National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

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.

Canada

UNIVERSITY OF ALBERTA

SUBCLONING AND SEQUENCING OF THE *bldD* GENE OF *STREPTOMYCES COELICOLOR*

BY

FARZANA DAMJI

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN

PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF MICROBIOLOGY

EDMONTON, ALBERTA

Fall, 1995

National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

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.

L'AUTEUR A ACCORDE UNE LICENCE IRREVOCABLE ET NON EXCLUSIVE PERMETTANT A LA BIBLIOTHEQUE NATIONALE DU CANADA DE REPRODUIRE, PRETER, DISTRIBUER OU VENDRE DES COPIES DE SA THESE DE QUELQUE MANIERE ET SOUS QUELQUE FORME QUE CE SOIT POUR METTRE DES EXEMPLAIRES DE CETTE THESE A LA DISPOSITION DES PERSONNE INTERESSEES.

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 CONSERVE LA PROPRIETE DU DROIT D'AUTEUR QUI PROTEGE SA THESE. NI LA THESE NI DES EXTRAITS SUBSTANTIELS DE CELLE-CI NE DOIVENT ETRE IMPRIMES OU AUTREMENT REPRODUITS SANS SON AUTORISATION.

ISBN 0-612-06458-1

Canada

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR: FARZANA DAMJI

TITLE OF THESIS: SUBCLONING AND SEQUENCING OF THE *bldD* GENE OF *STREPTOMYCES COELICOLOR*

DEGREE: MASTER OF SCIENCE

YEAR THIS DEGREE GRANTED: 1995

Permission is hereby 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 all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Farzana Damji

#101 10624 - 78 Avenue

Edmonton, Alberta

T6E 1P5

Date: Aug. 24, 1995

Struggle is the meaning of life.
Defeat or victory is in the hands of Allah.
But struggle itself is man's duty, and ought to be his joy.

Aga Khan III

THE 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 CLONING AND SEQUENCING OF THE *bldD* GENE OF *STREPTOMYCES COELICOLOR* submitted by Farzana Damji in partial fulfillment of the requirements for the degree of Master of Science.

Dr. B.K. Leskiw - Supervisor

Dr. K.L. Roy

Dr. S.E. Jensen

Dr. J. Bell

Date: July 24, 1995

ABSTRACT

The nucleotide sequence of a 3.4kb fragment of DNA that was able to restore both antibiotic production and sporulation to *bldD* mutants of *S. coelicolor* revealed two truncated open reading frames (ORFs) and one complete ORF capable of encoding a protein. When an 2.2kb *Sph*I/*Hind*III subclone from the 3.4kb fragment, containing the putative *bldD* ORF and one truncated ORF, was introduced into a *bldD* mutant on the *att*-site integrating *E. coli-Streptomyces* shuttle vector pSET152, the recombinant plasmid was able to complement the mutation. The ORF corresponding to the *bldD* gene is 504 nucleotides and translation of the ORF would result in a 18167 Dalton protein.

The predicted amino acid sequence of the ORF was compared to other proteins in the databases but did not show significant similarity to known proteins. However, analysis of the amino acid sequence revealed a potential helix-turn-helix motif similar to those found in the LysR family of regulators, suggesting it could be a DNA-binding protein.

ACKNOWLEDGMENT

I would like to thank Dr. Brenda Leskiw for her supervision. Her guidance and expertise in molecular biology helped me through some challenging problems. Thanks are also due to the remaining members of my supervisory committee, Dr. S.E. Jensen and Dr. K. Roy for their suggestions regarding my research project. Special thanks is owed to Dr. M. Pickard for his support and guidance through some difficult times and for the opportunity to complete my work in his laboratory. To Dr. Ashish Paradakar and Dr. Roy Mosher, I would like to extend my gratitude for many discussions and for sharing their expertise in molecular genetics. Special thanks are due to Richard Mah and Kevin Wenzel for their technical support.

My life has been greatly enriched by the friendship of Karen, Manisha, Linda, and Wendy, especially for support and encouragement during my most trying moments. Also, acknowledgments is due to many other friends, too numerous to mention individually.

To my parents, words cannot express my gratitude for all the sacrifices they have made on my behalf, without which none of what I achieved would have been possible. I would like also to thank my brother for his support and computer expertise, my brother-in-law for helpful discussions and pep talks, and my parents-in-law for their continued encouragement.

Lastly, I would like to thank my husband, Karim, for his support, understanding, empathy and patience. Without him, this degree would not have been possible.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF ABBREVIATIONS

CDA	Calcium-dependent antibiotic
CsCl	Cesium chloride
EDTA	ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide
LB	Luria-Bertani medium
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
RF	Replicative form
rpm	revolutions per minute
SDS	Sodium dodecyl sulfate
SSC	Standard Saline-Citrate
TAE	Tris-acetate-EDTA buffer
TBE	Tris-borate-EDTA buffer
TE	Tris-EDTA buffer
TSB	Trypticase soy broth
YEME	Yeast extract-malt extract

1.0 INTRODUCTION

1.1 Streptomycetes - General Overview

Streptomycetes are aerobic, soil-dwelling, Gram-positive bacteria that have DNA with approximately 70% G+C content (Chater & Merrick, 1979). These bacteria grow vegetatively as branching hyphae which form a complex substrate mycelium that utilizes extracellular hydrolytic enzymes to break down organic debris typically found in the soil (Chater, 1993). An unknown stimulus, possibly nutrient depletion, then causes the formation of aerial hyphae from which long chains of spores form. Usually, at a time coinciding with aerial mycelium formation, streptomycetes produce secondary metabolites including many useful antibiotics. Nearly 70% of the known antibiotics are produced by this one genus and these compounds exhibit great diversity in their chemical structure and in their specificity of action. These characteristics make these compounds valuable in medicine, veterinary medicine and agriculture and consequently commercially (Chater, 1993). Some examples include polyketide antibiotics, such as tetracyclines, macrolide antibiotics and anthracyclines, and ß-lactam antibiotics, such as the penicillins and cephamycins (Hopwood and Sherman, 1990). Other useful products include adriamycin, an antitumor drug, immunosuppressants like FK506 and rapamycin and the herbicide bialaphos (Champness and Chater, 1994). These applications stimulate an interest in streptomycete biology, including secondary metabolism and morphological differentiation (Hodgson, 1992).

1.2 Differentiation in *Bacillus subtilis*

Differentiation is not unique to *Streptomyces spp.* Other genera, such as *Bacillus* and *Clostridium*, also sporulate in response to nutrient stress. Sporulation in

the genus *Bacillus*, and more specifically in *Bacillus subtilis*, is the best studied example of prokaryotic differentiation. Although *Bacillus* and *Streptomyces* are similar in some aspects of differentiation, there are morphological differences between the two. Despite these differences, insightful analogies between the developmental processes of *Bacillus* and *Streptomyces* may prove useful.

B. subtilis is a Gram-positive, soil-inhabiting organism, which forms endospores in response to nutritional deprivation. Experiments have also shown that *B. subtilis* sporulated in a normal growth medium, containing excess ammonium ions, glucose and phosphate, if the synthesis of purine nucleotides was partially reduced either by a slow supply of purine precursors in purine auxotrophs (Freese *et al.*, 1979), or by the addition of inhibitors of purine synthesis (Lopez *et al.*, 1979; Mitani *et al.*, 1977). For example, decoyinine, an inhibitor of GMP synthetase, induced sporulation. Nutritional starvation was also accompanied by a decrease in GTP (and GDP) levels, resulting in a depletion of amino acids, which in turn produced a stringent response (i.e. a decrease in the rate of RNA synthesis, an increase in ppGpp and pppGpp, and a decrease in guanine nucleotides). Lopez *et al.* (1981) showed that the stringent response caused sporulation by demonstrating that a relaxed strain, lacking this response, was unable to sporulate.

A sigma factor is a single subunit component of the bacterial RNA polymerase holoenzyme that is required at the stage of initiation for recognizing the promoter (Lewin, 1990). The other component of RNA polymerase is the core enzyme which is responsible for elongating the RNA chain. The core enzyme can be directed to recognize promoters with different consensus sequences by alternative sigma factors. *B. subtilis* contains a single major sigma factor, and a variety of minor sigma factors, some of which are activated when sporulation is initiated. A cascade

of five σ factors have been implicated in the temporal and spatial regulation of gene expression during *Bacillus* sporulation. The earliest-acting sigma factor is σ^H, encoded by *spoOH*. This factor is present in vegetative cells and transcribes genes throughout vegetative growth, and at the onset of sporulation or slightly later (Tatti *et al.*, 1989).

The synthesis of the next σ factor, σ^F, appears to be controlled by σ^H (Wu *et al.*, 1989). σ^F is encoded by *spoIIAC*, one of three genes belonging to the *spoIIA* operon, with the other two genes, *spoIIAA* and *spoIIAB*, regulating *spoIIAC* (Schmidt *et al.*, 1990). The timing of transcription of *spoIIA* suggested that σ^F is synthesized prior to septation (Gholamhoseinian *et al.*, 1989). Despite the production of σ^F before septum formation, its activity is confined to the forespore (Margolis *et al.*, 1991).

Formation of the septum appears to coincide with the next σ factor, σ^E, which is synthesized as an inactive precursor protein called pro- σ^E (Labell *et al.*, 1987) that is encoded in the two cistron *spoIIG* operon (Jonas *et al.*, 1988). It should be noted however, that the *spoIIG* promoter is transcribed by σ^A RNA polymerase, the major form of RNA polymerase in vegetative cells (Kenney *et al.*, 1988) and that the active σ^E RNA polymerase is produced only in the mother cell. SpoIIGA is required for the accumulation of σ^E (Jonas *et al.*, 1988) and is thought to be an aspartic protease which activates σ^E (Straiger *et al.*, 1988). It has been suggested that the processing of σ^E by SpoIIGA is triggered by correct formation of the spore septum which requires *spoIIA* and *spoIIE*. Mutations in *spoIIE* and *spoIIA* have been shown to cause abnormal septal structures. The *spoIIE* products could be structural components of the septum, involved in the maintenance of its flexibility, which is needed for the subsequent engulfment step (Straiger *et al.*, 1988). Since mutations in

spoIIAC (which encodes SpoIIAC) prevent *spoIID* transcription but do not block *spoIIG* expression and since transcription of *spoIID* is dependent on the presence of σ^E, the *spoIIAC* gene product is probably necessary for pro-σ^E processing (Straiger *et al.*, 1988).

The *spoIIIG* gene product, σ^G, is the next σ factor in the cascade (Sun *et al.*, 1989; Karmazyn-Campelli *et al.*, 1989). The *spoIIIG* gene is controlled by σ^F and is induced shortly after septation, but the transcription of its target genes is delayed until the forespore is fully engulfed. It has been shown that σ^G activation requires the products of the *spoIIIA* operon (Kroos and Cutting, 1994), the *spoIID* (Straiger, 1992), *spoIIB* and *spoVG* (Margolis *et al.*, 1993) genes. Interestingly, *spoIID* and *spoIIIA* are transcribed by σ^E in the mother cell, but the products are involved in activating σ^G in the forespore, demonstrating that a signal transduction pathway exists from the mother cell to the forespore (Kroos and Cutting, 1994).

The final sigma factor, σ^K, is active in the mother cell compartment. σ^K is responsible for transcribing many genes involved in the synthesis and assembly of the spore coat, including the *cot* genes, which encode spore coat proteins, and a DNA binding protein encoded by the *gerE* gene that regulates the transcription of some of the *cot* genes. The structural gene of σ^K (*sigK*) is a composite gene made from two truncated genes called *spoIIIC* and *spoIVCB*, which are joined together by a site-specific recombination event into a single cistron (Straiger *et al.*, 1989). σ^K is activated from an inactive precursor form and this activation also appears to be delayed until a critical point in the developmental cycle. The signal for the accumulation of σ^K is directly under the control of σ^G(Cutting *et al.*, 1991). None of the signaling processes are fully understood, but since each sigma factor is produced

at critical points of the developmental cycle, it is reasonable that they play an important role in developmental regulation in *Bacillus*.

1.3 *Streptomyces* Genetics

Streptomyces coelicolor A3(2) is the genetically most studied streptomycete. The chromosome of *S. coelicolor* has been estimated to be approximately 8Mb by pulse-field gel electrophoresis (PFGE). It was previously determined that the physical map of *S. coelicolor* was in agreement with the circular genetic map (Kieser *et al.*, 1992). Recent evidence, however, has shown that although the genetic map is circular, the chromosomal DNA is linear with terminal proteins (Lin *et al.*, 1993). A more detailed analysis showed that the apparent linkage between the *AseI*-A and -J fragments of this chromosome was false. The contiguity of the two fragments, which was deduced from their hybridization to the same *Ssp*I fragment was probably caused by the presence of terminal inverted repeats on the A and J fragments.

S. coelicolor contains two plasmids, SCP1 and SCP2, which promote *in vivo* gene exchange. SCP1 is a giant linear plasmid of 350kb, which was identified by the infectious transfer of the ability to produce and be resistant to an antibiotic (Vivian, 1971), that was later identified as methylenomycin (Wright and Hopwood, 1976). SCP1 is also found to code for several spore-associated proteins which may be involved in spore formation, such as SapC, SapD and SapE (Willey *et al.*, 1991). This plasmid is known to exist in *S. coelicolor* in several states i.e. as an autonomously replicating plasmid, as an autonomously replicating plasmid containing a chromosomal fragment or integrated into the chromosome (Kinashi *et al.*, 1992). SCP2 is a 31kb covalently, closed, circular plasmid present in low copy number. It was shown to be a second fertility factor independent of SCP1 (Bibb *et al.*, 1977). SCP2* is a high fertility variant of SCP2 present in one or two copies per

chromosome. Several useful cloning vectors have been constructed from SCP2* that are able to maintain large DNA inserts with stable inheritance at a low copy number, and without structural rearrangements, in *Streptomyces* hosts (Lydiate *et al.*, 1985). A variety of other low and high copy number plasmid vectors and single copy bacteriophage vectors (e.g. vectors derived from the temperate phage, ØC31 (Lomovskaya *et al.*, 1980)) have also been useful for *in vitro* cloning. This, in conjunction with PEG induced transformation or transfection of protoplasts, provides a useful tool for genetic analysis (Hopwood, 1987).

The linkage map of *S. coelicolor* is circular and incorporates nearly 150 genes (Hopwood and Kieser, 1990), some of which are clustered. Every biosynthetic pathway for a secondary metabolite so far examined is organized in a single gene cluster in which regulatory and self-resistance genes are also contained. *S. coelicolor* produces four antibiotics, actinorhodin, undecylprodigiosin, methylenomycin and a calcium-dependent antibiotic. All of the biosynthetic genes for these antibiotics were found to be clustered. For example, in *S. coelicolor*, the actinorhodin biosynthetic genes, including a regulatory region, are clustered in a chromosomal region of about 26kb (Malpartida *et al.*, 1986). Clustering is also seen in the genes responsible for undecylprodigiosin production (Rudd and Hopwood, 1980) as the entire biosynthetic cluster has been cloned on a DNA fragment of approximately 35.7kb (Malpartida *et al.*, 1990). Methylenomycin genes have also been found to lie adjacent to each other (Wright and Hopwood, 1976). With the exception of methylenomycin, all of these gene clusters have been mapped to the chromosome, and the cluster for the calcium-dependent antibiotic is the only one that has not been cloned.

The production of secondary metabolites often coincides with the onset of morphological differentiation. The *bld* mutants (discussed in detail in section 1.6)

provide genetic evidence that secondary metabolite production is regulated, in part, by the same mechanisms that control differentiation, since almost all mutants were found to be deficient in both antibiotic production and formation of aerial mycelium (Champness and Chater, 1994). Although regulation of developmental genes is being studied in *Streptomyces*, it is not as well understood as in *B. subtilis* . Among several levels of regulation of differentiation that appear to be functioning in these bacteria, transcriptional regulation involving a heterogeneous set of sigma factors has been shown to play an important role. As in *Bacillus* the different sigma factors associate themselves with the DNA-dependent RNA polymerase, which can exist in many holoenzyme forms. This confers on a common-core RNA polymerase moiety, the ability to respond to a particular set of transcription initiation signals.

The majority of vegetatively expressed genes in *Streptomyces* possess promoters which resemble *E. coli* consensus sequences. Other classes of promoters required for the expression of specialized sets of genes are recognized by different RNA polymerase holoenzyme forms. Evidence for the heterogeneity of RNA polymerases was first provided by Westphaling *et al* (1985) when they characterized two alternative RNA polymerase σ factors, σ^{35} and σ^{49}, using two well characterized promoters from *B. subtilis* as templates for *in vitro* transcription studies. σ^{35} was found to be similar in its specificity to the predominant form of RNA polymerase in *B. subtilis*, σ^{55} (now σ^{43}), which is encoded by the *rpoD* (*sigA*) gene (Wang and Doi, 1986). σ^{49} on the other hand was similar to the holoenzyme form, $E\sigma^{B}$ (σ^{37}), only observed in *B. subtilis*, where it is involved in the transcription of developmentally regulated genes. Buttner *et al.* (1988) identified yet another σ factor, σ^{28}, which is responsible for the transcription of one of the four promoters found in the *dagA* gene, which encodes an extracellular agarase. The concept of RNA polymerase heterogeneity was advanced further when the *hrd* genes

(homologues of the *rpoD* gene of *E. coli*) were cloned and sequenced. Using an oligonucleotide probe designed from the major sigma factors of *E. coli* and *B. subtilis*, four *rpoD* homologues in *S. coelicolor* were cloned i.e. *hrdA, hrdB, hrdC* and *hrdD* (Tanaka *et al.*, 1988). Independent experiments in which the *Myxococcus xanthus rpoD* gene was used as a probe, also resulted in cloning *hrdB, hrdC* and *hrdD* genes from *S. coelicolor* (Buttner *et al.*, 1990). Gene disruption experiments revealed that while *hrdA, hrdC* and *hrdD* are dispensable for cell viability, *hrdB* is essential. These observations lead to the conclusion that the RNA polymerase containing σ^{HrdB} is the major form of RNA polymerase in *S. coelicolor* (Brown *et al.*, 1992), consistent with the observation that of the four *hrd* genes, only *hrdB* is present in all species of *Streptomyces* examined so far. SDS-PAGE analysis of RNA polymerase preparations, showing σ^{hrdB} RNA polymerase to be the most prevalent form of RNA polymerase in *S. coelicolor*, also supports this conclusion. Three *hrd* genes homologous to the *rpoD* gene of *E. coli* have also been found in the genome of *Streptomyces griseus* (Marcos *et al.*, 1995). Two of the *hrd* genes, *hrdB* and *hrdD*, are very similar to the homologous *hrdB* and *hrdD* genes of *S. coelicolor*. The *hrdD* gene is linked to a gene encoding a N-acetyltransferase that also occurs in *S. coelicolor*. Transcription studies showed that the *hrdB* gene of *S. griseus* is actively expressed during growth in phosphate-rich medium but not under sporulation conditions. This suggests that it may provide a σ factor for transcription of primary metabolism genes. On the other hand, *hrdD* was transcribed efficiently after nutritional shiftdown which triggers sporulation and lower expression of *hrdD* was observed in rich medium. This suggests that the *hrdD* gene might provide a σ factor required to trigger expression of sporulation genes. The third *hrd* gene, *hrdT*, was also found in *S. griseus* but it has shown no close homology to known *hrd* genes. Its

expression was found to be poor in both phosphate-rich medium and under sporulation conditions.

Consider the analogy between *Streptomyces* and *Bacillus*: the *rpoD* gene, a homolog of σ^{hrdB}, plays a role in vegetative transcription and is crucial in initiating sporulation in *B. subtilis*, suggesting an equivalent role for σ^{hrdB} in *S. coelicolor*. It would be interesting to determine, what function, if any, σ^{hrdB} has in controlling sporulation initiation or antibiotic synthesis in *Streptomyces*. The σ factor, σ^{35}, that Westphaling *et al.* (1985) described, may be a product of another *hrd* gene.

Genetic analysis in *S. coelicolor* has identified eight genes (*whiA* to *whiI*) involved in the development of aerial hyphae into spores. Of these genes, *whiG*, a gene dispensable for growth but needed for the earliest stages of spore formation, was determined to encode an RNA polymerase σ factor which was very similar to the motility factor of *B. subtilis* (σ^{28} or σ^{D}) (Chater *et al.*, 1989). The introduction of many copies of a σ^{28}-dependent promoter from *B. subtilis* into *S. coelicolor* reduced sporulation, suggesting the σ^{whiG} was being partially sequestered from its natural products (Chater *et al.*, 1989). It was also shown that a sporulation-deficient phenotype was caused by a high copy number of a σ^{D}-dependent promoter, $P_{28\text{-}1}$, of *B. subtilis* (Tan and Chater, 1993). Two small DNA fragments from *S. coelicolor*, both containing promoters with sequences resembling those of σ^{D} dependent promoters, also inhibited sporulation when introduced on a high copy number plasmid. Therefore, the amount of the *whiG* gene product appears to be a limiting factor for initiation of sporulation. The two promoters, identified as P_{TH4} and P_{TH270}, were found to activate the *xylE* reporter gene in a *whiG*$^{+}$ strain but failed to do so in a *whiG* mutant. In surface grown cultures, the promoters demonstrated little or no activity during early vegetative growth. Their maximum activities were

detected when aerial mycelium was abundant. Since *whiG* mutants are blocked in the transition of aerial hyphae into spores, with apparently normal vegetative growth, these results are conceivable (Tan and Chater, 1993).

Recently, two other sigma factors were identified from *S. coelicolor* (Lonetto *et al.*, 1994). One factor, σ^F, is encoded by the *sigF* gene and is required for normal spore maturation. The other factor, σ^E, is encoded by *sigE* . The sequence of σ^E showed greatest similarity to sequences of seven other proteins: *Myxococcus xanthus* CarQ, *Pseudomonas aeruginosa* AlgU, *Pseudomonas syringae* HrpL, *E. coli* σ^E, *Alcaligenes eutrophus* CnrH, *E. coli* FecI, and *B. subtilis* SigX, a protein of unknown function. These eight proteins belong to a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions.

Formation of endospores in *Bacillus* and sporulation in *Streptomyces*, despite differences, may be homologous in their underlying regulation (Westphaling *et al.*, 1985). However, this homology may prove to be limited. It appears that sporulation specific σ factors are involved in both processes, but σ^{whiG}, has the same promoter specificity as σ^D, which plays no part in sporulation in *B. subtilis*, and is found to be different from σ^H, which is needed at the earliest stages of sporulation in *B. subtilis*. A regulatory cascade similar to the one controlling sporulation in *Bacillus* has not been identified in *Streptomyces*, suggesting that the central regulation of sporulation may have evolved separately in these two genera and may prove to be very different from each other. On the other hand, it is interesting to note that as in *whiG* mutants of *S. coelicolor*, σ^D mutants of *B. subtilis* form filaments (Helmann *et al.*, 1988) and therefore it is possible that both σ factors are involved in transcribing genes for autolysins (Chater *et al.*, 1989). It is also possible that a progenitor of actinomycetes had a σ^{WhiG}/σ^D-like protein which fragmented these filaments into single motile

cells but evolution of aerially borne spores and of life in a dry environment led to the loss of motility characteristic of streptomycetes. Also, in streptomycetes, sporulation is a mechanism of dispersal and therefore may be linked to the motility factor (Chater *et al.*, 1989).

1.4 Life Cycle of *Streptomyces coelicolor* A3(2)

The developmental cycle of *S. coelicolor* begins with a spore, which germinates under appropriate conditions. Germ tubes growing out from the spores develop into young vegetative hyphae, which continue to grow by hyphal extension and branching, forming a matted substrate mycelium (Chater & Merrick, 1979). Vegetative septa occur infrequently and therefore hyphal compartments contain many copies of the genome.

As colonies age, the parts farthest from the advancing edge accumulate various storage materials such as glycogen, lipids and polyphosphate, which are produced in many bacteria when a nutrient limitation slows growth. After a day or two, aerial hyphae grow vertically into the air giving the colonies a white and hairy appearance (Chater, 1993). A reduction in macromolecular synthesis is observed preceding differentiation (Granozzi *et al.*, 1990) and seems to involve the reuse of material from the substrate mycelium. Therefore, as aerial hyphae form, many cells in the substrate mycelium die. The aerial layer is also hydrophobic enabling it to grow and function in a nonaqueous environment (Chater and Merrick, 1979). It requires osmotic pressure to drive extension into the air, which may be supplied by the solubilization of macromolecules such as glycogen (Chater, 1989a). Glycogen granules were observed in hyphae in the air-agar interface region of colonies that were undergoing aerial mycelium formation, but not in aerial hyphae themselves (termed phase I) (Plaskitt and Chater, 1995). In order to extend vertically away from

the surface of the colony, aerial hyphae also need directionality, and this may be influenced by spore-associated proteins (Saps) (Willey *et al.*, 1991). Five Saps, which appear with the onset of aerial mycelium formation, have been identified and are designated SapA, -B, -C, -D and -E. They evidently contribute to the surface layer of the mature spores and to the surface layer of the aerial hyphae from which the spores are derived. Consequently, with the possible aid of Saps and storage compounds, aerial hyphae emerge (Willey *et al.*, 1991; Willey *et al.*, 1993; Guijjaro *et al.*, 1988).

Eventually the extension of aerial hyphae ceases and unlike substrate mycelium, regularly spaced and synchronous septa are formed, which subdivide the hyphal tips into many unigenomic compartments. New glycogen deposits form in the sporulating compartment and may be a temporary sink for carbon metabolites, facilitating their transport from the lower part of aerial hyphae (Brana *et al.*, 1986). This accumulation of glycogen was specifically observed in aerial hyphal tips and was termed phase II (Plaskitt and Chater, 1995). In addition to providing a local source for carbon and energy for sporulation, the formation of glycogen during sporulation may reduce the osmotic pressure of the spore compartments, leading to efflux of water. This could balance the partial displacement of cytoplasmic volume by the developing sporulation septa. Other storage materials, such as trehalose, spherical electron-transparent bodies thought to contain lipids, and electron-dense granules thought to contain polyphosphate, also appear in the sporulating parts of aerial hyphae (Brana *et al.*, 1986). During the rounding up of spores, glycogen deposits disappear, causing an increase in osmotic pressure, thereby aiding in the shape change (Brana *et al.*, 1986; Plaskitt and Chater, 1995). A pigment is also deposited in the wall.

In *B. subtilis* there is evidence that nutrient limitation, which produces a stringent response, decreases the guanine nucleotide pool and triggers sporulation (Lopez *et al.*, 1981). Inhibitors of purine synthesis, such as decoyinine, also cause sporulation (Mitani *et al.*, 1977). In an unclassified *Streptomyces sp.*, it was observed that decoyinine also stimulated the formation of aerial mycelium(Ochi, 1986a; Ochi, 1986b). Furthermore in *S. griseus*, a downshift in nutrients, followed by a drop in GTP levels and an increase in ppGpp, induced sporulation (Ochi, 1987), and a stringent response also led to increased levels of ppGpp in *S. coelicolor* (Ochi, 1986b). Although evidence seems to support that sporulation is induced under conditions of nutrient limitation, other evidence seems to refute this. For example, when *S. coelicolor* substrate mycelium is placed on fresh medium, the formation of aerial mycelium is not delayed (Granozzi *et al.*, 1990). This observation is supported by Allan and Prosser (1987) who also found that the formation of aerial mycelium is not induced by the depletion of nutrients in the medium. It has been suggested that perhaps stress takes place within the population, which is not induced by nutrient depletion. For example, a limitation of the transport of nutrients to parts of the colony may induce stress, mimicking a stringent response, and thereby signaling the onset of aerial hyphae formation. Study of the stringent response and its correlation with major metabolic processes may be helpful in understanding the mechanisms that activate differentiation (Granozzi *et al.*, 1990).

1.5 Antibiotic Production

At about the same time as aerial mycelium emerge on a streptomycete colony, secondary metabolites, including antibiotics, are detectable. As mentioned in earlier, *S. coelicolor* A3(2) produces four antibiotics including actinorhodin, undecylprodigiosin, methylenomycin and a Ca^{2+}-dependent antibiotic. Actinorhodin

has been studied extensively because it is a polyketide, as many medically and commercially important antibiotics are. Study of the molecular genetics of actinorhodin biosynthesis has provided information on the structure and organization of antibiotic biosynthesis genes in streptomycetes, including the clustering of structural genes, and the similarity between certain structural genes within several different pathways i.e. the genes encoding polyketide synthases (PKS). Study of the actinorhodin PKS gene set will enable further analysis of the molecular basis for PKS programming. Another feature of actinorhodin is that it is pigmented and therefore can be identified by visual inspection.

Undecylprodigiosin was identified to be another secondary metabolite by Rudd and Hopwood (1980). The biosynthetic genes for this red-pigmented, highly non-polar compound were also found to map on the chromosome in a closely linked cluster. Unlike, the actinorhodin regulatory gene, which is positioned in the center of the cluster, the regulatory gene of undecylprodigiosin, *redD*, maps to the far right end of the cluster (Coco *et al.*, 1991). As with actinorhodin, this pigmented antibiotic is also useful in the identification of mutants affected in antibiotic production.

As mentioned above, studies have shown a correlation between nutrient depletion, which produces a stringent response causing a decrease in GTP levels, and an increase in ppGpp levels and aerial mycelium formation. In both an unclassified *Streptomyces sp.* and *S. coelicolor*, it was shown that ppGpp is needed to trigger antibiotic production (Ochi, 1986b; Strauch *et al.*, 1991). However ppGpp synthesis alone does not appear to be sufficient to initiate secondary metabolism in *S. coelicolor* A3(2). The synthesis of ppGpp was higher after a nutritional shiftdown than after a treatment with serine hydroxymate (provokes the stringent response by acting as a competitive inhibitor of seryl-tRNA synthetase thus leading to the

accumulation of uncharged $tRNA^{ser}$), but antibiotic production appeared to occur normally. Also, the suppression of ppGpp, by the addition of chloramphenicol before shiftdown did not prevent antibiotic production. Consequently it was concluded that, if ppGpp does play a role in antibiotic production, it is unable to do so alone (Strauch *et al.*, 1991).

It has also been shown that there is no obligatory relationship between initiation of secondary metabolism and the stringent response in *Streptomyces clavuligerus*, as cephalosporin biosynthetic activity increased during growth in defined medium, while the level of ppGpp remained very low and stable (Bascaran *et al.*, 1991). However, in complex medium, cephalosporin production was depressed or stimulated and a wide range of ppGpp formation was found. These results did not rule out a possible stimulating effect of ppGpp on antibiotic production, but they did show that there is no obligate relationship between this activity and the stringent response. Initiation of antibiotic biosynthesis may be a consequence of the transfer to nonrepressing conditions, rather than a result of the transient production of ppGpp. The impairment in antibiotic synthesis observed in thiostrepton-resistant mutants could be explained by a ribosomal alteration that interferes with ppGpp formation and also critically affects secondary metabolism initiation by a mechanism not involving ppGpp. It was similarly shown that although the stringent response initiates both sporulation and antibiotic synthesis in *B. subtilis* , it occurs by different mechanisms(Ochi and Ohsawa, 1984).

1.6 Developmental Mutants of *S. coelicolor* A3(2)

The identification of genes that play a role in development and differentiation in *Streptomyces* has been approached in various ways. Isolating mutations that block the life cycle at specific points has been used as one method to study streptomycete

development. In *S. coelicolor* A3(2), the model of streptomycete development, two classes of morphological mutants have been found which could easily be discerned by simple visual inspection of mutant colonies. The *bld* (for bald) mutants lack aerial mycelium under all or certain growth conditions, and in *whi* mutants aerial hyphae develop but remain white instead of acquiring a grey color typical of mature spores (Chater, 1989b). The *whi* genes have not been implicated in regulation of the transition from vegetative growth to aerial mycelium formation and antibiotic production and will not be discussed further.

In 1976, Merrick genetically mapped twelve *bld* mutations of *S. coelicolor*. These mutants were classified into three groups on the basis of colony morphology, production of antibiotics and morphology on different carbon sources. The mutants were mapped to four distinct map locations designated *bldA, bldB, bldC* and *bldD*. The *bldA* and *bldD* mutants are phenotypically alike (class 1), as they produce wrinkled, soft, fragmenting colonies with prostrate, sheathed, aerial hyphae. The colonies are nonpigmented and the production of antibiotics made by the wildtype strain is lacking. *bldA* and *bldD* colonies produce aerial mycelium and spores when arabinose, galactose, maltose, glycerol or mannitol are used as the sole carbon source instead of glucose, but are still unable to produce pigments or antibiotics. However, some *bldA* mutants can make actinorhodin on R2YE (Thompson *et al.*, 1980) medium or minimal medium supplemented with yeast extract (Chater, 1984) and undecylprodigiosin in low phosphate conditions such as R2 (Okanishi *et al.*, 1974; Hopwood and Wright, 1978) medium or minimal medium with a lowered phosphate concentration. It has been suggested that production of undecylprodigiosin is inhibited on R2YE medium by the phosphate in the yeast extract (Guthrie and Chater, 1990).

The colonies of *bldB* mutants exhibit another morphology (class 2). *bldB* mutants produces smooth, hard, non-fragmenting colonies. Weak aerial mycelium and antibiotic production develop only after prolonged incubation.

bldC mutants, which belong to the third class, produce smooth, non-fragmenting colonies with no evidence of aerial hyphae either. These colonies are initially nonpigmented at first, but eventually do produce actinorhodin and methylenomycin. They can also sporulate on galactose, maltose and mannitol.

Later, *bldE* (Hodgson, 1980) and *bldF* (Puglia *et al.*, 1984) mutants were identified. *bldE* mutants were isolated by selection for their ability to use agar as a sole carbon source in the presence of homoserine. They form smooth but sculpted colonies and are able to produce undecylprodigiosin, and can also sporulate on mannitol. *bldF* mutants have a similar phenotype as the *bldE* mutants, but cannot sporulate on any medium.

Two more *bld* genes were identified later, *bldG* and *bldH* (Champness, 1988). *bldG* mutants produce smooth, soft, fragmenting colonies and there is some evidence of aerial hyphae but no spores. *bldH* mutants produce smooth, hard, fragmenting colonies. Neither of these mutants produce any of the above four antibiotics when glucose is used as a carbon source. When other carbon sources, such as mannitol, are used, the developmental blocks are partially relieved for *bldG* mutants (i.e. they form aerial hyphae and spores), and fully relieved for *bldH* mutants, including antibiotic production(Champness, 1988).

The *bldI* mutant has similar colony morphology and properties to *bldB*. This mutant was found to map close to the *bldB* locus but is not complemented by DNA clones that complement true *bldB* mutants (Champness, 1988; Harasym *et al.*, 1990).

bldA (Piret and Chater, 1985), *bldB* (Champness, 1988), *bldD* and *bldG* (Passantino, R., Leskiw, B.K. and Chater, K.F., unpublished data) have been cloned

by complementation of mutants, and in each instance ØC31 bacteriophage based vectors were used. Of these, only *bldA* and *bldB* genes have been sequenced. The *bldA* gene was sequenced and shown to encode a tRNA that would recognize the leucine codon UUA, a codon found to be very rare in the GC-rich mRNA of *Streptomyces* (Lawlor *et al.*, 1987; Leskiw *et al.*, 1991b). Studies have verified that the *bldA* gene product is the principal means by which UUA codons can be translated (Chater *et al.*, 1988; Leskiw *et al.*, 1991a). The *bldA* mutant grows normally as substrate mycelium and no TTA codons have been identified in any genes essential for vegetative growth. Evidence has shown that TTA is present in very few genes and its occurrence is nonrandom with respect to the class of genes concerned. The TTA codon has mainly been found in several genes involved in antibiotic resistance and regulation, and in morphological and physiological differentiation suggesting a regulatory role in events that occur late in colony growth (Leskiw *et al.*, 1991b). Such a regulatory role is supported by the fact that representative actinorhodin and undecylprodigiosin genes are transcriptionally inactive in a *bldA* mutant. A TTA codon was identified in *actII-ORF4*, the pathway-specific activator gene for actinorhodin biosynthesis and in *actII-ORF2*, an actinorhodin export gene (Fernandez *et al.*, 1991). However, the *redD* pathway-specific activator gene, which is homologous to *actII-ORF4*, does not contain any TTA codons (Narva and Reitelson, 1990). To explain the *bldA*-dependence of the *red* structural genes, it was proposed that another TTA-containing activator is involved (Guthrie and Chater, 1990). Methylenomycin genes were also found to be transcriptionally blocked in *bldA* mutants (Champness and Chater, 1994). However, the role of *bldA* in the biosynthesis of CDA has not been determined. In *S. griseus bldA* mutants are also antibiotic negative, producing no streptomycin. As expected, *strR*, the pathway-specific activator gene for streptomycin biosynthesis, contains a TTA codon (Distler

et al., 1992). Another interesting finding was that mycelium-associated (phase I) glycogen accumulation was seldom observed in *bldA* mutants, irrespective of whether the carbon source was mannitol or glucose (Plaskitt and Chater, 1995). This phase of glycogen accumulation may require the action of one or more genes containing the TTA codon. However, phase II (sporulation-associated) glycogen synthesis in *bldA* mutants could be induced when mannitol was used to induce normal aerial mycelium formation and sporulation. This sporulation-associated glycogen accumulation indicates that genes for this phase of accumulation contain no TTA codons, since the dependence on *bldA* of TTA-containing genes is not alleviated by growth on mannitol (Leskiw *et al.*, 1991b).

The suggestion that the *bldA* gene product plays a role in events late in colony growth, such as antibiotic production and morphological differentiation, is supported by the finding that accumulation of the tRNA gene is temporally regulated. Although early analyses suggested that *bldA* gene transcription occurred late in growth, about the time of initiation of aerial hyphae formation (Lawlor *et al.*, 1987), subsequent promoter-probing and S1 nuclease protection experiments, have shown that the promoter is active even in very young cultures (Leskiw *et al.*, 1993). In contrast to the primary transcript however, accumulation of the 5'-processed form of the tRNA was temporally regulated, showing accumulation only just before the appearance of aerial hyphae. Evidence of temporal regulation of *bldA* expression supports the hypothesis that *bldA* may have a regulatory role.

Recent studies have shown that the *bldA* gene is expressed at significantly lower levels in a *bldI* mutant (Leskiw and Mah, 1995). The mature form of the *bldA*-encoded tRNA and the expression of a TTA-containing reporter gene was reduced in the *bldI* mutant. Also, when the *bldA* gene was introduced on a high copy number

plasmid into the *bldI* mutant, levels of tRNA were reduced. This suggests that *bldI* is required either directly or indirectly for *bldA* expression or function.

As mentioned earlier, the *bldB* gene has also been cloned (Piret *et al.*, 1988; Harasym *et al.*, 1990) and sequenced. It was found to encode a negatively charged 11kDa protein. Sequence similarity searches have not identified similarity to any known proteins and therefore its developmental role is still unknown (Champness and Chater, 1994).

In addition to the actinorhodin and undecylprodigiosin biosynthetic genes, SapB, a spore associated protein, is not produced by *bldA*, *B*, *C*, and *D* mutants (Willey *et al.*, 1991). Newly identified *bld* mutants, *bld217* and *bld221* were shown to restore their ability to form aerial hyphae by the juxtaposition of the bald colonies near colonies of a SapB producing bacterium, or by the application of purified SapB near mutant colonies. Willey *et al.* (1991) therefore suggested that SapB is a morphogenetic protein that either creates a scaffold for the erection of aerial hyphae or it enables the hyphae on the surface of colonies to break surface tension and grow into the air. Later reports (Willey *et al.*, 1993) indicated that most *bld* mutants are rescued for SapB production and aerial mycelium formation when grown near certain other *bld* mutants. These experiments indicated that morphological differentiation is governed by a hierarchical cascade of at least 4 intercellular signals which involve *bld261* (a newly identified *bld* gene (Willey *et al.*, 1993)), *bldA*, *bldH*, *bldG*, *bldC* and *bldD*. The *bldD* mutant was capable of complementing all of the other mutants, placing it at the top of the hierarchy and at the end of the cascade. SapB was shown not to be ribosomally synthesized and therefore, as with many peptide antibiotics, SapB was postulated to be synthesized by a large complex composed of peptide synthetases. Given this information, it was suggested that *bldD* may be the structural gene for a peptide synthetase involved in SapB production or that it could be involved

in the expression of this putative synthetase (Willey *et al.*, 1993). It is important to note that SapB and the *bld* genes are involved in only one of two or more potential pathways responsible for the formation of aerial hyphae. As mentioned above, it is well established that many of the *bld* mutants exhibit a conditional phenotype, showing restoration of aerial mycelium formation (but not antibiotic production) when grown on poorly utilized carbon sources. Although SapB is produced on rich medium, it is not made in detectable levels during sporulation when minimal medium containing mannitol or certain carbon sources, which suppressed the *bld* mutant phenotype, were used. It has therefore been postulated that *Streptomyces* may be able to erect aerial hyphae in at least two different ways, one that is dependent on, and the other independent of SapB (Willey *et al.*, 1991).

SapA, another spore associated protein, also appeared to be subject to temporal and developmental regulation since its transcription coincided with the appearance of aerial mycelium. This transcription was found to be reduced in *bldC* and *bldD* mutants but was comparable to wildtype levels in the other *bld* mutants studied. Even though morphological differentiation does not occur when *S. coelicolor* is grown in liquid medium, *sapA* could be strongly induced after cells entered stationary phase. Hence, although *sapA* is associated with aerial mycelium formation, morphological differentiation is not required for the activation of the *sapA* gene (Guijjaro *et al.*, 1988).

SapC, SapD and SapE, were recently shown to be encoded by SCP1, which is dispensable for the processes of aerial mycelium and spore formation. Therefore, these three Saps evidently do not play an indispensable role in the developmental process (Willey *et al.*, 1991; Willey *et al.*, 1993; Guijjaro *et al.*, 1988).

The phenotype of an early sporulation block and a pleiotropic effect on antibiotic production, displayed by the *bld* mutants, is similar to the pleiotropic

phenotype of the early *B. subtilis spoOA, spoOB* and *spoOF* mutants. The *spoOA, spoOB* and *spoOF* genes belong to a multicomponent phosphorelay. The *bld* genes have not yet been implicated in this type of signal transduction system. There have, however, been findings which associate signal transduction by two component systems with the regulation of secondary metabolism (Champness and Chater, 1994). The *afsR* gene is one example of this type of regulation. The *afsR* gene was cloned from *S. coelicolor* as a pleiotropic regulatory gene that globally controls production of A-factor (a butyrolactone compound essential for both sporulation and streptomycin production in *S. griseus*) and the pigmented antibiotics actinorhodin and undecylprodigiosin in this organism and its related species *S. lividans* (Horinouchi *et al.*, 1983; Horinouchi *et al.*, 1990). This global regulatory protein has recently been found to be phosphorylated by a membrane-associated phosphokinase, named AfsK (Matsumoto *et al.*, 1994). The N-terminal portion of AfsK showed significant sequence similarity to the catalytic domain of eukaryotic Ser/Thr protein kinases. This is one example of a signal transduction system similar to that found in eukaryotes.

A two-component regulatory system typical of prokaryotes, in which phosphorylation of sensor proteins on a His residue and regulatory proteins on an Asp residue is involved, may also play a role in the regulation of secondary metabolism in *Streptomyces*. This system involves two genes, namely *afsQ1* and *afsQ2*, which showed great similarity to response regulators and sensory histidine protein kinases, respectively. Disruption of these genes did not result in any detectable change in secondary metabolism or morphogenesis, which implies that *afsQ1* and *afsQ2* play no obligatory role in normal antibiotic synthesis. The idea that the *afsQ1-afsQ2* system is involved in regulation of secondary metabolism was supported by the finding that an *absA* mutation (Ishizuka *et al.*, 1992), which causes production of all four

antibiotics to be blocked but sporulates normally, was suppressed by multiple copies of *afsQ1*.

Two other groups of developmental genes, *aba* (Fernandez-Moreno *et al.*, 1992) and *abs* genes (Adamidis *et al.*, 1990; Adamidis and Champness, 1992), have been identified. Both groups affect antibiotic production but not sporulation. The only *aba* gene identified so far, *abaA*, abolishes actinorhodin production, strongly reduces undecylprodigiosin and CDA, but has no effect on methylenomycin when it is disrupted. Mutants of the two *abs* genes, *absA* and *absB*, however, produce none of the four antibiotics. These genes have not been implicated in a two-component system, as with the *afsR-afsK* system and the *afsQ1-afsQ2* system.

It is evident that there are several pleiotropically acting *S. coelicolor* genes that affect production of secondary metabolites. Recent findings have revealed the involvement of signal transduction systems in the regulation of secondary metabolism. The fact that other regulatory mechanisms are independent of the two-component system create a complex system of regulation of secondary metabolism, that is still unclear and needs to be investigated further.

Investigating the *bld* mutants, which are defective in both aerial mycelium and antibiotic production, will provide information on both morphological differentiation and secondary metabolism. The production of secondary metabolites appears to be regulated, at least in part, by different mechanisms, as implied by the existence of mutants defective only in antibiotic production. The *whi* mutants, blocked at various stages of sporulation, however, will aid us in the understanding of the late stages of morphological differentiation. Together, they will provide a better grasp of developmental processes in Streptomycetes.

1.7 Sequence Analysis of *bldD*

The *bldD* gene has been isolated (Passantino, R. and Leskiw, B.K., unpublished) and this research focuses on determination of the DNA sequence of the gene and on analysis of the gene product. It is hoped that such information will help in understanding the role that the *bldD* gene product plays in differentiation and antibiotic production.

2.0 MATERIALS AND METHODS

2.1 Materials

The computer software used for the analysis of DNA and protein sequences was DNA Strider, which was designed and written by C. Marck (Commissariat a l'Energie Atomique, France). A program for the analysis of protein coding r gions, adapted for the IBM microcomputer and based on the FRAME program of Bibb *et al.* (1984), modified by Uchiyama and Weisblum (University of Wisconsin) (1985), was adapted for our use on the Apple MacIntosh by C. Jensen. Protein similarities were determined through the use of the BLAST (Altschul *et al.*, 1990) and Fasta (Pearson and Lipman, 1988) programs which were accessed by electronic mail.

Restriction endonucleases, Klenow fragment of *E. coli* DNA polymerase I and T4 DNA ligase were obtained from Boehringer Mannheim. All enzymes were used according to the specifications of the manufacturer.

Deoxyribonucleoside triphosphates, dideoxyribonucleoside triphosphates and the modified nucleoside triphosphate 2'-deoxyribo-7-deazaguanosine-5'-triphosphate were purchased from Boehringer Mannheim. Radioactively labeled [α^{32}P]dATP was purchased from ICN Biomedicals Canada, Ltd. and [α^{35}S]dATP was purchased from New England Nuclear.

Specific oligonucleotide primers and universal M13 sequencing primers, were obtained from the Department of Biological Sciences DNA Synthesis Laboratory, University of Alberta, Edmonton, Alberta, Canada.

"Sequenase" sequencing kit was purchased from United States Biochemical Corporation.

Hygromycin B was purchased from Sigma and thiostrepton was a gift from S. Lucania, Squibb and Sons, Inc.

All other chemicals used in this study were of reagent grade.

2.2 Bacterial Strains, Plasmids and Bacteriophages

The *Streptomyces* strains and plasmids used are listed in Table 2.1, the *E. coli* strains are listed in Table 2.2 and *E. coli* plasmids and bacteriophages are listed in Table 2.3.

2.3 *Streptomyces* Techniques

2.3.1 Propagation and Culture

Glycerol and Lyophil Stocks

Glycerol stocks were prepared in the following manner: using a sterile spatula, spores of wildtype strains or mycelial fragments of *bld* mutants were scraped from the surface of agar plates. The spores or mycelial fragments were resuspended in water, homogenized to break up clumps, the suspension was filtered through non-absorbent cotton packed loosely into a filter tube, and the filtered suspension was centrifuged for 10 minutes at 2900 rpm to pellet the spores or mycelial fragments. Finally the pellet was resuspended in sterile 20% glycerol and stored at -20°C.

Lyophilized stocks were prepared by scraping the spores or mycelium into water, homogenizing the spores or mycelium and then centrifuging the homogenized mixture for 10 minutes at 2900 rpm to form a pellet. The pellet was then resuspended in 20% skim milk (Difco), dispensed into lyophilization vials and lyophilized.

Table 2.1. *Streptomyces* bacterial strains and plasmids

Streptomyces strain or plasmid	Relevant characteristic or genotype	Reference and Source
Strains		
Streptomyces coelicolor		
J1501	*hisA1, uraA1, strA1, pgl*, NF, SCP2$^-$	John Innes Institute; Chater *et al.*, 1982
1169	*bldD53, hisA1, mthB2, pheA1, strA1*, NF, SCP2	John Innes Institute; Merrick, 1976
J774	*mthB2, cysD18, pheA1, agaA7,bldD53, strA1*, NF, SCP2	John Innes Institute; Merrick, 1976
HU66	*bldD53, hisA1, uraA1, strA1, pgl*	Joanne Willey, Harvard University; Willey *et al.*, 1993
Streptomyces lividans		
1326	wildtype	John Innes Institute; Lomovskaya *et al.*, 1972
Plasmids		
pIJ486	high copy number cloning vector	John Innes Institute; Ward *et al.*, 1986
pIJ941	low copy number cloning vector	John Innes Institute; Lydiate *et al.*, 1985

Table 2.2. *E. coli* strains

E. coli strain	Relevant characteristic or genotype	Reference or Source
MV1193	*δ(lac-proAB), rpsL, thi, endA, spcB15, hsdR4, δ(srl-recA)306::Tn10(tet^r), F'[traD36, proAB$^+$, lacI^{q}lacZδM15]*	Zoller and Smith, 1987
XL1-Blue	*recA1, endA1, gyrA96, thi1, hsdR17, supE44, relA1, lac[F' proAB, lacI^{q}lacZδM15, Tn10(tet^r)]*	Bullock et al., 1987
DH5α	*supE44, δlacU169(ϕ8ϕlacZδM15), hsdR17, recA1, endA1, gyrA96, thi1, relA1*	Hanahan, 1983
ET12567	*F-dam13::Tn9, dcm6, hsdM, hsdR, recF143, zjj202::Tn10, galK2, galT22, ara14, lacY1, xyl5, leuB6, thi1, tonA31, rpsL136, hisG4, tsx78, mtl1, glnV44*	gift from Doug MacNeil; MacNeil, T., 1991, Merck Sharp & Dohme Research Laboratories

Table 2.3. *E. coli* plasmids and bacteriophages

E. coli plasmid or bacteriophage	Relevant characteristic or genotype	Reference or Source
Plasmids		
pBR322	pMB1-derived cloning vector	GIBCO-BRL; Bolivar *et al.*, 1977; Sutcliffe, 1978
pIJ2925	pUC derived cloning vector	John Innes Institute; Janssen, G. and Bibb, M., 1993
pSET152	shuttle vector	NRRL; Bierman, M. *et al.*, 1992
pUC118/119	phagemid cloning vector	gift from J. Vieira, Department of Biochemistry, University of Minnesota; Vieira and Messing, 1987
Bluescript SK+/-	phagemid cloning vector	Stratagene; Short *et al.*, 1988
Bacteriophages		
M13mp18/mp19	sequencing vector	Boehringer Mannheim; Yanisch-Perron *et al.*, 1985
M13K07	helper phage	gift from J. Vieira, Department of Biochemistry, University of Minnesota; Vieira and Messing, 1987

Growth of *Streptomyces* strain for Plasmid Isolation

S. coelicolor strains were grown for the purposes of plasmid isolation according to Hopwood *et al.* (1985). For small scale preparations, mycelia were grown at 30°C and 250 rpm in 5mL of trypticase soy broth (TSB) with thiostrepton (5ug/mL) or hygromycin (200ug/mL). For large scale plasmid isolation, seed cultures were started in 5mL TSB as described above, grown for 24-48 hours and used to inoculate 500mL of Yeast Extract-Malt Extract medium (YEME) supplemented with Tiger's milk (L-arginine at 75ug/mL, L-cystine, L-histidine, DL-homoserine, L-leucine, L-phenylalanine and L-proline at 56.25ug/mL, adenine and uracil at 11.25ug/mL and nicotinamide at 0.75ug/mL))(1.5mL/200mL)(Hopwood *et al.*, 1985), 0.5% glycine and 5mM $MgCl_2 \cdot 6H_20$ and further incubated for 48 hours at 30°C and 250 rpm. The appropriate antibiotic was also added to the same final concentrations described above.

Growth of *Streptomyces* strain for Protoplast Preparation

S. coelicolor bldD mutant strains were cultivated initially on R2YE (Hopwood *et al.*, 1985) plates containing the required supplements (see Table 2.1). The supplements were added in the form of Tiger's milk or individual amino acids at the following final concentrations: L-phenylalanine and DL-homoserine at 57ug/mL, L-histidine at 75ug/mL and uracil at 11ug/mL. A small patch of mycelium was scraped from plate cultures and inoculated into 25mL YEME medium (Hopwood *et al.*, 1985) containing the required amino acids at the concentrations mentioned above. The liquid cultures were incubated for 36-40 hours at 30°C and 250 rpm.

Growth of *Streptomyces* strain for Chromosomal DNA Preparation

The *Streptomyces* mycelium used to prepare chromosomal DNA was grown as for the preparation of protoplasts.

2.3.2 Ligation of DNA, Protoplast formation and Transformation of *S. coelicolor*

Ligation of DNA into a plasmid was done according to the methods outlined in Sambrook *et al.* (1989). The procedure for protoplast formation of *S. coelicolor* was outlined by Hopwood *et al.* (1985). Transformation of *Streptomyces* protoplasts with plasmid DNA was also done as described by Hopwood *et al.* (1985). Selection of plasmid-containing transformants was done by flooding the plates with an aqueous solution of the antibiotics thiostrepton, hygromycin or apralan to give final concentrations of 25ug/mL, 200ug/mL and 25ug/mL, respectively. The plates were flooded after approximately 16 hours incubation at 30°C to allow the protoplasts to regenerate.

2.3.3 Chromosomal DNA Preparation

Streptomyces chromosomal DNA was isolated using procedure 3 described by Hopwood *et al.* (1985).

2.3.4 Plasmid Preparation

Plasmid DNA was isolated using a standard alkaline lysis protocol (Sambrook *et al.*, 1989), except that lysozyme (2mg/mL) was added to Solution I and the mycelia were incubated in this solution for 1hr at 37°C.

2.4 *E. coli* Techniques

2.4.1 Propagation and Culture

E. coli MV1193, XL1-Blue and ET12567 were grown in 2xYT broth, (1.6% w/v tryptone, 1% w/v yeast extract, 0.5% w/v NaCl), or LB broth (1% w/v tryptone, 0.5% w/v yeast extract, 1% w/v NaCl). *E. coli* MV1193 was maintained on minimal medium plates and XL1-Blue was maintained on LB containing tetracycline (12.5ug/mL). Competent cells of these strains were prepared by the method of Morrison (1979). DH5α competent cells were obtained from GIBCO-BRL.

2.4.2 Ligation of DNA and Transformation of *E. coli* strains

Ligation of DNA into plasmids and bacteriophages was carried out using the methods outlined by Sambrook *et al.* (1989). Transformation of competent *E. coli* cells followed standard procedures outlined by Sambrook *et al.* (1989) with minor alterations. The ligated DNA was added to a 250uL aliquot of competent cells, prepared as described in 2.4.1, and incubated on ice for 30 minutes. The cells were then heat-shocked at 42°C for 90 seconds, placed on ice for 2 minutes, followed by one hour incubation at 37°C after adding 750uL LB broth. The cells were then plated on LB agar plates supplemented with either tetracycline (12.5ug/mL), apralan (50ug/mL) or ampicillin (100ug/mL) and incubated overnight at 37°C. The commercial DH5α competent cells were transformed according to the manufacturer's protocol. *E. coli* MV1193 or XL1-Blue competent cells were also transformed with M13 DNA using the procedure outlined by GIBCO-BRL for DH5α, with the exception that the heat shock was done for 90 seconds at 42°C.

2.4.3 Preparation of Template DNA for Sequencing

Single stranded template DNA was prepared as described by Sambrook *et al.* (1989) with the following modifications: the polyethylene glycol (PEG) solution contained 3.5M ammonium acetate instead of NaCl and the DNA was precipitated using 7.5M ammonium acetate instead of sodium acetate.

2.4.4 Plasmid Preparation

E. coli plasmid DNA was isolated using the standard alkaline lysis protocol outlined by Sambrook *et al.* (1989). Plasmid DNA isolated from a large scale culture was purified by CsCl-EtBr density gradient centrifugation also described by Sambrook *et al.* (1989).

2.5 Restriction Fragment Analysis

2.5.1 Polyacrylamide Gel Electrophoresis

DNA fragments less than 1kb were subjected to electrophoresis on 5% (w/v) polyacrylamide gels using a TBE buffer system (100mM Tris-HCl, pH 8.0, 1mM EDTA, 60mM borate). Appropriate molecular weight markers were *Pst*I fragments of λ phage.

2.5.2 Agarose Gel Electrophoresis

DNA fragments greater than 1kb were subjected to electrophoresis on 1% agarose gels using the TAE buffer system (40mM Tris-acetate, 1mM EDTA, pH 8.0). *Hind*III and *Pst*I fragments of λ phage were used as molecular weight markers.

2.6 Recovery of DNA fragments

2.6.1 Recovery of DNA fragments from agarose gels - Method I

The DNA fragments were excised from agarose gels with a scalpel and individual gel slices were placed in 0.5mL eppendorf tubes, which have a hole in the bottom and are packed with silanized glass wool. The 0.5mL eppendorf tubes were then placed inside 1.5mL eppendorf tubes and centrifuged at 4°C for 10 minutes. The aqueous solutions were collected in the outer 1.5mL eppendorf tubes. The aqueous solutions were then extracted with phenol:chloroform, followed by an ether extraction and then precipitated with ethanol. The final DNA pellets were then dissolved in dH_2O.

2.6.2 Recovery of DNA fragments from agarose gels - Method II

DNA fragments were isolated from agarose gels using the protocol outlined by Zhen and Swank (1993). In brief, DNA fragments were separated by electrophoresis in a 1% agarose gel and then stained in ethidium bromide. Using a scalpel, a rectangular trough was cut directly in front of the leading edge of the band of interest. The trough was filled with 15% PEG/TAE with ethidium bromide (0.5ug/mL) and electrophoresis was continued until the DNA band of interest moved into the center of the trough. The DNA-containing PEG/TAE solution was then removed and subjected to a phenol:chloroform and an ether extraction. The DNA was then precipitated and dissolved in dH_2O.

2.6.3 Recovery of DNA fragments from polyacrylamide gels

DNA fragments were isolated from polyacrylamide gels using the "Crush and Soak" method outlined by Sambrook *et al.* (1989).

2.7 Techniques Using Radiolabeled DNA

2.7.1 Transfer of DNA from agarose gels to nylon or nitrocellulose filters

The digested DNA of interest was separated by electrophoresis on a 1% agarose gel. The separated DNA fragments were transferred to a nylon (Amersham Hybond-N) or nitrocellulose (Biorad) filter by the method of Southern (1975) as described by Sambrook *et al.* (1989). The DNA was cross-linked to the nylon membrane by exposure to ultraviolet light using a Bio-Rad GS Gene Linker UV chamber following the manufacturer's suggested conditions for a damp Southern blot. On the other hand, the nitrocellulose filters were air dried and baked in an 80°C vacuum oven for 2 hours.

2.7.2 Generation of ^{32}P-labeled DNA probes

DNA fragments were labeled using the random primed DNA labeling protocol described by Boehringer Mannheim. In a screw cap tube, 9uL of denatured DNA (10-100ng), 2uL of hexanucleotide mixture (Boehringer Mannheim), 3uL of dNTP labeling mixture containing dCTP, dGTP and dTTP (all at 1.5mM), 50uCi of α-[^{32}P]-dATP and 1uL of Klenow (2U/uL) were mixed and incubated at room temperature overnight. The labeled probe was then purified from unincorporated α-[^{32}P]-dATP by filtration through a disposable syringe filter unit (Stratagene Nuctrap Push Columns).

2.7.3 Hybridization Analysis

Membrane filters (Southern blots) were prehybridized in a heat sealable, plastic bag containing sufficient prehybridization solution [0.45M NaCl and 0.045M sodium citrate (3xSSC), 4xDenhardt's (Denhardt, 1966) and denatured salmon sperm

DNA (100ug/mL)] to cover the filter when laid in a horizontal position. The bag was sealed and incubated at 70°C for at least one hour. The prehybridization solution was then removed and replaced with hybridization solution (prehybridization solutic containing approximately $5.0x10^6$ cpm of denatured α-^{32}P-labeled probe) and hybridization was carried out overnight. The filter was then removed and washed twice for 30 minutes at 70°C in 2xSSC+0.1% SDS and then twice in 0.2xSSC+0.1% SDS as described by Hopwood *et al.* (1985). Filters were air-dried and exposed to Kodak X-OMAT AR film at -70°C.

2.8 Filling Recessed 3' Termini

After digesting the DNA with the appropriate restriction enzyme, the DNA was extracted with phenol:chloroform and then with ether. The DNA was then precipitated with ethanol and the final pellet was dissolved in 27uL dH_20. To the DNA, 3uL of 10X Polymerase I buffer (50mM Tris, pH 7, 5mM $MgCl_2$, 10mM β-mercaptoethanol, 50ug/mL BSA), 2uL of dNTP mix (each at 0.125mM) and 1 unit of the Klenow fragment of *E. coli* DNA polymerase I was added and then the reaction was incubated at 37°C for 10 minutes. The incubation was followed by an extraction with phenol:chloroform and an extraction with ether. The DNA was then precipitated with ethanol and dissolved in dH_20.

2.9 DNA Sequence Analysis

DNA sequence analysis was performed using the chain termination method of Sanger *et al.* (1977) as modified by Tabor and Richardson (1987) for use with Sequenase. All reactions used α-^{35}S-dATP as the radioactive marker.

Single stranded template DNA was prepared as described in section 2.4.3.

Labeled fragments produced in the sequencing reactions were separated electrophoretically on 6% polyacrylamide (38:2, acrylamide:N,N'-methylene bisacrylamide) denaturing gels (8.3M urea) using a TBE buffer system (see 2.5.1). To resolve compressions in the sequence sequencing gels containing 40% formamide were used or the templates were sent for cycle sequencing on a Applied Biosystems 373 DNA Sequencer. After electrophoresis sequencing gels were placed in a fixing solution (10% methanol, 10% acetic acid) for 10-15 minutes, the gel was lifted onto Whatmann (3MM) filter paper and dried using a Biorad Model 583 Gel Drier. Radioactive bands were visualized by exposing the sequencing gels to Kodak X-OMAT AR film at room temperature. Films were developed using a FUJI RGII X-ray film processor.

3.0 RESULTS AND DISCUSSION

A 3.4kb fragment that was able to restore both antibiotic production and sporulation to *bldD* mutants of *S. coelicolor* was isolated from a library of *S. coelicolor* DNA fragments in the ØC31-based phage vector KC304 (Hopwood *et al.*, 1987). This clone was designated KC742 (Leskiw, B.K. and R. Passantino, unpublished; see Figure 3.1). The object of this thesis project was to subclone and sequence the *bldD* gene. Details of the *bldD* gene isolation follow.

3.1 *bldD* Gene Isolation

3.1.1 Subcloning the *bldD* gene

Because of the difficulty in isolating large amounts of insert DNA from a phage vector, attempts were made to transfer the insert from KC742 into the high-copy number *Streptomyces* plasmid vector pIJ486 (Ward, *et al.* 1986). KC742 was digested with *Eco*RV to release the insert DNA along with 1kb of flanking ØC31 vector DNA (see Figure 3.2). A 10 µl aliquot of the reaction was electrophoresed on a 1% agarose gel to confirm complete digestion. The remainder of the digested DNA was extracted with phenol:chloroform followed by an ether extraction and then precipitated with ethanol. Finally the pellet was redissolved in TE buffer.

Several unsuccessful attempts were then made to ligate the 4.5kb *Eco*RV fragment into pIJ486 that had been digested with *Bam*HI and treated with Klenow fragment of DNA polymerase to blunt the ends (see section 2.8). Since the possibility existed that the *bldD*-containing *Eco*RV fragment might not be stable in a high copy number vector, attempts were made to use the low copy number, SCP2*-based plasmid vector, pIJ941 (Lydiate *et al.*, 1985). The remainder of the digested KC742

Figure 3.1. Map of KC742. The 3.4kb insert which was a partial *Sau*3A fragment cloned into the *Bam*HI site is indicated by the shaded box.

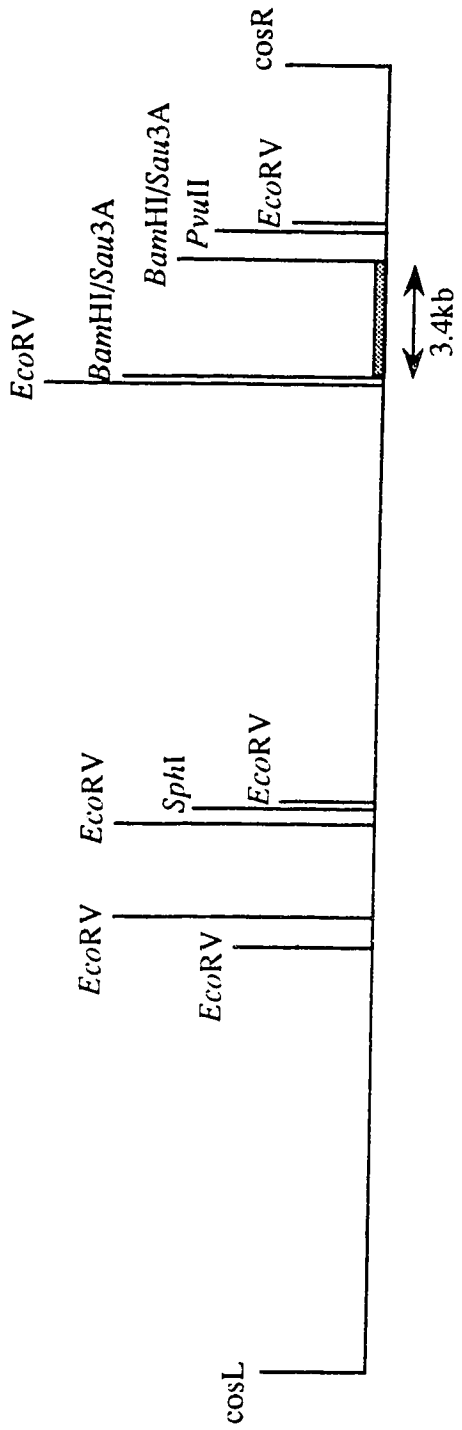
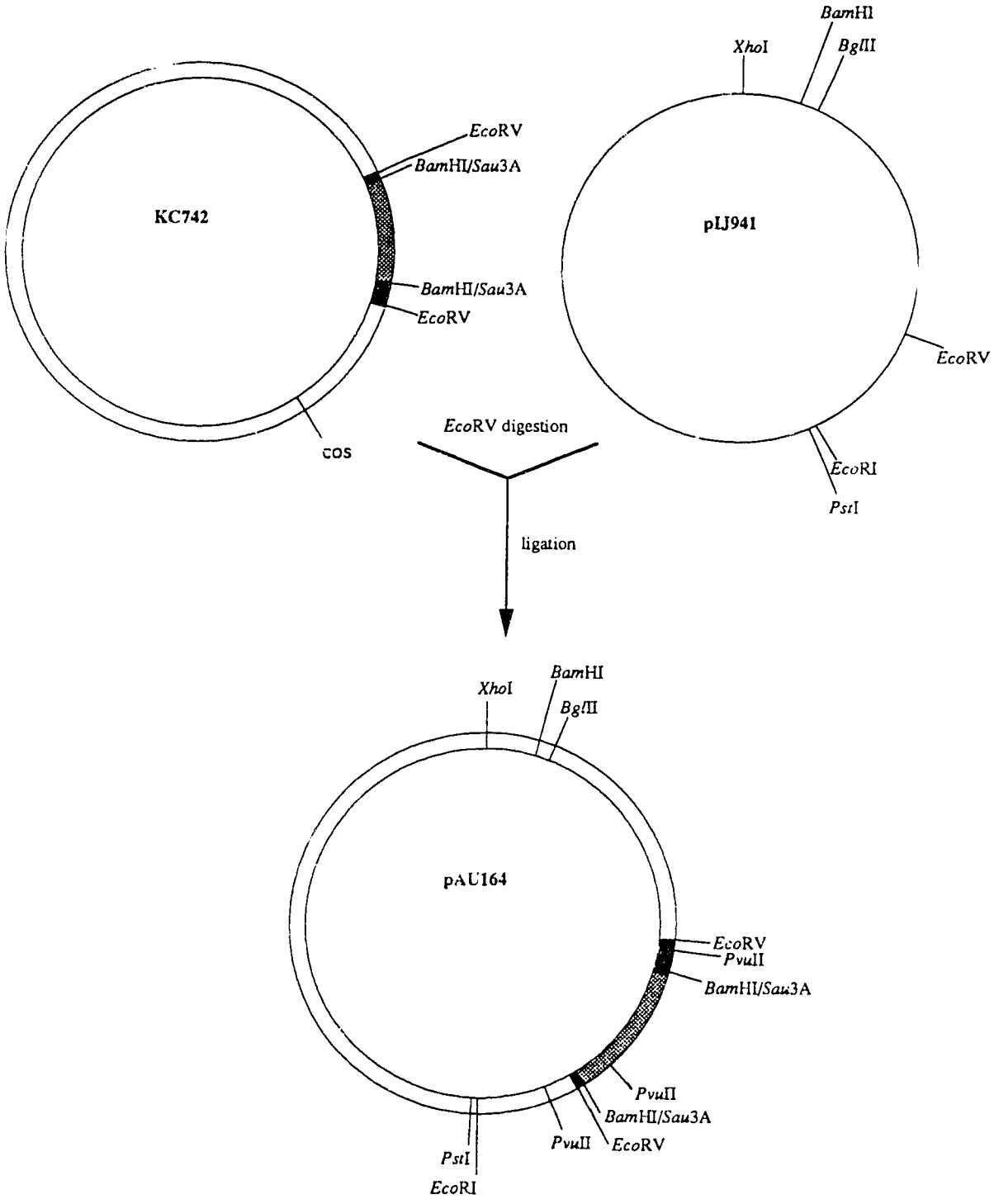

cosR
EcoRV
BamHI/Sau3A
PvuII
EcoRV
BamHI/Sau3A
EcoRV
3.4kb
EcoRV
SphI
EcoRV
EcoRV
EcoRV
cosL

Figure 3.2. Strategy for subcloning the 4.5kb *Eco*RV *bldD*-containing fragment from KC742 into pIJ941. The insert was a partial *Sau*3A fragment cloned into the *Bam*HI site (represented by the shaded box). The flanking ØC31 sequence that is removed after *Eco*RV digestion is indicated by the filled black boxes.

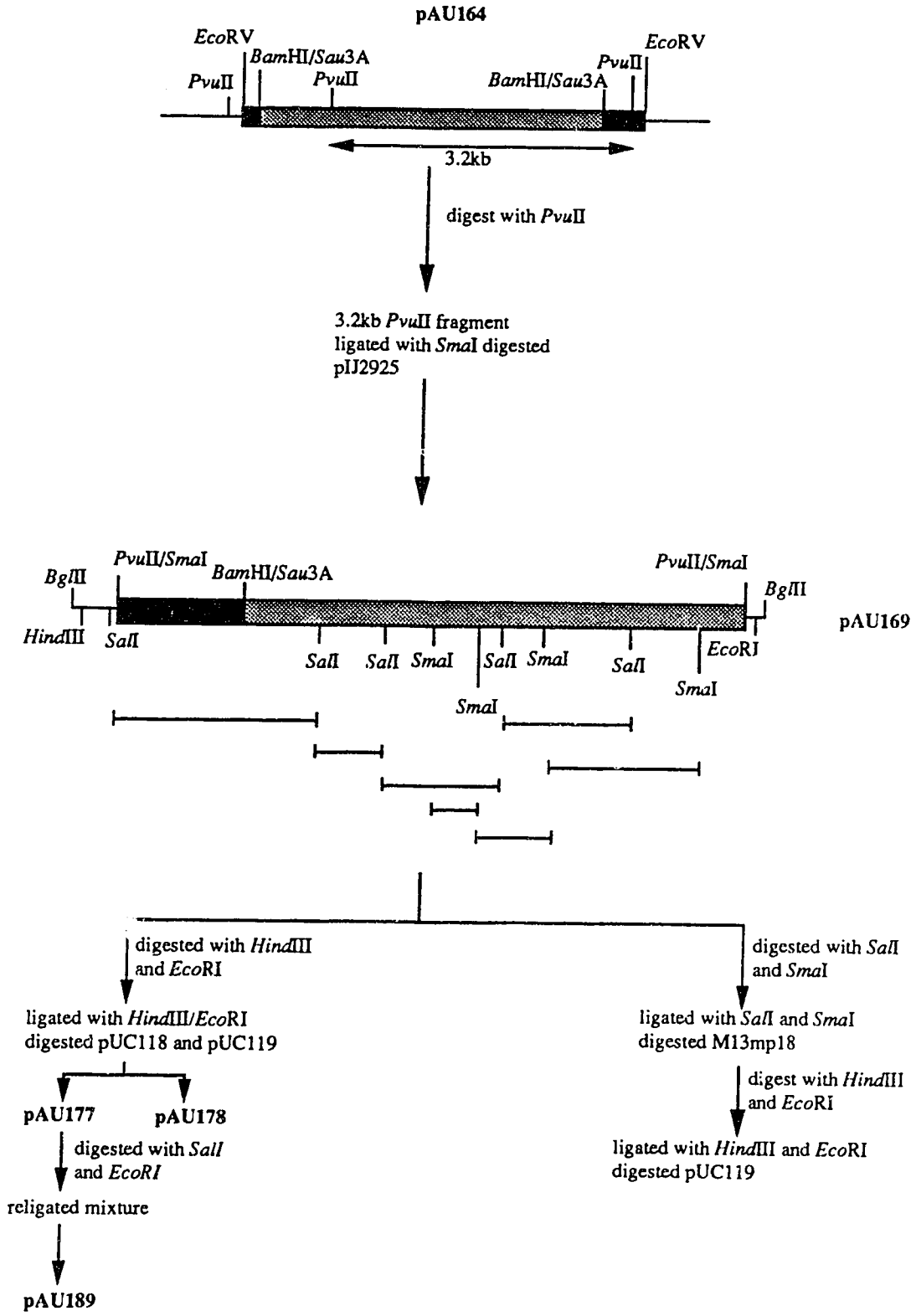

KC742
EcoRV
BamHI/Sau3A
BamHI/Sau3A
EcoRV
cos
pLJ941
BamHI
XhoI
BglII
EcoRV
EcoRI
PstI
EcoRV digestion
ligation
pAU164
XhoI
BamHI
BglII
EcoRV
PvuII
BamHI/Sau3A
PvuII
BamHI/Sau3A
EcoRV
PvuII
PstI
EcoRI

DNA was ligated to *Eco*RV-digested pIJ941 (see Figure 3.2). The ligated DNA was then used to transform protoplasts of the *bldD* mutant strain *S. coelicolor* HU66 and hygromycin resistant transformants were selected. Two hygromycin resistant transformant colonies were selected, both of which were able to sporulate and produce the pigmented antibiotics normally produced by wildtype *S. coelicolor*. Plasmid DNA was isolated from each transformant and both plasmids were shown, after digestion with *Eco*RV, to contain an insert of the expected size (4.5kb). This result confirmed that the *Eco*RV fragment was able to complement the *bldD* mutation in *S. coelicolor* HU66. An aliquot of plasmid DNA was also used to transform protoplasts of the *bldD* mutant strains, *S. coelicolor* 1169 and J774 to further confirm that the *Eco*RV fragment was able to complement the *bldD* mutation in each of the strains. These strains carry the same *bldD53* mutation but in a slightly different genetic background. Only one clone, designated pAU164, was chosen for further study.

3.1.2 Subcloning of a 3.2 kb *Pvu*II Fragment from pAU164.

In order to isolate a smaller complementing subclone of the 4.5kb *Eco*RV fragment, attempts were made to subclone a 3.2kb *Pvu*II fragment from the recombinant plasmid pAU164. Digestion of pAU164 with *Pvu*II generated a fragment containing 600bp of ØC31 sequence along with 2.6kb of the original *Sau*3A partial insert fragment from KC742 (see Figure 3.3). This *Pvu*II fragment was then purified from the agarose gel as described in section 2.6.1. After purification, the DNA was ligated to *Sma*I-digested pIJ2925 (an *E. coli* pUC18 plasmid derivative with *Bgl*II flanking both ends of the multiple cloning site; Janssen and Bibb, 1993). The ligated DNA was used to transform *E. coli* DH5α competent cells, and ampicillin resistant, lactose negative (white) transformant colonies were then selected. Plasmid

Figure 3.3. Restriction map and strategy for sequencing the 3.2kb *Pvu*II fragment from pAU164. The shaded box represents the region of DNA from the original *Sau*3A partial insert in KC742 The sequence from pIJ941 (complete vector not shown) is indicated by the thin black lines on pAU164 and the sequence from ØC31 is represented by filled black boxes. The thin black lines on pAU169 represent the multiple cloning site region of pIJ2925. The other parts of pIJ2925 vector are not shown. Bars below the map indicate the *Sal*I and *Sma*I fragments that were subcloned and sequenced.


pAU164
EcoRV
BamHI/Sau3A
PvuII
PvuII
BamHI/Sau3A
PvuII
EcoRV
3.2kb
digest with PvuII
3.2kb PvuII fragment
ligated with SmaI digested
pIJ2925
BglII
PvuII/SmaI
BamHI/Sau3A
PvuII/SmaI
BglII
pAU169
HindIII
SalI
SalI
SalI
SmaI
SalI
SmaI
SalI
EcoRI
SmaI
SmaI
digested with HindIII
and EcoRI
digested with SalI
and SmaI
ligated with HindIII/EcoRI
digested pUC118 and pUC119
ligated with SalI and SmaI
digested M13mp18
digest with HindIII
and EcoRI
pAU177
pAU178
digested with SalI
and EcoRI
ligated with HindIII and EcoRI
digested pUC119
religated mixture
pAU189

DNA was isolated from each of the transformants and insert-containing recombinant plasmids were identified after digestion of the plasmid DNA with *Hind*III and *Eco*RI. Only one plasmid contained an insert of the expected size (3.2kb) and was designated pAU169 (see Figure 3.3).

To verify that this shortened fragment was still able to complement *S. coelicolor bldD* mutants, the insert from pAU169 was then further subcloned into pIJ486. pAU169 was digested with *Bgl*II to release the 3.2kb insert and the DNA was purified from the agarose gel as described before. After purification, this DNA was ligated to *Bgl*II-digested pIJ486. The ligated DNA was used to transform protoplasts of *S. coelicolor* 1169. Two of the resulting thiostrepton resistant *S. coelicolor* 1169 transformants showed a "partial" complementation phenotype: they produced some spores and pigmented antibiotics, but quite obviously not at wildtype levels. Because the *bldD* mutation was complemented at least to some degree, and, since it was thought that the partial complement was due to structural instability of the plasmid (as had been seen with the larger 4.5kb *Eco*RV fragment in pIJ486), the *Pvu*II subclone was selected for DNA sequence determination.

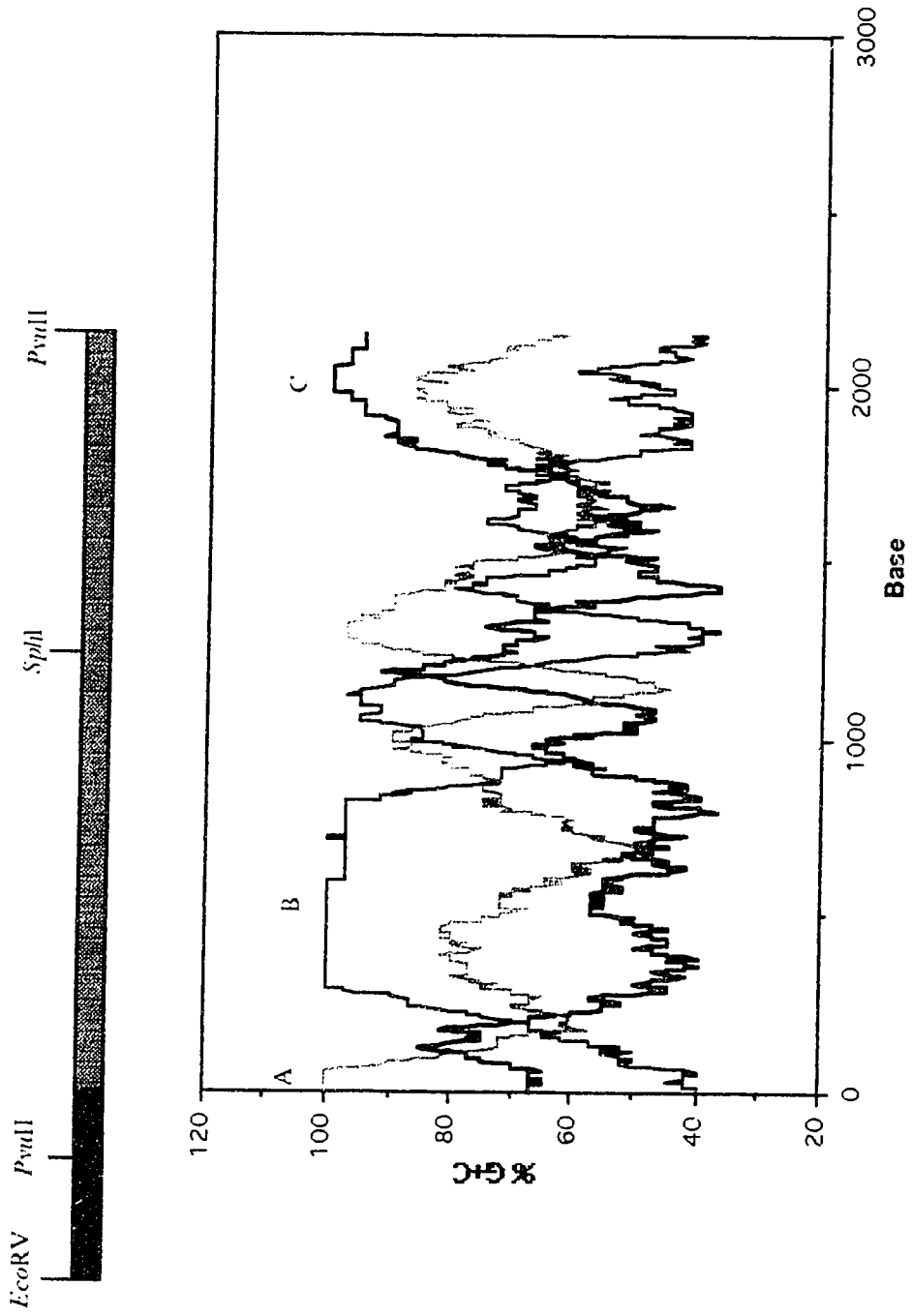
3.1.3 Subcloning the *Sal*I and *Sma*I fragments of pAU169 for DNA sequence analysis

To obtain fragments small enough to sequence, pAU169 was separately digested with *Sal*I and *Sma*I (Figure 3.3 shows the location of all the *Sal*I and *Sma*I sites), and each fragment was purified from a 5% polyacrylamide gel as described in section 2.6.3. Each fragment was subcloned into M13mp18 by shotgun cloning and appropriate subclones were identified after plasmid isolation. The M13mp18 recombinants containing the *Sal*I and *Sma*I fragments were then digested with *Hind*III and *Eco*RI and the digestion products were subcloned into pUC119 in order to obtain

the inserts in the opposite orientation. Appropriate subclones were again identified after plasmid isolation. The insert DNA from pAU169 was also transferred as an *Eco*RI/*Hind*III fragment into pUC118 and pUC119, in order to sequence the ends of the fragment. These clones were designated pAU177 and pAU178 respectively (see Figure 3.3). To sequence the end of the *Pvu*II fragment that corresponded to the *hldD*-containing sequence, rather than the ØC31 sequence, pAU177 was digested with *Sal*I and *Eco*RI and then religated. [This generated a mixture of *SalI* fragments and a *SalI*/*Eco*RI fragment which would be the only fragment capable of ligating back into the *SalI*/*Eco*RI -digested vector; see the location of the restriction sites in Figure 3.3.] The ligation mixture was used to transform *E. coli* XL1-Blue competent cells and ampicillin resistant, lactose negative transformant colonies were screened as described before. The resulting subclone, designated pAU189, was then sequenced using the method outlined in section 2.9.

Computer analysis of the nucleotide sequence using a modification of the FRAME (Bibb *et al.*, 1984) program was used to identify protein-coding sequences. Protein-coding sequences with a high overall G+C composition, as is the case in *Streptomyces*, have a nonrandom G+C distribution at each of the three positions, where the third position shows a high G+C bias and the second position always has a lower G+C content than the first codon position. Inspection of the resulting nucleotide sequence indicated at least two potential open reading frames (ORFs) (see Figure 3.4). Although the ORFs were not free of sequencing errors, the DNA sequence was translated into protein sequence and a database search (see section 2.1) was conducted to look for similarity to other known proteins. The partial ORFs, labeled A and B in Figure 3.4, showed a high degree of similarity to aspartate carbamoyltransferase suggesting that the two partial ORFs originated from a single ORF, the sequence of which contains several frameshift sequencing errors. Since

Figure 3.4. Frame analysis of the sequence 3.2kb *Pvu*II fragment from pAU169. The G+C base composition at the first (black), second (blue), and third (red) nucleotides of a window of 120 nucleotides was scanned and the resulting computer-generated frame analysis profile is shown. Above the profile is a schematic diagram of the 3.2kb *Pvu*II fragment represented by the shaded box. The ØC31 sequence is indicated by the filled black box. The labels A, B and C represent the open reading frames discussed in the text.


EcoRV
PvuII
SphI
PvuII
A
B
C
%GC
Base
120
100
80
60
40
20
0
1000
2000
3000

bldD was unlikely to encode an enzyme involved in pyrimidine biosynthesis, it was presumed that the partial ORF transcribed in the rightward direction corresponded to *bldD* (labeled C in Figure 3.4). The partial complementation observed with this *Pvu*II fragment probably resulted from a recombination event in the chromosome. This conclusion was supported by the fact that attempts to isolate plasmid DNA from the pAU169-containing transformants were not successful.

3.1.4 Complementation of a *bldD* mutant using the right-hand region of the 4.5kb *Eco*RV fragment

As discussed above, it was previously shown that the *Pvu*II fragment from pAU169, when cloned into pIJ486, was able to partially complement a *bldD* mutant. Sequencing of the *Pvu*II fragment revealed that a truncated ORF at the righthand end of the fragment probably corresponded to *bldD*. In order to verify this, a fragment which contained this entire putative *bldD* ORF was isolated to see if it was able to fully complement a *bldD* mutant.

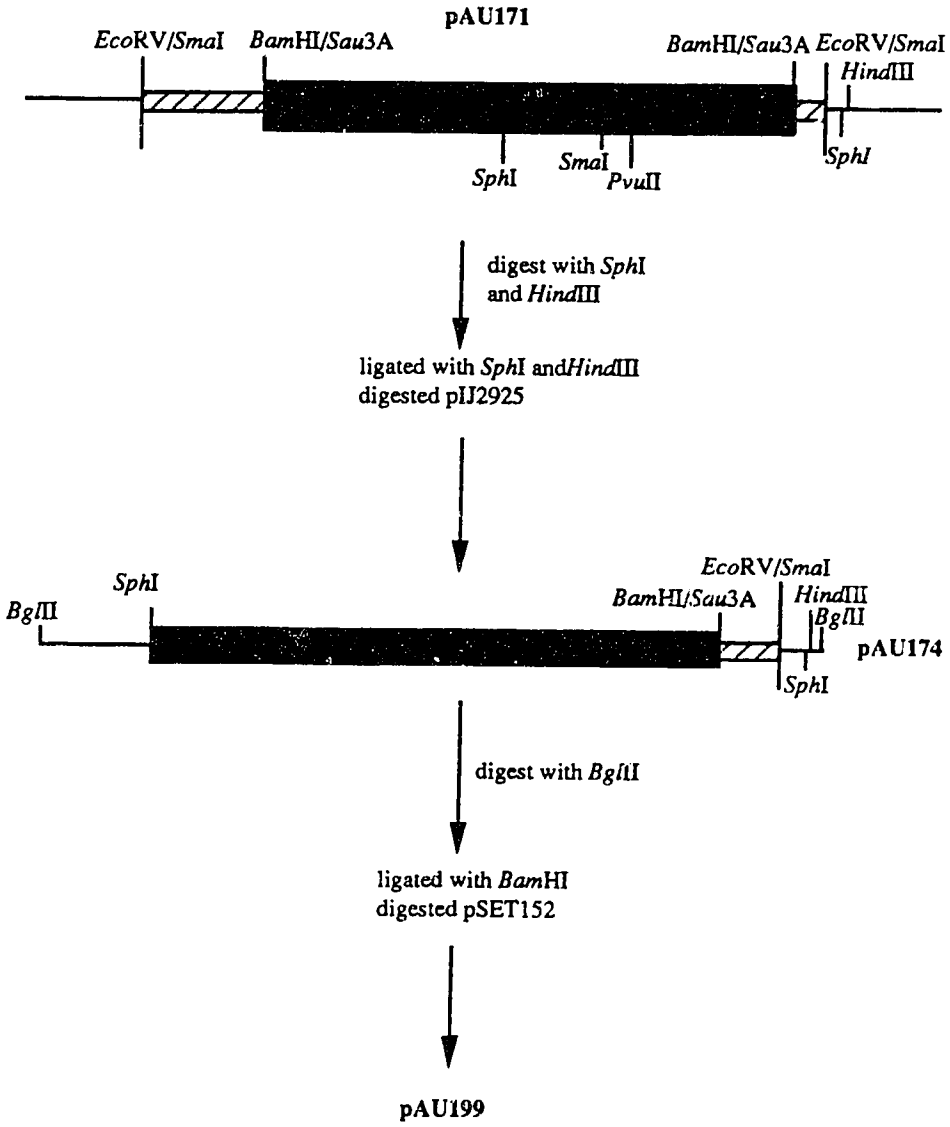
Because of the difficulty in isolating large amounts of insert DNA from the low copy number vector, pIJ941, attempts were made to transfer the entire 4.5kb *Eco*RV fragment from the pIJ941 recombinant pAU164 into the high copy number *E. coli* vector, pIJ2925. pAU164 was digested with *Eco*RV to release the insert, electrophoresed on a 1% agarose gel, and the 4.5kb fragment was purified from the gel as described in section 2.6.2. After purification, the DNA was ligated to *Sma*I-digested pIJ2925. The ligated DNA was then used to transform *E. coli* DH5α competent cells. Ampicillin resistant, lactose negative transformant colonies were selected and plasmid DNA was isolated from each of the transformants. Insert containing recombinant plasmids were identified after digestion of the plasmid DNA with *Bgl*II, an enzyme that flanks both ends of the multiple cloning site. Only one

plasmid contained the insert of the expected size (4.5kb), and it was designated pAU171.

To obtain a fragment which encompassed the putative *bldD* ORF, pAU171 was digested with *Sph*I and *Hind*III to release a 2.2kb fragment from the righthand end of the cloned insert (see Figure 3.5). This digestion should have released a 2.2kb *SphI* fragment but the *SphI* probably did not cleave at the *SphI* site immediately adjacent to the *HindIII* site and therefore resulted in the formation of a 2.2kb *SphI/HindIII* fragment. It is thought that the plasmid was partially digested because it was digested using both enzymes [illegible]sly. This fragment was purified in the same manner described above a[illegible] [illegible]s then ligated to the vector pIJ2925 digested with the same enzym[illegible] [illegible]a DNA was used to transform *E. coli* DH5α competent cells as above. Recombinant plasmids were screened by digesting the isolated plasmid with *Hind*III and *Eco*RI. Ten plasmids contained the 2.2kb insert. One of the ten, designated pAU174, was chosen for further study.

Once the *SphI/Hind*III fragment (2.2kb) was subcloned into pIJ2925, it was easier to transfer this fragment into a vector suitable for complementation in a *S. coelicolor bldD* mutant. pAU[illegible]4 was digested with *Bgl*II to release the insert and then ligated to a *Bam*HI-digested *E. coli-Streptomyces* shuttle vector pSET152 (Bierman *et al.*, 1992). The vector pSET152 consists of an *E. coli* colE1 replicon for replication in *E. coli* and apramycin resistance gene (apralan can be used as an alternative to apramycin) for selection in *E. coli* and for selection in *Streptomyces.* The vector lacks a *Streptomyces* replicon but contains the ØC31 phage integration and *att* site regions for integration into the *Streptomyces* chromosome. The ligated DNA was used to transform *E. coli* DH5α competent cells. Insert-containing plasmids were identified as described above, and one clone, designated pAU199,

Figure 3.5. Strategy for subcloning the 2.2kb *Sph*I/*Hind*III fragment from pAU171 into pIJ2925 and then into pSET152. The filled black box represents the *bldD*-containing fragment. The hatched box represents ØC31 sequence. pIJ2925 vector DNA is represented by the thin line. The entire pIJ2925 vector is not shown.


pAU171
EcoRV/SmaI
BamHI/Sau3A
BamHI/Sau3A
EcoRV/SmaI
HindIII
SphI
SmaI
PvuII
SphI
digest with SphI
and HindIII
ligated with SphI andHindIII
digested pIJ2925
SphI
BglII
EcoRV/SmaI
BamHI/Sau3A
HindIII
BglII
pAU174
SphI
digest with BglII
ligated with BamHI
digested pSET152
pAU199

which contained the 2.2kb insert was then used to transform a *dam*⁻ *dcm*⁻ *E. coli* strain, ET12567. Plasmid DNA was isolated and used to transform protoplasts of the *bldD* mutant *S. coelicolor* 1169. Apralan resistant transformants which complemented the *bldD* mutant developed on the R2YE plates, proving that the *bldD* gene was on the *Sph*I and *Hind*III fragment and that the beginning of an open reading frame found on pAU169 likely corresponded to the 5' region of the *bldD* gene. The partial complementation seen after transformation with the plasmid containing the 3.2kb *Pvu*II fragment (pAU169) could have resulted from partial function of a truncated version of the protein containing two-thirds of the putative helix-turn-helix motif. By sequencing the mutant *bldD* gene, the location and number of mutations could be determined.

3.1.5 Hybridization analysis to confirm the existence of the *bldD* gene at the *att* site in the chromosome of a recombinant pSET152-containing transformant

Before proceeding to determine the DNA sequence of the remaining insert DNA, attempts were made to confirm that the insert from the pAU199 had integrated into the chromosome at the *att* site and had complemented the *bldD* mutation *in trans* rather than by recombinational repair. To begin with, a probe internal to the putative *bldD* ORF was obtained. pAU174 was digested with *Sma*I and *Pvu*II and the DNA fragments (refer to Figure 3.5 for the location of the sites) were separated on a 5% polyacrylamide gel as described in section 2.5.1. A 215bp fragment was then purified from the gel using the method outline in section 2.6.3. The fragment was then labeled with ^{32}P as described in section 2.7.2.

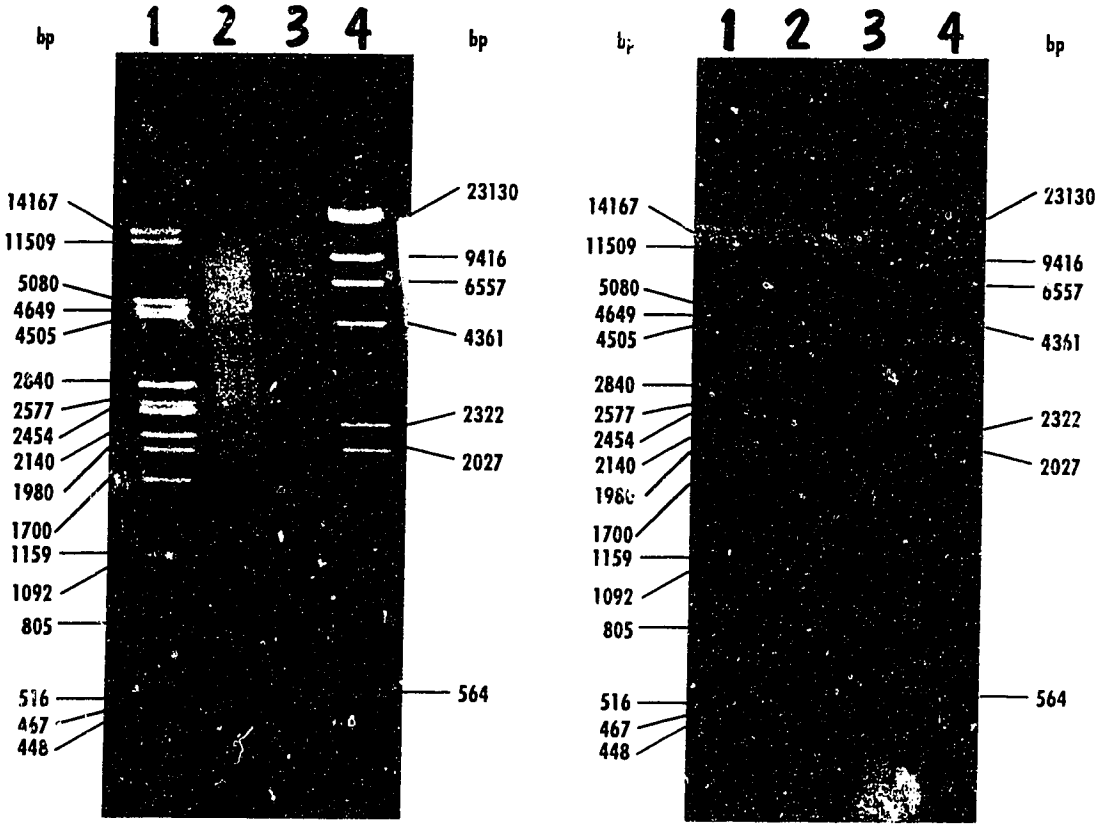
Chromosomal DNA of *S. coelicolor* 1169 and *S. coelicolor* 1169 transformed with the pAU199, containing the 2.2kb *SphI/Hind*III fragment, described in section

3.1.4, was digested with *Sph*I. Following separation by agarose gel electrophoresis, the DNA fragments were transferred to a nylon filter according to the method described in section 2.7.1 and the nylon filter was hybridized with the ^{32}P-labeled probe using the hybridization procedure described in section 2.7.3. In Plate 3.1 a single weak hybridizing band in the lane containing only *S. coelicolor* 1169 chromosomal DNA is seen, as well as a single weak band and a strong band in the lane containing the *S. coelicolor* 1169+pAU199 total DNA. To interpret these results, it should be noted that plasmids that can integrate site-specifically at the ØC31 attachment site (*att* site), such as pSET152, can do so very efficiently. Such plasmids that contain homologous DNA also integrate at the *att* site, with no detectable integration by insert-directed homologous recombination (Bierman *et al.*, 1992). Given this, it is most likely that the pSET152 clone integrated into the *S. coelicolor* 1169 chromosome at the *att* site, resulting in two copies of the *bldD* gene, one wildtype and the other mutated. The 2.2kb *Sph*I/*Hind*III fragment in pSET152 contains an additional *Sph*I site immediately before the *Hind*III site obtained by subcloning from pAU171 (4.5kb *Eco*RV fragment in pIJ2925) (see Figure 3.5). Therefore, digestion with *Sph*I would result in the release of the 2.2kb insert. The results from the hybridization analysis confirm the presence of a 2.2kb fragment, only present in the transformed *S. coelicolor* chromosomal DNA, shown by the strong band in lane 3. Both *S. coelicolor* 1169 and *S. coelicolor* 1169 transformed with pAU199 showed a weak band 4.5kb in size, which represents the *Sph*I fragment containing the chromosomal copy of the *bldD* gene. This analysis showed that pAU199 had integrated into the chromosome at the *att* site and had complemented the *bldD* mutation.

Plate 3.1: Agarose gel electrophoresis and autoradiogram of the *S. coelicolor* chromosomal DNA blot after hybridization with the ^{32}P-labeled, 215bp *bldD* probe.

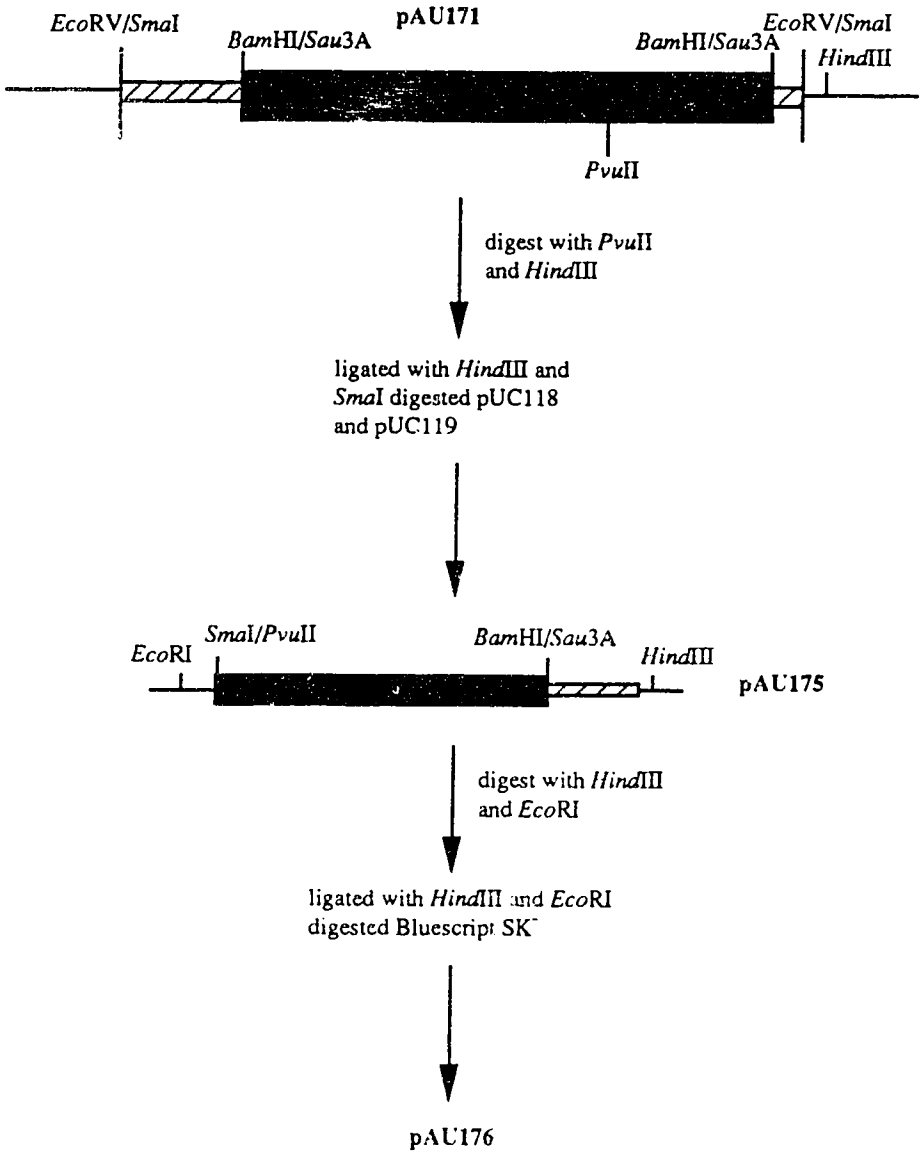
a. Aliquots of *S. coelicolor* 1169 and *S. coelicolor* 1169 + pAU199 total DNA were digested with *SphI* and then subjected to electrophoresis on a 1% agarose gel. Lanes 1 and 4 contained λ DNA digested with *HindIII* and *PstI* respectively as size markers (sizes are shown). Lane 2 contained *S. coelicolor* 1169 DNA and lane 3 contained *S. coelicolor* 1169 + pAU199 DNA.

b. Autoradiogram of the Southern transfer of the DNA from the agarose gel after hybridization with the ^{32}P-labeled, 215bp *bldD* probe. Following electrophoresis the DNA fragments were transferred to a nylon filter and the filter was exposed to the ^{32}P-labeled probe. The autoradiogram is in the same orientation as the gel in 3.1a.


bp
1
2
3
4
bp
14167
11509
5080
4649
4505
2840
2577
2454
2140
1980
1700
1159
1092
805
516
467
448
23130
9416
6557
4361
2322
2027
564
bp
1
2
3
4
bp
14167
11509
5080
4649
4505
2840
2577
2454
2140
1700
1159
1092
805
516
467
448
23130
9416
6557
4361
2322
2027
564

fragment contained the 3' end of the *bldD* gene. Figure 3.6 outlines the strategy for subcloning the 1kb *Pvu*II/*Hind*III fragment for sequencing. Initially pAU171 was digested with *Pvu*II and *Hind*III and the 1kb fragment was purified from the agarose gel as described in section 2.6.2. The DNA was ligated to the *Sma*I/*Hind*III-digested vectors pUC118 and pUC119 in order to obtain clones containing the insert in both orientations. The ligation mixture was used to transform *E. coli* XL1-Blue competent

Figure 3.6. Strategy for subcloning the 1kb *Pvu*II/*Hin*dIII fragment from pAU171 into pUC118 and Bluescript SK$^-$. The filled black box represents the *bldD*-containing fragment and the hatched box represents ØC31 sequence. pIJ2925 DNA in pAU171 and pUC118 DNA in pAU175 are indicated by the thin black lines. The entire pIJ2925 and pUC118 vectors are not shown.

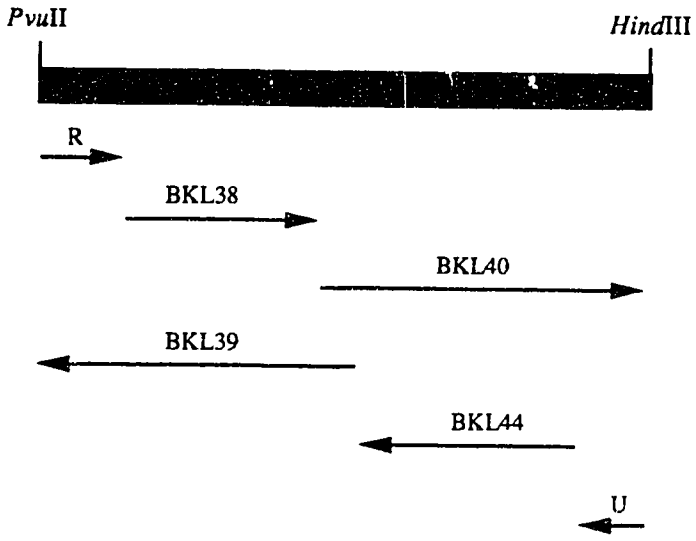

pAU171
EcoRV/SmaI
BamHI/Sau3A
BamHI/Sau3A
EcoRV/SmaI
HindIII
PvuII
digest with PvuII
and HindIII
ligated with HindIII and
SmaI digested pUC118
and pUC119
SmaI/PvuII
BamHI/Sau3A
EcoRI
HindIII
pAU175
digest with HindIII
and EcoRI
ligated with HindIII and EcoRI
digested Bluescript SK-
pAU176

pAU175 and pAU176 were then sequenced using the method described in section 2.9. The strategy for sequencing the 1kb *Pvu*II/*Hind*III fragment is outlined in Figure 3.7. The entire fragment was sequenced in both directions using either universal primer, reverse primer or specific oligonucleotide primers shown in Figure 3.7.

3.1.7 Subcloning the *Sal*I fragment spanning across the *Pvu*II junction

In order to verify that the original 4.5kb *Eco*RV fragment did not contain an additional small internal *Pvu*II fragment between the 3.2kb *Pvu*II and the 1kb *Pvu*II fragments, a subclone spanning the *Pvu*II junction was required. The subclone spanning the *Pvu*II junction was obtained by digesting pAU174 (containing the 2.2kb *Sph*I/*Hind*III fragment) with *Sal*I and ligating the two resulting 800bp *Sal*I fragments to the *Sal*I-digested M13mp18. The ligated DNA was transformed into *E. coli* XL1-Blue competent cells. Lactose negative (white) plaques were screened by digesting the double-stranded RF (replicative form) DNA with *Hind*III and *Eco*RI. To distinguish between the two *Sal*I fragments, the resulting recombinant plasmids were screened by sequencing. The recombinant mp18 derivative containing the fragment of interest was designated pAU179. The RF DNA from pAU179 was then digested with *Hind*III and *Eco*RI and transferred to pBluescript SK$^-$ digested with the same enzymes in order to obtain a clone with the fragment in the opposite orientation. Recombinant plasmids were screened by digesting the isolated plasmid DNA with *Hind*III and *Eco*RI and the recombinant plasmid which contained an insert of the expected size was designated pAU180. Both recombinant plasmids were then sequenced as outlined in section 2.9. The sequence analysis revealed that a small fragment of DNA was present between the 3.2kb and 1kb *Pvu*II fragments and this DNA fragment contained an additional *Pvu*II site.

Figure 3.7. Strategy for sequencing the *Pvu*II/*Hind*III fragment. Arrows below the map indicate the direction and extent of the sequence information determined by each primer. U: universal primer, R: reverse primer. Other specific primers are designated BKL.


*Pvu*II
*Hind*III
R
BKL38
BKL40
BKL39
BKL44
U

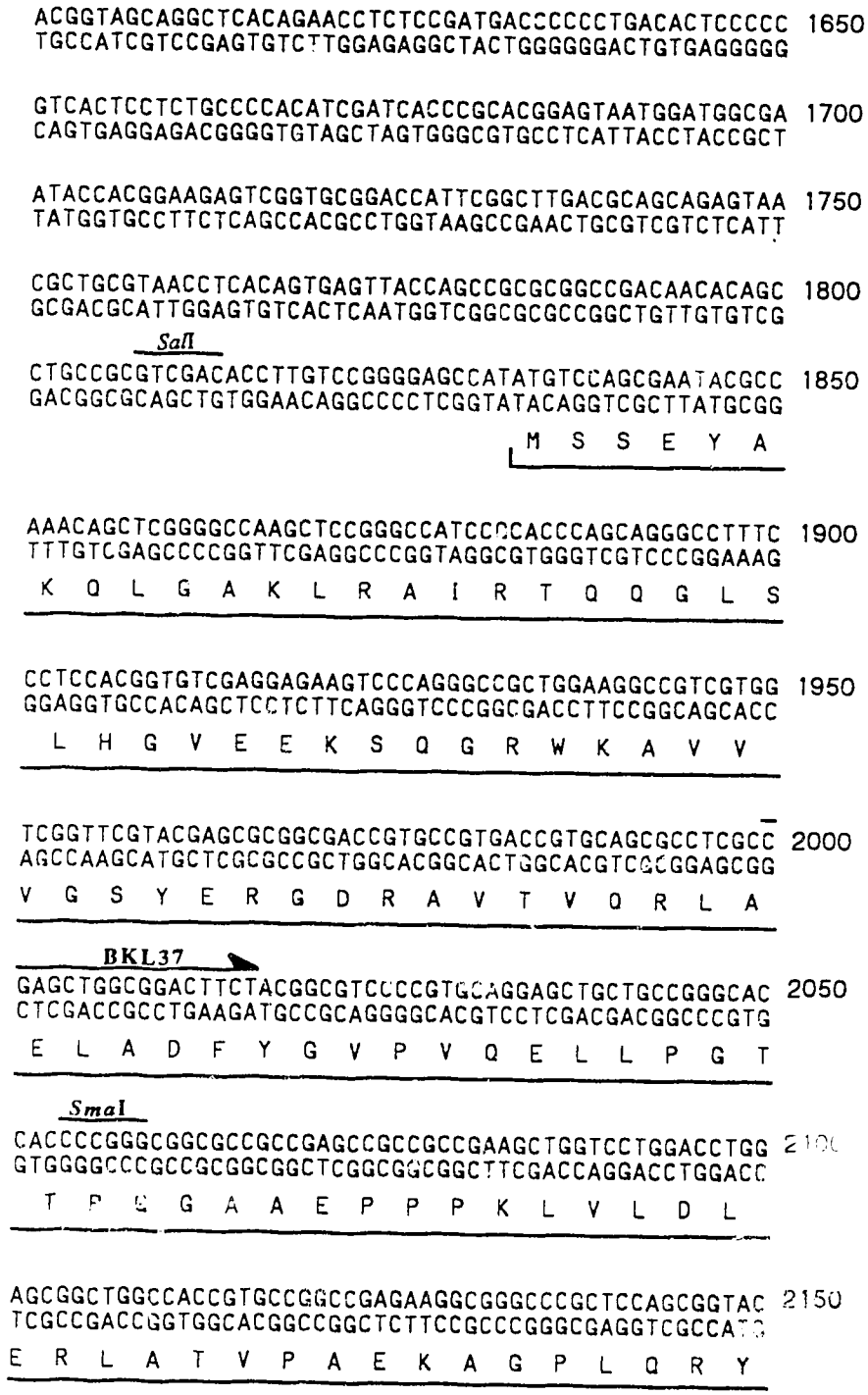
3.2 Sequence Analysis of the *bldD* gene

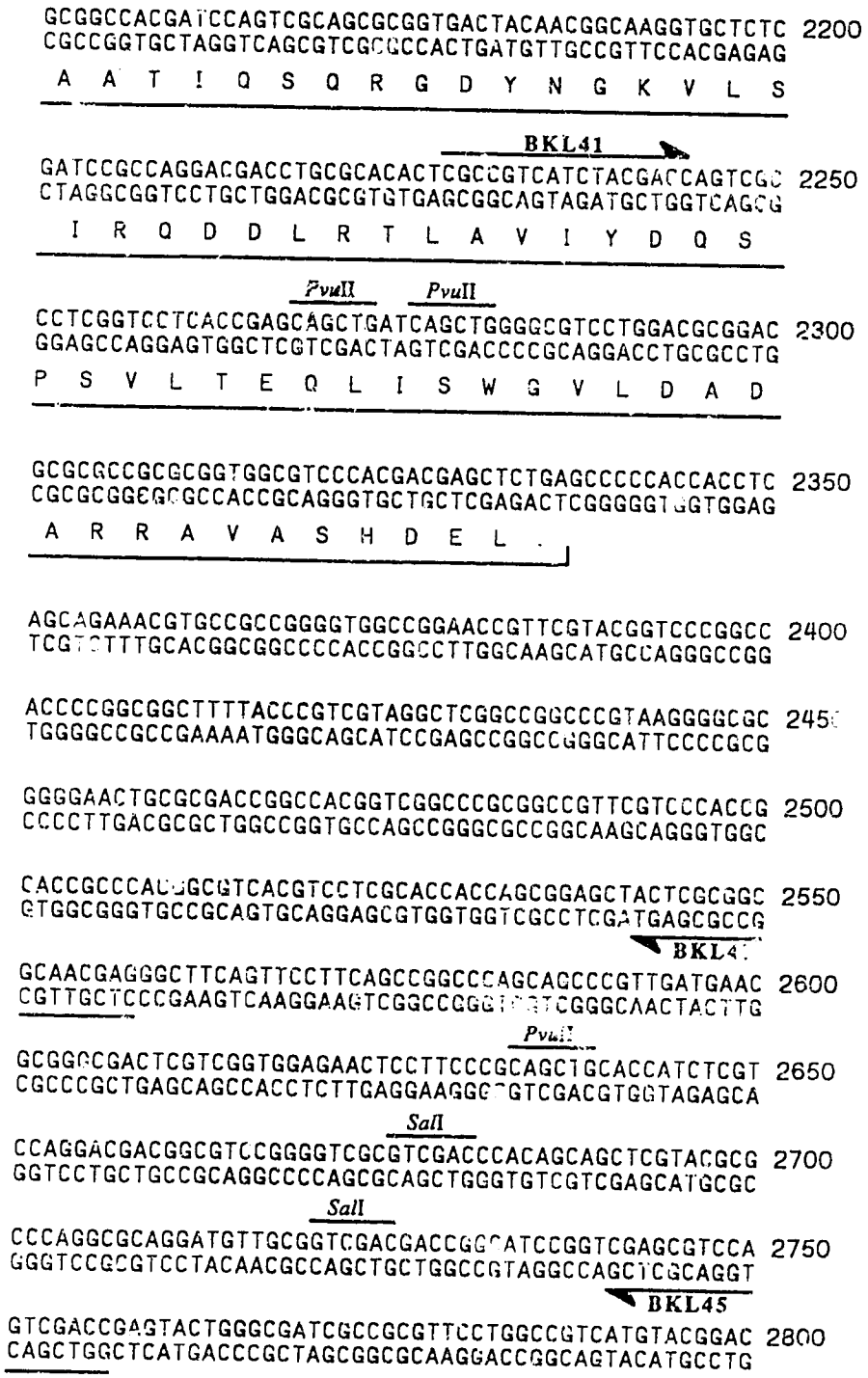
The entire nucleotide sequence of the 3.4kb of insert DNA obtained from the original 4.5kb *Eco*RV fragment which contained 1kb of flanking ØC31 vector DNA (not shown), is shown in Figure 3.8. A final computer analysis of the nucleotide sequence using the FRAME program was used to identify potential protein-coding sequences (Figure 3.9). This program identified one complete ORF, which on the basis of the *in trans* complementation analysis was shown to be the *bldD* gene, and at least two more incomplete ORFs. Although the region corresponding to the putative *bldD* gene was not subcloned on its own for the complementation analysis, the truncated ORF facing rightwards is not likely to be the *bldD* gene because integration of pAU199 (*Sph*I/*Hind*III fragment containing the *bldD* gene) at the ØC31 *att* site, would not have allowed recombinational repair of that particular ORF. As discussed above, the incomplete ORF transcribed in the leftward direction shows a high degree of similarity to aspartate carbamoyltransferase, while the truncated ORF transcribed in the rightward direction did not show similarity to any proteins in the database. The complete ORF corresponding to the *bldD* is 504 nucleotides extending from nt 1833-2336 (Figure 3.10). A potential ribosome binding site is found 4nt upstream from the ATG start codon. The coding sequence ends with a TGA and is followed by an inverted repeat and a stretch of T residues that could presumably function in transcription termination. In contrast to the sequence in general, note that the sequence of the inverted repeat could only be obtained for one strand. Figure 3.10 also shows the putative -10 and -35 regions identified by sequence gazing. Whether or not the putative promoter functions *in vivo* remains to be determined. Utilizing the DNA Strider program, translation of the ORF would result in a 18167 Dalton protein that was found to be very hydrophilic. The putative protein contains 22 basic and 21

Figure 3.8. Nucleotide sequence of the entire 3.4kb fragment and deduced amino acid sequence of the *bldD* gene. *Sal*I, *Sma*I, *Sph*I, and *Pvu*II restriction sites are shown. Arrows indicate the primer sequences and the direction the fragment was sequenced.

```
TCGTTGCCGCCGAGCAGCAGGTAGAGGACGGCCATCCGGATGGAGACGCC 50
AGCAACGGCGGCTCGTCGTCCATCTCCTGCCGGTAGGCCTACCTCTGCGG

GTTGGTGACCTGCTCGACGACGGTGCAGCGGTCGGAGTCGGCGACCTCGG 100
CAACCACTGGACGAGCTGCTGCCACGTCGCCAGCCTCAGCCGCTGGAGCC

CGGTGATCTCCATGCCGCGGACCATGGGGCCGGGGTGCATCACGATGGCG 150
GCCACTAGAGGTACGGCGCCTGGTACCCCGGCCCCACGTAGTGCTACCGC

TGGTCGGGCATCTTCGCCATCGGTCGCCGTCGAGGCCGTAGCGGCGGGAG 200
ACCAGCCCGTAGAAGCGGTAGCCAGCGGCAGCTCCGGCATCGCCGCCCTC

TACTCGCGCTCGGTGGGGAAGAACGCGGCGTTCATCGCTCGCGCTGCACG 250
ATGAGCGCGAGCCACCCCTTCTTGCGCCGCAAGTAGCGAGCGCGACGTGC

CGCAGCACGTCACCGCAGTCAGGACTTGGGCAGCAAGTGGAGTCGAGGTC 300
GCGTCGTGCAGTGGCGTCAGTCCTGAACCCGTCGTTCACCTCAGCTCCAG

GTAGGAGACCTCGCAGGGCCAGGTCTCGACGCCGACCCGCAGCAGGGTGG 350
CATCCTCTGGAGCGTCCCGGTCCAGAGCTGCGGCTGGGCGTCGTCCCACC
                                              SalI
GCGGGGCGACGAGGGTGACCTCGGCGCCGAGGGTGTGCAGCAGGTCGACG 400
CGCCCCGCTGCTCCCACTGGAGCCGCGGCTCCCACACGTCGTCCAGCTGC

TTGGAGCGGGCGACGCGGCTGTGCAGGATGTCGCCGACGAGGGTGATCCG 450
AACCTCGCCCGCTGCGCCGACACGTCCTACAGCGGCTGCTCCCACTAGGC

CCGGCCGTCCAGGTCCTTGCCGAGCCCGGCGTCGCGGCCGACCACCAGGC 500
CGCCGGCAGGTCCAGGAACGGCTCGGGCCGCAGCGCCGGCTGGTGGTCCG

GGCGGCGCATGGTGAAGGCGTCGAGCAGGGCCTGGGTGGGGTGCTGGTGG 550
CCGCCGCGTACCACTTCCGCAGCTCGTCCCGGACCCACCCCACGACCACC

GTGCCGTCCCCGGCGTTGACCACGGCGGCGTCGATCCAGCCGGAGGTGGC 600
CACGGCAGGGGCCGCAACTGGTGCCGCCGCAGCTAGGTCGGCCTCCACCG
                                              SalI
CAGGCGGTAGGGGGGCCCGGAGGCGCTGTGCCGGATGAGCACGGCGTCGA 650
GTCCGCCATCCCCCCGGGCCTCCGCGACACGGCCTACTCGTGCCGCAGCT

CGCCCATCGCCTCCAGGGTCTGGGCGGTGTCCTTGAGGGACTCGCCCTTG 700
GCGGGTAGCGGAGGTCCCAGACCCGCCACAGGAACTCCCTGAGCGGGAAC

GAGACGCTGGATCCCTTGGCGGTGAAGTTGATGACGTCCGCGGACAGCCG 750
CTCTGCGACCTAGGGAACCGCCACTTCAACTACTGCAGGCGCCTGTCGGC

CTTCTCGGCGGCCTCGAAGGAGATCCGGGTCCGGGTGGAGTCCTCGAAGA 800
GAAGAGCCGCCGGAGCTTCCTCTAGGCCCAGGCCCACCTCAGGAGCTTCT
```

```
AGAGGTTGACGACGGTGCGGCCGCGCAGGGTCGGCAGTTTCTTGATCGGC 850
TCTCCAACTGCTGCCACGCCGGCGCGTCCCAGCCGTCAAAGAACTAGCCG
          SmaI
CGGTCCGCGACCCGGGCCATCTCCTCGGCGGTGTCGAGGATCAGGACGGC 900
GCCAGGCGCTGGGCCCGGTAGAGGAGCCGCCACAGCTCCTAGTCCTGCCG

GTCGTCGCGGGTGAGGTCGGCGGCCGAGATGAGATGACGCTGCATCTGTC 950
CAGCAGCGCCCACTCCAGCCGCCGGCTCTACTCTACTGCGACGTAGACAG

AGGCTCCGTAAGGCGATTCATGCGGAAGAGCGGGATAATTCGGGCGGCCG 1000
TCCGAGGCATTCCGCTAAGTACGCCTTCTCGCCCTATTAAGCCCGCCGGC

GGCGCGTGCGGGGGCGCGCCGAGCACGCGTACGGCACGAAGGGGTGCGTA 1050
CCGCGCACGCCCCCGCGCGGCTCGTGCGCATGCCGTGCTTCCCCACGCAT
                    SmaI
CGCGGGGTGCTAGGGGTGGGCGCCCGGGGCGGCCGGCTTGGCACCGAGCA 1100
GCGCCCCACGATCCCCACCCGCGGGCCCCGCCGGCCGAACCGTGGCTCGT

GCACGGTGTCGCGACCGTCCTCCTCGCCGAGCTGGACCTTGACCGTCTCC 1150
CGTGCCACAGCGCTGGCAGGAGGAGCCGCTCGACCTGGAACTGGCAGAGG

CGCAGCGACGTCGGGAGGTTCTTGCCGACGTAGGATGGGCAGTTCGCGG( 1200
GCGTCGCTGCAGCCCTCCAAGAACGGCTGCATCCTACCCGTCAAGCGCCA
     SalI
GGCCGCGGTCGACGAGGACGACGAGGCGGCCGTCGATGCCGTCACCGGGG 1250
CCGGCGCCAGCTGCTCCTGCTGCTCCGCCGGCAGCTACGGCAGTGGCCCC
                                  SphI
ATCTCGGTGCGGGCCAGGGCGGCGCGGCGGATGCATGCGCAGGTCGTCGC 1300
TAGAGCCACGCCCGGTCCCGCCGCGCCGCCTACGTACGCGTCCAGCAGCG

GGTACATGGTGATGTCCAGCGAGCCGACCGGCATCTTGCGTTCGGTGATC 1350
CCATGTACCACTACAGGTCGCTCGGCTGGCCGTAGAACGCAAGCCACTAG
                                  SmaI
TGCTCCAGCTTGTCGGCGAGCCGCCGGGCGAGGAAGACGCCCCGGGTCGG 1400
ACGAGGTCGAACAGCCGCTCGGCGGCCCGCTCCTTCTGCGGGGCCCAGCC

GATGCCGAGGAGCACCACGTCGTCGGCGCCCTTGGCGTTCGACGATCTCG 1450
CTACGGCTCCTCGTGGTGCAGCAGCCGCGGGAACCGCAAGCTGCTAGAGC

TGGGCGATGCGGGTCAGCACCCGCGCGATGTCGGGCCCTTCGAGAACGGG 1500
ACCCGCTACGCCCAGTCGTGGGCGCGCTACAGCCCGGGAAGCTCTTGCCC

CCGGGCTTCCTGCTGCTGGTCCTGTTGCTTGTCCATACGAAACGGACCCC 1550
GGCCCGAAGGACGACGACCAGGACAACGAACAGGTATGCTTTGCCTGGGG

CTTCTCCGCCTCACGGGACGGACCTTAAAGGACGTCGGATATGCGCCACT 1600
GAAGAGGCGGAGTGCCCTGCCTGGAATTTCCTGCAGCCTATACGCGGTGA
```


ACGGTAGCAGGCTCACAGAACCTCTCCGATGACCCCCCTGACACTCCCCC 1650
TGCCATCGTCCGAGTGTCTTGGAGAGGCTACTGGGGGGACTGTGAGGGGG
GTCACTCCTCTGCCCCACATCGATCACCCGCACGGAGTAATGGATGGCGA 1700
CAGTGAGGAGACGGGGTGTAGCTAGTGGGCGTGCCTCATTACCTACCGCT
ATACCACGGAAGAGTCGGTGCGGACCATTCGGCTTGACGCAGCAGAGTAA 1750
TATGGTGCCTTCTCAGCCACGCCTGGTAAGCCGAACTGCGTCGTCTCATT
CGCTGCGTAACCTCACAGTGAGTTACCAGCCGCGCGGCCGACAACACAGC 1800
GCGACGCATTGGAGTGTCACTCAATGGTCGGCGCGCCGGCTGTTGTGTCG
SalI
CTGCCGCGTCGACACCTTGTCCGGGGAGCCATATGTCCAGCGAATACGCC 1850
GACGGCGCAGCTGTGGAACAGGCCCCTCGGTATACAGGTCGCTTATGCGG
M S S E Y A
AAACAGCTCGGGGCCAAGCTCCGGGCCATCCCCACCCAGCAGGGCCTTTC 1900
TTTGTCGAGCCCCGGTTCGAGGCCCGGTAGGCGTGGGTCGTCCCGGAAAG
K Q L G A K L R A I R T Q Q G L S
CCTCCACGGTGTCGAGGAGAAGTCCCAGGGCCGCTGGAAGGCCGTCGTGG 1950
GGAGGTGCCACAGCTCCTCTTCAGGGTCCCGGCGACCTTCCGGCAGCACC
L H G V E E K S Q G R W K A V V
TCGGTTCGTACGAGCGCGGCGACCGTGCCGTGACCGTGCAGCGCCTCGCC 2000
AGCCAAGCATGCTCGCGCCGCTGGCACGGCACTGGCACGTCGCGGAGCGG
V G S Y E R G D R A V T V Q R L A
BKL37
GAGCTGGCGGACTTCTACGGCGTCCCCGTGCAGGAGCTGCTGCCGGGCAC 2050
CTCGACCGCCTGAAGATGCCGCAGGGGCACGTCCTCGACGACGGCCCGTG
E L A D F Y G V P V Q E L L P G T
SmaI
CACCCCGGGCGGCGCCGCCGAGCCGCCGCCGAAGCTGGTCCTGGACCTGG 2100
GTGGGGCCCGCCGCGGCGGCTCGGCGGCGGCTTCGACCAGGACCTGGACC
T P G G A A E P P P K L V L D L
AGCGGCTGGCCACCGTGCCGGCCGAGAAGGCGGGCCCGCTCCAGCGGTAC 2150
TCGCCGACCGGTGGCACGGCCGGCTCTTCCGCCCGGGCGAGGTCGCCATG
E R L A T V P A E K A G P L Q R Y


GCGGCCACGATCCAGTCGCAGCGCGGTGACTACAACGGCAAGGTGCTCTC 2200
CGCCGGTGCTAGGTCAGCGTCGCGCCACTGATGTTGCCGTTCCACGAGAG
A A T I Q S Q R G D Y N G K V L S
BKL41
GATCCGCCAGGACGACCTGCGCACACTCGCCGTCATCTACGACCAGTCGC 2250
CTAGGCGGTCCTGCTGGACGCGTGTGAGCGGCAGTAGATGCTGGTCAGCG
I R Q D D L R T L A V I Y D Q S
PvuII PvuII
CCTCGGTCCTCACCGAGCAGCTGATCAGCTGGGGCGTCCTGGACGCGGAC 2300
GGAGCCAGGAGTGGCTCGTCGACTAGTCGACCCCGCAGGACCTGCGCCTG
P S V L T E Q L I S W G V L D A D
GCGCGCCGCGCGGTGGCGTCCCACGACGAGCTCTGAGCCCCCACCACCTC 2350
CGCGCGGCGCCGCCACCGCAGGGTGCTGCTCGAGACTCGGGGGTGGTGGAG
A R R A V A S H D E L .
AGCAGAAACGTGCCGCCGGGGTGGCCGGAACCGTTCGTACGGTCCCGGCC 2400
TCGTCTTTGCACGGCGGCCCCACCGGCCTTGGCAAGCATGCCAGGGCCGG
ACCCCGGCGGCTTTTACCCGTCGTAGGCTCGGCCGGCCCGTAAGGGGCGC 245
TGGGGCCGCCGAAAATGGGCAGCATCCGAGCCGGCCGGGCATTCCCCGCG
GGGGAACTGCGCGACCGGCCACGGTCGGCCCGCGGCCGTTCGTCCCACCG 2500
CCCCTTGACGCGCTGGCCGGTGCCAGCCGGGCGCCGGCAAGCAGGGTGGC
CACCGCCCACGGCGTCACGTCCTCGCACCACCAGCGGAGCTACTCGCGGC 2550
GTGGCGGGTGCCGCAGTGCAGGAGCGTGGTGGTCGCCTCGATGAGCGCCG
BKL47
GCAACGAGGGCTTCAGTTCCTTCAGCCGGCCCAGCAGCCCGTTGATGAAC 2600
CGTTGCTCCCGAAGTCAAGGAAGTCGGCCGGGTCGTCGGGCAACTACTTG
PvuII
GCGGCCGACTCGTCGGTGGAGAACTCCTTCCCGCAGCTGCACCATCTCGT 2650
CGCCCGCTGAGCAGCCACCTCTTGAGGAAGGCGTCGACGTGGTAGAGCA
SalI
CCAGGACGACGGCGTCCGGGGTCGCGTCGACCCACAGCAGCTCGTACGCG 2700
GGTCCTGCTGCCGCAGGCCCCAGCGCAGCTGGGTGTCGTCGAGCATGCGC
SalI
CCCAGGCGCAGGATGTTGCGGTCGACGACCGGCATCCGGTCGAGCGTCCA 2750
GGGTCCGCGTCCTACAACGCCAGCTGCTGGCCGTAGGCCAGCTCGCAGGT
BKL45
GTCGACCGAGTACTGGGCGATCGCCGCGTTCCTGGCCGTCATGTACGGAC 2800
CAGCTGGCTCATGACCCGCTAGCGGCGCAAGGACCGGCAGTACATGCCTG

BKL38 →

AAGTCGCTCTGGTCGCGCATGACATGGCCCATCGTCAGGTGTTCCGTCGC 2850
TTCAGCGAGACCAGCGCGTACTGTACCGGGTAGCAGTCCACAAGGCAGCG

← BKL39

CGTCGGGCCAGCGAGCTGTCCGGACCGATCGCCGGCGCGTCGATCGGCAT 2900
GCAGCCCGGTCGCTCGACAGGCCTGGCTAGCGGCCGCGCAGCTAGCCGTA

GAGTTACGGCTGGTGGCAGGACAAGCACACCCGTCACCATGCCAACCCCA 2950
CTCAATGCCGACCACCGTCCTGTTCGTGTGGGCAGTGGTACGGTTGGGGT

ACACCGAGGACCTCGGACATCGGCCCCGACCTGCTCGTCTGGTCC 3000
TGTGGCTCCTCGAGCTGTGTAGCCGGGGCTGGACGAGCAGACCAGG

CCGGACCAGCCGTGCCGCCACCGGACTGCCCCGCCTCCTCGGCCGCTG 3050
GGCCTGGTCGGCACGGCGGTGGCCTGACGGGGCGGAGGAGCCGGCGAC

GCAGGCGTTCCTGGCTTTCCCCTGCTCACGCTGGAGGGCTTCAATCTGC 3100
CGTCCGCAAGGACAAGAAAGGGGACGAGTGCGACCTCCCGAAGTTAGACG

BKL40 →

ACGTTGCGAGCGGCAGGGCCATGGCCAACCGTCGGCTCAAACGCCCGCA 3150
TGCAACGCTCGCCGTCCCGGTACCGGTTGGCAGCCGAGTTTGCGGCCCGT

CTGGACGGCGCTCTGCTCCTCGCGCACTGCGCCGTCTACCTGACCGCCCT 3200
GACCTGCCGCGAGACGAGGAGCGCGTGACGCGGCAGATGGACTGGCGGGA

GTTCTGGGCTCCTGCCGCCCGGCATGGCCATCGCCTTCCTCGCCGTCCAC 3250
CAAGACCCGAGGACGGCGGGCCGTACCGCTAGCGGAAGGAGCGGCAGGTG

CAGTGCCTGTTCGGCGTCTACCTCGGTTCGGCCTTCGCCCCCAACCACAA 3300
GTCACGGACAAGCCGCAGATGGAGCCAAGCCGGAAGCGGGGGTTCGTGTT

GGGGATGCCGATCCTGACGGCCGACGACCGCCCCGACTTCCTCCGCCGCC 3350
CCCCTACGGCTAGGACTGCCGGCTGCTGGCGGGGCTGAAGGAGGCGGCGG

AGGTGCTCACCTCACGCAACGTCAACGGCGGCCTGTTCACCGACCTGGCG 3400
TCCACGAGTGGAGTGCGTTGCAGTTGCCGCCGGACAAGTGGCTGAACCGC

← BKL44

CTCGGCGGCCTGAACCACCAGATCC 3425
GAGCCGCCGGACTTGGTGGTCTAGG

Figure 3.9. Frame analysis of the entire 3.4kb sequence. The G+C base composition at the first (black), second (blue), and third (red) nucleotides of a window of 120 nucleotides was scanned and the resulting computer-generated frame analysis profile is shown. Arrows represent the direction of each ORF. Above the profile is a schematic diagram of the 3.4kb fragment represent by the shaded box. The hatched box represents the *bldD* gene, on which the ATG and TGA are indicated and the ØC31 sequence is represented by the filled black box.

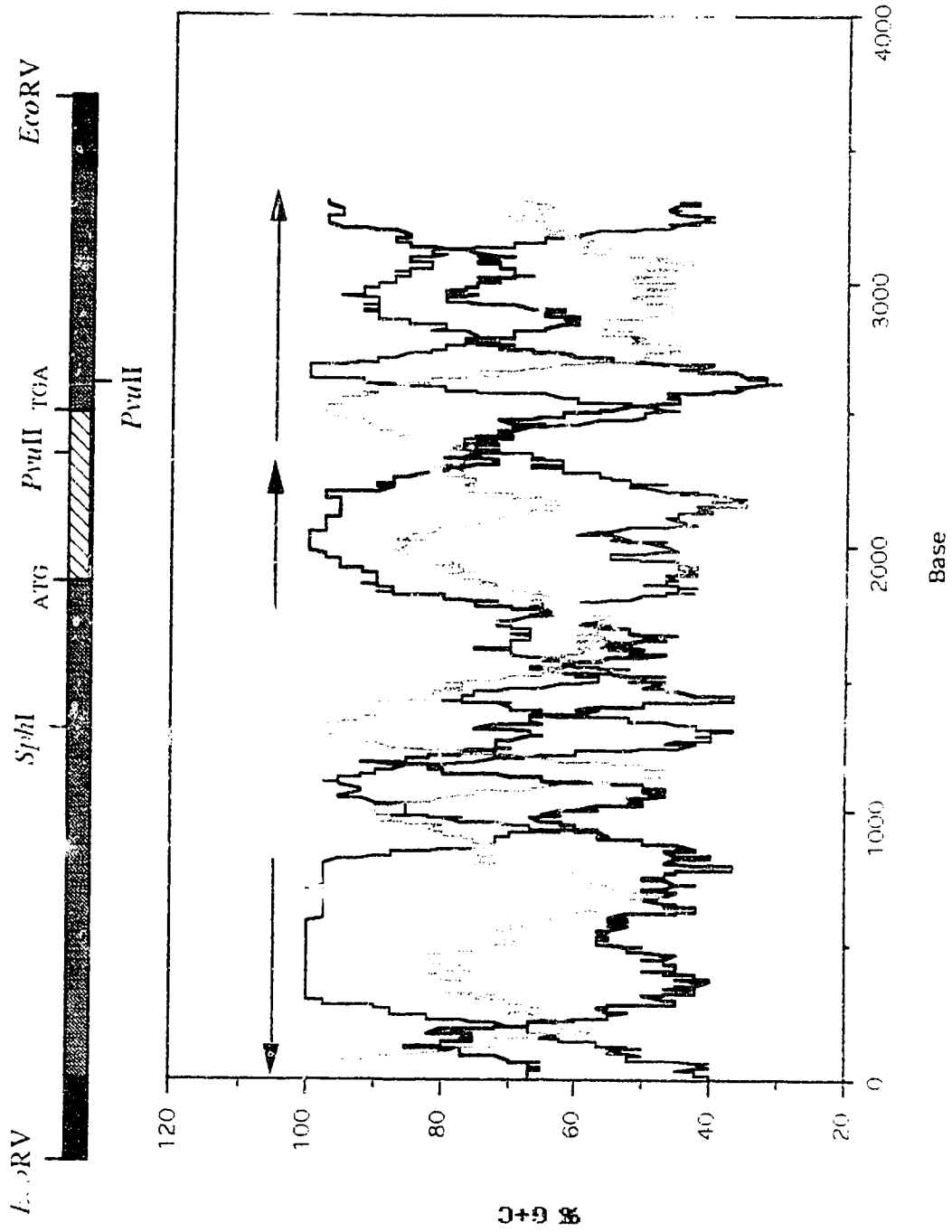
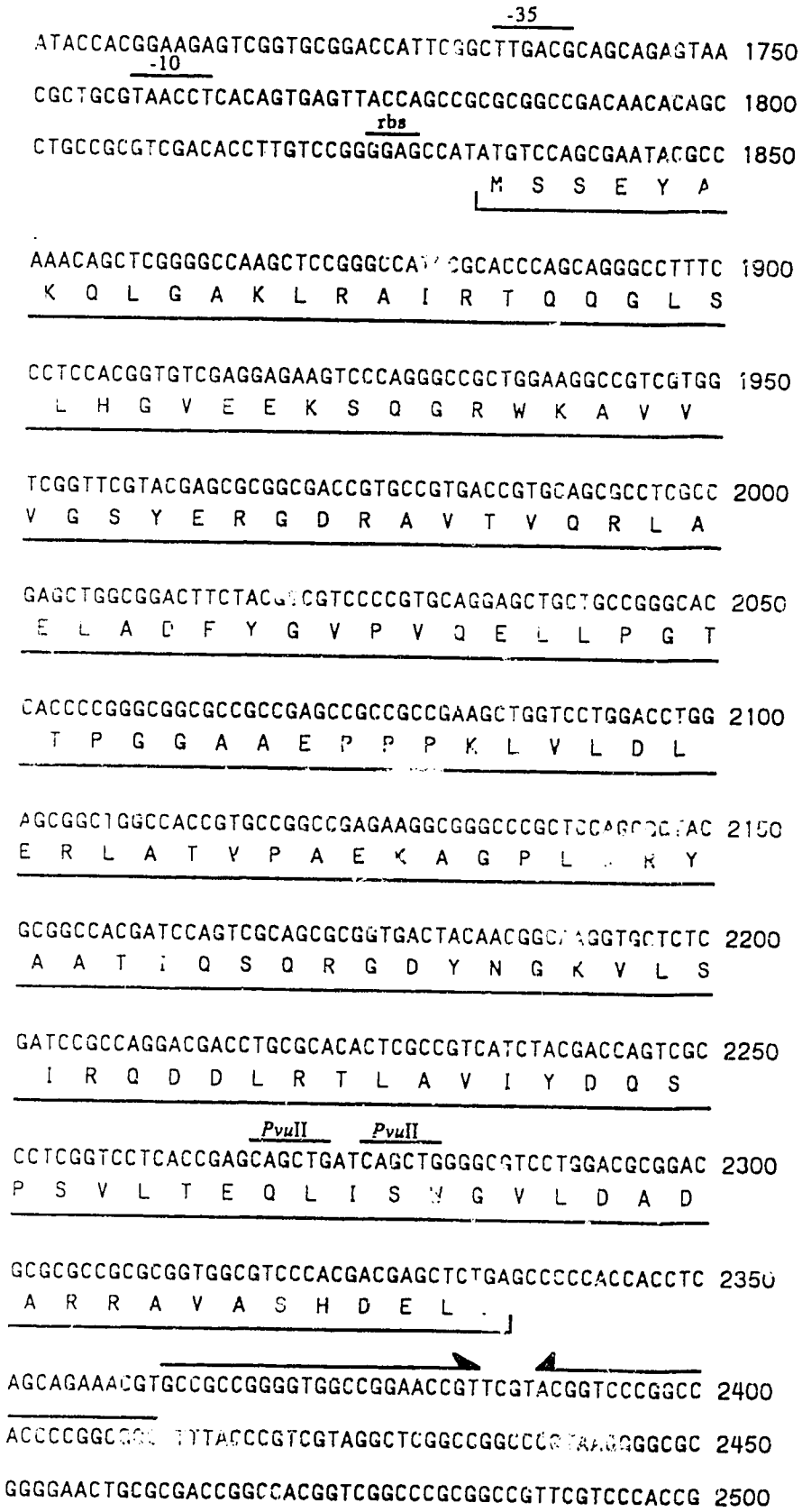

EcoRV
PvuII
TGA
ATG
SphI
EcoRV
PvuII
% G+C
120
100
80
60
40
20
0
1000
2000
3000
4000
Base

Figure 3.10. Nucleotide sequence and amino acid sequence of the *bldD* gene, along with approximately 100 nucleotides upstream and downstream of the protein-coding region. The putative -10 and -35 regions and the potential ribosome binding site (rbs) are indicated on the sequence. Convergent arrows mark an inverted repeat. The predicted amino acid sequence is given directly below. The one letter amino acid code was used and is as follows: A, alanine; R, arginine; N, asparagine; D, aspartate; C, cysteine; E, glutamate; Q, glutamine; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T threonine; W, tryptophan; Y, tyrosine; V, valine; .; stop codon.


-35
-10
rbs
M S S E Y A
K Q L G A K L R A I R T Q Q G L S
L H G V E E K S Q G R W K A V V
V G S Y E R G D R A V T V Q R L A
E L A D F Y G V P V Q E L L P G T
T P G G A A E P P P K L V L D L
E R L A T V P A E K A G P L R Y
A A T I Q S Q R G D Y N G K V L S
I R Q D D L R T L A V I Y D Q S
PvuII
PvuII
P S V L T E Q L I S W G V L D A D
A R R A V A S H D E L .
1750
1800
1850
1900
1950
2000
2050
2100
2150
2200
2250
2300
2350
2400
2450
2500

acidic amino acids and no stretches of hydrophobic amino acids. This suggests that it is a cytoplasmic rather than a membrane-bound protein.

The predicted amino acid sequence of the ORF was compared to other proteins in the databases but did not show significant similarity to known proteins. However, an interesting feature has been identified. There may be a helix-turn-helix motif at the C-terminus of the *bldD*-encoded protein. The potential DNA-binding domain was similar to those found in the LysR family of regulators (see Figure 3.11). The LysR family is a group of bacterial transcription regulation proteins which possess a helix-turn-helix motif with sequence similarities. Most of these proteins seem to be transcriptional activators, the majority of which are known to negatively regulate their own expression (Henikoff *et al.*, 1988; Sung and Fuchs, 1992). The potential helix-turn-helix motif lies within a region containing basic amino acid residues typical of many proteins known to interact directly with DNA. The motif found in the *bldD*-encoded product differs from the LysR signature sequence by two amino acids at the end and does not have the glycine commonly found at position 9 of known DNA-binding domains, but does have the alanine commonly found at position 5 and the leucine found at position 15 (Dodd and Egan, 1987). However, unlike the bldD protein, which has a helix-turn-helix DNA-binding motif in the C-terminal region, these proteins all possess a potential motif in their N-terminal region. The proteins which belong to the LysR family also share a long stretch of similarity at the N-terminus, of which the helix-turn-helix is a small part , and a small domain in the C-terminal region is also well conserved (Gyorgypal and Kondorosi, 1991). Other than the helix-turn-helix, the bldD protein showed no other similarity to those proteins and was not identified by the database as being similar to the LysR family. Therefore, the bldD protein may not be part of the LysR family, but does contain a helix-turn-helix motif suggesting it could be a DNA-binding protein.

Figure 3.11. Amino acid sequence of the *bldD* gene product indicating the potential DNA-binding helix-turn-helix domain, contained by the box, and the LysR family signature sequence. The asterisks represent amino acids which match the signature sequence and the # represents amino acids which do not match the signature sequence. The bracketed amino acids are the group from which these specific amino acids are chosen. The dashes represent any amino acid (Xn where n is the number of amino acids).

bldD Gene Product

MSSEYAKQLGAKLRAIRTQQGLSLHGVEEKSQGRWKAVVVGSYERGDRAVTVQRLAELAD
FYGVPVQELLPGTTPGGAAEPPPKLVLDLERLATVPAEKAGPLQRYAATIQSQRGDYNGK
VLSIRQDDLRTLAVIYDQSPSVLTEQLISWGVLDADARRAVASHDEL

LysR Family Helix-turn-helix Signature Sequence

[LIVMFYT]-X(2)-[STGALV]-[STA]-X(5)-[PSTA]-[PNQHKR]-X(2)-
[LIVMA]-[STA]-X(2)-[LIVMFW]-X(2)-[LIVMFW]-[RKEQA]-X(2)-
[LIVMFYNT]

Willey *et al.* (1991) reported that SapB appears to be directly involved in erecting aerial hyphae and that the production of SapB is impaired by mutations in various *bld* genes. Certain pairs of *bld* mutants however, are able to restore aerial mycelium formation by extracellular complementation in which their ability to produce SapB is also restored (Willey *et al.*, 1993). This involved a hierarchical cascade of at least four kinds of signals exchanged between cells by diffusion, and these signals are either products or under control of *bld* genes. Since the *bldD* mutant was capable of complementing all of the other mutants, it was thought that the *bldD* mutant produced all four signals and was placed at the top of the hierarchy. The *bldD* mutant was also unable to synthesize SapB. Therefore, they reasoned that *bldD* could be the structural gene for a peptide synthetase that is responsible for the synthesis of SapB or a regulator involved in controlling the expression of a peptide synthetase. Peptide synthetases that have so far been identified have been very large ranging from 100000 to >600000 Daltons (Zuber *et al.*, 1993), which would rule out *bldD* since it codes for a small protein. However, having a potential DNA binding domain, it may be speculated that the *bldD*-encoded protein is a DNA binding protein that activates transcription, perhaps of many genes only one of which may be the putative peptide synthetase involved in SapB production.

3.3 Summary and Future Directions

In summary, sequencing the *bldD* gene from *S. coelicolor* has not revealed an obvious role for the *bldD*-encoded protein. Further investigation of the *bldD* gene is required, beginning with S1 mapping to determine the transcriptional start point and the promoter sequences. Also, the *bldD53* mutation should also be sequenced to provide information on where the mutation lies and how this mutation is affecting the function of the protein. Preliminary Northern blot analysis (Leskiw, B.K.,

unpublished data) has shown that the *bldD* transcript is present in RNA from a *bldA* and *bldD* mutant. The analysis also showed that the *bldD* transcript is overexpressed in a *bldD* mutant, suggesting that the *bldD* gene product negatively autoregulates its own transcription. It would also be interesting to see if the *bldD* transcripts exists in RNA of other *bld* mutants to identify if *bldD* is regulated by any of the other *bld* genes. The preliminary analysis also showed that the *bldD* transcript is present in RNA isolated at different time points from *S. coelicolor* 1501 (wildtype) and wildtype *S. clavuligerus*. Smaller transcripts were observed in both *S. coelicolor* 1501 and *S. clavuligerus* at later time points in colony growth suggesting that the *bldD* transcript is being processed, and therefore would be of interest to study more carefully. These studies would be the next steps in identifying the role of *bldD*.

4.0 BIBLIOGRAPHY

Adamidis, T. and W. Champness. 1992. Genetic analysis of *absB*, a *Streptomyces coelicolor* locus involved in global antibiotic regulation. J. Bacteriol. 174: 4622-4628.

Adamidis, T., P. Riggle, and W. Champness. 1990. Mutations in a new *Streptomyces coelicolor* locus which globally block antibiotic biosynthesis but not sporulation. J. Bacteriol. 172: 2962-2969.

Allan, E J., and J.I. Prosser. 1987. Colony growth of *Streptomyces coelicolor* A3(2) under conditions of continuous nutrient supply. FEMS Microbiol. Lett. 43: 139-142.

Altschul, S.F., W. Gesh, W. Miller, E.W. Myers, d D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Bascaran, V., L. Sanchez, C. Hardisson, and A. Brana. 1991. Stringent response and initiation of secondary metabolism in *Streptomyces clavuligerus*. J. Gen. Microbiol. 137: 1625-1634.

Bibb, M., R. 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., 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.

Bierman, M., R. Logan, K. O'Brien, E.T. Seno, R. Nagaraja Rao, and B.E. Schoner. 1992. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene 116: 43-49.

Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heynecker, H.W. Boyer, J.H. Crosa, and S. Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2: 95-113.

Brana, A.F., C. Mendez, L.A. Diaz, M.B. Manzanal, and C. Hardisson. 1986. Glycogen and trehalose accumulation during colony development in *Streptomyces antibioticus*. J. Gen. Microbiol. 132: 1319-1326.

Brown, K.L., S. Wood, and M.J. Buttner. 1992. Isolation and characterization of the major vegetative RNA polymerase of *Streptomyces coelicolor* A3(2); renaturation of a sigma subunit using GroEL. Mol. Microbiol. 6: 1133-1139.

Bullock, W.O., J.M. Fernandez, and J.M. Short. 1987. XL1-Blue: A high efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase selection. Biotechniques 5: 376-379.

Buttner, M.J., A.M. Smith, and M.J. Bibb. 1988. At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). Cell 52: 599-607.

Buttner, M.J., K.F. Chater, and M.J. Bibb. 1990. Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). J. Bacteriol. 172: 3367-3378.

Champness, W.C. 1988. New loci required for *Streptomyces coelicolor* morphological and physiological differentiation. J. Bacteriol.. 170: 1168-1174.

Champness, W.C. and K.F. Chater. 1994. Regulation and integration of antibiotic production and morphological differentiation in *Streptomyces* spp. p. 61-93. *In* P. Piggot *et al.* (ed.), Regulation of Bacterial differentiation. American Society for Microbiology, Washington, D.C.

Chater, K. F., and M.J. Merrick. 1979. Streptomycetes. p. 93-114. *In* J.H. Parish (ed.), Developmental Biology of Prokaryotes. Blackwell, Oxford, U.K.

Chater, K.F. 1984. Morphological and physiological differentiation in *Streptomyces*. p. 89-115. *In* R. Losick, and L. Shapiro (ed.), Microbial Development. Cold Spring Harbor Lab, Cold Spring Harbor.

Chater, K.F. 1989a. Multilevel regulation of *Streptomyces* differentiation. TIG 5: 372-377.

Chater, K.F. 1989b. Sporulation in *Streptomyces*. p. 277-299. *In* I. Smith, R.A. Slepecky and P. Setlow (ed.), Regulaion of Prokaryotic Development. American Society for Microbiology, Washington, D.C.

Chater, K.F. 1990. Genetics of Differentiation in *Streptomyces*. Ann. Rev. Microbiol. 47: 685-713.

Chater, K.F., C.J. Bruton, A.A. King, and J.E. Suarez. 1982. The expression of *Streptomyces* and *Escherichia* drug resistance determinants cloned into the *Streptomyces* phage ØC31. Gene 19: 21-32.

Chater, K.F., C.J. Bruton, K.A. Plaskitt, M.J. Buttner, C. Mendez, and J.D. Helmann. 1989. The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility σ factor of *B. subtilis*. Cell 59: 133-143.

Chater, K.F., E.J. Lawlor, C. Mendez, C.J. Bruton, N.K. Davis, K. Plaskitt, E.P. Guthrie, B.L. Daly, H.A. Baylis, and K.V. Trong. 1988. Gene expression during *Streptomyces* development. p. 64-70. *In* T. Beppu, and H. Ogawara (ed.), Biology of Actinomycetes '88. Japan Scientific Societies Press, Tokyo.

Coco, E.A., K.E. Narva, and J.S. Feitelson. 1991. New classes of *Streptomyces coelicolor* A3(2) mutants blocked in undecylprodigiosin (Red) biosynthesis. Mol. Gen. Genet. 227: 28-32.

Harasym, M., L. Zhang, K.F. Chater, and J. Piret. 1990. The *Streptomyces coelicolor* A3(2) *bldB* region contains at least two genes involved in morphological development. J. Gen. Microbiol. 136: 1543-1550.

Helmann, J., L. Marquez, and M. Chamberlin. 1988. Cloning, sequencing and disruption of the *Bacillus subtilis* σ^{28} gene. J. Bacteriol. 170: 1568-1574.

Henikoff, S., G. Haughn, J. Calvo, and J. Wallace. 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA. 85: 6602-6606.

Hodgson, D.A. 1980. Carbohydrate utilization in *Streptomyces coelicolor* A3(2). PhD. thesis. Univ. East Anglia, Norwich, UK.

Hodgson, D.A. 1992. Differentiation in actinomycetes. p. 407-440. *In* Mohan, S., C. Dow, J.A. Cole, J.A. (ed.), Prokaryotic Structure and Function: A New Perspective. SGM Symposium Volume 47. Society for General Microbiology, Cambridge, U.K.

Hopwood, D.A. 1987. The Leuwenhoek Lecture: Towards an understanding of gene switching in *Streptomyces*, the basis of sporulation and antibiotic production. Proc. R. Soc. London. B. p.1-19.

Hopwood, D.A. and H.M. Wright. 1978. Bacterial protoplast fusion: recombination in fused protoplasts of *Streptomyces coelicolor*. Mol. Gen. Genet. 162: 307-317.

Hopwood, D.A., and D.H. Sherman. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Ann. Rev. Genet. 24: 37-66.

Hopwood, D.A., and T. Kieser. 1990. The *Streptomyces* Genome. p. 147-162. *In* Drlica, K. and M. Riley (ed.).The Bacterial Chromosome. American Society for Microbiology, Washington, D.C.

Hopwood, D.A., Bibb, M.J., Chater, K.F. and T. Kieser. 1987. Plasmid and phage vectors for gene cloning and analysis in *Streptomyces*. Methods Enzymol. 153: 116-166.

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

Horinouchi, S., M. Kito, M. Neshiyama, K. Furuya, S. Hong, K. Miyake, and T. Beppu. 1990. Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). Gene 95: 49-56.

Horinouchi, S., O. Hara, and T. Beppu. 1983. Cloning of a pleiotropic gene that positively controls biosynthesis of A-factor, actinorhodin, and prodigiosin in *Streptomyces coelicolor* A3(2) and *Streptomyces lividans*. J. Bacteriol. 155: 1238-1248.

Ishizuka, H., S. Horinouchi, H. Kieser, D.A. Hopwood, and T. Beppu. 1992. A putative two-component regulatory system involved in secondary metabolism in *Streptomyces* spp. J. Bacteriol. 174: 7585-7594.

Janssen, G.R., and M.J. Bibb. 1993. Derivatives of pUC18 that have *Bgl*II sites flanking a modified multiple cloning site and that retain the ability to identify recombinant clones by visual screening of *Escherichia coli* colonies. Gene 124: 133-134.

Jonas, R.M., E.A. Weaver, T.J. Kenney, C.P. Moran Jr., and W.G. Haldenwang. 1988. The *Bacillus subtilis spoIIG* operon encodes both σ^E and a gene necessary for σ^E activation. J. Bacteriol. 170: 507-511.

Karmazyn-Campelli, C., C. Bonamy, B. Savelli, and P. Straiger. 1989. Tandem genes encoding σ-factors for consecutive steps of development in *Bacillus subtilis*. Genes Dev. 3: 150-157.

Kenney, T.J., P.A. Kirchman, and C.P. Moran Jr. 1988. Gene encoding σ^E transcribed from a σ^A-like promoter in *Bacillus subtilis*. J. Bacteriol. 170: 3058-3064.

Kieser, H.M., T. Keiser, and D.A. Hopwood. 1992. A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. J. Bacteriol. 174: 5496-5507.

Kinashi, H., M. Shimaji-Murayama, and T. Hanafusa. 1992. Integration of SCP1, a giant linear plasmid, into the *Streptomyces coelicolor* chromosome. Gene 115: 35-41.

Kroos, L., and S. Cutting. 1994. Intercellular and intercompartmental communication during *Bacillus subtilis* sporulation in regulation of bacterial differentiation. p.155-180. *In* Piggot, P. *et al.* (ed.), Regulation of Bacterial Differentiation. American Society for Microbiology. Washington, D.C.

LaBell, T.L., J.E. Trempy, and W.G. Haldenwang. 1987. Sporulation-specific σ factor σ^{29} of *Bacillus subtilis* is synthesized from a precursor protein, P^{31}. Proc. Natl. Acad. Sci. USA 84: 1784-1788.

Lawlor, E.J., H.A. Baylis, and K.F. Chater. 1987. Pleiotropic morphological and antibiotic deficiences result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). Genes Dev. 1: 1305-1310.

Leskiw, B.K. and R. Mah. 1995. The *bldA*-encoded tRNA is poorly expressed in the *bldI* mutant of *Streptomyces coelicolor* A3(2). Microbiology (UK). In press.

Leskiw, B.K., E.J. Lawlor, J.M. Fernandez-Abalos, and K.F. Chater. 1991a. TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. Proc. Natl. Acad. Sci. USA 88: 2461-2465.

Leskiw, B.K., M.J. Bibb, and K.F. Chater. 1991b. The use of a rare codon specially during development. Mol. Microbiol. 5: 2861-2867.

Leskiw, B.K., R. Mah, E.J. Lawlor, and K.F. Chater. 1993. Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3(2). J. Bacteriol. 175: 1995-2005.

Lewin, B. 1990. Genes IV. Oxford University Press, Oxford and Cell Press, Cambridge, Mass.

Lin, Y., H.M. Kieser, D.A. Hopwood, and C.W. Chen. 1993. The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol. Microbiol. 10 : 923-933.

Lomovskaya, N.D., K.F. Chater, and N. Mkrtimian, N. 1980. Genetics and molecular biology of *Streptomyces* bacteriophages. Microbiol. Rev. 44: 206-229.

Lomovskaya, N.D., N.M. Mkrtumian, N.L. Gostimskaya, and V.N. Danilenko. 1972. Characterization of temperate bacteriophage ØC31 isolated from *Streptomyces coelicolor* A3(2). J. Virol. 9:258-262.

Lonetto, M.A., K.L. Brown, K.E. Rudd, and M.J. Buttner. 1994. Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. Proc. Natl. Acad. Sci. USA 91: 7573-7577.

Lopez, J.M., A. Dromerick, and E. Freese. 1981. Response of guanosine 5'-triphosphate concentration to nutritional changes and its significance for *Bacillus subtilis* sporulation. J. Bacteriol. 146: 605-613.

Lopez, J.M., C.L. Marks, and E. Freese. 1979. The decrease of guanine nucleotides initiates sporulation of *Bacillus subtilis*. Biochim. Biophys. Acta 587: 238-252.

Lydiate, D.J., F. Malpartida, and D.A. Hopwood. 1985. The *Streptomyces* plasmid SCP2*: its functional analysis and development into useful cloning vectors. Gene 35: 223-235.

Malpartida, F., and D.A. Hopwood. 1986. Physical and genetic characterization of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. 205: 66-73.

Malpartida, F., J. Niemi, R. Navarette, and D.A. Hopwood. 1990. Cloning and expression in a heterologous host of the complete set of genes for biosynthesis of the *Streptomyces coelicolor* antibiotic undecylprodigiosin. Gene 93: 91-99.

Marcos, A.J., S. Gutierrez, B. Diez, F. Fernandez, J. Oguiza, and J. Martin. 1995. Three genes *hrdB, hrdD* and *hrdT* of *Streptomyces griseus* IMRU 3570, encoding sigma factor-like proteins, are differentially expressed under specific nutritional conditions. Gene 153: 41-48.

Margolis, P., A. Driks, and R. Losick. 1991. Establishment of cell type by compartmentalized activation of a transcription factor. Science 254: 562-565.

Margolis, P.S., A. Driks, and R. Losick. 1993. Sporulation gene *spoIIB* from *Bacillus subtilis*. J. Bacteriol. 175: 528-540.

Matsumoto, A., S. Hong, H. Ishizuka, S. Horinouchi, and T. Beppu. 1994. Phosphorylation of the AfsR protein involved in secondary metabolism in *Streptomyces* species by a eukaryotic-type protein kinase. Gene 146: 47-56.

Merrick, M.J. 1976. A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor*. J. Gen. Microbiol. 96: 299-315.

Mitani, T., J.E. Heinze, and E. Freese. 1977. Induction of sporulation in *Bacillus subtilis* by decoyinine or hadacidin. Biochem. Biophys. Res. Commun. 77: 1118-1125.

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

Narva, K. and J. Feitelson. 1990. Nucleotide sequence and transcriptional analysis of the *redD* locus of *Streptomyces coelicolor* A3(2). J. Bacteriol. 172: 326-333.

Ochi, K. 1986a. A decrease in GTP content is associated with aerial mycelium formation in *Streptomyces* MA406-A-1. J. Gen. Microbiol. 132: 299-305.

Ochi, K. 1986b. Occurrence of the stringent response in *Streptomyces* sp. and its significance for the initiation of morphological and physiological differentiation. J. Gen. Microbiol. 132: 2621-2631.

Ochi, K. 1987. Metabolic initiation of differentiation and secondary metabolism by *Streptomyces griseus*: Significance of the stringent response (ppGpp) and GTP content in relation to A factor. J. Bacteriol. 169: 3608-3616.

Ochi, K., and S. Ohsawa. 1984. Initiation of antibiotic production by the stringent response of *Bacillus subtilis* Marburg. J. Gen. Microbiol. 130: 2473-2482.

Okanishi, M., K. Suzuki, and H. Umezawa. 1974. Formation and reversion of streptomycete protoplasts: cultural conditions and morphological study. J. Gen. Microbiol. 80: 389-400.

Pearson, W.R., and D.J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85: 2444-2448.

Piret, J., and K.F. Chater. 1985. Phage-mediated cloning of *bldA*, a region involved in *Streptomyces coelicolor* morphological development, and its analysis by genetic complementation. J. Bacteriol. 163: 965-972.

Piret, J., V. Bernan, M. Harasym, and M. Brandt. 1988. Characterization of cloned bald genes of *Streptomyces coelicolor* A3(2). p. 321-323. *In* Y. Okami, T. Beppu, and H. Ogawara (ed.), Biology of Actinomycetes '88. Japan Scientific Societies Press, Tokyo.

Plaskitt, K.A., and K.F. Chater. 1995. Influences of developmental genes on localized glycogen deposition in colonies of a mycelial prokaryote, *Streptomyces coelicolor* A3(2): a possible interface between metabolism and morphogenesis. Phil. Trans. R. Soc. Lond. B. pp. 105-121.

Puglia, A.M., and E. Cappelletti 1984. A bald ultrafertile UV-resistant strain in *Streptomyces coelicolor* A3(2). Microbiologica 7: 263-266.

Rudd, B., and D.A. Hopwood. 1980. A pigmented mycelial antibiotic in *Streptomyces coelicolor*: Control by a chromosomal gene cluster. J. Gen. Microbiol. 119: 333-340.

Sambrook, J. E.F. Fritsch, and T. Maniatis. 1989. *In* N. Ford, C. Nolan, and M. Ferguson (ed.), Molecular Cloning: A Laboratory Manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

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

Schmidt, R., P. Margolis, L. Duncan, R. Coppolecchia, C.P. Moran Jr., and R. Losick. 1990. Control of developmental transcription factor σ^F by sporulation regulatory proteins SpoIIAA and SpoIIAB in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 87: 9221-9225.

Short, J.M., J.M. Fernandez, J.A. Sorge, and W.D. Huse. 1988. λ Zap: A bacteriophage λ expression vector with in vivo excision properties. Nucleic Acids Res. 16: 7583-7600.

Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517.

Straiger, P. 1992. Establishment of forespore-specific gene expression during sporulation of *Bacillus subtilis*. p. 297-310. *In* C. Dow, and J.A. Cole (ed.).Prokaryotic Structure and Function: A New Perspective, vol. 47. Society for General Microbiology of Cambridge University Press, Cambridge.

Straiger, P., B. Kunkel, L. Kroos, and R. Losick. 1989. Chromosomal rearrangement generating a composite gene for a developmental transcription factor. Science 243: 507-512.

Straiger, P., C. Bonamy, and C. Karmazyn-Campelli. 1988. Processing of a sporulation sigma factor in *Bacillus subtilis*: How morphological structure could control gene expression. Cell 52: 697-704.

Strauch, E., E. Takano, H.A. Baylis, and M.J. Bibb. 1991. The stringent response in *Streptomyces coelicolor* A3(2). Mol. Microbiol. 5: 289-298.

Sun, D., P. Straiger, and P. Setlow. 1989. Identification of a new σ-factor involved in compartmentalized gene expression during sporulation of *Bacillus subtilis*. Genes Dev. 3: 141-149.

Sung, Y. and J. Fuchs. 1992. The *Escherichia coli* K-12 *cyn* operon is positively regulated by a member of the LysR family. J. Bacteriol. 174: 3645-3650.

Sutcliffe, J.G. 1978. pBR322 restriction map marked from the DNA sequence: Accurate DNA size markers up to 4361 nucleotide pairs long. Nucleic Acids Res. 5: 2721-2728.

Tabor, S. and C.C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84: 4767-7771.

Tan, H., and K.F. Chater. 1993. Two developmentally controlled promoters of *Streptomyces coelicolor* A3(2) that resemble the major class of motility-related promoters in other bacteria. J. Bacteriol. 175: 933-940.

Tanaka, K., T. Shiina, and H. Takahashi. 1988. Multiple principle sigma factor homologs in eubacteria: Identification of the "rpoD box". Science 242: 1040-1042.

Tatti, K.M., H.L. Carter III, A. Moir, and C.P. Moran Jr.. 1989. Sigma H-directed transcription of *citG* in *Bacillus subtilis*. J. Bacteriol. 171: 5928-5932.

Thompson, C.J., J.M. Ward, and D.A. Hopwood. 1980. DNA cloning in *Streptomyces*: resistance genes from antibiotic-producing species. Nature 286: 525-527.

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.

Wang, L.F., and R.H. Doi. 1986. Nucleotide sequence and organization of *Bacillus subtilis* RNA polymerase major sigma-43 operon. Nucleic Acid Res. 14: 4293-4307.

Ward, J.M., G.R. Janssen, T. Kieser, M.J. Bibb, M.J. Buttner, and M.J. Bibb. 1986. Construction and characterization of a series of multi-copy promoter-probe plasmid vectors for *Streptomyces* using the aminoglycoside phosphotransferase gene from Tn5 as indicator. Mol. Gen. Genet. 203: 468-478.

Westpheling, J., M. Reines, and R. Losick. 1985. RNA polymerase heterogeneity in *Streptomyces coelicolor*. Nature 313: 22-27.

Willey, J., J. Schwedock, and R. Losick. 1993. Multiple extracellular signals govern the production of a morphogenetic protein involved in aerial mycelium formation by *Streptomyces coelicolor*. Genes Dev. 7: 895-903.

Willey, J., R. Santamaria, J. Guijjaro, M. Geistlich, and R. Losick. 1991. Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by *S. coelicolor*. Cell 65: 641-650.

Wright, L.F., and D.A. Hopwood. 1976. Identification of the antibiotic determined by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. 95: 96-106.

Wu, J., M.G. Howard, and P.J. Piggot. 1989. Regulation of transcription of the *Bacillus subtilis spoIIA* locus. J. Bacteriol. 171: 692-698.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

Zhen, L. and R.T. Swank. 1993. A simple and high yield method for recovering DNA from agarose gels. BioTechniques 14: 894-898.

Zoller, M.J. and M. Smith. 1987. Oligonucleotide-directed mutagenesis: A simple method using two oligonucleotide primers and a single-stranded DNA template. Methods Enzymol. 154: 329-349.

Zuber, P., M. Nakano, and M. Marahiel. 1993. Peptide Antibiotics. p. 897-916. *In* A. Sonenshein, J. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other Gram positive bacteria. American Society for Microbiology, Washington, D.C.

APPENDIX

During the course of this work a fragment of DNA from the ØC31 vector (KC304) was mistakenly subcloned from the recombinant KC742 instead of the *bldD*-containing fragment. Before realizing that the fragment did not correspond to *bldD*, the 3.5kb fragment was subcloned and the DNA sequence determined. A map showing the location on the ØC31 KC304 vector of the *Eco*RV partial fragment is shown below (Figure A.1) along with the double stranded sequence (Figure A.2).

Figure A.1.

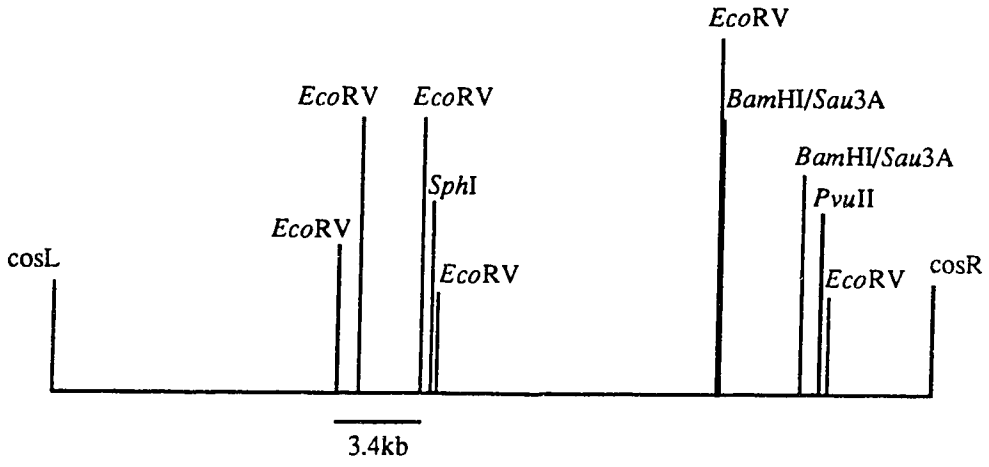


Figure A.2

SmaI

CCCGGGAAGTCCGGCTTCACGGCCGTAGCGACGCCGACAAGGGCGACGAC 50
GGGCCCTTCAGGCCGAAGTGCCGGCATCGCTGCGGCTGTTCCCGCTGCTG

GACGGACGTGACGACCTTCGGGTGAGCCTTCGCCCAACTCAACGCGGCGC 100
CTGCCTGCACTGCTGGAAGCCCACTCGGAAGCGGGTTGAGTTGCGCCGCG

GTGCGTAGTTGCCGCCCGGCTTGCTGTGGTCGCCCATGGCGCCCCCTTAC 150
CACGCATCAACGGCGGGCCGAACGACACCAGCGGGTACCGCGGGGGAATG

AGAGAGTGAGATTTGAGCCGGAGCCGGCATGCCCGGAACCCGTACAGAAT 200
TCTCTCACTCTAAACTCGGCCTCGGCCGTACGGGCCTTGGGCATGTCTTA

BKL16 → ← BKL17

CGGACATACCGGCTTCCGCGGAACGAACAGAGAGTGATCGACGCTCAGAC 250
GCCTGTATGGCCGAAGGCGCCTTGCTTGTCTCTCACTAGCTGCGAGTCTG

CTTCGGGACCTTCAGCGCGTCCCACGTGGTCTTACCGGGCCAACCGTCAG 300
GAAGCCCTGGAAGTCGCGCAGGGTGCACCAGAATGGCCCGGTTGGCAGTC

CGTCGGCGCCGGAGTAACCGCGCTTGCGCTGCCACTTCGCGTAAGAAGCC 350
GCAGCCGCGGCCTCATTGGCGCGAACGCGACGGTGAAGCGCATTCTTCGG

TTGTCGGCGTTCGTCCACTGGGCACCCGGACCGGACCGGTACGCGGAGCA 400
AACAGCCGCAAGCAGGTGACCCGTGGGCCTGGCCTGGCCATGCGCCTCGT

CCCTTCAGCGACAAGGCGCTTGCCCATGGCCGTAACGATCGGGCTCTTCG 450
GGGAAGTCGCTGTTCCGCGAACGGGTACCGGCATTGCTAGCCCGAGAAGC

SmaI

GGTTCTTCTTGAACCATGCGGCACCCGGGAACGGCTCATACGACGGCTTC 500
CCAAGAAGAACTTGGTACGCCGTGGGCCCTTGCCGAGTATGCTGCCGAAG

GGCTTCGCCTTCGCGGGAAGCTTGCCGAGAAGCTTCGTCAGCGACCCTTC 550
CCGAAGCGGAAGCGCCCTTCGAACGGCTCTTCGAAGCAGTCGCTGGGAAG

GCCCGGCACGCCGTCAGCGTCCTTGCCGGAGTAGCCCAACGACCGCTGAA 600
CGGGCCGTGCGGCAGTCGCAGGAACGGCCTCATCGGGTTGCTGGCGACTT

ACGCGGCATAGTTGGTCGTGTCGGCGTCGGACCAATTCGGCCCCGGACCG 650
TGCGCCGTATCAACCAGCACAGCCGCAGCCTGGTTAAGCCGGGGCCTGGC

TCCTTGTACGCGTCGCCGAAGCCCTTCGCGACAAGCGCCTTACCGACCTG 700
AGGAACATGCGCAGCGGCTTCGGGAAGCGCTGTTCGCGGAATGGCTGGAC

GGTGACCTGGGCGCCACGTGCCCCGTAGCCGTACTCAAGACCGTTGATGG 750
CCACTGGACCCGCGGTGCACGGGGCATCGGCATGAGTTCTGGCAACTACC

TCACCTTGTACCGGGCGACGGCTCCGGACGGCTCAGCGGGCTTGTCAGGC 800
AGTGGAACATGGCCCGCTGCCGAGGCCTGCCGAGTCGCCCGAACAGTCCG

BKL22 →

GTGGTCGAGCCGCCAACGTCGCCAACGTCGCCGGTCGCCTTGTGCTTGTA 850
CACCAGCTCGGCGGTTGCAGCGGTTGCAGCGGCCAGCGGAACACGAACAT

ACCGAACTTCGAAGCGTCCGGGTCGGCGGAGACGGAGCCGCCCGCAACGT 900
TGGCTTGAAGCTTCGCAGGCCCAGCCGCCTCTGCCTCGGCGGGCGTTGCA

← BKL30

CTGGGTAGCCGTAGCCGTACAGGTACGAATCCCGACGCGCACGCTTGCGC 950
GACCCATCGGCATCGGCATGTCCATGCTTAGGGCTGCGCGTGCGAACGCG

AGATACACGCCGTCGCCCTCAGCGCTGCCGTTCGTGTTCGTGTTGCCGCC 1000
TCTATGTGCGGCAGCGGGAGTCGCGACGGCAAGCACAAGCACAACGGCGG

GACCGTGTACGCATAGTCAGCGTCGTACGCGTAGCAAATCTCCGTGTGCG 1050
CTGGCACATGCGTATCAGTCGCAGCATGCGCATCGTTTAGAGGCACACGC

BKL25 →

AGCCGCCCCCGTTGCCGAAGAACACCTGAGCGCCGACCGCCGGATACGCC 1100
TCGGCGGGGGCAACGGCTTCTTGTGGACTCGCGGCTGGCGGCCTATGCGG

GACCAACGCCCCTTGTTCCGGAACCAGTTCACGCCCGTCGCACACGAAGC 1150
CTGGTTGCGGGGAACAAGGCCTTGGTCAAGTGCGGGCAGCGTGTGCTTCG

CGTGCGCGGGTAGTGCGCCGACTCCCCAGCCTGAAGAGCAACCCAACTCA 1200
GCACGCGCCCATCACGCGGCTGAGGGGTCGGACTTCTCGTTGGGTTGAGT

← BKL26

CGAAGGTCGCGCACCAAGCCTGATTCTGCGACCACTCAAGCCCCGGCACG 1250
GCTTCCAGCGCGTGGTTCGGACTAAGACGCTGGTGAGTTCGGGGCCGTGC

GCAGGCGAATACTTCTGGTGGTTGTTCCAATGTCCGCCCGACCGGCCTTC 1300
CGTCCGCTTATGAAGACCACCAACAAGGTTACAGGCGGGCTGGCCGGAAG

GTGATAGCCGACCTCGGCCTTCGCGACCGACACAACCTTTGCGAGACTCA 1350
CACTATCGGCTGGAGCCGGAAGCGCTGGCTGTGTTGGAAACGCTCTGAGT

BKL29 →

AAACGCGCTCCGTTTCATGGCATGGAAAAGCCCTGGGGGCGGGATTGCCC 1400
TTTGCGCGAGGCAAAGTACCGTACCTTTTCGGGACCCCCGCCCTAACGGG

CCAGGGCGCTACCTACTCACGTGAGTAGGTTGAGACTCAGACGTTGGCAA 1450
GGTCCCGCGATGGATGAGTGCACTCATCCAACTCTGAGTCTGCAACCGTT

GCGCCATGATCGGAATGTCATGGTTGGCAAGCACGTCGGACGCTGCCTGA 1500
CGCGGTACTAGCCTTACAGTACCAACCGTTCGTGCAGCCTGCGACGGACT

← BKL23

TTCAGCGTGTCGGGAAGTGCCGTCTTATCGGACACGTACCATGCGCGCGC 1550
AAGTCGCACAGCCCTTCACGGCAGAATAGCCTGTGCATGGTACGCGCGCG

CCAGGGCGAATCCGTCATGAAGAAGTTTGCCGTCGAAATCGTGGCTTCGT 1600
GGTCCCGCTTAGGCAGTACTTCTTCAAACGGCAGCTTTAGCACCGAAGCA

```
TCTGCACGTGGTAGAACGCGAAGTCGGCCGTGCCCCCATTCACCTGAAGC 1650
AGACGTGCACCATCTTGCGCTTCAGCCGGCACGGGGGTAAGTGGACTTCG

                                          SalI
CACGTCACCCAGTAGCGCCCAGGCTGAAGCACGGTCGACGCAATGGGCAG 1700
GTGCAGTGGGTCATCGCGGGTCCGACTTCGTGCCAGCTGCGTTACCCGTC

CGGAACGGCGCCGACATGCGTTGACTGCATAGCGGCAAGCGCGCCGGTTT 1750
GCCTTGCCGCGGCTGTACGCAACTGACGTATCGCCGTTCGCGCGGCCAAA

CCTGCCCAGCCTGGGGCAGTGCGACGGAATCCGACTTCACAACTACCGCG 1800
GGACGGGTCGGACCCCGTCACGCTGCCTTAGGCTGAAGTGTTGATGGCGC

CCCCTGCCGGTAGTGCTTTCCTTGTAAATGCCCGCCATGAATCGGTCAGC 1850
GGGGACGGCCATCACGAAAGGAACATTTACGGGCGGTACTTAGCCAGTCG

CGAAACGCCGCCCCAACCACGCGCGAAGATGACGATTCGGTTTACGGTCG 1900
GCTTTGCGGCGGGGTTGGTGCGCGCTTCTACTGCTAAGCCAAATGCCAGC
BKL19
TCGGCTCCGTGATGTTGAAGCCCGTCACGTAAAGACGCTGCGGAGTCAGA 1950
AGCCGAGGCACTACAACTTCGGGCAGTGCATTTCTGCGACGCCTCAGTCT
                                  SmaI  SalI
TACTTCGCAACCGGGTTCACCACGCCGCCCGGGTCGACCGACCACGCTTC 2000
ATGAAGCGTTGGCCCAAGTGGTGCGGCGGGCCCAGCTGGCTGGTGCGAAG

GAAGCCCAGCGCTTGCGGAGTCCAACTGTTCTTCGGAACGGCGCTGGGAA 2050
CTTCGGGTCGCGAACGCCTCAGGTTGACAAGAAGCCTTGCCGCGACCCTT

CGTTGGCAATCGGCAGGCGCTTCGCGGAGTCGAGCGGGGCAACTCCGTCC 2100
GCAACCGTTAGCCGTCCGCGAAGCGCCTCAGCTCGCCCCGTTGAGGCAGG
                                        BKL18
GCAGCGTTCTTCGTTGAAGTGGCGAGCGCCCCAACGTCGGCAGCGGTAAG 2150
CGTCGCAAGAAGCAACTTCACCGCTCGCGGGGTTGCAGCCGTCGCCATTC

CGTGACCGTCGGACCCGGCTTCGTGTTGACCGACGTCACGGCGTTCGGGT 2200
GCACTGGCAGCCTGGGCCGAAGCACAACTGGCTGCAGTGCCGCAAGCCCA

CGGCAAGCTGGGGGAGTTGGTCAGTCGGAACAAGCCCGTCGGTGCCCAGC 2250
GCCGTTCGACCCCCTCAACCAGTCAGCCTTGTTCGGGCAGCCACGGGTCG

GTGGCAACGCCGGACGCAACTCCCTTGTCGGCAACGGGAATCGCCCCGAC 2300
CACCGTTGCGGCCTGCGTTGAGGGAACAGCCGTTGCCCTTAGCGGGGCTG

GTCGGCAGCGGCAAGGGTGATATCCGGCCCTAGGTCACCGTTCACAGAGT 2350
CAGCCGTCGCCGTTCCCACTATAGGCCGGGATCCAGTGGCAAGTGTCTCA
                  SmaI
TGACGGAGCCCGTACCCGGGTCACCCTTCGGACCCTGGGGACCCTCCGGA 2400
ACTGCCTCGGGCATGGGCCCAGTGGGAAGCCTGGGACCCCTGGGAGGCCT
```

CCGGCCGGACCCTCCGGACCCTCCGGCCCTTCCGGACCCTGGGGACCGGG 2450
GGCCGGCCTGGGAGGCCTGGGAGGCCGGGAAGGCCTGGGACCCCTGGCCC

AACCGTCGAGTCGGCGCCCGTCGCGCCCGTGTCGCCCTTGTCGCCCTTCG 2500
TTGGCAGCTCAGCCGCGGGCAGCGCGGGCACAGCGGGAACAGCGGGAAGC

CGCCCTTGATGTTCCCCTTCAAGTCCCAGCCGGACGCACCGCGTTGGTAC 2550
GCGGGAACTACAAGGGGAAGTTCAGGGTCGGCCTGCGTGGCGCAACCATG

SalI
ACGTCGCCGGAGTCGACCCGAAGGAGCATGTCACCAACGGGCACGTCGGC 2600
TGCAGCGGCCTCAGCTGGGCTTCCTCGTACAGTGGTTGCCCGTGCAGCCG

CGACGGGGTGCCCGTGTTGTTCGTGTACCACGCTGCGCCACGAACGCCGT 2650
GCTGCCCCACGGGCACAACAAGCACATGGTGCGACGCGGTGCTTGCGGCA

CCGACACCTTCGACCACACGCCGCCCGACTTCGCCCACATAGTGACCGTC 2700
GGCTGTGGAAGCTGGTGTGCGGCGGGCTGAAGCGGGTGTATCACTGGCAG

SalI
GTTGAGCTGACGCCGAGTGTGGTCGACACAGTGAACTGGGCGTACGTGTC 2750
CAACTCGACTGCGGCTCACACCAGCTGTGTCACTTGACCCGCATGCACAG

GCCGTCGAGCCCCAGCGAAGCCGCCGGAACCGTCGAGCCGTTGAACGCGT 2800
CGGCAGCTCGGGGTCGCTTCGGCGGCCTTGGCAGCTCGGCAACTTGCGCA

GAGCGCCCCGCCCAACAAGGGAGTCGAGCCACTGAGCCTCAGTGCCGGTG 2850
CTCGCGGGGCGGGTTGTTCCCTCAGCTCGGTGACTCGGAGTCACGGCCAC

BKL28
AAGCCTTCCGCAACAGCGACTTCGTACGCGCTGGGACCCGGGACGGCAAC 2900
TTCGGAAGGCGTTGTCGCTGAAGCATGCGCGACCCTGGGCCCTGCCGTTG

SmaI
SalI
GTACGTCGGCGTAGTCGGGTCGGCAGGGGCGACGTCGGCAAGGTCGACGG 2950
CATGCAGCCGCATCAGCCCAGCCGTCCCCGCTGCAGCCGTTCCAGCTGCC

AGTTGTTCAGCGTGTCTTCGGCAGCACCATGGAATAGGTGCGCGCGCCAA 3000
TCAACAAGTCGCACAGAAGCCGTCGTGGTACCTTATCCACGCGCGCGGTT

BKL27
CGACGCCGGTCAGATTCTCCTTAACCGTGTACGTCCATTCCGACGGGTTC 3050
GCTGCGGCCAGTCTAAGAGGAATTGGCACATGCAGGTAAGGCTGCCCAAG

ATGTCGGGGGAGTCGGTAGCGGGAAGACGAACGCCGACGTTGCCGTTGGC 3100
TACAGCCCCCTCAGCCATCGCCCTTCTGCTTGCGGCTGCAACGGCAACCG

BKL20
GTCAATGATCTGCCCCGTCTCATCGAGCGTGCAGACGACCGGACCGGCAA 3150
CAGTTACTAGACGGGGCAGAGTAGCTCGCACGTCTGCTGGCCTGGCCGTT

TGAACAAGTCCGACTCAGGGAACGTCACAAGGCCCGGACCCGTGAAGGTA 3200
ACTTGTTCAGGCTGAGTCCCTTGCAGTGTTCCGGGCCTGGGCACTTCCAT

ACCGTGCCCTTCAGCGCTCGACCGTCCCAACCCAGGTAAGTACCCGTGAC 3250
TGGCACGGGAAGTCGCGAGCTGGCAGGGTTGGGTCCATTCATGGGCACTG

CCGTACGGTAGGAATCTCATTCGGAATCGCCACGGGCGCCCCCAATCAAG 3300
GGCATGCCATCCTTAGAGTAAGCCTTAGCGGTGCCCGCGGGGGTTAGTTC
BKL24

TTGTCAATGTGGTCGCGAAGCTCACGATTCTCACGCCGCAGCTCCGCCAC 3350
AACAGTTACACCAGCGCTTCGAGTGCTAAGAGTGCGGCGTCGAGGCGGTG

TTCAGCGCGCAACCCGGCGTTCTCATCGCGCAGATTCCGCACTTCCGTGA 3400
AAGTCGCGCGTTGGGCCGCAAGAGTAGCGCGTCTAAGGCGTGAAGGCACT
BKL34

TGAGAAGCGTTAGGTCGTCCGACAAGCGCTGTGAGCGCGCAATCTGTGCT 3450
ACTCTTCGCAATCCAGCAGGCTGTTCGCGACACTCGCGCGTTAGACACGA

TCGGCTTCGTCGCGCCATACGTTGCCCGTGCCGGTACGCACGCGCTGATA 3500
AGCCGAAGCAGCGCGGTATGCAACGGGCACGGCCATGCGTGCGCGACTAT

GACGAGAAACCCACCGATGAGAAGCGCGCCGATTACGTCGGCGTGACCGG 3550
CTGCTCTTTGGGTGGCTACTCTTCGCGCGGCTAATGCAGCCGCACTGGCC

TGATA 3555
ACTAT