lydiate, D.J., D.J. Henderson, A.M. Ashby, and D.A. Hopwood. 1987. Transposable elements of *Streptomyces coelicolor* A3(2). *In* Genetics of industrial microorganisms. Part B. *Edited by* M. Alacevic, D. Hranueli, and Z. Toman. Ogenjen Prica Printing Works, Karlovac, pp. 49-56.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning - A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Marck, C. 1988. "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids. Res. **16**: 1829-1836.

Mark, D.F., and C.C. Richardson. 1976. *Escherichia coli* thioredoxin: a subunit of bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **73**: 780-784.

Maxam, A.M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. **65**: 499-560.

Myers, C.J., A.J.F. Griffiths, and H. Bertrand. 1989. Linear *kalilo* DNA is a *Neurospora* mitochondrial plasmid that integrates into the mitochondrial DNA. Mol. Gen. Genet. **220**: 113-120.

Messing, J. 1983. New M13 vectors for cloning . Methods Enzymol. **101**: 20-78.

Miyashita, S., H. Hirochika, J.E. Ikeda, and T. Hasiba. 1990. Linear plasmid DNAs of the plant pathogenic fungus *Rhizoctonia solani* with unique terminal structures. Mol. Gen. Genet. **220**: 165-171.

Miyoshi, Y.K., S. Ogata, and S. Hayashida. 1986. Multicopy derivative of pock-forming plasmid pSA1 in *Streptomyces azureus*. J. Bacteriol. **168**: 452-454.

Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucl. Acids Res. **14**: 1319-1324.

Modrich, P., and C.C. Richardson. 1975. Bacteriophage T7 deoxyribonucleic acid replication *in vitro*. J. Biol. Chem. **250**: 5515-5522.

Mohan, M., R.J. Meyer, J.B. Anderson, and P.A. Horgen. 1984. Plasmid-like DNAs in the commercially important mushroom genus *Agaricus*. Curr. Genet. **8**: 615-619.

Muller, M.M., S. Ruppert, W. Schaffner, and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature **336**: 544-551.

Morrison, D.A. 1979. Transformation and preservation of competent bacterial cells by freezing. Methods Enzymol. **68**: 326-331.

Nagata, K., R.A. Guggenheimer, and J. Hurwitz. 1983. Adenovirus DNA replication *in vitro*: synthesis of full-length DNA with purified proteins. Proc. Natl. Acad. Sci. USA **80**: 4266-4270.

Niwa, O., K. Sakaguchi, and N. Gunge. 1981. Curing of the killer deoxyribonucleic acid plasmids of *Kluyveromyces lactis*. J. Bacteriol. **148**: 988-990.

Oeser, B. 1988. S2 plasmid from *Zea mays* probably encodes a specific RNA polymerase: an alternative alignment. Nucl. Acids Res. **16**: 8729.

Oeser, B., and P. Tudzynski. 1989. The linear mitochrondrial plasmid pClK1 of the phytopathogenic fungus *Claviceps purpurea* may code for a DNA polymerase and an RNA polymerase. Mol. Gen. Genet. **217**: 132-140.

Ogata, S., Y. Koyama, Y. Sakaki, and S. Hayashida. 1983. Isolation of a linear DNA associated with pock formation in *Streptomyces azureus*. Agric. Biol. Chem. **47**: 2127-2129.

Okayama, H., and P. Berg. 1982. High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2**: 161-170.

Olson, E.R., and S.T. Chung. 1988. Transposon Tn*4556* of *Streptomyces fradiae*: mucleotide sequence of the ends and the target sites. J. Bacteriol. **170**: 1955-1957.

Omer, C.A., and S.N. Cohen. 1984. Plasmid formation in *Streptomyces* - excision and integration of the SLP1 replicon at a specific chromosomal site. Mol. Gen. Genet. **196**: 429-438.

Omer, C.A., and S.N. Cohen. 1986. Structural analysis of plasmid and chromosomal loci involved in site-specific excision and integration of the SLP1 element of *Streptomyces coelicolor*. J. Bacteriol. **166**: 999-1006.

Pabo, C.O., and R.T. Sauer. 1984. Protein-DNA recognition. Ann. Rev. Biochem. **53**: 293-321.

Paillard, M., R.R. Sederoff, and C.S. Levings III. 1985. Nucleotide sequence of the S1 mitochondrial DNA from the S cytoplasm of maize. EMBO J. **4**: 1125-1128.

Pernodet, J.L., J.M. Simonet, and M. Guerineau. 1984. Plasmids in different strains of *Streptomyces ambofaciens*: free and integrated form of plasmid pSAM2. Mol. Gen. Genet. **198**: 34-41.

Pring, D.R., C.S. Levings III, W.W.L. Hu, and D.H. Timothy. 1977. Unique DNA associated with mitochondria in the 'S'-type cytoplasm of male-sterile maize. Proc. Natl. Acad. Sci. USA **74**: 2904-2908.

Raden, B., and L. Rutberg. 1984. Nucleotide sequence of the temperate *Bacillus subtilis* bacteriophage SPO2 DNA polymerase gene L. J. Virol. **52**: 9-15.

Rao, J.K.M., and P. Argos. 1986. A conformational preference parameter to predict helices in integral membrane proteins. Biochim. Biophys. Acta **869**: 197-214.

Reed, C.K., and D.A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res. **13**: 7207-7221.

Rekosh, D.M.K., W.C. Rusell, A.J.D. Bellet, and A.J. Robinson. 1977. Identification of a protein linked to the ends of adenovirus DNA. Cell **11**: 283-295.

Rhodes, P.M., I.S. Hunter, E.J. Friend, and M. Warren. 1984. Recombinant DNA methods for the oxytetracycline producer *Streptomyces rimosus*. Biochem. Soc. Trans. **12**: 586-588.

Rigaud, G, T. Grange, and R. Pictet. 1987. The use of NaOH as transfer solution of DNA onto nylon membrane decreases the hybridization efficiency. Nucl. Acids Res. **15**: 857.

Robinson, A.J., H.B. Younghusband, and A.J.D. Bellett. 1973. A circular DNA-protein complex from adenoviruses. Virology **56**: 54.

Ruan, K.S., and S. W. Emmons. 1984. Extrachromosomal copies of transposon Tc1 in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **81**: 4018-4022.

Ruben, M., S. Bacchetti, and F.L. Graham. 1982. Integration and expression of viral DNA in cells transformed by host range mutants of adenovirus. J. Virol. **41**: 674-685.

Ruben, M., S. Bacchetti, and F.L. Graham. 1983. Covalently closed circles of adenovirus 5 DNA. Nature **301**: 172-174.

Rubin, C.M., and C.W. Schmid. 1980. Pyrimidine-specific chemical reactions useful for DNA sequencing. Nucleic Acid Res. **8**: 4613-4619.

Saiki, R.K., D.H. Gelfrand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**: 487-494.

Sakaguchi, K. 1990. Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. Microbiol. Rev. **54**: 66-74.

Salas, M. 1988. Phages with protein attached to the DNA ends. *In* The Bacteriophages. Vol. 1. *Edited by* R. Calendar. Plenum Press, New York, pp. 169-191.

Salas, M., R.P. Mellado, E. Vinuela, and J.M. Sogo. 1978. Characterization of a protein covalently linked to the 5' termini of the DNA of *Bacillus subtilis*. J. Mol. Biol. **119**: 269-291.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463-5467.

Saunders, G., M.E. Rogers, M.W. Adlard, and G. Holt. 1984. Chromatographic resolution of nucleic acids: application to organisms of industrial importance. Biochem. Soc. Trans. **12**: 694-695.

Schrempf, H. 1985. Genetic instability: amplification, deletion and rearrangement within *Streptomyces* DNA. *In* Microbiology - 1985. *Edited by* L. Leive. American Society for Microbiology, Washington, DC, pp. 436-440.

Seno, E.T., and R.H. Baltz. 1989. Structural organization and regulation of antibiotic biosynthesis and resistance genes in actinomycetes. *In* Regulation of secondary metabolism in actinomycetes. *Edited by* S. Shapiro. CRC Press, Boca Raton, Florida, pp. 1-48.

Skurray, R.A., and P. Reeves. 1973. Characterization of lethal zygosis associated with conjugation in *Escherichia coli* K-12. J. Bacteriol. **113**: 58-70.

Sladkova, I.A. 1982. Circular permutation of DNA molecules of *Streptomyces griseus* Kr actinophage Pg2. Mol. Biol. (Moscow) **16**: 98-103.

Sladkova, I.A. 1986. Physical mapping of IS*281* of *Streptomyces*. Mol. Biol. (Moscow) **20**: 1079-1083.

Smith, C.L., J.G. Econome, A. Schutt, S. Klco, and C.R. Cantor. 1987. A physical map of the *Escherichia coli* K12 genome. Science **236**: 1448-1453.

Stam, J.C., J. Kwakman, M. Meijer, and A.R. Stuitje. 1986. Efficient isolation of the linear DNA killer plasmid of *Kluyveromyces lactis*: evidence for location and expression in the cytoplasm and characterization of their terminally bound proteins. Nucl. Acids Res. **14**: 6871-6884.

Suss, F., and S. Klaus. 1981. Transduction in *Streptomyces hygroscopicus* mediated by the temperate bacteriophage SH10. Mol. Gen. Genet. **181**: 552-555.

Southern, E.M. 1979. Gel electrophoresis of restriction fragments. Methods Enzymol. **69**: 152-176.

Tabor, S., and C.C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**: 4767-4771.

Tabor, S., H. Huber, and C.C. Richardson. 1987. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. J. Biol. Chem. **262**: 16212-16223.

Tamanoi, F. 1986. On the mechanism of adenovirus DNA replication. *In* Adenovirus DNA: the viral genome and its expression. *Edited by* W. Doerfler. Martinus Nijhoff Publishing, Boston, pp. 97-128.

Tamanoi, F., and B.W. Stillman. 1982. Function of the adenovirus terminal protein in the initiation of DNA replication. Proc. Natl. Acad. Sci. USA **79**: 2221-2225.

Thomas, P.S. 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol. **100**: 255-266.

Thompson, J., S. Rae, and E. Cundliffe. 1984. Coupled transcription-translation in extracts of *Streptomyces lividans*. Mol. Gen. Genet. **195**: 39-43.

Tommasino, M., S. Ricci, and C.L. Galeotti. 1988. Genome organization of the killer plasmid pGKl2 from *Kluyveromyces lactis*. Nucl. Acids Res. **16**: 5863-5878.

Turmel, M., G. Bellemare, R.W. Lee, and C. Lemieux. 1986. A linear DNA molecule of 5.9 kilobase-pairs is highly homologous to the chloroplast DNA in the green algae *Chlamydomonas moewusii*. Plant Mol. Biol. **6**: 313-319.

Twigg, A.J., and D. Sherratt. 1980. Trans-complementable copy-number mutants of plasmid ColE1. Nature **283**: 216-218.

Vieira, J., and J. Messing. 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**: 3-11.

Vivian, A. 1971. Genetic control of fertility in *Streptomyces coelicolor* A3(2). Plasmid involvement in the interconversion of UF and IF strains. J. Gen. Microbiol. **69**: 353-364.

Vlcek, C., and V. Paces. 1986. Nucleotide sequence of the late region of *Bacillus* phage ϕ29 completes the 19285 bp sequence of the ϕ29 genome. Comparison with the homologous sequence of phage PZA. Gene **46**: 215-225.

Vosberg, H-P. 1989. The polymerase chain reaction: an improved method for the analysis of nucleic acids. Hum. Genet. **83**: 1-115.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene **33**: 103-119.

Walker, J.E., M. Saraste, M.J. Runswick, and N.J. Gay. 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin kinase and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**: 945-951.

Wharton, R.P., and M. Ptashne. 1986. An α-helix determines the DNA-binding specificity. Trends Biochem. Sci. **11**: 71-73.

Williams, J.G., and P.J. Mason. 1985. Hybridization in the analysis of RNA. *In* Nucleic acid hybridization - A Practical Approach. *Edited by* B.D. Hames and S.J. Higgins. IRL Press, Oxford, Washington DC, pp. 139-178.

Wilson, D.W., and P.A. Meacock. 1988. Extranuclear gene expression in yeast: evidence for a plasmid-encoded RNA polymerase of unique structure. Nucl. Acids Res. **16**: 8097-8112.

Winship, P.R. 1989. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucl. Acids Res. **17**: 1266.

Zheng, Y., R. Zhao, and Y. Zhang. 1982. Genetic evidence for plasmid SQP1 controlling fertility in *Streptomyces qingfengmyceticus*. Acta Genetica Sinica **9**: 8-13.

6. APPENDIX

Streptomyces Codon Usage Table [c]

Amino Acid	Codon	Total	% [a]	% [b]	Amino Acid	Codon	Total	% [a]	% [b]
Phe	TTT	6	0.06	1.8	Tyr	TAT	13	0.13	4.8
	TTC	319	3.3	98.2		TAC	260	2.7	95.2
Leu	TTA	3	0.03	0.3	His	CAT	26	0.27	11.3
	TTG	24	0.25	2.7		CAC	203	2.1	88.7
	CTT	22	0.23	2.5	Gln	CAA	24	0.25	9.1
	CTC	339	3.5	38.4		CAG	239	2.5	90.9
	CTA	4	0.04	0.5	Asn	AAT	16	0.16	5.4
	CTG	491	5	55.6		AAC	282	2.9	94.6
Ile	ATT	7	0.07	2.4	Lys	AAA	6	0.06	2.8
	ATC	268	2.8	93.4		AAG	211	2.2	97.2
	ATA	12	0.12	4.2	Asp	GAT	36	0.37	6.2
Met	ATG	165	1.7	100		GAC	543	5.6	93.8
Val	GTT	17	0.18	2.2	Glu	GAA	93	0.96	17.5
	GTC	434	4.5	57.8		GAG	439	4.5	82.5
	GTA	14	0.14	1.9	Cys	TGT	8	0.08	9.8
	GTG	286	2.9	38.1		TGC	74	0.76	90.2
Ser	TCT	4	0.04	0.7	Trp	TGG	184	1.9	100
	TCC	214	2.2	38.3	Arg	CGT	64	0.66	8.2
	TCA	18	0.18	3.2		CGC	331	3.4	42.2
	TCG	169	1.7	30.3		CGA	48	0.49	6.1
	AGT	16	0.16	2.9		CGG	264	2.7	33.6
	AGC	137	1.4	24.6		AGA	16	0.16	2
Pro	CCT	26	0.27	4.3		AGG	62	0.64	7.9
	CCC	258	2.7	43.3	Gly	GGT	82	0.84	8.6
	CCA	13	0.13	2.2		GGC	612	6.3	64.2
	CCG	299	3.1	50.2		GGA	84	0.86	8.8
Thr	ACT	13	0.13	2.2		GGG	175	1.8	18.4
	ACC	408	4.2	68.6					
	ACA	21	0.22	3.5	Stop	TAA	1	0.01	3.7
	ACG	153	1.6	25.7		TAG	8	0.08	29.6
Ala	GCT	34	0.35	2.9		TGA	18	0.18	66.7
	GCC	672	6.9	56.7					
	GCA	50	0.51	4.2					
	GCG	429	4.4	36.2					

a. The percentage of usage of a codon among the total codons.

b. The percentage of usage of a codon among the total codons for a particular amino acid.

c. This table is a summary of the total codons used in the following *Streptomyces* genes: *act* III (Hallam *et al.*, 1988); *afsB* (Horinouchi *et al.*, 1986); *aml* (Long *et al.*, 1987); *aph* (Thompson and Gray, 1983); APH(6) (Distler *et al.*, 1987); *amy* (Hoshiko *et al.*, 1987); *bapA* (Anzai *et al.*, 1987); *bar* (Davies *et al.*, 1987); β-galactosidase (Eckhardt *et al.*, 1987); *carB* (Epp *et al.*, 1987); *casA* (Nakai *et al.*, 1988); *choA* (Ishizaki *et al.*,

1989); *dagA* (Buttner *et al.*, 1987); *endoH* (Robbins *et al.*, 1984); *ermE* 1 and *ermE* 2 (Uchiyama and Weisblum, 1985; Bibb *et al.*, 1985a, 1986); *galT*, *galE* and *galK* (Adams *et al.*, 1988); *hyg* (Zalacain *et al.*, 1986); IPNS (Leskiw *et al.*, 1988); IPNS (Weigel *et al.*, 1988); *mel* (Bernan *et al.*, 1985); *mmr* (Neal and Chater, 1987); *tet* 347 (Reynes *et al.*, 1988); tyrosinase (Huber *et al.*, 1985); *tsr* (Bibb *et al.*, 1985b); *tylF* (Fishman *et al.*, 1987); *vph* (Bibb *et al.*, 1985b); xp55 and p49 proteins (Burnett *et al.*, 1987; Eckhardt, 1987). The sequences were obtained from the GenBank data base through the Bionet National Computing Resource (Intelligenetics, Mountain View, California, USA) and analyzed with the Bionet computer network system.

REFERENCES

Adams, C.W., J.A. Fornwald, F.J. Schmidt, M. Rosenberg, and M.E. Brawner. 1988. Gene organization and structure of the *Streptomyces lividans gal* operon. J. Bacteriol. **170**: 203-212.

Anzai, H., T. Murakami, S. Imai, A. Satoh, D. Nagaoka, and C.J. Thompson. 1987. Transcriptional regulation of bialaphos biosynthesis in *Streptomyces hygroscopicus*. J. Bacteriol. **169**: 3482-3488.

Bernan, V., D. Filpula, W. Herber, M. Bibb, and E. Katz. 1985. The nucleotide sequence of the tyrosinase gene from *Streptomyces antibioticus* and characterization of the gene product. Gene **37**: 101-110.

Bibb, M.J., G.R. Janssen, and J.M. Ward. 1985a. Cloning and analysis of the promoter region of the erythromycin-resistance gene (*ermE*) of *Streptomyces erythraeus*. Gene **38**: 215-226.

Bibb, M.J., G.R. Janssen, and J.M. Ward. 1986. Cloning and analysis of the promoter region of the erythromycin-resistance gene (*ermE*) of *Streptomyces erythraeus*. Gene **41**: E357-E368.

Bibb, M.J., J.M. Ward, and S.N. Cohen. 1985b. Nucleotide sequences encoding and promoting expression of three antibiotic resistance genes indigenous to *Streptomyces*. Mol. Gen. Genet. **199**: 26-36.

Burnett, W.V., J. Henner, and T. Eckhardt. 1987. The nucleotide sequence of the gene coding for xp55, a major secreted protein from *Streptomyces lividans*. Nucl. Acids Res. **15**: 3926.

Buttner, M.J., I.M. Fearnley, and M.J. Bibb. 1987. The agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2): nucleotide sequence and transcriptional analysis. Mol. Gen. Genet. **209**: 101-109.

Distler, J., C. Braun, A. Ebert, and W. Piepersberg. 1987. Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: analysis of a central region including the major resistance gene. Mol. Gen. Genet. **208**: 204-210.

Eckhardt, T. 1987. Unpublished. The sequence data was obtained from the GenBank data base through the Bionet National Computing Resource, Intelligenetics, Mountain View, California, USA.

Eckhardt, T., J. Strickler, L. Gorniak, W.V. Burnett, and L.R. Fare. 1987. Characterization of the promoter, signal sequence, and amino terminus of a secreted β-galactosidase from "*Streptomyces lividans*". J. Bacteriol. **169**: 4249-4256.

Fishman, S.E., K. Cox, J.L. Larson, P.A. Reynolds, E.T. Seno, W.K. Yeh, R. van Frank, and C.L. Hershberger. 1987. Cloning genes for the biosynthesis of a macrolide antibiotic. Proc. Natl. Acad. Sci. USA **84**: 8248-8252.

Epp, J.K., S.G. Burgett, and B.E. Schoner. 1987. Cloning and nucleotide sequence of a carbomycin-resistance gene from *Streptomyces thermotolerans*. Gene **53**: 73-83.

Hallam, S., F. Malpartida, and D.A. Hopwood. 1988. DNA sequence, transcription and deduced function of a gene involve in polyketide synthesis in *Streptomyces*. Gene **74**: 305-320.

Horinouchi, S., H. Suzuki, and T. Beppu. 1986. Nucleotide sequence of *afsB*, a pleiotropic gene involved in secondary metabolism in *Streptomyce coelicolor* A3(2) and "*Streptomyces lividans*". J. Bacteriol. **168**: 257-269.

Hoshiko, S., O. Makabe, C. Nojiri, K. Katsumata, E. Satoh, and K. Nagaoka. 1987. Molecular cloning and characterization of the *Streptomyces hygroscopicus* α-amylase gene. J. Bacteriol. **169**: 1029-1036.

Huber, M., G. Hintermann, and K. Lerch. 1985. Primary structure of tyrosinase from *Streptomyces glaucescens*. Biochem. **24**: 6038-6044.

Ishizaki, T., N. Hirayama, H. Shinkawa, O. Nimi, and Y. Murooka. 1989. Nucleotide sequence of the gene for cholesterol oxidase from a *Streptomyces* sp. J. Bacteriol. **171**: 596-601.

Leskiw, B.K., Y. Aharonowitz, M. Mevarech, S. Wolfe, L.C. Vining, D.W.S. Westlake, and S.E. Jensen. 1988. Cloning and nucleotide sequence determination of the isopenicillin N synthetase gene from *Streptomyces clavuligerus*. Gene **62**: 187-196.

Long, C.M., M.J. Virolle, S.Y. Chang, S. Chang, and M.J. Bibb. 1987. α-amylase gene of *Streptomyces limosus*: nucleotide sequence, expression motifs, and amino acid sequence homology to mammalian and invertebrate α-amylase. J. Bacteriol. **169**: 5745-5754.

Neal, R.J., and K.F. Chater. 1987. Nucleotide sequence analysis reveals similarities between proteins determining methylenomycin A resistance in *Streptomyces* and tetracycline resistance in eubacteria. Gene **58**: 229-241.

Reynes, J.P., T. Calmels, D. Drocourt, and G. Teraby. 1988. Cloning, expression in *Escherichia coli* and nucleotide sequence of a tetracycline-resistance gene from *Streptomyces rimosus*. J. Gen. Microbiol. **134**: 585-598.

Robbins, P.W., R.B. Trimble, D.F. Wirth, C. Hering, F. Maley, G.F. Maley, F. Das, B.W. Gibson, N. Royal, and K. Biemann. 1984. Primary structure of the *Streptomyces* enzyme endo-β-N-acetylglucosaminidase H. J. Biol. Chem. **259**: 7577-7583.

Thompson, C.J., and G.S. Gray. 1983. Nucleotide sequence of a streptomyete aminoglycoside phosphotransferase gene and its relationship to phosphotransferase encoded by resistance plasmids. Proc. Natl. Acad. Sci. USA **80**: 5190-5194.

Thompson, C.J., N.R. Movva, R. Tizard, R. Crameri, J.E. Davies, M. Lauwereys, and J. Botterman. 1987. Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygroscopicus*. EMBO J. **6**: 2519-2523.

Uchiyama, H., and B. Weiblum. 1985. N-methyl transferase of *Streptomyces erythraeus* that confers resistance to the macrolide-lincosamide-streptogramin B antibiotics: amino acid sequence and its homology to cognate R-factor enzymes from pathogenic bacilli and cocci. Gene **38**: 103-110.

Weigel, B.J., S.G. Burgett, V.J. Chen, P.L. Skatrud, C.A. Frolik, S.W. Queener, and T.D. Ingolia. 1988. Cloning and expression in *Escherichia coli* of isopenicillin N synthetase genes from *Streptomyces lipmanii* and *Aspergillus nidulans*. J. Bacteriol. **170**: 3817-3826.

Zalacain, M., A. Gonzalez, M.C. Guerrero, R.J. Mattaliano, F. Malpartida, and A. Jimenez. 1986. Nucleotide sequence of the hygromycin B phosphotransferase gene. Nucl. Acids Res. **14**: 1565-1581.