University of Alberta

Identification and analysis of genes involved in 5*S* clavam biosynthesis in *Streptomyces clavuligerus*

by

Nathan Zelyas ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Microbiology and Biotechnology

Department of Biological Sciences

Edmonton, Alberta

Fall, 2007

Your file Votre référence
ISBN: 978-0-494-33380-8
Our file Notre référence
ISBN: 978-0-494-33380-8

Abstract

The upstream and downstream regions of the paralogue gene cluster of *Streptomyces clavuligerus* were examined for importance in clavam biosynthesis. The gene upstream of *ceaS1*, *orfA*, was sequenced, and mutants in *orfA*, *orfB*, *orfC*, *orfD*, and *orfE* were reanalyzed. *orfF*, *orfG*, and *orfH* mutants were prepared. Among these mutants, *orfA*, *orfB*, *orfC*, and *orfD* mutants did not produce alanylclavam. A 7.7 kb region downstream of *snk* in the paralogue cluster was sequenced, revealing six open reading frames. Mutagenesis demonstrated that *res1*, *res2*, and *orf4* play roles in 5*S* clavam production. Phosphotransfer experiments with purified proteins showed that Snk, Res1, and Res2 comprise an atypical two-component system in which Snk transfers phosphate molecules to Res1 and Res2.

An in vivo method for generating unmarked in-frame *Streptomyces* mutants using FLP recombinase was devised and used to mutate *c7p* in the paralogue cluster and *redJ* in the *red* gene cluster of *Streptomyces coelicolor*.

Table of Contents

List of Tables

page

List of Figures

page

List of Abbreviations

A	Adenine
aa	amino acids
AC	alanylclavam
ACVS	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase
amp	Ampicillin
apra	Apramycin
ARE	Autoregulatory element
ATP	Adenosine triphosphate
β-LS	β-lactam synthetase
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
C2C	Clavam-2-carboxylate
CA	Clavulanic acid
cAMP	Cyclic adenosine monophosphate
CAD	Clavaldehyde dehydrogenase
CAS	Clavaminate synthase
CcaR	Cephamycin C and clavulanic acid regulator
CeaS	Carboxyethylarginine synthase
cGMP	Cyclic guanosine monophosphate
CHAP	Cysteine, histidine-dependent amidohydrolase/peptidases
CL	Cleared lysate
DAOCS	Deacetoxycephalosporin C synthase
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
egfp	Enhanced green fluorescent protein
FRT	FLP recombination target
FT	Flow-through
G	Guanine
His	Histidine
HMC	2-hydroxymethylclavam
HPLC	High performance liquid chromatography
hyg	Hygromycin
I	Induced
IC	Intermediate clavam
IPNE	Isopenicillin N epimerase
IPNS	Isopenicillin N synthase
IPTG	Isopropyl-β-D-thiogalactopyranoside
ISP-4	International *Streptomyces* project medium

kan	Kanamycin
kb	Kilobase
LAT	Lysine ε-aminotransferase
LB	Lennox broth
LBA	Lennox broth agar
LC-MS	Liquid chromatography-mass spectrometry
M	Molecular weight marker
Mb	Megabase
ME	Mosaic end
MS	Mannitol soy agar
MYM	Maltose yeast extract malt extract medium
NA	Nutrient broth agar
NADPH	Nicotinamide adenine dinucleotide
NC	Novel clavam
NI	Non-induced
OAT	Ornithine acetyltransferase
ORF	Open reading frame
oriT	Origin of transfer
PAGE	Polyacrylamide gel electrophoresis
PAH	Proclavaminate amidinohydrolase
PBP	Penicillin-binding protein
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
ppGpp	Guanosine tetraphosphate
^{r}	Resistant
RBS	Ribosome binding site
RNA	Ribonucleic acid
rpm	Rotations per minute
rRNA	Ribosomal ribonucleic acid
RT	Reverse transcriptase
^{s}	Sensitive
SARP	*Streptomyces* antibiotic regulatory protein
SDS	Sodium dodecyl sulfate
t	Transcriptional terminator
T	Thymine
TAE	Tris-acetate-EDTA buffer
TEMED	Tetramethylethylenediamine
thio	Thiostrepton
tRNA	Transfer ribonucleic acid
ts	Temperature sensitive
TSBA	Trypticase soy broth agar
TSBS	Trypticase soy broth starch medium
UV	Ultraviolet
w/v	Weight/volume
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
YT	Yeast extract tryptone medium

I. INTRODUCTION

I.1 *Streptomyces*

Members of the *Streptomyces* genus are filamentous, Gram-positive aerobic bacteria that form spores and live in the soil. The *Streptomyces* genus belongs in the *Streptomycetaceae* family, which is placed in the *Actinomycetales* order of the *Actinobacteria* class (Waksman and Henrici 1943; Stackebrandt et al. 1997). One of the notable aspects of the *Streptomyces* genus is the unique and complex life-cycle in which members engage. This morphological development involves the progression of groups of cells from a substrate mycelium to an aerial mycelium, and, finally, to mature spores. Interestingly, as cells start to differentiate into an aerial mycelium, the secondary metabolism of the culture begins (Chater 1989). Secondary metabolites include a number of biologically active compounds, such as antibiotics, antihelminthics, immunosuppressants, antitumor agents, and growth factors, many of which have important medical and economic value (Demain 2000). The interest in the *Streptomyces* life-cycle and in producing high levels of beneficial secondary metabolites has spurred research of the genetic aspects of these organisms. Genetic analysis of *Streptomyces* spp. has revealed a number of novel characteristics within this group of microbes and links between morphological differentiation and secondary metabolite production.

I.1.2 The *Streptomyces* life-cycle

The complex life-cycle starts when a *Streptomyces* spore germinates with the formation of a germ tube, and continues to grow vegetatively to eventually form a network of branching multinucleoid hyphae (Hardisson et al. 1978). This mycelium

grows on the surface of and into the medium, and is thus called the substrate mycelium. As the substrate mycelium ages and uses up the surrounding nutrients, it undergoes morphological differentiation to form the aerial mycelium. Aerial hyphae grow upwards and away from the surface of the colony; growth occurs from the tops of the aerial hyphae and from their branching inner areas (Miguelez et al. 1994). The aerial hyphae, separated from a source of nutrients because they grow away from the surface of the medium, obtain macromolecules for growth from the lysing substrate mycelium (Mendez et al. 1985). Over time the aerial mycelium forms cross walls between individual nuclei (Wildermuth 1970). These hyphae of divided compartments further differentiate to produce uni-nucleoid chains of spores which can separate to begin the process anew to generate more *Streptomyces* colonies (Hardisson and Manzanal 1976). Typically, the appearance of aerial hyphae is accompanied by the production of secondary metabolites, which, for some *Streptomyces* spp., include β-lactam metabolites.

I.1.3 β-lactam antibiotic production

Penicillin, the first β-lactam antibiotic discovered, was found by Fleming (1929) in broth filtrates produced by cultures of a *Penicillium* species originally identified through its ability to lyse surrounding staphylococcal colonies. Since then, it has been found that penicillin inhibits the cross-linking reaction required for the synthesis of the bacterial cell wall by acting as an analog of acyl-D-alanyl-D-alanine and binding the transpeptidase that catalyzes the reaction (Tipper and Strominger 1965; Wise and Park 1965). This has allowed penicillin and other compounds with a β-lactam structure to be used commonly in the treatment of diseases caused by pathogenic bacteria. Miller et al.

(1962) reported the first instance of a bacterium (a member of the *Streptomyces* genus) capable of producing a β-lactam. A number of bacterial species have since been found to produce β-lactam antibiotics.

β-lactam antibiotics include hydrophobic and hydrophilic penicillins, cephalosporins, clavams, carbapenems, and monocyclic β-lactams (Demain and Elander 1999). All β-lactam metabolites contain the four-membered β-lactam ring; different types of β-lactam antibiotics have different additional ring structures fused to the β-lactam ring. Penicillins all contain the five-membered thiazolidine ring; cephalosporins contain the six-membered dihydrothiazine ring; clavams carry the five-membered oxazolidine ring; carbapenems contain a five-membered, unsaturated ring that lacks the sulfur atom present in the thiazolidine ring; and monocyclic β-lactams do not have an additional ring (Demain and Elander 1999). Further variation between members of specific classes of β-lactams often occurs within side-chains attached to the β-lactam ring or to the class-specific ring.

The development of resistance in many pathogenic organisms to β-lactam antibiotics is caused by the presence of β-lactamases. β-lactamases cleave the β-lactam ring of penicillins and cephalosporins to render them inactive (Baggaley et al. 1997). There are four recognized classes of β-lactamases that fall into two groups according to their modes of action: class A, C, and D β-lactamases are all serine hydrolases, while class B β-lactamases are zinc-containing metalloenzymes (Ambler 1980). With the increasing spread of genes encoding β-lactamases to pathogenic organisms, β-lactam antibiotics resistant to hydrolysis and agents capable of inhibiting the reaction catalyzed

by β-lactamases have become more relevant than ever before. One avenue of research used to find or engineer such new β-lactams has been the genetic study of *Streptomyces*.

I.1.4 *Streptomyces* genetics

Streptomyces chromosomes are large compared to those of most other bacteria, from 8 to 9 Mb, and have a high G+C content of about 70% (Kieser et al. 1992; Kieser et al. 2000; Leblond et al. 1990; Leblond et al. 1993; Lezhava et al. 1995). Unlike most other bacteria, the *Streptomyces* chromosome is linear and both linear and circular plasmids can be found in *Streptomyces* (Hayakawa et al. 1979; Lin et al. 1993; Schrempf et al. 1975). The *Streptomyces* linear chromosome has a centrally located origin of replication (Musialowski et al. 1994). Terminal inverted repeats can be found at the ends of the chromosome and these telomeres have proteins covalently attached (Leblond et al. 1996; Lin et al. 1993). Both linear chromosome and linear plasmid replication proceeds bidirectionally from the central origin of replication to the telomeres so that 3' overhangs are generated (Chang and Cohen 1994; Fischer et al. 1998). There is evidence that the 3' ends fold back by virtue of the repetitive sequences in the telomeres, leading to the formation of secondary structure in these terminal regions (Qin and Cohen 1998). Proteins required for chromosome and plasmid replication, such as Tpg (terminal protein gene), Tap (telomere-associated protein), DNA polymerase, and topoisomerase I, bind to this telomere complex presumably to prime and catalyze DNA synthesis to fill in the 5' recessed ends (Bao and Cohen 2001; Bao and Cohen 2003; Bao and Cohen 2004).

Two *Streptomyces* genomes have been sequenced, annotated, and made available to the research community so far: those of *Streptomyces coelicolor* and *Streptomyces*

avermitilis (Bentley et al. 2002; Ikeda et al. 2003). The *S. coelicolor* chromosome is 8.7 Mb in length and contains 7,825 genes, while the *S. avermitilis* chromosome is 9.0 Mb and is comprised of 7,574 genes (Bentley et al. 2002; Ikeda et al. 2003). The chromosomes are structured such that a highly similar 6.5 Mb core region carrying most of the essential genes can be found in the middle of both; however, the outer regions of the chromosomes, the arms, are not similar to each other and contain genes with non-essential functions, such as secondary metabolite production (Bentley et al. 2002; Ikeda et al. 2003).

The genome sequences have revealed that both organisms are capable of producing a large number of secondary metabolites, such as antibiotics, pigments, lipids, and siderophores. *S. coelicolor* carries 22 gene clusters for secondary metabolite biosynthesis, while *S. avermitilis* has 30 (Bentley et al. 2002; Ikeda et al. 2003; Omura et al. 2001). It is clear that *Streptomyces* spp. have enormous secondary metabolite-producing potential and that cloning these gene clusters will not only lead to the discovery of novel mechanisms of biosynthesis but also to the finding of new biologically-active compounds of industrial or medical significance.

Interest in the *Streptomyces* life-cycle and secondary metabolite production has initiated studies into the unique genetics of these organisms. The organism of interest of this study, *Streptomyces clavuligerus*, has been placed under scientific scrutiny for a number of years because of its antibiotic-producing capabilities. Much of this work has focused on the genetics and biochemistry behind *S. clavuligerus* antibiotic biosynthesis.

I.2 Clavam biosynthesis by *Streptomyces clavuligerus*

Streptomyces clavuligerus was originally isolated from South American soil over 35 years ago based on its ability to produce β-lactam antibiotics (Higgens and Kastner 1971). Initial analyses indicated that *S. clavuligerus* is capable of producing penicillin N, cephamycin C, and *o*-carbamoyl-deacetoxycephalosporin (Nagarajan et al. 1971). It was later found that penicillin N and *o*-carbamoyl-deacetoxycephalosporin are intermediates in cephamycin C biosynthesis. Besides these antibiotics, *S. clavuligerus* is also able to produce five different clavam metabolites, as well as holomycin and tunicamycin (Kenig and Reading 1979; Figure I.1).

One of the clavams produced by *S. clavuligerus* is clavulanic acid, a weak but broad-spectrum antibiotic with strong β-lactamase inhibitory activity (Brown et al. 1976). Organisms resistant to β-lactam antibiotics typically produce one or more β-lactamases; these enzymes render β-lactams inactive by hydrolyzing the β-lactam ring. Compounds capable of inhibiting β-lactamases, such as clavulanic acid, have obvious medical relevance in that they render β-lactamase-bearing pathogens sensitive to β-lactam antibiotics. Clavulanic acid is normally marketed in a mixture with a β-lactam antibiotic. One such mixture is called Augmentin (a combination of clavulanic acid and amoxicillin) and was the second-best selling antibacterial in 1995 with sales of US$1.3 billion (Demain and Elander 1999; Elander 2003). Clavulanic acid, like other clavams, has a β-lactam ring fused to an oxazolidine ring with a carboxyl group attached at the C3 position and a side chain at C2 (Howarth et al. 1976; Figure I.1). It is believed that the 3*R*, 5*R* stereochemistry of clavulanic acid is responsible for its inhibition of β-lactamases

(Baggaley et al. 1997). In agreement with this, the other clavams produced by *S. clavuligerus* with 5*S* stereochemistry do not exhibit β-lactamase inhibitory activity. These clavams are thus called the 5*S* clavams and are clavam-2-carboxylate, 2-hydroxymethylclavam, 2-formyloxymethylclavam, and alanylclavam. Although the 5S clavams do not inhibit β-lactamases, they are biologically active. When clavam-2-carboxylate, 2-hydroxymethylclavam, and 2-formyloxymethylclavam were first isolated from *S. clavuligerus* culture broths, they were found to be antifungal (Brown et al. 1979). Alanylclavam is bacteriostatic to a number of microbes through the inhibition of methionine biosynthesis (Pruess and Kellett 1983).

The structural similarities between clavulanic acid and the 5*S* clavams indicate that they likely share elements of a common biosynthetic pathway. The shared steps in their biosynthetic pathways are referred to as the "early stages," while the steps specific to either clavulanic acid or 5*S* clavam production are called the "late stages."

I.2.1 Early stages in clavam biosynthesis

The genes responsible for clavulanic acid biosynthesis by *S. clavuligerus* are clustered together on the chromosome adjacent to the cephamycin C biosynthetic gene cluster (Ward and Hodgson 1993; Figure I.2). The clavulanic acid gene cluster contains genes encoding enzymes involved in the early reactions required for both clavulanic acid and 5*S* clavam biosynthesis.

The first reaction dedicated to clavam production in *S. clavuligerus* is the condensation of L-arginine and glyceraldehyde-3-phosphate to produce N^2-(2-carboxyethyl)arginine (Khaleeli et al. 1999; Valentine et al. 1993; Figure I.3). This

reaction is catalyzed by carboxyethylarginine synthase 2 (CeaS2) which requires thiamine pyrophosphate and Mg^{2+} for activity. Mutation of *ceaS2*, the first gene in the clavulanic acid gene cluster and the gene encoding CeaS2, severely disables clavam production (Jensen et al. 2000; Perez-Redondo et al. 1999). The crystal structure of CeaS2 reveals that it exists as two dimers that associate loosely with each other to form a tetramer (Caines et al. 2004).

After this first reaction, N^2-(2-carboxyethyl)arginine undergoes cyclization to form a β-lactam ring-containing intermediate called deoxyguanidinoproclavaminate (Figure I.3). This reaction is catalyzed by β-lactam synthetase 2 (β-LS2), an ATP and Mg^{2+}-requiring enzyme encoded by *bls2* (Hodgson et al. 1995; Bachmann et al. 1998; McNaughton et al. 1998). Disruption of *bls2* also results in greatly decreased clavam metabolite yields (Bachmann et al. 1998; Jensen et al. 2000). Structural studies have shown that β-LS2 exists in solution as a dimer and is similar to a class of asparagine synthetases (Miller et al. 2001).

Clavaminate synthase 2 (CAS2) catalyzes the hydroxylation of deoxyguanidinoproclavaminate to form guanidinoproclavaminate (Baldwin et al. 1993; Figure I.3). CAS2 is encoded by *cas2*, the fourth gene in the clavulanic acid gene cluster (Marsh et al. 1992). This enzyme requires α-ketoglutarate, O_2, and Fe^{2+} to carry out catalysis (Salowe et al. 1990). Mutation of *cas2* causes mutants to produce lower amounts of clavams (Jensen et al. 2000). The crystal structure of clavaminate synthase reveals a β-barrel core in which the active site resides (Zhang et al. 2000).

The guanidino group of guanidinoproclavaminate is then removed by hydrolysis through the action of proclavaminate amidino hydrolase 2 (PAH2) (Wu et al. 1995;

Figure I.3). This reaction results in the formation of proclavaminic acid. The gene encoding PAH2, *pah2*, was first identified by Aidoo et al. (1994) and was found to be necessary for wild type levels of clavulanic acid production. Crystal studies of PAH2 show that it is a hexameric protein that requires two Mn^{2+} atoms in its active site for its hydrolytic activity (Elkins et al. 2002).

The next two reactions are both catalyzed by CAS2. First, proclavaminic acid undergoes cyclization to produce dihydroclavaminic acid; second, dihydroclavaminic acid is desaturated to produce clavaminic acid (Salowe et al. 1991; Figure I.3).

Although its exact role in clavam biosynthesis is unknown, the next gene in the clavulanic acid gene cluster, *oat2*, is categorized as an early gene because it is expressed in the same pattern as the other early genes (Jensen et al. 2000). This gene encodes an ornithine acetyltransferase (OAT2) that catalyzes the transfer of an acetyl group from *N*-acetylornithine to glutamate (Kershaw et al. 2002). Like other ornithine acetyltransferases, OAT2 undergoes autoproteolysis to produce α and β fragments which associate to form an $\alpha_2\beta_2$ heterotetramer (Kershaw et al. 2002). Insertional mutagenesis of *oat2* does result in decreased levels of clavam biosynthesis (Jensen et al. 2000). It has been hypothesized by Kershaw et al. (2002) that OAT2 may have a role in increasing cellular pools of arginine for use in clavam biosynthesis.

Clavaminic acid is believed to be the branch-point intermediate between the clavulanic acid and the 5*S* clavam biosynthetic pathways. Egan et al. (1997) demonstrated through feeding experiments with ^{13}C-labeled clavaminic acid that both clavulanic acid and other clavams incorporate the radioactive carbons into their final structures. Although it is thought that clavaminic acid is the last common intermediate to

both the clavulanic acid and 5*S* clavam pathways, the subsequent steps leading to the final compounds are unknown. However, the rest of the clavulanic acid gene cluster has been entirely sequenced, and two other gene clusters involved in clavam production by *S. clavuligerus* have also been identified, thus providing hints as to the nature of the steps leading from clavaminic acid to the final clavam metabolites.

I.2.2.1 Late stages in clavulanic acid biosynthesis

The next gene in the clavulanic acid gene cluster is *oppA1* (Hodgson et al. 1995; Jensen et al. 2000; Figure I.2). Interestingly, *oppA1* has significant similarity (47% identity at the amino acid level) to the *oppA2* gene found downstream in the clavulanic acid gene cluster (Mellado et al. 2002). Both of these genes are predicted to encode oligopeptide permeases and both contain a domain found in bacterial extracellular solute-binding proteins (Mellado et al. 2002). Mutants in either gene cannot produce clavulanic acid and *oppA2* mutants are unable to produce alanylclavam (Lorenzana et al. 2004). Analysis of *oppA1* and *oppA2* mutants revealed evidence that both are involved in peptide transport as their predicted amino acid sequences suggest, but the roles that the genes play in clavulanic acid production are still unclear (Lorenzana et al. 2004).

In the first step devoted specifically to producing clavulanic acid, clavaminic acid is converted to *N*-glycylclavaminic acid by ORF17 (Arulanantham et al. 2006; Figure I.3). Mutants in *orf17* were unable to produce clavulanic acid (Jensen et al. 2004a). In-depth analysis revealed that ORF17 is a member of the ATP-grasp fold superfamily and catalyzes a carboxylate-amine ligation reaction using ATP and glycine in the presence of Mg^{2+} and K^{+} (Arulanantham et al. 2006).

The last reaction leading to the production of clavulanic acid is carried out by CAD (clavaldehyde dehydrogenase), encoded by *cad* (Perez-Redondo et al. 1998; Figure I.3). CAD reduces clavaldehyde to form clavulanic acid (Nicholson et al. 1994). Unsurprisingly, mutation of *cad* prevents clavulanic acid production and overexpression of *cad* increases clavulanic acid production (Perez-Redondo et al. 1998; Jensen et al. 2000). Crystal structure studies reveal that CAD acts as a tetramer to reduce clavaldehyde using NADPH (MacKenzie et al. 2007).

As of yet, any other reactions leading to clavulanic acid production have not been characterized. However, all of the genes in the clavulanic acid gene cluster have been sequenced and subjected to mutagenic analysis, thus revealing their importance to clavam biosynthesis.

cyp resides in the ninth position in the clavulanic acid gene cluster (Figure I.2). This gene encodes a protein with high similarity to cytochrome P-450s and disruption of *cyp* blocks the production of clavulanic acid (Jensen et al. 2000; Li et al. 2000). It is believed that the protein encoded by *cyp* may carry out an oxidative reaction in cooperation with a ferredoxin protein encoded by the adjacent gene, *fd*, as is seen in other organisms (Li et al. 2000; Figure I.2). Mutation of *fd* results in significantly decreased clavulanic acid biosynthesis (Jensen et al. 2004a). Despite the suggestive proximity of *cyp* and *fd*, the reaction catalyzed by this putative enzymatic complex remains unclear. It is believed that CYP and FD may perform a double enantiomerization in the conversion of the 3*S*, 5*S* stereochemistry of clavaminic acid to the 3*R*, 5*R* stereochemistry of clavulanic acid. However, Jensen et al. (2004a) noted that mutants in either *cyp* or *fd*

produce decreased levels of 5*S* clavams as well as clavulanic acid, making it unclear where CYP and FD act in the clavam biosynthetic pathway.

Further downstream in the gene cluster is *orf12* (Figure I.2). The predicted amino acid sequence encoded by *orf12* shows that it has similarity to β-lactamases (Li et al. 2000). Although the similarity of *orf12* to β-lactamases is not indicative of a role in clavulanic acid biosynthesis, mutation of this gene causes the complete cessation of clavulanic production (Jensen et al. 2004a; Li et al. 2000).

The remainder of the genes in the clavulanic acid gene cluster (*orf13* to *orf19*) have been sequenced by both Mellado et al. (2002) and Jensen et al. (2004a). All of these genes have also been subjected to insertional mutagenesis by Jensen et al. (2004a). The groups' findings are discussed below.

orf13 encodes a protein that shows considerable similarity to amino acid metabolite efflux pumps from a number of bacterial species. A disruption mutant in *orf13* produces only low levels of clavulanic acid and the 5*S* clavams (5% relative to wild type), indicating that this gene is somehow involved in the production of all clavams. It is thought that the *orf13* protein product may be involved in the transport of clavam metabolites out of the cell.

Because *orf14* shares significant similarity with a number of acetyltransferases, it may be involved in the inactivation of antibiotics via chemical modification. The clustering of antibiotic resistance genes with genes encoding the biosynthesis of that same antibiotic is a common phenomenon. However, mutation of *orf14* results in the almost complete loss of clavulanic acid production. Such a phenotype for an *orf14*

mutant suggests that this gene may have a different role in the cell than as an antibiotic resistance gene.

orf16 shows only a small degree of similarity to hypothetical proteins from different microorganisms, such as *Roseovarius nubinhibens*, *Rhodococcus* sp. RHA1, and *Parvularcula bermudensis*. Although the gene product of *orf16* carries a conserved amino acid sequence believed to be required for attaching proteins to the cell wall, little else can be deduced from the gene's nucleotide sequence. Mutants in *orf16* are not only completely disabled in clavulanic acid production, but they produce two different clavams: *N*-acetylglycylclavaminic acid and trace amounts of *N*-glycylclavaminic acid. It should be mentioned that *oppA2* mutants also produce these extra clavam metabolites. While *N*-glycylclavaminic acid is now believed to be an intermediate in clavulanic acid biosynthesis, it is possible that *N*-acetylglycylclavaminic acid may represent an intermediate in clavulanic acid biosynthesis that has accumulated due to the absence of an enzyme in the pathway responsible for its conversion.

Besides *N*-glycylclavaminic acid and *N*-acetylglycylclavaminic acid, *N*-acetylclavaminic acid has also been detected in culture broths of an *S. clavuligerus* mutant unable to produce clavulanic acid (Elson et al. 1988). Since clavaminic acid is converted to *N*-glycylclavaminic acid by ORF17, the next step in the pathway may be the *N*-acetylation of *N*-glycylclavaminic acid by ORF14 to produce *N*-acetylglycylclavaminic acid. Because *orf16* mutants accumulate appreciable amounts of *N*-acetylglycylclavaminic acid, ORF16 may convert this intermediate to another metabolite, possibly *N*-acetylclavaminic acid.

The final two genes in the clavulanic acid gene cluster both encode proteins with high similarity to penicillin-binding proteins (PBPs). *orf18*'s deduced product has 76% identity to an *S. coelicolor* PBP while that of *orf19* has 74% identity to an *S. avermitilis* PBP. Mutational analyses of *orf19* and *orf18* suggested that they have no role in clavulanic acid production, but *orf18* does appear to be an essential gene. Even though these two genes do not seem to be involved in the making of clavulanic acid, they are considered to be part of the gene cluster since it does not seem likely that proteins often targeted by β-lactams would be found immediately adjacent to the cluster simply by chance.

Analysis of the region downstream of these two genes reveals an additional gene, *orf20*, which appears to encode a cytochrome P-450 (S. E. Jensen, personal communication). However, mutation of this gene does not appear to affect clavulanic acid production, so it is not yet certain whether *orf20* should be considered part of the clavulanic acid gene cluster.

I.2.2.2 Late stages in 5*S* clavam biosynthesis

While the late stages in clavulanic acid biosynthesis are unclear, the late stages in 5*S* clavam biosynthesis are even less certain. As mentioned earlier, Egan et al. (1997) determined that clavaminic acid is probably the branch-point intermediate between the pathways leading to clavulanic acid and the 5*S* clavams. The same group also proposed a mechanism in which clavaminic acid is converted into an aldehyde intermediate via decarboxylation and deamination in reactions involving a pyridoxal phosphate cofactor. This aldehyde intermediate could then be successively converted into 2-

formyloxymethylclavam, 2-hydroxymethylclavam, and clavam-2-carboxylate by oxidation, hydrolysis, and a further oxidation, respectively. Genes specifically involved in the production of 5*S* clavams have been sought with some success.

Preliminary evidence that separate gene clusters are involved in the biosynthesis of clavulanic acid and the 5*S* clavams was found when two separate *cas* genes (*cas1* and *cas2*) were isolated by Marsh et al. (1992). When the two separate CAS enzymes were originally purified, it was assumed that the smaller protein was a proteolytic fragment of the larger one (Salowe et al. 1990). However, once it was found that the two forms of CAS were actually encoded by two separate genes, it was hypothesized that one of the genes encodes the CAS enzyme involved in clavulanic acid biosynthesis while the other gene encodes the CAS enzyme responsible for 5*S* clavam production (Marsh et al. 1992).

Since the genes for clavulanic acid production are clustered together around *cas2*, a logical step in looking for genes involved in 5*S* clavam biosynthesis was to sequence the region around *cas1*. Mosher et al. (1999) undertook this endeavor and found a number of new genes (*cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, and *cvm6*) in what is now called the clavam gene cluster (Figure I.4). More recently, Tahlan et al. (2007) sequenced further outward to extend the length of the clavam gene cluster and carried out mutagenesis of the genes found in the region. Analysis of these genes has revealed a putative pathway of 5*S* clavam biosynthesis (Figure I.5).

The first one of these genes, *cvm1*, is located immediately upstream of *cas1* and is oriented in the opposite direction (Figure I.4). *cvm1* shows strong similarity to aldo and keto reductases. Mutation of *cvm1* does not affect clavulanic acid production but does eliminate clavam-2-carboxylate, 2-hydroxymethylclavam, and alanylclavam production

(Mosher et al. 1999). The importance of *cvm1* may lie in the reduction of an unidentified aldehyde or ketone clavam intermediate leading to the production of 5*S* clavams.

Another gene in the clavam gene cluster, *cvm2*, is found downstream of *cvm1* and its predicted amino acid sequence shows limited similarity to isomerases (Mosher et al. 1999; Tahlan et al. 2007; Figure I.4). Mutants in *cvm2* were severely compromised in 5*S* clavam production (Tahlan et al. 2007). It is unknown what role an isomerase may play in 5*S* clavam biosynthesis.

cvm3 is a putative flavin reductase which may provide reduced flavin for the protein product of *cvm5* (a gene downstream of *cas1*), a predicted mono-oxygenase (Mosher et al. 1999;Tahlan et al. 2007; Figure I.4). While mutagenesis of *cvm3* did not affect clavam production, mutants in *cvm5* were unable to produce normally observed 5*S* clavams and accumulated a possible intermediate in the pathway, 2-carboxymethylideneclavam (Tahlan et al. 2007; Figure I.5). It is possible that a flavin reductase other than Cvm3 can supply Cvm5 with reduced flavin in the absence of *cvm3*. It is currently thought that the protein encoded by *cvm5* may carry out a Baeyer-Villiger oxidation in the eventual conversion of 2-carboxymethylideneclavam to 2-formyloxymethylclavam (Figure I.5).

A number of genes have been identified even further upstream of *cas1*. These include *cvm7*, *cvm11*, *cvm12*, *cvm13*, *cvmH*, *cvmP*, and *cvmG* (Tahlan et al. 2007; Figure I.4). All of these genes, with the exception of *cvmH*, were subjected to mutagenesis and found not to be required for clavam production.

Several genes downstream of *cas1* have also been discovered and analyzed. Of *cvm4*, *cvm5*, *cvm6*, *cvm9*, and an incompletely sequenced gene, *cvm10*, only *cvm5* is

essential for 5*S* clavam production (Figure I.4). Included in this downstream region is a $tRNA^{ala}$ gene not yet subjected to mutagenic studies.

Although it appears that *cvm1*, *cvm2* and *cvm5* are all involved in 5*S* clavam biosynthesis, the roles of at least *cvm1* and *cvm2* remain unclear. As well, it seems unlikely that these few genes were responsible for the production of all of the 5*S* clavams, especially since knocking out these genes affected the production of all 5*S* clavams rather than the production of any single clavam.

Another line of investigation in the search for late 5*S* clavam biosynthetic genes involves the existence of other duplicate genes (besides *cas1* and *cas2*) for the early stages of clavam production. The finding that mutations in the early genes in the clavulanic acid gene cluster (*ceaS2*, *bls2*, *pah2*, *cas2*, and *oat2*) still allowed the production of clavulanic acid hinted at the existence of such duplicate genes (Jensen et al. 2000). It was already known that an additional *cas* gene was present in *S. clavuligerus*, so it was hypothesized that paralogous genes may also exist for *ceaS2*, *bls2*, *pah2*, and *oat2*. Indeed, a paralogue of *pah2* with 71% identity at the amino acid level, called *pah1*, was eventually found (Jensen et al. 2004b). Although a *pah1* mutant was not greatly affected in clavam production, a *pah1/pah2* double mutant was incapable of producing any clavams. As was performed for the clavam gene cluster, the region surrounding *pah1* was sequenced in the hopes of finding more paralogues or late 5*S* clavam biosynthetic genes. The sequencing and mutational analyses of the genes paralogous to those found in the clavulanic acid gene cluster discussed below were performed by Tahlan et al. (2004b).

Three genes sharing remarkable similarity to early genes in the clavulanic acid gene cluster were found in the region surrounding *pah1*. The first gene in this cluster, termed the paralogue gene cluster (Figure I.6), was named *ceaS1* and has 66% identity to *ceaS2* at the amino acid level. Downstream of this gene is *bls1*, which shares 49% identity to *bls2* at the amino acid level. The next gene found in the cluster is *pah1*. Sequencing further downstream revealed *oat1*, a gene with 47% identity to *oat2* at the amino acid level. While mutation of *ceaS1* did decrease clavulanic acid production, mutation of either *bls1* or *oat1* did not appear to affect clavam production. However, *ceaS1/ceaS2* and *bls1/bls2* double mutants were completely impaired in clavam biosynthesis, indicating that other paralogues of these genes do not exist. *oat1/oat2* mutants, though still able to produce clavams, produced lower amounts of clavulanic acid. This result supports the non-essential role proposed for the *oat* genes in increasing precursor flux into the clavam biosynthetic pathways.

Further downstream in the paralogue gene cluster are more paralogues, but not paralogues of genes in the clavulanic acid gene cluster. Instead, the two next genes are paralogues of genes found in the clavam gene cluster. *c6p* and *c7p* are paralogues of *cvm6* and *cvm7* (Figure I.6). Unlike *cvm6*, mutation of *c6p* (encoding a putative aminotransferase) does abolish 5*S* clavam production (Tahlan et al. 2007). It is predicted that *c6p* performs the deamination of clavaminic acid to produce the aldehyde intermediate predicted as predicted by Egan et al. (1997)(Figure I.5).

In contrast to *cvm7*, mutation of *c7p* eliminates 5*S* clavam biosynthesis without affecting clavulanic acid production (Tahlan et al. 2007). A gene upstream of *c7p*, *snk*, is also essential for 5*S* clavam biosynthesis (Tahlan 2005; Figure I.6). Since both of the

protein products encoded by these genes resemble regulatory proteins, they are discussed in more detail below (Section I.3).

A portion of the chromosome upstream of *ceaS1* has also been sequenced and most of the genes discovered there have been subjected to insertional mutagenesis (Cai 2003). Four of the 12 genes found in this upstream region were found to have possible roles in 5*S* clavam biosynthesis.

The four genes that appear to have roles in 5*S* clavam biosynthesis are *orfA*, *orfB*, *orfC*, and *orfD* (Cai 2003; Kwong and Jensen, unpublished; Figure I.6). *orfA*, an incompletely sequenced gene, is predicted to encode a glycine/serine hydroxymethyltransferase and preliminary study of mutants disrupted in *orfA* indicates that they may not produce alanylclavam but can produce clavulanic acid and other 5*S* clavams (Cai 2003). While *orfB* encodes a member of the YjgF/YER057c/UK114 family of proteins (involved in purine and amino acid biosynthesis in prokaryotes), *orfC* and *orfD* encode proteins with similarities to aminotransferases and threonine dehydratases, respectively (Cai 2003; Kwong and Jensen, unpublished). Mutation of these genes also appears to abolish alanylclavam production but does not affect the production of other clavams (Cai 2003; Kwong and Jensen, unpublished). Interestingly, mutants in *orfC* and *orfD* produce an additional unidentified clavam not produced by wild type cells (Cai 2003). This suggests that *orfC* and *orfD* mutants are deficient in a biosynthetic step leading to alanylclavam and accumulate an intermediate produced in the normal pathway. Naturally, it is expected that determining the identity of this clavam metabolite will shed some light on the alanylclavam biosynthetic pathway.

Unfortunately, much of the data indicating the importance of these genes in alanylclavam biosynthesis were inconsistent. During the cultivation of the mutant strains, the wild type cultures frequently also did not produce 5*S* clavams, thus throwing the implication of these genes in alanylclavam biosynthesis into doubt.

Other genes found in this upstream region include *orfL*, *orfK*, *orfJ*, *orfI*, and *orfE* (Cai 2003; Figure I.6). *orfL* and *orfI* are both predicted to encode transcriptional regulators, *orfK* and *orfJ* encode proteins without similarity to proteins with known functions, and *orfE* encodes a potential AzlD type amino acid permease (Cai 2003). Disruption of *orfI*, *orfJ*, *orfK*, and *orfL* does not affect clavam biosynthesis (Cai 2003). Although *orfE* mutants have been generated, their clavam-production capabilities have not yet been ascertained (Kwong and Jensen, unpublished).

Finally, the other genes in the paralogue gene cluster, *orfH*, *orfG*, and *orfF*, have not yet been subjected to mutagenesis (Figure I.6). *orfH* encodes a predicted multidrug resistance transporter, *orfG* encodes a putative protein kinase, while *orfF* encodes a possible AzlC type amino acid permease (Cai 2003). Until mutations are generated in these genes, their potential roles in 5*S* clavam biosynthesis remain unknown.

Recent analysis has revealed that the clavulanic acid gene cluster, the clavam gene cluster, and the paralogue gene cluster are unlinked and separated on the chromosome by at least tens to hundreds of kilobases of DNA (Tahlan et al. 2004c). Thus, it appears that the three gene clusters are genuine, individual groups of genes that contribute to the production of a group of biosynthetically related metabolites. This is quite interesting in view of the usual clustering of secondary metabolite biosynthetic genes.

I.3 Regulation of clavam biosynthesis in *S. clavuligerus*

A number of different factors regulate antibiotic biosynthesis in *Streptomyces* spp. In liquid cultures (in which only substrate mycelia typically develop), antibiotic production normally begins with the onset of stationary phase, as growth slows down (Demain 2000). During growth on surface cultures, secondary metabolism is coordinated with morphological differentiation, as the onset of antibiotic biosynthesis coincides with the development of aerial hyphae (Demain 2000; Chater 1989).

The molecular mechanisms behind the regulation of antibiotic biosynthesis are of considerable interest due to economic and medical benefits underlying increased antibiotic production. In *S. clavuligerus*, a variety of regulatory factors have been identified that play important roles in the biosynthesis of all clavams, only clavulanic acid, or only the 5*S* clavams.

One type of protein that has been found in a number of antibiotic regulatory circuits belongs to the *Streptomyces* Antibiotic Regulatory Protein (SARP) family (Wietzorrek and Bibb 1997). SARPs activate the transcription of genes involved in antibiotic biosynthesis by virtue of carrying an OmpR-like winged helix domain responsible for DNA-binding (Wietzorrek and Bibb 1997). Some of the best-characterized SARPs include ActII-ORF4 (required for actinorhodin production) and RedD (required for undecylprodigiosin production) from *S. coelicolor*, and DnrI (required for daunorubicin production) from *S. peucetius* (Fernandez-Morena et al. 1991; Narva and Feitelson 1990; Stutzman-Engwall et al. 1992). In *S. clavuligerus*, another SARP called CcaR (cephamycin and clavulanic acid regulator) is encoded in the

cephamycin C gene cluster and is essential for the production of cephamycin C, clavulanic acid, and the 5*S* clavams (Alexander and Jensen 1998; Perez-Llarena et al. 1997).

All of the aforementioned SARPs are involved in activating the transcription of structural genes that catalyze steps in antibiotic biosynthesis. ActII-ORF4 activates the transcription of the actinorhodin biosynthetic genes *act*III, *act*VI-ORF1 and *act*VI-ORFA (Arias et al. 1999; Gramajo et al. 1993). Likewise, mutants in *redD* exhibited decreased expression of genes required for undecylprodigiosin production, and the same was observed for the transcription of daunorubicin biosynthetic genes in a *dnrI* mutant (Madduri and Hutchinson 1995; Narva and Feitelson 1990). *ccaR* mutants are unable to produce a number of cephamycin biosynthetic enzymes, including LAT, ACVS, IPNS, IPNE and DAOCS (Alexander and Jensen 1998). As well, CcaR is required for the expression of a polycistronic transcript including *ceaS2*, *bls2*, *pah2*, and *cas2*, the early genes required for clavulanic acid biosynthesis (Paradkar and Jensen 1995; Tahlan et al. 2004a; Figure I.2 and Figure I.3). Interestingly, CcaR also appears to activate its own transcription by binding to its promoter region (Santamarta et al. 2002).

Because the SARPs described so far are required for the expression of antibiotic biosynthetic genes, they must be expressed prior to the appearance of antibiotic in cultures. Indeed, such has been demonstrated for ActII-ORF4 and RedD (Gramajo et al. 1993; Takano et al. 1992). Expression of *ccaR* also occurs before the production of cephamycin C and the clavams (Bignell et al. 2005; Perez-Llarena et al. 1997).

The consensus binding site of SARPs in the promoters of the actinorhodin and daunorubicin gene clusters is believed to be 5'-TCGAGC(G/C)-3' (Wietzorrek and Bibb

1997). DNase I footprint analyses indicated that ActII-ORF4 and DnrI bind to 5'-TCGAG-3' sequences in the promoters of regulated genes (Arias et al. 1999; Tang et al. 1996). Sequence analyses of a number of the promoters upstream of genes transcriptionally regulated by CcaR, such as *cefD*, *cmcI* and *ceaS2*, reveals that they contain sequence repeats that are very similar to those described above (Santamarta et al. 2002; Tahlan et al. 2004a).

Another of the genes regulated by CcaR is *claR*, which encodes a LysR-type transcriptional regulator found in the clavulanic acid gene cluster (Pradkar et al. 1998; Perez-Redondo et al. 1998; Figure I.2). ClaR contains a helix-turn-helix DNA-binding domain in both the N-terminus and the C-terminus (Pradkar et al. 1998; Perez-Redondo et al. 1998). Mutants in *claR* are unable to produce clavulanic acid (but can still produce cephamycin C, clavaminic acid, and the 5*S* clavams) while strains carrying multiple copies of *claR* overproduce clavulanic acid and alanylclavam (Paradkar et al. 1998; Perez-Redondo et al. 1998). ClaR is necessary for the transcription of *oppA1*, *cad*, and *cyp*, which are all late genes required for clavulanic acid production (Paradkar et al. 1998). CcaR mutants cannot express *claR* transcripts; thus, CcaR not only affects the early stages of clavulanic acid biosynthesis by activating the expression *ceaS2*, *bls2*, *pah2*, and *cas2* from a promoter upstream of *ceaS2*, but it is also necessary for the late stages of clavulanic acid production through *claR* (Perez-Redondo et al. 1998; Tahlan et al. 2004a).

A global regulator of *S. coelicolor* morphological differentiation and antibiotic production is called *bldA* (Merrick 1976). *bldA* mutants, while unaffected in vegetative growth, are unable to erect aerial hyphae or produce antibiotics (Merrick 1976). It was

later found that *bldA* encodes a leucyl tRNA for translation of UUA codons (Lawlor et al. 1987). It has been demonstrated that a number of genes containing TTA codons are severely dependent on *bldA* for their expression (Leskiw et al. 1991b). This is unsurprising, considering that *bldA* is the only leucyl tRNA specifying UUA translation in the *S. coelicolor* genome (Bentley et al. 2002). However, it has been found that some TTA-containing genes in *bldA* mutants do undergo some expression (and are thus only partially dependent on *bldA* for translation), possibly due to translation of UUA codons by a tRNA with an imperfectly-matched anticodon (Leskiw et al. 1991a; Leskiw et al. 1991b).

It is believed that *bldA* may regulate the expression of genes that are needed only in the late stages of growth, when growth slows down and secondary metabolism begins. Although *bldA* is transcribed throughout the growth of *S. coelicolor*, its active, processed form becomes available primarily during the later stages of growth, upon the formation of the aerial mycelium (Leskiw et al. 1993). This accumulation of mature *bldA* was found to coincide with the expression of a TTA-containing gene placed in *S. coelicolor* (Leskiw et al. 1993). Furthermore, TTA codons are found preferentially in genes involved in aerial mycelium formation, antibiotic regulation, and antibiotic resistance (Leskiw et al. 1991a).

In *S. coelicolor*, both actinorhodin and undecylprodigiosin biosynthetic pathways are governed by *bldA*. Interestingly, none of the genes encoding enzymes responsible for the enzymatic conversions in either pathway contain TTA codons. TTA codons are instead found in the genes regulating the expression of the biosynthetic genes. As described earlier, actinorhodin production is regulated by the transcriptional activator,

ActII-ORF4 (Fernandez-Moreno et al. 1991). *act*II-ORF4 contains a single TTA codon. *bldA* mutants cannot produce actinorhodin; when the TTA codon in *act*II-ORF4 was changed to a TTG codon (still coding for leucine), actinorhodin production was restored in a *bldA* mutant (Fernandez-Moreno et al. 1991). On the other hand, *redD*, encoding a direct transcriptional regulator of the undecylprodigiosin biosynthetic genes, does not contain a TTA codon (Narva and Feitelson 1990). However, *redD* transcription requires *bldA* (White and Bibb 1997). It was found that *redD* transcription is activated by RedZ, which is encoded by a TTA-containing gene and is translationally controlled by *bldA* (White and Bibb 1997).

CcaR, the cephamycin C and clavulanic acid regulator in *S. clavuligerus*, also carries a TTA codon in its gene (Perez-Llarena et al. 1997). This makes it a likely regulatory target for *bldA*. However, although *S. clavuligerus bldA* mutants do not form aerial hyphae, they are still capable of producing both cephamycin C and clavulanic acid (Trepanier et al. 2002). CcaR is still produced in *bldA* mutants even though it was demonstrated that TTA-containing genes may still normally be dependent upon *bldA* in *S. clavuligerus* as seen in *S. coelicolor* (Trepanier et al. 2002). It was hypothesized that CcaR could still be produced through the mistranslation of its single TTA codon by a non-cognate leucyl tRNA (Trepanier et al. 2002).

Even though *ccaR* expression was found to be independent of *bldA*, another gene, *bldG*, is required for the transcription of *ccaR* (Bignell et al. 2005). BldG, first characterized in *S. coelicolor*, shows similarity to anti-anti-sigma factors in *Bacillus subtiltis* and *Staphylococcus aureus* (Bignell et al. 2000). Sigma factors are components of an RNA polymerase that determine the DNA-binding specificity of the holoenzyme.

Anti-sigma factors bind to specific sigma factors and prevent the sigma factor from interacting with an RNA polymerase. When unphosphorylated, anti-anti-sigma factors will bind to certain anti-sigma factors and therefore allow the sigma factor to attach to RNA polymerases and direct the specificity of transcription. However, phosphorylated anti-anti-sigma factors are unable to interact with anti-sigma factors, thus allowing the inhibition of sigma factor activity. *S. coelicolor bldG* mutants cannot form aerial hyphae and do not produce either actinorhodin or undecylprodigiosin (Bignell et al. 2000). It was later found that BldG can also be phosphorylated and that mutants expressing non-phosphorylatable BldG could not produce aerial hyphae or antibiotics (Bignell et al. 2003).

In *S. clavuligerus*, a *bldG* homologue was also found to be required for aerial hyphae formation and for the production of cephamycin C and the clavams (Bignell et al. 2005). BldG is needed for the transcription of *ccaR*, which is presumably how BldG directs the production of cephamycin C and clavulanic acid (Bignell et al. 2005). However, since CcaR may not necessarily regulate 5*S* clavam biosynthesis, it is likely that BldG is responsible for the expression of at least one other gene involved in clavam production (Tahlan et al. 2004a).

For a number of *Streptomyces* spp., microbial hormones play an important role in regulating secondary metabolism. These microbial hormones have a γ-butyrolactone structure and are produced, secreted, and then taken up by surrounding cells, influencing the transcription of genes regulating antibiotic production and morphological development. One of the best-studied systems involves the A-factor hormone produced by *S. griseus* and required for aerial hyphae formation and streptomycin production (Hara

and Beppu 1982; Khokhlov et al. 1967). A-factor is synthesized using primary metabolites by the key enzyme encoded by the *afsA* gene (Ando et al. 1997; Horinouchi et al. 1989). As A-factor molecules enter cells, ArpA (A-factor receptor protein) binds to them and releases the DNA to which it is normally bound (Onaka et al. 1995; Onaka and Horinouchi 1997). Because ArpA is a repressor, once it is released from bound operator DNA, the transcription of downstream genes ensues (Ohnishi et al. 1999). One of the genes regulated by ArpA, *adpA* (A-factor dependent protein), is a transcriptional regulator that activates the expression of the *strR* gene (Ohnishi et al. 1999; Vujaklija et al. 1993). StrR is a positive regulator for streptomycin biosynthesis through transcriptional activation of the biosynthetic genes (Retzlaff and Distler 1995). Besides *strR*, a number of other genes involved in secondary metabolite production, antibiotic resistance, and morphological differentiation are regulated by *adpA* (Yamazaki et al. 2003; Yamazaki et al. 2004).

The binding sites of ArpA and other γ-butyrolactone receptors have been identified and characterized as AREs (autoregulatory elements)(Kinoshita et al. 1999; Onaka and Horinouchi 1997). Interestingly, an ARE has been identified upstream of *ccaR* in *S. clavuligerus* (Folcher et al. 2001). Additionally, *S. clavuligerus* has been found to produce a γ-butyrolactone autoregulator (Hashimoto et al. 1992). *S. clavuligerus* carries a gene called *scaR* that encodes a protein that binds to certain types of γ-butyrolactones (Kim et al. 2004). ScaR binds not only to the ARE upstream of *ccaR* but also to an ARE upstream of its own gene, *scaR* (Santamarta et al. 2005). Since a *scaR* mutant overproduces both cephamycin C and clavulanic acid, it is probable that

ScaR, like ArpA, acts as a repressor and inhibits the expression of *ccaR* by binding the upstream ARE (Santamarta et al. 2005).

During times of amino acid starvation, the polyphosphorylated nucleotide (p)ppGpp accumulates and greatly decreases the production of tRNA and rRNA. The buildup of (p)ppGpp and the subsequent decrease in stable RNA production is called the stringent response. (p)ppGpp is synthesized by RelA, a ribosome-associated protein, and by SpoT, which is independent of ribosomes. As well, the ribosomal protein L11 is important for RelA activity. Mutants with defects in the genes encoding these proteins typically produce low levels of (p)ppGpp and are called relaxed mutants.

Since *Streptomyces* spp. produce antibiotics upon entrance into stationary phase (a phase associated with the lack of essential nutrients), it would not be surprising if a connection existed between (p)ppGpp production and antibiotic biosynthesis. Ochi (1986) found that relaxed mutants of *Streptomyces* sp. MA406-A-1 produced low levels of ppGpp and were unable to produce the nucleoside antibiotic formycin. An *S. antibioticus* relaxed mutant produced low levels of ppGpp and actinomycin; in this case, it was confirmed that antibiotic production was hindered by the lack of actinomycin biosynthetic enzymes (Kelly et al. 1991). However, examination of the stringent response in *S. coelicolor* indicated that while (p)ppGpp may play a role in antibiotic production, it is unable to initiate antibiotic production by itself (Strauch et al. 1991). Initially, transcription of *redD* appeared to occur independently of cellular (p)ppGpp content in *S. coelicolor* (Takano et al. 1992). On the other hand, an *S. coelicolor relA* mutant incapable of producing ppGpp did not transcribe *redD* and only produced low levels of *act*II-ORF4 transcript during nitrogen limitation (Chakraburtty and Bibb 1997).

In *S. clavuligerus*, the accumulation of ppGpp precedes clavulanic acid production (Jones et al. 1997). Jin et al. (2004) isolated two *relA/spoT* homologues from the *S. clavuligerus* chromosome: *relA* and *rsh*. Both RelA and Rsh appear to be involved in (p)ppGpp synthesis; *relA* mutants were completely unable to produce (p)ppGpp and *rsh* mutants produced less than wild type (Jin et al. 2004). *relA* and *rsh* mutants could not produce cephamycin C and produced only very little clavulanic acid (Jin et al. 2004). Production of a mutant form of the L11 protein in *S. clavuligerus* resulted in a decrease in (p)ppGpp, cephamycin C, and clavulanic acid biosynthesis (Gomez-Escribano et al. 2006). The mutant produced only low levels of the *ccaR* transcript and low levels of transcripts of genes positively regulated by CcaR, indicating that secondary metabolite production of the mutant was affected through *ccaR* transcription (Gomez-Escribano et al. 2006). However, in wild type *S. clavuligerus* the expression of the cephamycin C and clavulanic acid biosynthetic genes appears to precede significant (p)ppGpp production (Gomez-Escribano et al. 2006; Jones et al. 1997). Despite the importance of RelA, Rsh and L11 in *S. clavuligerus* antibiotic production, it is unclear what kind of role, if any, is played by (p)ppGpp.

There also exists a fairly new class of antibiotic regulatory proteins that has an N-terminal SARP-like domain and a C-terminal ATPase domain. The biosynthetic gene cluster for the macrolide antifungal agent pimaricin in *S. natalensis* is transcriptionally regulated by PimR, the first example found of this type of regulator (Anton et al. 2004). Soon after the discovery of *pimR*, a very similar gene called *sanG* was found in *S. ansochromogenes* (Liu et al. 2005). *sanG* is required for normal morphological development and pigment production by *S. ansochromogenes* and for the biosynthesis of

the peptidyl nucleoside antibiotic nikkomycin (Liu et al. 2005). Further analysis revealed that *sanG* is also necessary for the transcription of two out of the three putative transcriptional units present in the nikkomycin biosynthetic gene cluster (Liu et al. 2005). A similar gene to *pimR* and *sanG* has been found close to the extreme downstream end of the paralogue gene cluster in *S. clavuligerus*. Disruption of this gene, called *c7p*, abolishes 5*S* clavam production without affecting morphological differentiation or clavulanic acid production (Tahlan et al. 2007; Figure I.6). The genes transcriptionally regulated by *c7p* have yet to be identified. Interestingly, *cvm7*, found in the clavam gene cluster, also appears to belong in this new class of antibiotic regulatory genes (Tahlan et al. 2007; Figure I.4). Mutation of *cvm7* does not affect clavam biosynthesis; *cvm7* may therefore regulate the production of other antibiotics produced by *S. clavuligerus* (Tahlan et al. 2007).

Immediately downstream of *c7p* is *snk*, a gene encoding a putative sensor kinase (Figure I.6). Sensor kinases are proteins that detect a specific stimulus and autophosphorylate in response to the stimulus. The phosphate molecule is then transferred to another protein called a response regulator. Response regulators are typically DNA-binding proteins that, upon phosphorylation, activate or repress the transcription of genes needed to deal with the original stimulus. A sensor kinase will usually recognize and phosphorylate its cognate response regulator. Sensor kinase/response regulator pairs are referred to as two-component systems and allow bacterial cells to sense and respond appropriately to varied stimuli.

Genes encoding cognate sensor kinases and response regulators tend to be immediately adjacent and are often transcribed within a single operon. For this reason, it

is expected that a response regulator will be found downstream of *snk*. Mutagenesis of *snk* abolishes 5*S* clavam production by *S. clavuligerus*, but does not affect clavulanic acid biosynthesis (Tahlan 2005).

In *Streptomyces*, it appears that two-component systems occur in abundance. The genome sequence of *S. coelicolor* revealed the presence of 84 genes encoding sensor kinases and 79 encoding response regulators, with 67 sensor kinase genes adjacent to response regulator genes (Bentley et al. 2002; Hutchings et al. 2004). Detailed analyses have revealed a number of *Streptomyces* two-component systems that, like *snk*, are involved in antibiotic production. The CutRS system, the first *Streptomyces* two-component system identified, represses actinorhodin production by *S. coelicolor* (Chang et al. 1996; Tseng and Chen 1991). The PhoPR system of *S. lividans* is not only required for growth in low phosphate media and the metabolism and transport of phosphate, but also represses actinorhodin and undecylprodigiosin production (Sola-Landa et al. 2003). These findings indicate that PhoPR undergoes phosphorylation and is active in times of phosphate starvation to ensure survival and that it serves to repress antibiotic production when phosphate is scarce.

An *S. coelicolor* two-component system, AfsQ1/Q2, activates actinorhodin and undecylprodigiosin production in *S. lividans*, but *S. coelicolor* mutants in either *afsQ1* (encoding a response regulator) or *afsQ2* (encoding a sensor kinase) did not display any defects in antibiotic production (Ishizuka et al. 1992). This likely reflects complex and subtle regulatory networks involved in the production of these antibiotics. The *S. coelicolor* OsaAB system appears to be involved in adaptation to high concentrations of salt (Bishop et al. 2004). Mutants in *osaB* (the response regulator gene) do not erect

aerial hyphae in media supplemented with high salt levels and produce elevated levels of actinorhodin and undecylprodigiosin (Bishop et al. 2004). The coordination of aerial hyphae development and secondary metabolism seems to be disrupted in *osaB* mutants; in this case, it is likely that *osaB* does not directly regulate the transcription of biosynthetic or regulatory genes for antibiotic synthesis.

One of the best-studied *Streptomyces* two-component systems is the AbsA1/A2 system in *S. coelicolor*. Although the original mutants in *absA1* (the sensor kinase of the system) did not produce any antibiotics, AbsA1/A2 is actually involved in repressing antibiotic biosynthesis (Adamidis et al. 1990; Brian et al. 1996). Those original mutations in *absA1* were likely gain of function mutations that caused constitutive phosphorylation of AbsA2 (the response regulator). Supporting the negative regulatory role of AbsA1/A2, Aceti and Champness (1998) found that mutants lacking *absA1/A2* expression produced high transcript levels of *act*II-ORF4 and *redD* (the pathway-specific transcriptional activators of the actinorhodin and undecylprodigiosin biosynthetic gene clusters, respectively). Further work has revealed that AbsA2 binds specifically to the promoter regions of *act*II-ORF4, *redZ* (a transcriptional activator of *redD*), and *cdaR* (a putative calcium-dependent antibiotic regulatory gene)(McKenzie and Nodwell 2007).

It is clear that a large number of factors influence clavam production in *S. clavuligerus*. While some of these regulatory factors are involved in both secondary metabolism and morphological differentiation, others are involved only in antibiotic production. This reinforces the apparent connection between the distinct *Streptomyces* life-cycle and the physiological changes that accompany morphological transformations. It must also be kept in mind that these regulatory circuits do not occur in isolation from

each other. For example, *ccaR* expression is likely regulated by BldG, a γ-butyrolactone autoregulator, and the L11 protein; CcaR is required for the transcription of the early genes in clavam biosynthesis in the clavulanic acid gene cluster, *claR*, and its own gene. In turn, ClaR is a positive regulator of some of the late genes of clavulanic acid biosynthesis. As more information is gathered, further interrelations between the known regulatory elements of clavam biosynthesis will be uncovered. A complete picture of the regulatory circuit governing clavam production will probably reveal a complex interdependence between morphological, nutritional, and antibiotic biosynthetic elements. One of the major ways in which these elements are investigated is through the production of gene-specific mutants. In recent years, this has been facilitated by understanding and taking adventage of molecular systems found in nature that are capable of carrying out otherwise difficult DNA manipulations, such as the FLP recombinase system found in yeast.

I.4 FLP recombinase in *Streptomyces* genetics

The yeast 2-μm plasmid carries a gene involved in recombination called *flp* (Broach et al. 1982). While the product of this gene, FLP recombinase, is necessary for the inversion of a segment of the 2-μm plasmid between two 599 bp inverted repeats, FLP has been recognized as a useful tool in bacterial genetic manipulation (Broach et al. 1982; Schweizer 2003).

Although FLP catalyzes site-specific recombination between the yeast 2-μm plasmid 599 bp inverted repeats, the minimal sequence needed for FLP-catalyzed recombination, called the FRT (FLP recombinase target) site, is only 34 bp in length

(Senecoff et al. 1985). FLP can act to either invert or excise DNA flanked by FRT sites. When two FRT sites are in the same orientation, the intervening DNA is excised; however, when the FRT sites are in the opposite orientation, the flanked DNA is inverted (Figure I.7). FLP is the only factor required for the site-specific recombination between FRT sites and is functional in prokaryotes (Cox 1983).

One of the common uses of FLP is to remove FRT-flanked segments of DNA that have been inserted into bacterial chromosomes. Usually, the mutation of a gene involves the insertion of an antibiotic resistance determinant into the gene of interest or the partial replacement of the gene with the marker. However, when releasing a bacterium into the environment it is desirable that it does not carry antibiotic resistance, as the gene responsible for resistance may be passed to other, possibly pathogenic, organisms. As well, if the mutated gene is transcribed in an operon, disruption of the gene with an antibiotic resistance determinant could result in polar effects on downstream genes. Removal of an antibiotic resistance gene is particularly useful when an unmarked mutant is desired or when polar effects on downstream genes must be avoided. Since a single FRT site does not appear to exert polar effects on downstream genes, excision of an antibiotic resistance gene flanked by FRT sites using FLP has been recognized as a useful method of generating unmarked nonpolar mutations (Merlin et al. 2002).

Earlier work revealed that FLP is functional in *E. coli* (Cox 1983). Later, FLP was used to excise antibiotic resistance genes replacing genes from the *E. coli* chromosome (Cherepanov and Wackernagel 1995). Further work revealed FLP can also be used to produce unmarked inframe mutations in other prokaryotes, such as *Pseudomonas aeruginosa*, *Mycobacterium smegmatis* and *Mycobacterium bovis* (Hoang

et al. 1998; Song and Niederweis 2007; Stephan et al. 2004). As of yet, FLP has not been demonstrated to be functional in members of the *Streptomyces* genus.

A recently-developed method of mutagenesis of *E. coli* makes use of the recombination functions of λ Red to delete genes of interest, replacing them with antibiotic resistance genes, and then uses FLP to remove unwanted sequences (Datsenko and Wanner 2000). The first step in this method is the amplification by PCR of an antibiotic resistance gene flanked by FRT sites. The primers used in the PCR carry 36-nt regions at their 5' ends that are identical to regions flanking the gene to be mutated. Resulting antibiotic resistance cassettes are transformed into *E. coli* cells expressing the λ Red function which greatly enhances the rate of recombination between chromosomal DNA and small linear pieces of DNA (Murphy 1998). This allows the replacement of genes with the FRT-flanked antibiotic resistance gene. The λ Red system is encoded on a temperature-sensitive plasmid and can thus be easily lost.

Once the initial mutant has been obtained, the antibiotic resistance cassette can be removed by the expression of FLP from a temperature-sensitive plasmid. Resulting mutants are unmarked and do not exhibit polar effects on downstream genes (Datsenko and Wanner 2000). The success of this method has led to its adaptation to *Streptomyces* genetics (Gust et al. 2003). This technique for the production of *Streptomyces* mutants is called REDIRECT©.

The main difference between REDIRECT© and the method developed by Datsenko and Wanner (2000) is that REDIRECT© requires that the λ Red- and FLP-mediated mutagenic events take place in another organism (*E. coli*) rather than the *Streptomyces* sp. of interest. The replacement of the gene of interest with the amplified

antibiotic resistance cassette using the λ Red function takes place on a cosmid carrying *Streptomyces* genomic DNA in *E. coli*. Since the cassette used for mutagenesis contains the RK2 *oriT*, the mutated cosmid can be delivered into the *Streptomyces* sp. of interest by intergeneric conjugation. Subsequent homologous recombination allows the mutation to be introduced into the *Streptomyces* chromosome.

To generate an unmarked nonpolar mutant, the mutant cosmid carrying the antibiotic resistance cassette can be transformed into an *E. coli* strain expressing FLP. This leads to the excision of the cassette with the resulting mutant cosmid carrying only an 81 bp "scar" flanked by start and stop codons in place of the original gene. The new mutant cosmid lacking the antibiotic resistance gene is then transformed into the original *Streptomyces* mutant. At this point, conjugation can no longer be used to transfer the cosmid into *Streptomyces* cells because the *oriT*, originally located on the antibiotic resistance cassette, has been removed, and so a more laborious transformation process is required. Once again, homologous recombination allows the "scar" to be introduced into the chromosome, replacing the antibiotic resistance cassette. The REDIRECT© technique is depicted in Figure I.8.

Although REDIRECT© facilitates the generation of marked mutants in *Streptomyces*, there are difficulties in producing unmarked nonpolar mutants. Some *Streptomyces* spp., such as *S. clavuligerus*, cannot be transformed with large (cosmid-sized) DNA molecules. For this reason, it would be desirable to express FLP directly within the *Streptomyces* mutant bearing the antibiotic resistance gene to produce the unmarked mutant. As well, in vivo expression of FLP would reduce the number of steps required to produce final mutants in any *Streptomyces* sp. Currently, the only way to

produce unmarked nonpolar mutants using REDIRECT© in *S. clavuligerus* requires that the region of the cosmid carrying the antibiotic resistance cassette be cloned into a smaller replicon that can be transformed into *S. clavuligerus*, or into another replicon bearing *oriT*. This new plasmid carrying the cassette is then introduced into *E. coli* cells expressing FLP to replace the cassette with the 81 bp "scar." The scar-bearing plasmid is then introduced into the original *S. clavuligerus* mutant so that homologous recombination can take place. This method was used to generate an unmarked in-frame Δ*ceaS2* mutant (Tahlan 2005).

Because difficulties have arisen with the expression of FLP in *Streptomyces*, other methods have been developed to generate unmarked nonpolar mutants. One such technique uses the same general steps as REDIRECT©, but uses antibiotic resistance cassettes that are flanked by *loxP* sites rather than FRT sites (Khodakaramian et al. 2006). Expression of the Cre recombinase (which catalyzes site-specific recombination between *loxP* sites) within the *Streptomyces* mutant allows the antibiotic resistance cassette to be removed directly from the chromosome (Khodakaramian et al. 2006). Although this method is faster than using FLP in *E. coli*, it requires that researchers who have already used REDIRECT© create mutants in the same way but with the new *loxP*-flanked cassettes.

Another method for generating unmarked in-frame mutants in *Streptomyces* spp. involves the use of antibiotic resistance genes flanked by *attR* and *attL* sites (Raynal et al. 2006). In this case, recombination between the *attR* and *attL* sites to excise intervening DNA is catalyzed by the products of the *xis* and *int* genes (Raynal et al. 2006). The cassettes can replace entire genes or portions of genes on a replicon through restriction

digest followed by ligation. The resulting mutant plasmid is then transferred into the *Streptomyces* spp. of interest and homologous recombination allows the mutation to be introduced onto the chromosome. An unstable plasmid expressing the Xis and Int proteins is then delivered into the mutant and an in-frame deletion mutant lacking the antibiotic resistance gene can be produced (Raynal et al. 2006). This method can be changed so that the λ Red functions are used to replace the gene of interest with the antibiotic resistance cassette to bypass the need to clone the antibiotic resistance cassette into the gene of interest.

Although alternative methods exist for the generation of unmarked inframe mutants, it is still desirable to develop a method to express FLP in *Streptomyces* cells. This is because many researchers still use REDIRECT© on a regular basis, are familiar with the methods and already possess the materials used in the process. As well, it is likely that a number of laboratories have produced the initial antibiotic resistant mutants, but have encountered difficulties in producing the final unmarked mutants similarly to what has been found in studies with *S. clavuligerus*.

I.5 Thesis objectives

Although the region upstream of *ceaS1* in the paralogue gene cluster had been mostly sequenced and subjected to preliminary mutagenic analysis, there were still some unanswered questions regarding its importance to clavam production (Figure I.6). *orfA* was not completely sequenced, and the phenotypes of *orfA*, *orfB*, *orfC*, and *orfD* mutants had to be confirmed. Initial work found that mutants in any of these genes are unable to produce alanylclavam, but inconsistencies in the data made these mutants worth

reexamining. As well, mutants had not yet been generated in *orfF*, *orfG*, and *orfH*. To obtain a complete set of mutants in the genes of the upstream region of the paralogue gene cluster, *orfF*, *orfG*, and *orfH* should be subjected to mutagenesis. Such mutants had to be acquired to assess the importance of the entire paralogue gene cluster in clavam biosynthesis.

To find more genes involved in 5*S* clavam biosynthesis, the downstream region of the paralogue gene cluster was sequenced and analyzed. Since a sensor kinase required for 5*S* clavam production is located at the far known downstream end of the cluster, it was likely that a response regulator with the same importance for clavam production was proximal (Figure I.6). Once the additional genes were sequenced, they were mutated and mutants were assessed for clavam biosynthesis. The putative two-component system was further analyzed for phosphotransfer between the proteins.

To facilitate the production of unmarked nonpolar mutants, plasmids were constructed for high levels of FLP expression in *Streptomyces* cells. These plasmids were tested for their ability to generate unmarked inframe mutations in initial antibiotic resistant REDIRECT© *S. clavuligerus* and *S. coelicolor* mutants.

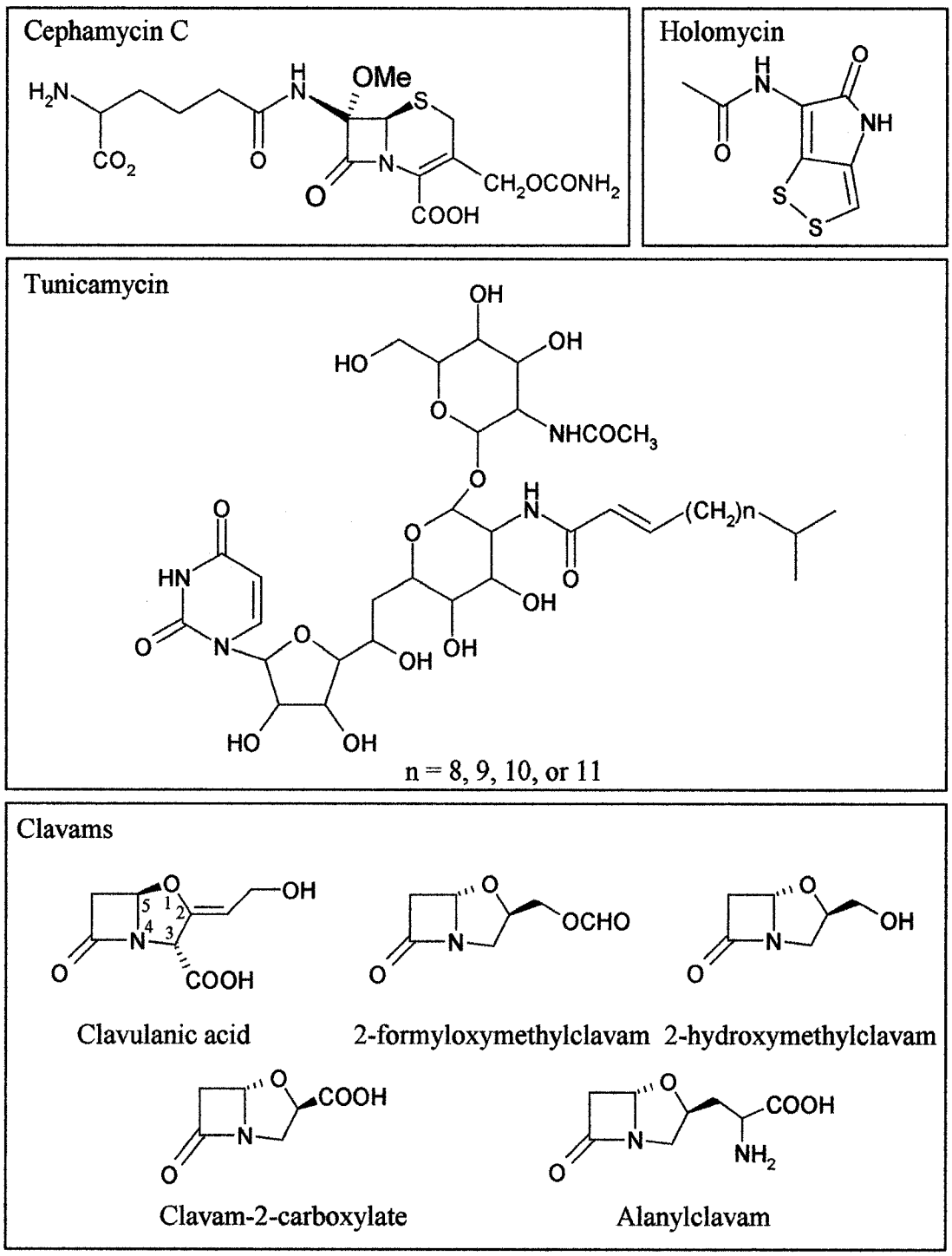


Figure I.1. Antibiotics produced by *S. clavuligerus*.

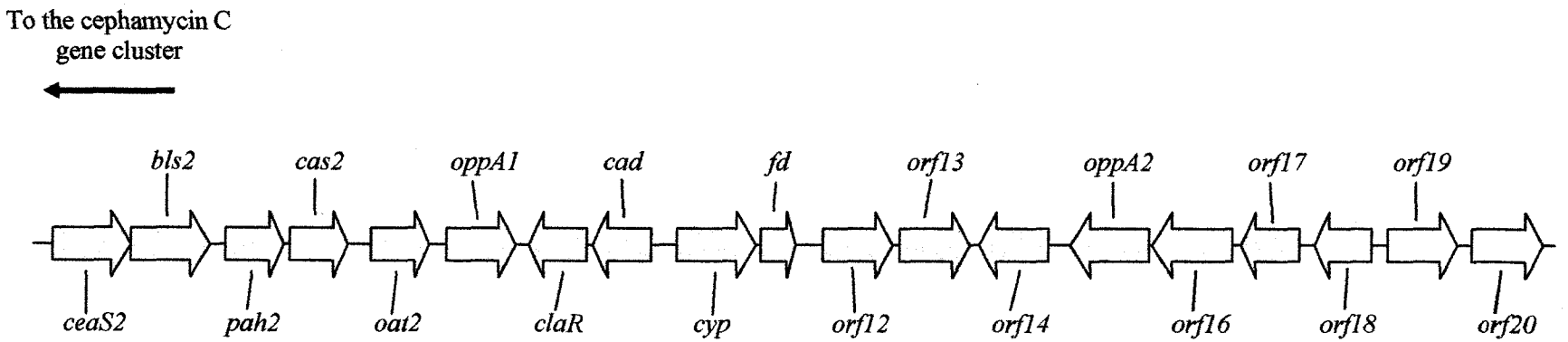


Figure I.2. The clavulanic acid gene cluster of *S. clavuligerus*. Genes of confirmed importance for clavulanic acid biosynthesis are gray and genes that are not required for clavam production are white.

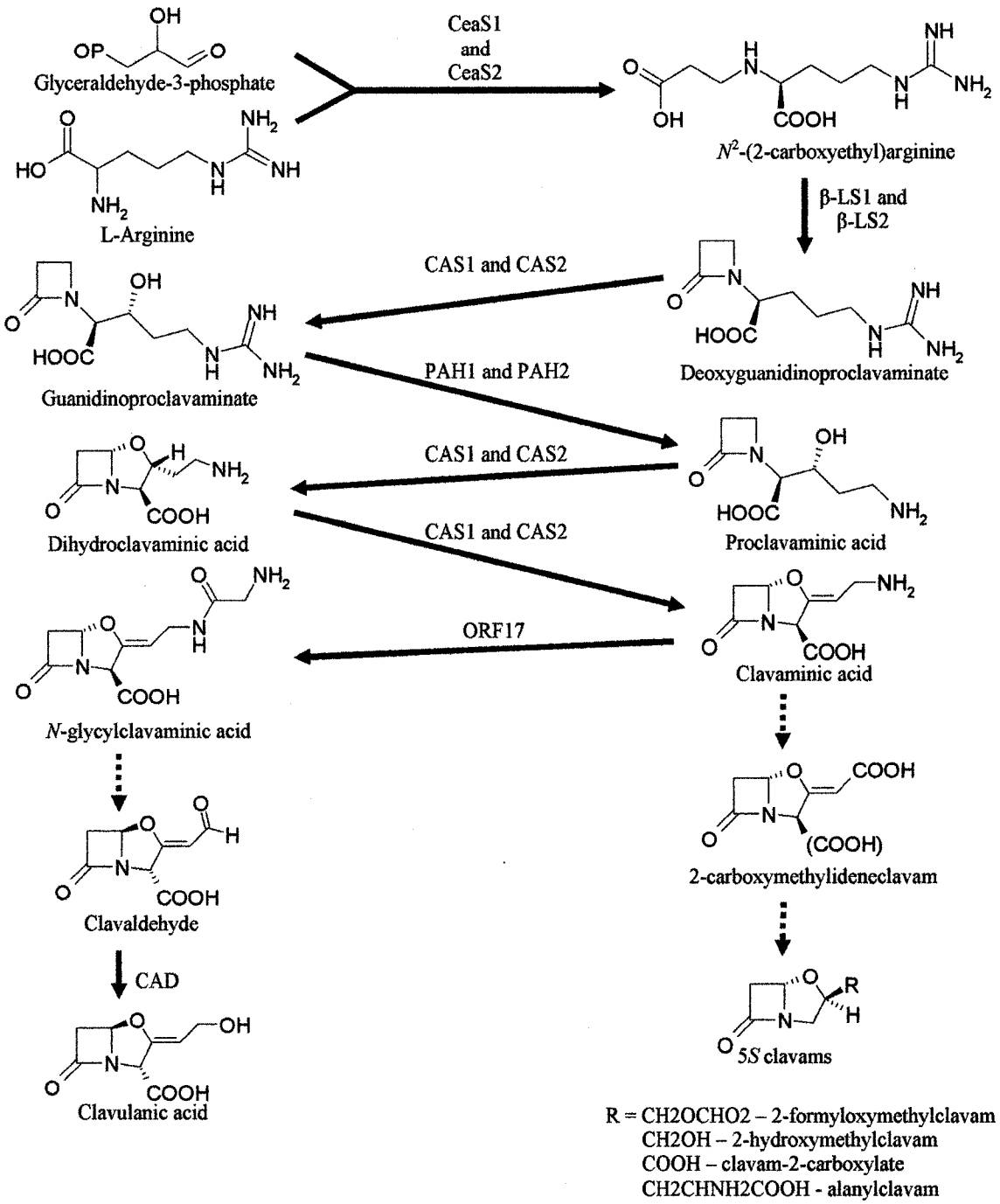


Figure I.3. The clavam biosynthetic pathway leading to the production of clavulanic acid and the 5*S* clavams in *S. clavuligerus*. Protein names adjacent to arrows indicate involvement in reactions Characterized reactions are indicated by full arrows and uncharacterized reactions are indicated by dotted arrows.

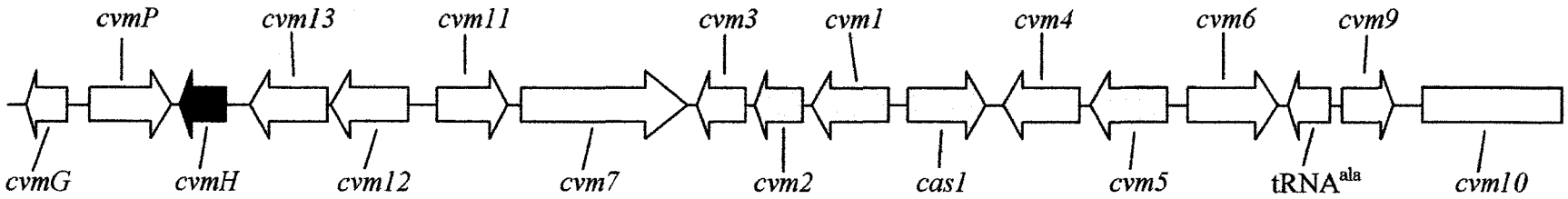


Figure I.4. The clavam gene cluster of *S. clavuligerus*. Genes of confirmed importance for clavam acid biosynthesis are gray, genes that are not required for clavam production are white, and genes of unknown importance to clavam biosynthesis are black.

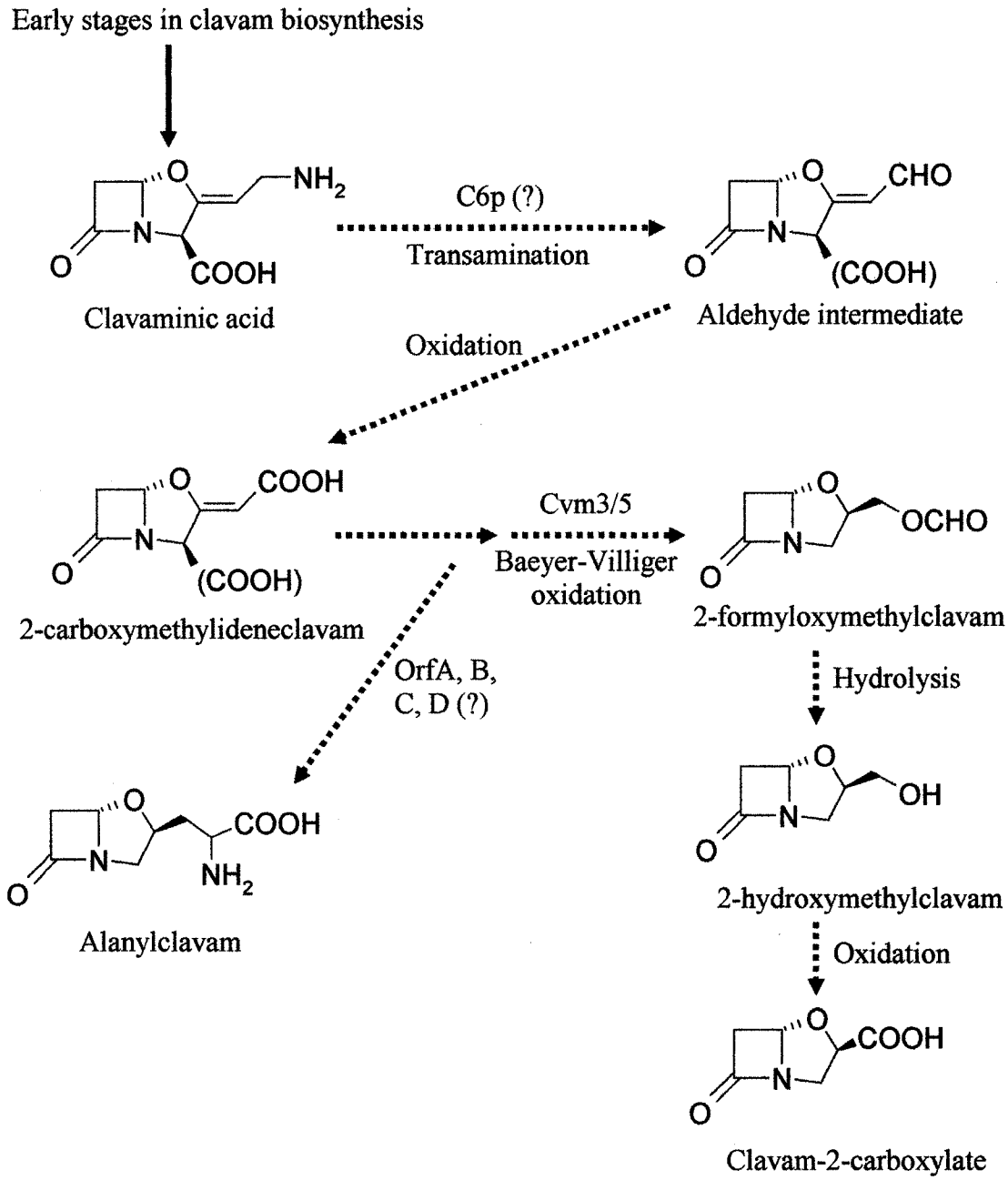


Figure I.5. The hypothetical pathway leading from clavaminic acid to the 5*S* clavams of *S. clavuligerus*. Uncharacterized reactions are indicated by dotted arrows. Protein names adjacent to arrows indicate involvement in reactions, and question marks indicate a protein's hypothetical role in a reaction. The C3 carboxylic acid group is in brackets to

indicate that decarboxylation occurs at some unknown point in the pathway before the appearance of 2-formyloxymethylclavam and alanylclavam. The aldehyde intermediate is a hypothetical clavam intermediate.

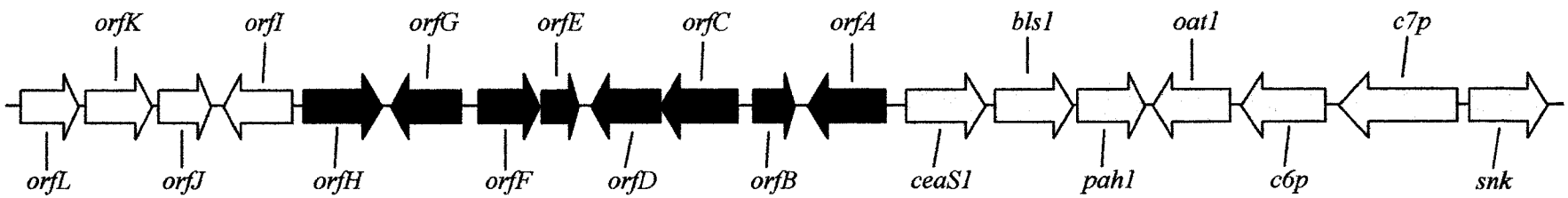


Figure I.6. The paralogue gene cluster of *S. clavuligerus*. Genes of confirmed importance for clavam acid biosynthesis are gray, genes that are not required for clavam production are white, and genes of unknown importance to clavam biosynthesis are black.

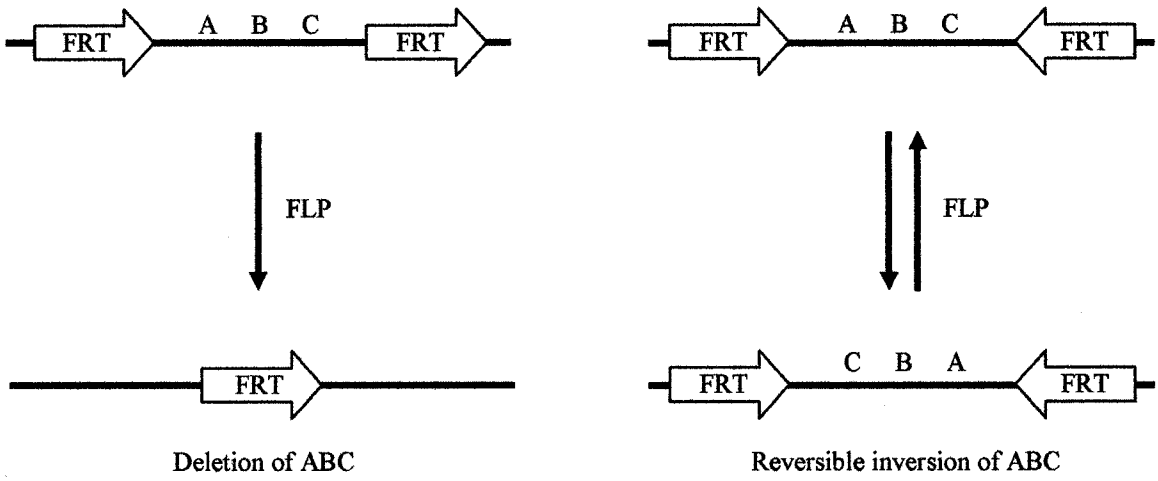


Figure I.7. The action of FLP recombinase in the presence of FRT sites. When FRT sites are oriented in the same direction, DNA between the sites is removed by site-specific recombination. When FRT sites are oriented in the opposite direction, DNA between the sites undergoes reversible inversion. Adapted from Schweizer (2003).

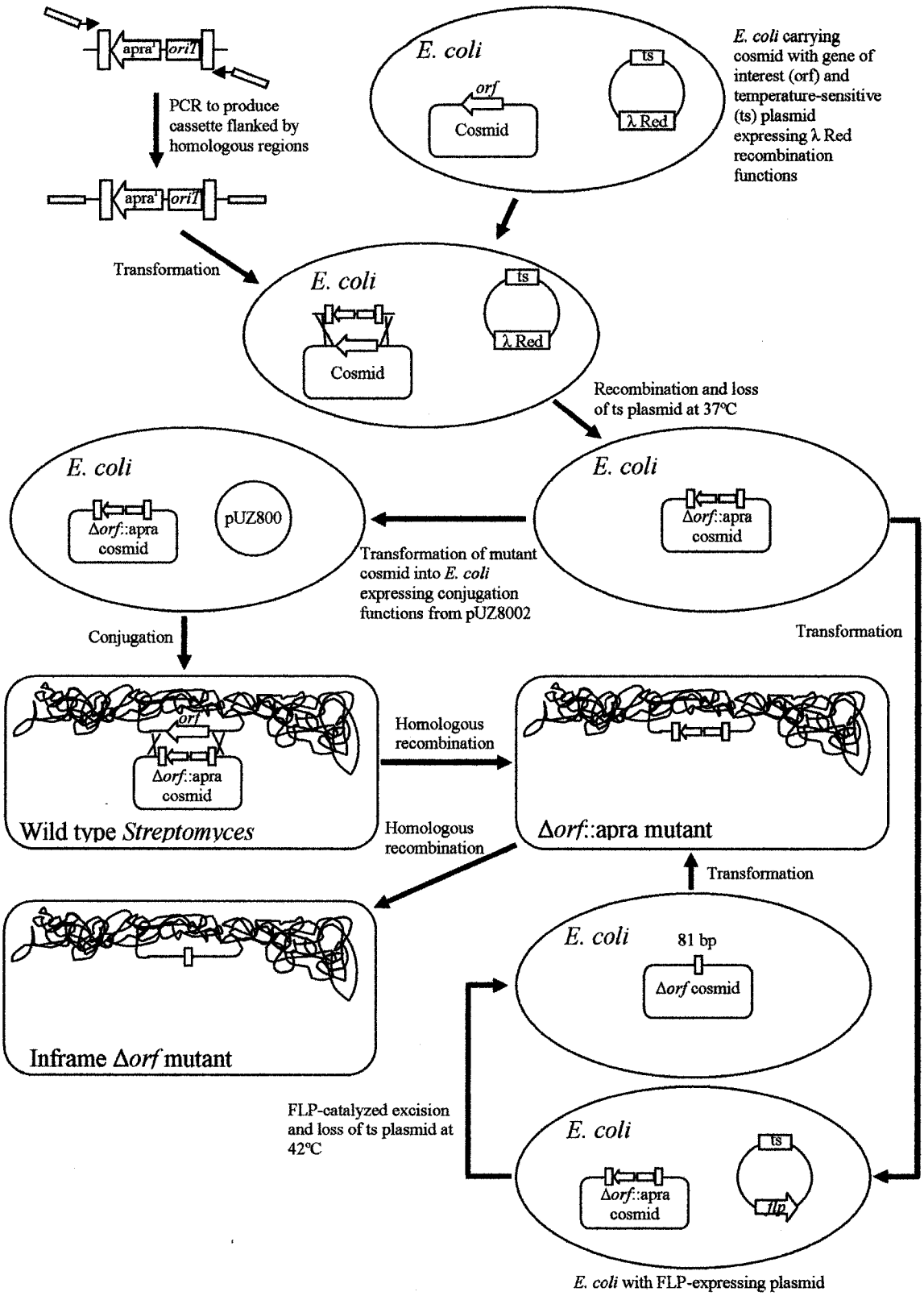


Figure I.8. The REDIRECT© protocol for producing unmarked inframe deletion mutants

in *Streptomyces* spp. Adapted from Gust et al. (2003).

II. MATERIALS AND METHODS

II.1 Bacterial strains, culture media, and culture conditions

All *Escherichia coli*, *S. clavuligerus*, and *S. coelicolor* strains used in this study are described in Table II.1, Table II.2, and Table II.3, respectively.

E. coli strains were typically grown in Lennox broth (LB; 1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 0.5% NaCl) at 37°C on a Cell Production Roller Drum (Bellco Biotechnology, Vineland, NJ) or a G-24 Environmental incubator shaker (New Brunswick Scientific Co., Edison, NJ) unless otherwise indicated. Solid LB (LBA) was used with 1.5% agar to produce cultures on a solid medium. Antibiotic-resistant strains of *E. coli* were supplemented with ampicillin (100 μg/ml), kanamycin (50 μg/ml), apramycin (50 μg/ml), or chloramphenicol (25 μg/ml).

Seed cultures of *S. clavuligerus* strains were grown by inoculating 10^8 spores into 25 ml of TSBS [3% (w/v) trypticase soy broth and 1% (w/v) soluble starch] and shaking cultures in a Model G-25 rotary shaker (New Brunswick Scientific Co., Edison, NJ) at 250 rpm and 28 °C. After 40 hours of growth, a 2% (v/v) inoculum was used to inoculate 100 ml of soy medium [15 g/L soy flour, 47 g/L soluble starch, 0.1 g/L KH_2PO_4, and 0.2 g/L $FeSO_4 \cdot 7H_2O$, (pH6.8)](Salowe et al. 1990). Culture supernatants were typically isolated from soy cultures after 72 and 96 hours of growth in the same conditions as seed cultures. Solid cultures of *S. clavuligerus* were grown on MYM [0.4% (w/v) maltose, 0.4% (w/v) yeast extract, 1% (w/v) malt extract, and 1.8% (w/v) agar](Stuttard 1982), ISP-4 (DIFCO), or solid soy medium [soy liquid medium with 1.5% (w/v) agar added] at 28°C. Antibiotic-resistant strains of *S. clavuligerus* were grown with apramycin (25 μg/ml), kanamycin (50 μg/ml), or thiostrepton (2.5-10 μg/ml).

S. coelicolor liquid cultures were grown in 25 ml of R5 [103 g/L sucrose, 0.25 g/L K_2SO_4, 10.12 g/L $MgCl_2 \cdot 6H_2O$, 10 g/L glucose, 0.1 g/L Casaminoacids, 2 ml/L trace element solution (40 mg/L $ZnCl_2$, 200 mg/L $FeCl_3 \cdot 6H_2O$, 10 mg/L $CuCl_2 \cdot 2H_2O$, 10 mg/L $MnCl_2 \cdot 4H_2O$, 10 mg/L $Na_2B_4O_7 \cdot 10H_2O$, 10 mg/L $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$) 5 g/L yeast extract, 5.73 g/L TES buffer, 0.05 g/L KH_2PO_4, 20 mM $CaCl_2 \cdot 2H_2O$, 0.03 g/L L-proline, and 7 mM NaCl](Kieser et al. 2000) supplemented with 5% polyethylene glycol (PEG) 8000 and in flasks with metal springs, in the same growth conditions as *S. clavuligerus*. Solid cultures of *S. coelicolor* strains were propagated on Nutrient Broth Agar (NA; 31 g/L Difco Nutrient Agar) or mannitol soy agar (MS; 2% mannitol, 2% soy flour, and 2% agar)(Hobbs et al. 1989). *S. coelicolor* strains resistant to antibiotics were grown with apramycin (50 μg/ml), kanamycin (50 μg/ml), or thiostrepton (50 μg/ml).

II.2 Preparation and storage of bacterial glycerol stocks

E. coli strains were grown in LB overnight at 37°C and 1.5 ml of culture was harvested in a 1.5 ml microfuge tube. The medium was aspirated and the resulting cell pellet was resuspended in 200 μl 20% (v/v) glycerol.

Streptomyces spores were scraped from ISP-4 (for *S. clavuligerus*) or MS (for *S. coelicolor*). *S. clavuligerus* spores were resuspended in 1 ml dH_2O and placed in a sonication bath for five minutes to break up clumps of spores. Four millilitres of dH_2O was added to the spore suspension which was then passed through a syringe containing non-absorbent cotton. This allowed mycelia to be separated from spores to obtain stocks containing only spores. Spores were harvested from the filtered suspensions via

microcentrifugation into 1.5 ml microfuge tubes and were resuspended in 500 µl 20% (v/v) glycerol. All microcentrifugation was carried out in an Eppendorf 5415 C microcentrifuge at 14,000 rpm unless otherwise stated (Brinkmann Instruments Inc., Mississauga, ON). *S. coelicolor* spore stocks were prepared by placing the spores in 500 µl 20% (v/v) glycerol and sonicating them for 5 minutes.

All glycerol stocks were preserved at -80°C.

II.3 Plasmids and cosmids

Plasmids and cosmids used in this study are summarized in Table II.4.

II.4 DNA methods

II.4.1 Isolation of plasmid and cosmid DNA from *E. coli*

Plasmid and cosmid DNA was isolated via the alkaline lysis method described by Birnboim and Doly (1979) and Sambrook et al. (1989).

II.4.2 Isolation of genomic DNA from *S. coelicolor*

S. coelicolor cultures grown at 28°C for 48 hours in R5 supplemented with 5% PEG 8000 had genomic DNA isolated using the GenElute™ Bacterial Genomic DNA kit as specified by the manufacturer (Sigma-Aldrich).

II.4.3 DNA quantification

Purified DNA was quantified using a UNICAM UV/Vis Spectrophotometer UV3 (ATI Unicam, Cambridge, UK) or with a NanoDrop® ND-1000 Spectrophotometer

(NanoDrop Technologies, Inc., Wilmington, DE). When using the UV/Vis Spectrophotometer, the machine was calibrated with the solvent in which the DNA was dissolved and then absorbance at 260 nm was measured for each sample. Concentration of DNA was calculated by multiplying the A_{260} by the dilution factor and then by 50 µg/ml DNA. After blanking the machine with Milli-Q™ water, the NanoDrop® scanned the absorbance of the DNA sample from 220 to 350 nm, provided the concentration of the sample, and calculated the A_{260}:A_{280} ratio for estimations of purity.

II.4.4 Introduction of DNA into *E. coli*

II.4.4.1 Preparation of electrocompetent *E. coli* cells

Strains of *E. coli* to be made electrocompetent were inoculated in 10 ml LB with the appropriate antibiotics. This culture was grown overnight at 37°C or another suitable growth temperature (such as 30°C for *E. coli* BW25113/pIJ790 and DH5α/BT340). This culture was used to inoculate 250 ml SOB [2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM $MgSO_4$] at 1% (v/v). This culture was then shaken at 37 °C until the OD_{600} reached 0.4-0.6. At this point, the cells were harvested by centrifugation at 4°C and then washed once with an equal volume of ice-cold Milli-Q™ water. Cells were then washed with a half-volume of ice-cold Milli-Q™ water and finally washed once more with 8 ml ice-cold Milli-Q™ water. The cells were resuspended in 1ml 10% (v/v) glycerol and 40 µl aliquots were transferred into 1.5 ml microfuge tubes as they were flash-frozen in 95% ethanol containing dry ice. These 40 µl samples of electrocompetent cells were stored at -80°C and used for single electroporations.

When performing the PCR-targeting REDIRECT© protocol, some changes were made to the above protocol. In this case, the *E. coli* strain was BW25113/pIJ790 carrying either 14E10 or St3f7. The SOB for such cultures also contained 10 mM L-arabinose to induce the expression of the λ Red genes required for recombination carried on pIJ790. SOB cultures were grown to an OD_{600} of ~0.6 and then harvested at 4°C. Cells were washed twice: once with an equal volume and once with a half-volume of 10% (v/v) glycerol. These cells were resuspended in the residual fluid and used immediately in 50 µl aliquots for the electroporation of PCR product.

II.4.4.2 Electroporation of DNA into electrocompetent *E. coli* cells

Electrocompetent *E. coli* cells were thawed on ice prior to electroporation. DNA was mixed with 40 µl electrocompetent cells. This mixture was placed in an ice-cold 2 mm gap BTX® Disposable Cuvette Plus™ (Genetromics, Inc., San Diego, CA). The electroporation cuvette containing the mixture was then pulsed using a Bio-Rad GenePulser II device set to 200 Ω, 25 µF and 2.5 kV. After pulsing, 1 ml LB was added to the cells and this suspension was incubated at 37°C for one hour. Cells were then plated onto LB agar containing appropriate antibiotics at 1/10 and 9/10 dilutions. When performing blue/white screening to detect DNA insertion into vectors, 100 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 40 µg/ml 5-bromo-4-chloro-3-β-D-galactopyranoside (X-gal) were added to the LB agar.

II.4.5 Introduction of DNA into *Streptomyces*

DNA was introduced into *S. clavuligerus* and *S. coelicolor* by inter-generic conjugation following the procedure outlined by Kieser et al. (2000). *E. coli* ET12567/pUZ8002 carrying the DNA molecule to be transferred was grown overnight in 10 ml LB at 37°C. One hundred microlitres of the overnight culture was inoculated into 10 ml of fresh LB and antibiotics. This culture was grown at 37°C until it reached an OD_{600} of 0.4-0.6. Cells were then harvested, washed twice with LB lacking antibiotics and resuspended in ~0.6ml residual fluid. *Streptomyces* spores were suspended in 2xYT (16 g/L tryptone, 10 g/L yeast extract, and 5 g/L NaCl) (Sambrook et al. 1989) and germinated by incubating them at 50°C for 10 minutes. After cooling *Streptomyces* spores to room temperature, they were combined with the washed *E. coli* cells carrying the DNA to be conjugated. For *S. clavuligerus*, the mixture was mixed and plated onto AS-1 [0.1% (w/v) yeast extract, 0.02% (w/v) L-alanine, 0.02% (w/v) L-arginine, 0.05% (w/v) L-asparagine, 0.5% (w/v) soluble starch, 0.25% (w/v) NaCl, 1% (w/v) Na_2SO_4, and 2% (w/v) agar (pH 7.5)] (Baltz 1980) supplemented with 10mM $MgCl_2 \cdot 6H_2O$. When conjugating DNA into *S. coelicolor*, the cell mixture was plated onto MS (Hobbs et al. 1989). Appropriate dilutions were plated after an initial attempt at spreading the entire mixture of cells onto a single plate. The resulting plates were incubated at 28°C for 16-20 hours. They were then overlaid with 1 ml of water containing nalidixic acid (25μg/ml, final concentration in the plate) and apramycin (25 μg/ml for *S. clavuligerus* and 50μg/ml for *S. coelicolor*) or thiostrepton (5 μg/ml for *S. clavuligerus* and 8 μg/ml for *S. coelicolor*), or both apramycin and thiostrepton. After 7-10 days of growth at 28°C, *S. clavuligerus* exconjugant colonies were then patched onto MYM (Stuttard 1982)

containing the appropriate antibiotics, and, after further growth on MYM, streaked onto ISP-4 with suitable antibiotics to generate spores.

II.4.6 Construction of *Streptomyces* mutants

Most *Streptomyces* mutants were made using the REDIRECT© PCR-targeting protocol (Gust et al. 2003). Briefly, this method involves the electroporation of a PCR-amplified cassette flanked by 39-nucleotide regions that also flank the gene to be mutated, into *E. coli* BW25113/pIJ790/14E10 or St3f7. The PCR-amplified cassette, carrying an apramycin resistance gene *aac3(IV)*, the RP4 origin of transfer, *oriT*, also has FRT (FLP recognition target) sites flanking these two elements. Once the cassette is introduced into the BW25113/pIJ790 strain carrying the cosmid of interest, the λ Red proteins (encoded by genes present on pIJ790) catalyze recombination between the DNA flanking the gene of interest on the cosmid and the 39-nucleotide flanking regions on the cassette. Cells having undergone the electroporation were plated onto LB agar supplemented with 50 μg/ml apramycin. Successfully mutagenized cosmids were confirmed by restriction digest and were then conjugated into *S. clavuligerus* NRRL 3858 or *S. coelicolor* M145. Apramycin resistant colonies were selected and used to generate mutant strains.

To generate a mutation in *orf2/3*, a mutant cosmid carrying a Tn*5062* insertion in *orf2/3*, II-G10, was conjugated into *S. clavuligerus*. Tn*5062* could be conjugated because it carries *oriT* and also contains *aac3(IV)*. In this way, apramycin resistant exconjugant colonies were isolated and used as mutant strains.

II.4.7 Agarose gel electrophoresis

DNA molecules were run in 1.5, 0.8, or 0.5% (w/v) agarose gels in 1xTAE buffer [40 mM Tris-Acetate and 1 mM EDTA (pH 8.0)] at 50-100 V. Gels were stained in 2 μg/ml ethidium bromide for 5-10 minutes and were viewed under UV light. DNA samples were mixed with 5 x loading buffer [60% (w/v) sucrose, 100 mM EDTA, 0.25% (w/v) bromophenol blue, 0.25% (w/v) ficoll orange, and 0.25% (w/v) xylene cyanol]. λ phage DNA was digested with either *Bst*EII or *Pst*I as molecular weight markers.

II.4.8 Purification of DNA from agarose gels

DNA was purified from agarose gels by using the QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA). DNA fragments, produced either by restriction digests or PCR, were run onto an agarose gel, the gels were ethidium bromide-stained, bands were visualized under UV light and excised. Purification was carried out according to the manufacturers' instructions.

II.4.9 Digestion and cloning of DNA

Restriction digests were carried out according to the manufacturers' specifications (New England Biolabs, Ipswich, MA; Roche, Basel, Switzerland). Typically, DNA molecules were digested for 3-6 hours and complete digests were confirmed via agarose gel electrophoresis. Digested DNA was either run onto an agarose gel to purify bands or ethanol-precipitated to concentrate the DNA samples.

Digested DNA fragments were inserted into vectors by using 1 unit of T4 DNA ligase and 10 x ligation buffer (Roche, Basel, Switzerland). Typical reactions contained 20-40 ng of vector DNA combined with insert DNA at a 1-10x greater molar concentration in a total volume of 10-50 µl. Ligations were carried out overnight at 16°C. Reactions were precipitated overnight and then resuspended in 10 µl Milli-Q™ water before electroporation into *E. coli* DH5α.

TA cloning was performed by mixing 4 µl of PCR product with 1 µl of the provided salt solution and 1 µl of the pCR®2.1-TOPO® vector (Invitrogen) in a total volume of 6 µl. After the mixture was incubated at room temperature for 5 minutes, 2 µl of it was mixed into a vial of One Shot® TOP10 *E. coli* chemically competent cells (Table II.1). The cells and the added ligation mix were incubated on ice for 25 minutes and then heat-shocked for 30 seconds at 42°C. The tube was shaken at 37°C for 1 hour after the addition of 250 µl SOC [20 g/l tryptone, 5 g/l yeast extract, 0.5 g/l NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, and 20 mM glucose (pH 7.0)]. From each transformation, 10 and 50 µl were spread onto LBA supplemented with 100 µg/ml ampicillin and 0.008% (w/v) X-gal. White colonies produced after an overnight incubation at 37°C were chosen for further analysis.

II.4.10 Polymerase chain reaction (PCR)

Polymerase chain reactions were performed using 5-10 ng of cosmid or plasmid DNA, 500 ng of genomic DNA, or 2.5 µl of mycelium (washed twice with 10.3% sucrose) from a 40-hour *S. clavuligerus* TSBS culture. Reactions typically contained 50 pmol of each oligonucleotide primer (Table II.5), 50 mM Tris-HCl (pH 9.0), 1.5 mM

$MgCl_2$, 0.4 mM β-mercaptoethanol, 0.1 mg/ml purified bovine serum albumin (BSA)(New England Biolabs, Ipswich, MA), 10 mM $(NH_4)_2SO_4$, 0.2 mM of each dNTP, and 5% (v/v) DMSO in a total volume of 50 µl. Reactions in which absolute sequence fidelity was not a concern used 0.5 µl of Taq DNA polymerase with or without 1 µl of Pfu DNA polymerase at a 1/10 dilution (both supplied by the Fermentation Service Unit, Department of Biological Sciences, University of Alberta). In cases where sequence fidelity was an issue, the EXPAND High Fidelity PCR system (Roche, Basel, Switzerland) was used according to the manufacturers' instructions with or without 5% (v/v) DMSO. The following cycling conditions were used: an initial 2 minute 96°C stage, 30-35 cycles of 96°C for 30-60 seconds, 55-65°C for 30 seconds, and 72°C for 1-3.5 minutes, followed by a single 5 minute 72°C stage. Reactions were carried out in thin-walled 0.2 ml PCR tubes in a TGradient thermocycler (Biometra, Goettingen, Germany).

When amplifying REDIRECT© cassettes, 50 ng of template DNA excised from pIJ773 was used with the following cycling conditions: 95°C for 3 minutes, followed by 10 cycles of 94°C for 45 seconds, 55°C for 45 seconds, and 72°C for 1 minute, followed by 15 cycles of 94°C for 45 seconds, 65°C for 45 seconds, and 72°C for 1 minute, followed by 72°C for 5 minutes.

II.4.11 DNA sequencing

The DYEnamic ET Terminator Cycle Sequencing System (Roche, Basel, Switzerland) was used with some modifications. Template DNA was used at different concentrations depending on the nature of the molecule: PCR products were used at 50-

100 ng, plasmids were used at 400 ng, and cosmids were used at 800 ng. Circular DNA molecules (plasmids and cosmids) were linearized with restriction endonucleases to facilitate DNA polymerase read-through. Reactions contained 5% (v/v) DMSO, 5-10 pmol of the chosen primer, and 8 µl ET sequence reagent premix in a total volume of 20 µl. Sequencing reactions were carried out in thin-walled 0.2 ml PCR tubes in a TGradient thermocycler (Biometra, Goettingen, Germany).

Tubes were exposed to the following cycling conditions: 35 cycles of 96 °C for 30 seconds and 60 °C for 1 or 2 minutes, or 30 cycles of 95 °C for 30 seconds, 50-55 °C for 30 seconds, and 60 °C for 1 minute. Once a reaction was complete, 2 µl of sodium acetate/EDTA buffer was added and 80 µl of 95% (v/v) ethanol was then added. After mixing, the tube was incubated on ice for 15 minutes and then microcentrifuged for an additional 15 minutes. Pellets were washed with 200 µl 70% ethanol, dried for 5 minutes, and sequence information was determined by the Molecular Biology Service Unit (Department of Biological Sciences, University of Alberta).

Whenever novel genomic DNA sequences or sequences for protein expression were determined, double-stranded sequence was obtained to ascertain sequence authenticity.

II.4.12 Software used for sequence analysis

GeneTools version 2.0 (BioTools Inc., Edmonton, AB) was used to produce assemblies of DNA sequences, perform sequence alignments, and to identify restriction sites. DNA Strider version 1.2 (Marck 1988) was also used to localize restriction sites. FramePlot 3.0beta (http://watson.nih.go.jp/~jun/cgi-bin/frameplot-3.0b.pl) was used to

identify putative ORFs by finding possible start and stop codons and by analyzing the biased usage of G+C in the third-letter position of codons (Ishikawa and Hotta 1999). Similarities within putative nucleotide and amino acid sequences were found by using the BLAST programs (Altschul et al. 1997) made available by the National Institute for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). Protein domains were found using the InterProScan Sequence Search (Zdobonov and Apweiler 2001) provided by the European Molecular Biology Laboratory – European Bioinformatics Institute (http://www.ebi.ac.uk/InterProScan/). Amino acid alignments were performed using CLUSTAL W and Jalview (Clamp et al. 2004; Higgens et al. 1994). The presence of signal peptides was predicted using the SignalP 3.0 Server (Nielsen et al. 1997) supplied by the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/SignalP/). Transmembrane domains of proteins were predicted using TMpred (Hoffman and Stoffel 1993) provided by the Swiss EMBnet node server (http://www.ch.embnet.org/software/TMPRED_form.html). Kyte-Doolittle hydropathy plots were also used to predict the locations of transmembrane domains (Kyte and Doolittle 1982).

II.5 Protein methods

II.5.1 Overexpression of Snk-His and Res2-His

Ten milliliter LB cultures of *E. coli* BL21(DE3) carrying either pT7-7His-*snk* or pT7-7His-*res2* were grown overnight at 37°C and then completely inoculated into 200 ml LB. These larger cultures were grown at 37°C for 1-1.5 hours until the OD_{600} reached ~0.6. At this point 0.4 mM IPTG was added to induce the expression of the T7 RNA

polymerase to enable the expression of Snk-His or Res2-His. Cultures were grown at 37°C for 5-6 hours and then harvested by centrifugation at 4000 x *g* for 20 minutes. Cell pellets were frozen overnight at -20°C.

II.5.2 Purification of Snk-His and Res2-His

Pellets were thawed on ice and then resuspended to a total 5 ml volume in lysis buffer [50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole (pH 8.0)] supplemented with 1 mg/ml lysozyme. After incubating on ice for 30 minutes, the preparations were sonicated 6-8 times for 10 seconds with a 2.5 mm probe on setting 1 attached to a Branson Sonifier 450 (Branson Ultrasonics Corporation, Garner, NC). Lysates were then centrifuged at 10,000 x g for 25 minutes at 4°C to pellet insoluble cellular debris and to obtain clarified lysates.

Clarified lysates were then combined with 1 ml of 50% nickel-nitrilotriacetic acid (Ni-NTA) slurry (Qiagen, Valencia, CA) and mixed on a rotary shaker (Bellco Biotechnology, Vineland, NJ) at 4°C for 1 hour to bind His-tagged proteins to the resin. This mixture was then loaded onto a chromatographic column for protein separation. All Ni-NTA chromatography was carried out at 4°C. Columns were made by filling the bottom of 5 ml syringes (Becton, Dickinson and Company, Franklin Lakes, NJ) with non-absorbent cotton and attaching tubing to the outlet. Once the flow-through was collected, the column was washed twice with 4 ml of wash buffer [50 mM NaH_2PO_4, 300 mM NaCl, and 20 mM imidazole (pH 8.0)]. His-tagged proteins were eluted by applying 0.5 ml of elution buffers with increasing concentrations of imidazole [50 mM NaH_2PO_4, 300 mM NaCl, and 50-250 mM imidazole (pH 8.0)]. The protein content of each fraction

was estimated through SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Section II.5.6).

Once elution fractions carrying most of the desired protein were identified, they were pooled and concentrated by centrifuging in Amicon® Centricon®-10 concentrators (Millipore, Billerica, MA) at 5000 x g for 2 hours at 4°C. After obtaining the retentate in ~50 µl final volume (containing any proteins above 10 kDa in size), the sample was resuspended in 850 µl storage buffer [10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 mM DTT, 0.3 M NaCl, 0.1% CHAPS, and 30% glycerol]. Protein samples were divided into 50 µl aliquots and stored at -80°C.

II.5.3 Overexpression of Res1-His

A 10 ml LB culture of E. coli BL21(DE3) carrying pET-19b-res1 was grown overnight at 37°C and then completely inoculated into 200ml LB. This larger culture was grown at 37°C for 1-1.5 hours until the OD_{600} reached ~0.6. At this point 1 mM IPTG was added to induce Res1-His expression and the culture was shaken at room temperature (~22°C) for 6.5-7 hours. The cell pellet was harvested by centrifugation at 4000 x g for 20 minutes and frozen overnight at -20°C.

II.5.4 Purification of Res1-His

The cell pellet was thawed on ice and resuspended to a total 5 ml volume in lysis buffer 2 [50 mM NaH_2PO_4, 2 M NaCl, 10mM imidazole, 20 mM β-mercaptoethanol, 2% Triton X-100 (pH 8.0)] to discourage nonspecific interactions between the His-tagged protein and unwanted proteins. Lysozyme was added to 1 mg/ml and the cell suspension

was incubated at 4°C for 30 minutes. Samples were then sonicated 6-8 times for 10 seconds with a 2.5 mm probe on setting 1 (Branson Sonifier 450). A clarified lysate was obtained by centrifuging at 10,000 x g at 4°C for 25 minutes.

The supernatant (carrying all soluble cellular material) was then combined with 1 ml of 50% Ni-NTA slurry and mixed for 1 hour on a rotary shaker (Bellco Biotechnology). The mixture was then loaded onto a syringe column and chromatography was carried out. After obtaining the flow-through, the column was washed once with 4 ml of a 20 mM imidazole solution and twice with 4 ml of a 125 mM imidazole solution [each solution consisted of 50 mM NaH_2PO_4, 300 mM NaCl, and either 20 or 125 mM imidazole (pH 8.0)]. His-tagged protein was then eluted from the column using four 0.5 ml fractions of elution buffer [50 mM NaH_2PO_4, 300 mM NaCl, and 250 mM imidazole (pH 8.0)]. Protein samples at different stages in the purification process were analyzed with SDS-PAGE to assess which fractions contained the most pure and concentrated Res1-His (Section II.5.6).

Those identified fractions were pooled and loaded onto an Amicon® Centricon®-10 concentrator (Millipore, Billerica, MA) and centrifuged at 5000 x g for 2 hours at 4°C. The ~90 µl retentate was brought to 600 µl with storage buffer [10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 mM DTT, 0.3 M NaCl, 0.1% CHAPS, and 30% glycerol] and divided into 30 µl aliquots. Samples were stored at -80°C.

II.5.5 Protein quantification

Protein levels were measured by using the Bio-Rad protein assay (Bio-Rad Laboratories, Inc., Mississauga, ON). This procedure is based on the protein-dye binding

assay of Bradford (1976). A standard curve using BSA (New England Biolabs, Ipswich, MA) was constructed and the A_{595} of samples was determined using a UNICAM UV/Vis Spectrophotometer UV3 (ATI Unicam, Cambridge, UK).

II.5.6 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were analyzed via discontinuous SDS-polyacrylamide gels. The stacking gels consisted of 4% (v/v) acrylamide/bis (37.5:1), 125 mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (v/v) TEMED, and 0.05% (w/v) ammonium persulfate. Separating gels were composed of 12 or 15% (v/v) acrylamide/bis (37.5:1), 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.05% (v/v) TEMED, and 0.5% (w/v) ammonium persulfate. Gels were prepared using the Mini-Protean II gel apparatus (Bio-Rad Laboratories, Inc., Mississauga, ON). Gels were run in running buffer (15 g/l Tris base, 72 g/l glycine, and 5 g/l SDS) at 200 V for 1 hour. Prior to loading, protein samples were combined with 5 x sample buffer [62.5 mM Tris-HCl (pH 6.8), 10% (v/v) glycerol, 2% (w/v) SDS, 720 mM β-mercaptoethanol, and 0.00125% (w/v) bromophenol blue] and heated to 95°C for 5 minutes.

After running, gels were stained with Phast Coomassie stain [0.1% (w/v) PhastGel Blue R Stain, 30% (v/v) methanol, and 10% (v/v) glacial acetic acid] for at least one hour and then destained [30% (v/v) methanol and 10% (v/v) glacial acetic acid]. After gels were satisfactorily destained, they were soaked in 10% (v/v) glacial acetic acid and 5% (v/v) glycerol for at least 24 hours and preserved in BioDesignGelWrap (BioDesign Inc., Carmel, NY).

II.5.7 In vitro phosphorylation assays

To detect the phosphorylation of Snk-His, Res1-His and Res2-His, the proteins were placed in reactions with each other in different combinations. Seven reactions were undertaken: Snk-His, Res1-His, and Res2-His by themselves, Snk-His with Res1-His, Snk-His with Res2-His, Res1-His with Res2-His, and all three proteins together. In each reaction, Snk-His was added to 1 μM and Res1-His and Res2-His were added to 3 μM. The only exception to this was the reaction containing all three proteins, in which 1 μM Snk-His and 1.5 μM of both Res1-His and Res2-His were used. Each reaction consisted of the appropriate protein amounts, 50 mM Tris-HCl (pH 8.0), 2 mM DTT, 50 mM KCl, 10% (v/v) glycerol, 0.1 mM $MgCl_2$, 60 μM dATP, and 2 μCi [γ-^{32}P]ATP. Reactions were carried out at 30°C for 30 minutes.

Once the reactions were complete, they were placed on ice and combined with sample buffer. They were then entirely run onto a 15% SDS-polyacrylamide gel (Section II.5.6) and the resulting gel was dried with a Model 583 Gel Dryer (Bio-Rad Laboratories, Inc., Mississauga, ON). After the gel was completely dried, it was exposed to a BAS cassette 4043 phosphor screen (Fujifilm, Tokyo, Japan) overnight. The image was developed in an Image Reader FLA-5000 phosphorimager (Fujiflim, Tokyo, Japan).

II.6 Antibiotic analysis

II.6.1 Clavulanic acid bioassay

The presence of clavulanic acid in culture filtrates was determined using *Klebsiella pneumonia* ATCC 15380 as an indicator organism (Mosher et al. 1999). Two separate bioassay plates containing 150 ml TSBA (3% (w/v) trypticase soy broth and

1.5% (w/v) agar) inoculated with 250 µl of the indicator organism glycerol stock were prepared: one with, and one without 6 µg/ml penicillin G. Discs of filter paper were placed on the TSBA surface and 25 µl of *S. clavuligerus* culture filtrates was deposited onto the discs. Plates were incubated overnight at 37°C. Inhibition zones surrounding discs on plates containing penicillin G were assumed to be due to the presence of clavulanic acid; these zones were measured and compared to those generated on plates lacking penicillin G.

II.6.2 Alanylclavam and 2-hydroxymethylclavam bioassay

The presence of alanylclavam or 2-hydroxymethylclavam was determined using *Bacillus* sp. 27860 as an indicator organism (Pruess and Kellett 1983). Two separate bioassay plates containing 150 ml Davis minimal medium [3 g/l KH_2PO_4, 7 g/l K_2HPO_4, 0.47 g/l di-hydrated sodium citrate, 0.1 g/l $MgSO_4 \cdot 2H_20$, 1 g/l $(NH_4)_2SO_4$, and 15 g/l agar (pH 7.0)] were inoculated with the indicator organism: one plate contained 200 µg/l L-methionine and the other did not contain L-methionine. The indicator organism was grown on NA supplemented with 2% glucose for 48 hours at 28°C. Filter paper discs were placed on the Davis minimal medium surface and 25 µl of *S. clavuligerus* culture filtrates was deposited onto the discs. Inhibition zones arising after a 28°C overnight incubation on the plates containing L-methionine were assumed to be due to the presence of alanylclavam or 2-hydroxymethylclavam. These zones were measured and compared to those produced on plates lacking L-methionine.

II.6.3 High performance liquid chromatography (HPLC)

High performance liquid chromatography was used to quantify the amounts of clavams present in *S. clavuligerus* culture filtrates (Foulstone and Reading 1982; Paradkar and Jensen 1995). Typically, culture supernatants were obtained from 72 and 96 hour soy cultures. Culture samples were microcentrifuged for 5 minutes to pellet mycelia and other insoluble material. Twenty microliters of each supernatant was added to 80 µl Milli-Q™ water. Two sets of samples were prepared: an underivatized set and an imidazole-derivatized set. Each culture supernatant had an underivatized and an imidazole-derivatized sample made. The underivatized samples had 25 µl Milli-Q™ water added while imidazole-derivatized samples had 25 µl of 25% (w/v) imidazole added. Samples were then incubated at room temperature for 15 minutes and microfuged for 5 minutes to pellet any insoluble material. From each sample, 100 µl was transferred to polypropylene spring inserts (Fisher Scientific, Pittsburgh, PA) which were then placed into HPLC screw neck vials.

Fifty microliters of each sample was injected using a Waters 2690 Separations Module and analyzed with a Waters 996 Photodiode Array Detector (Waters, Milford, MA) to determine the absorbance spectrum. A Phenomenex® Bondclone 10µ C18 (100 x 800 mm 10 micron) column (Phenomenex, Torrance, CA) was used to separate the components of samples. The mobile phase consisted of 100 mM NaH_2PO_4 + 6% (v/v) methanol in an isocratic elution run at 2 ml/minute. Each run was 15 minutes in length and derivatized clavams were detected by viewing the absorbance of samples at 311 nm.

Waters Millenium[32] Software (version 3.20; Waters, Milford, MA) was used to analyze and process data.

II.6.4 Liquid chromatography-mass spectrometry (LC-MS)

To confirm the absence of certain clavams in different *S. clavuligerus* culture filtrates, LC-MS was used as described by Jensen et al. (2004b). Five microliters of derivatized samples was run onto an XTerra column (0.21 x 10 cm, Waters, Milford, MA) at 0.25 ml/minute. Each run was 45 minutes in length and the running buffer consisted of solvent A [10 mM ammonium bicarbonate (pH 10)] and solvent B (acetonitrile). A binary gradient system was run in the following manner: 100% solvent A for 5 minutes, linear gradient to 85% solvent A and 15% solvent B over 20 minutes, 85% solvent A and 15% solvent B for 5 minutes, linear gradient to 100% solvent A over 1 minute, and 100% solvent A for 14 minutes. Imidazole-derived clavams were monitored by determining the A_{311} of samples run off the column (Section II.5.4) and electrospray mass spectra were observed using a ZMD-2 single quadrupole instrument (Waters, Milford, MA). Data were collected and analyzed using the MassLynx software (version 3.4; Waters, Milford, MA).

II.6.5 Analysis of undecylprodigiosin production by *S. coelicolor*

II.6.5.1 Extraction of undecylprodigiosin from *S. coelicolor* cultures

Unecylprodigiosin was solvent-extracted from *S. coelicolor* R5 cultures supplemented with 5% PEG 8000. Seventy-five microliters of *S. coelicolor* cultures was mixed with 1.45 ml of methanol and microfuged for 5 minutes to pellet mycelia. These

mixtures then had 6 µl HCl (1N) added and incubated at room temperature overnight. The resulting supernatant contained extracted undecylprodigiosin.

II.6.5.2 Quantification of undecylprodigiosin

The amount of solvent-extracted undecylprodigiosin was quantified by measuring the A_{530} of samples with a UNICAM UV/Vis Spectrophotometer UV3 (ATI Unicam, Cambridge, UK). To take the growth level of cultures into account, the OD_{450} was also measured for the cultures from which the undecylprodigiosin was extracted.

TableII.1. *E. coli* strains used in this study.

Strain	Description	Reference or Source
BL21(DE3)	Carries the bacteriophage T7 RNA polymerase encoded on its chromosome, used for expression of proteins at high levels under the control of the T7 promoter	Stratagene
BW25113/pIJ790	Expresses recombination promoting proteins from pIJ790, used for the PCR-targeting REDIRECT© mutagenesis system	Gust et al. 2003
DH5α	General cloning host	Gibco BRL
DH5α(BT340)	Expresses FLP recombinase from the temperature-sensitive BT340, used for excision of PCR-derived cassettes flanked by FRT sites	Gust et al. 2003
ET12567/pUZ8002	Methylation-deficient host carrying the plasmid mobilization functions on pUZ8002, used for inter-generic conjugation from *E. coli* to *Streptomyces*	Kieser et al. 2000
One Shot® TOP10	Host for TOPO™ vectors, used for TA cloning of PCR products	Invitrogen
TransforMax™ EC100™	Highly transformable, used as the transformation host of transposon-containing cosmids (Δ*endA1*, Δ*recA1*)	Epicentre

Table II.2. *S. clavuligerus* strains used in this study.

Strain	Description	Reference or Source
NRRL 3585	Wild type strain	Northern Regional Research Laboratory, Peoria, Ill. USA
IC4	Mutant strain with a Tn*5062* insertion in *orfA*	Cai 2003
Δ*orfB::apra*	Mutant strain with a REDIRECT© cassette replacing *orfB*	Kwong and Jensen, unpublished
B50	Mutant strain with a Tn*5062* insertion in *orfC*	Cai 2003
B50+pSET-AT	B50 mutant carrying pSET-AT integrated in its chromosome	This study
B50+pSET-AT-*orfC*	B50 mutant carrying pSET-AT-*orfC* integrated in its chromosome	This study
D41	Mutant strain with a Tn*5062* insertion in *orfD*	Cai 2003
Δ*orfE::apra*	Mutant strain with a REDIRECT© cassette replacing *orfE*	Kwong and Jensen, unpublished
Δ*orfF::apra*	Mutant strain with a REDIRECT© cassette replacing *orfF*	This study
Δ*orfG::apra*	Mutant strain with a REDIRECT© cassette replacing *orfG*	This study
Δ*orfH::apra*	Mutant strain with a REDIRECT© cassette replacing *orfH*	This study
Δ*orf1::apra*	Mutant strain with a REDIRECT© cassette replacing *orf1*	This study

Table II.2. (continued).

Strain	Description	Reference or Source
Δ*orf2/3::apra*	Mutant strain with a REDIRECT© cassette replacing *orf2/3*	This study
Δ*orf4::apra*	Mutant strain with a REDIRECT© cassette replacing *orf4*	This study
Δ*res2::apra*	Mutant strain with a REDIRECT© cassette replacing *res2*	This study
Δ*res1::apra*	Mutant strain with a REDIRECT© cassette replacing *res1*	This study
Δ*res1/res2::apra*	Mutant strain with a REDIRECT© cassette replacing *res1* and *res2*	This study
Δ*snk::apra*	Mutant strain with a REDIRECT© cassette replacing *snk*	Tahlan 2005
Δ*snk*	Mutant strain with an in-frame 81 bp scar replacing *snk*, generated using FLP recombinase	This study
Δ*c7p::apra*	Mutant strain with a REDIRECT© cassette replacing *c7p*	Tahlan 2005
Δ*c7p*	Mutant strain with an in-frame 81 bp scar replacing *c7p*, generated using FLP recombinase	This study
WT+pSET152	Wild type carrying pSET152 integrated in its chromosome	This study
Δ*snk::apra* + pSET-AT	Δ*snk::apra* mutant carrying pSET-AT integrated in its chromosome	This study

Table II.2. (continued).

Strain	Description	Reference or Source
Δ*snk::apra*+ pSET-AT-*snk*	Δ*snk::apra* mutant carrying pSET-AT-*snk* integrated in its chromosome	This study
Δ*res2::apra*+ pSET-AT	Δ*res2::apra* mutant carrying pSET-AT integrated in its chromosome	This study
Δ*res2::apra*+ pSET-AT-*res2*	Δ*res2::apra* mutant carrying pSET-AT-*res2* integrated in its chromosome	This study
WT+pSET-P_E*res1*	Wild type carrying pSET-P_E*res1* integrated in its chromosome	This study
WT+pSET-P_E*res2*	Wild type carrying pSET-P_E*res2* integrated in its chromosome	This study

Table II.3. *S. coelicolor* strains used in this study.

Strain	Description	Reference or Source
M145	Prototrophic strain, SCP1-, SCP2-, Pgl+	John Innes Institute
Δ*redJ::apra*	Mutant strain with a REDIRECT© cassette replacing *redJ*	This study
Δ*redJ*	Mutant strain with an in-frame 81 bp scar replacing *redJ*, generated using FLP recombinase	This study

Table II.4. Plasmids and cosmids used in this study.

Replicon	Description	Reference or Source
***E. coli* cosmids**		
14E10	pWE15-derived cosmid carrying the *S. clavuligerus* paralogue gene cluster and flanking regions	Jensen et al. 2004b
14E10-Δ*orf1::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *orf1*	This study
14E10-Δ*orf4::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *orf4*	This study
14E10-Δ*orf5::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *orf5*	This study
14E10-Δ*orfF::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *orfF*	This study
14E10-Δ*orfG::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *orfG*	This study
14E10-Δ*orfH::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *orfH*	This study
14E10-Δ*res1::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *res1*	This study

Table II.4. (continued).

Replicon	Description	Reference or Source
14E10-Δ*res1/res2::apra*	14E10-derived mutant cosmid carrying a REDRECT© cassette replacing *res1* and *res2*	This study
14E10-Δ*res2::apra*	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *res2*	This study
14E10-SKN	14E10-derived mutant cosmid carrying a REDIRECT© cassette replacing *snk*	Tahlan 2005
I-A7, I-B12, I-C9, I-C10, I-D3, I-D8, I-D12, I-F6, I-G3, I-G10, II-A6, II-B4, II-B8, II-C1, II-D4, II-E8, II-F4, II-F7, II-F8, II-H1, II-H5	14E10-derived mutant cosmids carrying a Tn*5062* insertion in the downstream region of the paralogue gene cluster, used for sequencing	This study
II-G10	14E10-derived mutant cosmid carrying a Tn*5062* insertion in the downstream region of the paralogue gene cluster, used for sequencing and for generating an insertional *orf2/3* mutant	This study
St3F7	SuperCos 1-derived cosmid carrying bases 6438307-6477847 of the *S. coelicolor* genome	John Innes Foundation (Redenbach et al. 1996)
St3F7- Δ*redJ::apra*	St3F7-derived mutant cosmid carrying a REDIRECT© cassette replacing *redJ*	This study

Table II.4. (continued).

Replicon	Description	Reference or Source
***E. coli* plasmids**		
P2.8-18	pUC18 with a 2.8 kb *Eco*RI fragment carrying part of *ceaS1* and all of *orfA*	Tahlan 2005
pCR®2.1-TOPO®	Phagemid for TA-cloning of PCR products	Invitrogen
pDA504	pSL1180 containing a 1.1 kb *Bcl*I fragment carrying *tsr*	Alexander 1998
pEAW115	*E. coli* cloning vector carrying *flp*	Cox, unpublished
pET-19b	Expression vector conferring an N-terminal 10x His-tag; protein expression is dependent on IPTG and T7 RNA polymerase	Novagen
pET-19b-*res1*	pET-19b derivative carrying *res1* as an *Nde*I/*Bam*HI insert	This study
pIJ773	Plasmid carrying the 1398 bp REDIRECT© cassette [*aac(3)IV* and *oriT* flanked by FRT sites] as an *Eco*RI/*Hind*III insert	Gust et al. 2003
pSL1180	*E. coli* cloning vector containing a large polylinker	Pharmacia
pSL-Δ*snk::apra*	pSL1180 derivative with a 10.4 kb *Kpn*I/*Nde*I fragment carrying the Δ*snk::apra* region along with flanking sequences	This study

Table II.4. (continued).

Replicon	Description	Reference or Source
pSL-dpar	pSL1180 derivative with a 7.7 kb *Not*I/*Nco*I fragment from 14E10 carrying *res2* (and most of *res1*) to *orf1*	This study
pSL-*orfC*	pSL1180 containing a 1.9 kb *Age*I/*Eco*RI fragment carrying *orfC* and its upstream region	This study
pT7-7His	Expression vector derivative of pT7-7 conferring a C-terminal 7x His-tag; protein expression is dependent on IPTG and T7 RNA polymerase	J. Manchak, personal communication
pT7-7His-*res1*	pT7-7His derivative with *res1* inserted as an *Nde*I/*Bam*HI fragment from TOPO-*res1*	This study
pT7-7His-*res2*	pT7-7His derivative with *res2* inserted as an *Nde*I/*Bam*HI fragment from TOPO-*res2*	This study
pT7-7His-*snk*	pT7-7His derivative with *snk* inserted as an *Nde*I/*Bam*HI fragment from TOPO-*snk*	This study
pUZ8002	Carries the RK2/RP4 transfer functions for *in trans* conjugation of *oriT*-bearing replicons	Kieser et al. 2000
TOPO-*flp*	pCR®2.1-TOPO® derivative carrying a 1.5 kb *Xho*I/*Xba*I fragment encompassing *flp* from pEAW115	This study
TOPO-*res1*	pCR®2.1-TOPO® derivative with *res1* inserted as a PCR amplified *Nde*I/*Bam*HI fragment	This study

Table II.4. (continued).

Replicon	Description	Reference or Source
TOPO-*res2*	pCR®2.1-TOPO® derivative with *res2* inserted as a PCR-amplified *NdeI*/*Bam*HI fragment	This study
TOPO-*snk*	pCR®2.1-TOPO® derivative with *snk* inserted as a PCR-amplified *NdeI*/*Bam*HI fragment	This study
***E. coli*/*Streptomyces* shuttle plasmids**		
pHJ401	Also called pJV326; pHJL400 derivative carrying *oriT*; segregationally unstable	He et al. 2001
pHJ401-*flp*	pHJ401 carrying a 5.1 kb *Bgl*II fragment in its *Bam*HI site including *flp* under the control of P_E from pHM8a-*flp*	This study
pHM8a	*Streptomyces* conjugative and integrative plasmid carrying the *ermE** promoter (P_E) and *E. coli* cloning vector	Motamedi et al. 1995
pHM8a-*flp*	pHM8a with a 1.5 kb *NdeI*/*Bam*HI fragment encompassing *flp* inserted downstream of P_E	This study
pHM8a-*res1*	pHM8a derivative with a 0.4 kb fragment carrying *res1* from TOPO-*res1*	This study
pHM8a-*res2*	pHM8a derivative with a 0.7 kb fragment carrying *res2* from TOPO-*res2*	This study

Table II.4. (continued).

Replicon	Description	Reference or Source
pSET152	*Streptomyces* conjugative and integrative plasmid and *E. coli* cloning vector	Bierman et al. 1992
pSET-AT	pSET152 derivative with a 1.1 kb *Bam*HI/*Bgl*II fragment carrying *tsr* from pDA504 inserted into its *Bam*HI site	This study
pSET-AT-*orfC*	pSET-AT derivative with a 1.9 kb *Eco*RI/*Not*I fragment carrying *orfC* and its upstream region from pSL-*orfC*	This study
pSET-AT*-*res2*	pSET-AT derivative with a 1.2 kb *Eco*RI/*Sac*II fragment from pSL-dpar including *res2* and an upstream region partially replacing *tsr*	This study
pSET-AT-*res2*	pSET-AT*-*res2* derivative with a 1.1 kb *Bam*HI/*Xba*I fragment from pDA504 carrying *tsr*	This study
pSET-AT-*snk*	pSET-AT derivative with a 2 kb *Nhe*I fragment carrying *snk* and its upstream region	This study
pSET-P_E*res1*	pSET152 derivative with a 4 kb *Bgl*II fragment carrying *res1* under the control of the *ermE** promoter (P_E) inserted into its *Bam*HI site	This study
pSET-P_E*res2*	pSET152 derivative with a 4.2 kb *Bgl*II fragment carrying *res2* under the control of the *ermE** promoter (P_E) inserted into its *Bam*HI site	This study

Table II.4. (continued).

Replicon	Description	Reference or Source
pUWL-KS	Cloning vector for blue-white screening with plJ101 *ori*	Wehmeier 1995
pUWL-Δ*snk*	pUWL- Δ*snk::apra* derivative with the REDIRECT© cassette replaced by an 81 bp "scar" produced by FLP recombinase	This study
pUWL-Δ*snk::apra*	pUWL-KS derivative with a 10.4 kb *Eco*RI/*Hind*III fragment carrying the Δ*snk::apra* region from 14E10-SKN	This study
pUWL-*oriT*	pUWL-KS derivative with a 0.7 kb *Pst*I fragment carrying *oriT* from pSET152	This study
pUWL-*oriT*-Δ*snk*	pUWL-*oriT* derivative with a 9 kb *Eco*RI/*Hind*III fragment carrying the 81 bp "scar" replacing *snk* from pUWL-Δ*snk*	This study
pUWL-*oriT*-*flp*	pUWL-*oriT* carrying a 5.1 kb *Bgl*II fragment in its *Bam*HI site including *flp* under the control of P_E from pHM8a-*flp*	This study

Table II.5. Oligonucleotide primers used in this study.

Primer	Sequence (5'-3')[a]	Use
2213F	GGCGGCGGACGGACTCAACG	Mycelial PCR to confirm Δ*orfF::apra* and Δ*orfG::apra* mutations
4243F	CGCGCACACCATGACGACGGACT	Mycelial PCR to confirm Δ*orfG::apra* and Δ*orfH::apra* mutations
5226R	GGCGGTGGTGGCGTTCGGTGTC	Mycelial PCR to confirm Δ*orfF::apra* and Δ*orfG::apra* mutations
ANW22	GTCCGCCGCGAGCAGCCAC	Mycelial PCR to confirm the Δ*c7p* mutation
ANW23	CCCCGGTCTGGATGTCGAGCC	Mycelial PCR to confirm the Δ*snk* mutation
ANW32	CCGACTCTCCGCGACGCACTTCC	Mycelial PCR to confirm the Δ*snk* mutation
ANW36	GCGAACCGGCGAAGGAGCTGC	Mycelial PCR to confirm the Δ*c7p* mutation
ANW40	GCCCGATCAGCCGCTCCCG	Mycelial PCR to confirm the Δ*snk* mutation
ANW41	GCCCAACGGAACCGCCCTGTATG	Mycelial PCR to confirm Δ*res1::apra* and Δ*res1/res2::apra* mutations
FWD-redJ	TCGACTCGCCCGAGGGACGCGCCCATGTCGCCCG CTGACATTCCGGGGATCCGTCGACC	Forward primer for *redJ* REDIRECT© mutagenesis

Table II.5. (continued).

Primer	Sequence (5'-3')[a]	Use
HCA1	CGGTGAACAGCTCCTCGCCCTTG	Reverse primer used to sequence outwards from Tn*5062* insertions
HCA2	CGACGAGCAAGGCAAGACCGATC	Forward primer used to sequence outwards from Tn*5062* insertions
NZE1	GGGCGGTGGCGACGGTGAG	Sequencing of the region downstream of the paralogue gene cluster
NZE2	GGCCGACCGGACCCGCTC	Sequencing of the region downstream of the paralogue gene cluster
NZE3	GCGCGGTCCGGTGGGTGC	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm the Δ*orf4::apra* mutation
NZE4	CGGCGGCGCTGTGGACCT	Sequencing of the region downstream of the paralogue gene cluster
NZE5	CGCCGTCGGGGAACTCG	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm the Δ*orf5::apra* mutation

Table II.5. (continued).

Primer	Sequence (5'-3')[a]	Use
NZE6	CCGGCCCGACCTGGTGATC	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm Δ*res1::apra*, Δ*res2::apra* and Δ*res1/res2::apra* mutations
NZE7	CCGGCAGGGGGACTACCAGCTA	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm Δ*res2::apra* and Δ*orf5::apra* mutations
NZE8	GGGCGAAGGTGTACGGGGAC	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm *orf2/3::*Tn*5062* and Δ*orf1::apra* mutations
NZE9	GGGCGGCGGAGCGGTGGT	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm the Δ*orf1::apra* mutation
NZE10	CGGCGTTAAGGTTGGCGGTGAA	Sequencing of the region downstream of the paralogue gene cluster
NZE12	CGCGCCGGTCAAGTCCTGGG	Sequencing of the region downstream of the paralogue gene cluster
NZE13	GGGCACCTCGCAGATGGGCAC	Sequencing of the region downstream of the paralogue gene cluster

Table II.5. (continued).

Primer	Sequence (5'-3')[a]	Use
NZE14	CCACCGGCAGGGCGTAGCAG	Sequencing of the region downstream of the paralogue gene cluster
NZE15	TCCGGGCCGTGAACGACTCC	Sequencing of the region downstream of the paralogue gene cluster
NZE16	CGGTGCCGGTCGAGACGTTG	Sequencing of the region downstream of the paralogue gene cluster
NZE17	CCCGCACCACGTCCCTCAAGC	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm *orf2/3::*Tn*5062* and Δ*orf1::apra* mutations
NZE18/NZE11	CGGCAAGACGGACGTGGGTG	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm the Δ*orf4::apra* mutation
NZE19	TCGCTGCCGGTGCTGGTGAG	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm the *orf2/3::*Tn*5062* mutation
NZE20	CGCGCCGGTCAAGTCCTGG	Sequencing of the region downstream of the paralogue gene cluster
NZE21	CGATGCCAGGCCGGTGACC	Sequencing of the region downstream of the paralogue gene cluster

Table II.5. (continued).

Primer	Sequence (5'-3')[a]	Use
NZE22	CCTCCGACGGCACCAACAAGACC	Sequencing of the region downstream of the paralogue gene cluster; mycelial PCR to confirm Δ*res2::apra* and Δ*orf5::apra* mutations
NZE23	CGGAGGACGGCGGCCACTAC	Sequencing of *orfA*
NZE24	GGGGCTGCGGCTCGGGAC	Sequencing of *orfA*
NZE26	CCGGGCGAGTTCCACGACC	Sequencing of *orfA*
orf1(PEN)FWD	CATCCGACAAACGCACCACATTTCTCAGGA GATTCCATGATTCCGGGGATCCGTCGACC	Forward primer for *orf1* REDIRECT© mutagenesis
orf1(PEN)REV	TCGGTGGCGGACCGGGCGGGGCCGGTCCGTCCCG GGTCATGTAGGCTGGAGCTGCTTC	Reverse primer for *orf1* REDIRECT© mutagenesis
orf4(SDE2)FWD	ATTCCGGGGATCCGTCGACC	Forward primer for *orf4* REDIRECT© mutagenesis
orf4(SDE2)REV	TGTAGGCTGGAGCTGCTTC	Reverse primer for *orf4* REDIRECT© mutagenesis
orf5(Oxid.)FWD	GCGTACTTGGCGCACCATCGACGACAGGGGTGCA G TATGATTCCGGGG ATCCGTCGACC	Forward primer for *orf5* REDIRECT© mutagenesis
orf5(Oxid.)REV	CCGGAGCCGGCACACGGCCGTTCGCCCGGGG CCGGGTCATGTAGGCTGGAGCTGCTTC	Reverse primer for *orf5* REDIRECT© mutagenesis

Table II.5. (continued).

Primer	Sequence (5'-3')[a]	Use
orf6(RRlong)FWD	GTCCGTTGCGTAGTCGATAATGTCGTCATGGCAG GAGTGATTCCGGGGATCCGTCGACC	Forward primer for *res2* REDIRECT© mutagenesis
orf6(RRlong)REV	TGCTCGACGGCGCCGCTACTGGGCCTCCA GAAAGGTGATTGTAGGCTGGAGCTGCTTC	Reverse primer for *res2* REDIRECT© mutagenesis
orf7(RRshort)FWD	GGTCGGCACCCGGCAGGGGGACTACCAGCTAG CGGCATGATTCCGGGGATCCGTCGACC	Forward primer for *res1* REDIRECT© mutagenesis
orf7(RRshort)REV	GGGTACTCGCCGTGATCACCTTTCTGGAGGCCCAG TAGCTGTAGGCTGGAGCTGCTTC	Reverse primer for *res1* and *res1/res2* REDIRECT© mutagenesis
orfF-FWD	ACTCGCCGGGATTTCGCGTCGACGAGAGGGCGG GTCATGATTCCGGGGATCCGTCGACC	Forward primer for *orfF* REDIRECT© mutagenesis
orfF-INT	CGGCTCGCCTGGGTGCTCGTGTT	Mycelial PCR to confirm Δ*orfF::apra* REDIRECT© mutagenesis
orfF-REV	CGAGGACCACCAGCACGGTCAGTGTCACGGCCG CCCTCCTGTAGGCTGGAGCTGCTTC	Reverse primer for *orfF* REDIRECT© mutagenesis
orfG-FWD	GCACGAGAGGCGGGTGGCCCGACGATGGAGGCA CAGGTGATTCCGGGGATCCGTCGACC	Forward primer for *orfG* REDIRECT© mutagenesis

Table II.5. (continued).

Primer	Sequence (5'-3')[a]	Use
orfG-REV	GCCCGGGCCCGGGCCCCGGCTCGGGAGTCCGCCG GTCTATGTAGGCTGGAGCTGCTTC	Reverse primer for *orfG* REDIRECT© mutagenesis
orfH-FWD	TTGTTCCGGCACCGGGGCGTTGATTGGTTTGTTG TCATGATTCCGGGGATCCGTCGACC	Forward primer for *orfH* REDIRECT© mutagenesis
orfH-REV	ACTTACCCGCGGACGCCGTACGCGCGGGTGCCG CTCTCATGTAGGCTGGAGCTGCTTC	Reverse primer for *orfH* REDIRECT© mutagenesis
RED-DN	CGAAGCAGCTCCAGCCTAC	Mycelial PCR to produce amplicons downstream of genes replaced by the REDIRECT© cassette
redJ-DN	GCGCCGCAGGTCCTCGTCGTC	Binds downstream of *redJ*; genomic DNA PCR to confirm *redJ* mutations
redJ-INT	CGGCGAGGGGTTCCATGGTGTC	Binds within *redJ*; genomic DNA PCR to confirm *redJ* mutations
redJ-UP	CCGCCGGGCTGCTCTACGTCTC	Binds upstream of *redJ*; genomic DNA PCR to confirm *redJ* mutations
RED-UP	CTGCAGGTCGACGGATCC	Mycelial or genomic DNA PCR to produce amplicons upstream of genes replaced by the REDIRECT© cassette

Table II.5. (continued).

Primer	Sequence (5'-3')[a]	Use
res1/2-REV	GTCCGTTGCGTAGTCGATAATGTCGTCATGGCAG GAGTGTGTAGGCTGGAGCTGCTTC	Reverse primer for *res1/res2* REDIRECT© mutagenesis
res1-pfwd	CATATGGCGGCCGGGGAGGCGCACTG	Forward primer for cloning of *res1*
res1-prev	GGATCCCCTTTCTGGAGGCCCAGTAGC	Reverse primer for cloning of *res1*
res2-pfwd	CATATGAGGGTAGTACTGGCCGAGGAC	Forward primer for cloning of *res2*
res2-prev	GGATCCCTGGGCCTCCAGAAAGGTGATC	Reverse primer for cloning of *res2*
REV-redJ	GGCTGCCGACGCTCTTGGCCAGGCTCAGAATGTC CATGTTGTAGGCTGGAGCTGCTTC	Reverse primer for *redJ* REDIRECT© mutagenesis
snk-pfwd	CATATGTCGTGCGAACAGATCGTC	Forward primer for cloning of *snk*
snk-prev2	GGCCGGGATCCGTCGGTGTGCTC	Reverse primer for cloning of *snk*

[a] Engineered restriction sites are underlined

III. RESULTS

III.1 Analysis of the upstream portion of the paralogue gene cluster

III.1.1 Sequencing of *orfA*

Original studies of the upstream region of the paralogue gene cluster revealed that certain genes may have roles in alanylclavam biosynthesis (Cai 2003). However, much of this work was incomplete and required further confirmation due to inconsistencies observed during fermentations. For this reason, the upstream region of the paralogue gene cluster was once again scrutinized.

One unfinished aspect of the original work was the sequencing of *orfA*. Although the past study had partially sequenced *orfA*, much of it remained unsequenced. The entire *orfA* sequence was obtained by sequencing from one end of p2.8-18 (Table II.4) and then bridging the sequence using a primer that binds near the furthest known 3' end of *orfA* (NZE 23; Table II.5). To obtain double-stranded sequence, two further primers (NZE 24 and NZE 26) that bind within *orfA* were used (Table II.5). FramePlot analysis revealed that *orfA* is 1173 bp in length and has a third-letter G+C content of 93.4%, which is in accordance with *Streptomyces* ORFs (Figure III.1; Appendix; Bibb et al. 1984).

Analysis of the *orfA* sequence revealed that it is most similar to archaeal serine hydroxymethyltransferases (Table III.1).

III.1.2 Re-analysis of *orfA*, *orfB*, *orfC*, *orfD*, and *orfE* mutants

Although earlier studies implicated *orfA*, *orfB*, *orfC*, and *orfD* in the biosynthesis of alanylclavam, the data were inconsistent. Because wild type cultures did not produce

alanylclavam in liquid soy medium in these fermentations, the involvement of the above genes in alanylclavam production could not be ascertained (Cai 2003). Wild type cultures grown on solid soy medium did produce alanylclavam, however, and mutants in *orfA*, *orfC*, and *orfD* could not produce alanylclavam in either liquid or solid conditions (Cai 2003). As well, a preliminary study analyzing the role of *orfB* and *orfE* in 5*S* clavam biosynthesis suggested that *orfB* is required for alanylclavam production (Kwong and Jensen, unpublished). Since there was a chance that these genes are implicated in alanylclavam biosynthesis, mutants in each of the genes were analyzed once more.

An updated sequence analysis of *orfA* – *orfE* indicates that each gene shows similarity to a gene involved in amino acid metabolism and transport (Table III.1). While one earlier study generated insertional transposon mutants in *orfA*, *orfC*, and *orfD* (Cai 2003), another produced REDIRECT© mutants in *orfB* and *orfE* (Kwong and Jensen, unpublished). Mutant strains in each gene (three strains bearing the same mutation were analyzed) were fermented in soy medium and culture filtrates were examined for the presence of clavams. Each mutant strain was fermented in liquid medium once. Examination of HPLC chromatograms of imidazole-derivatized culture filtrates reveals that neither *orfA* nor *orfB* mutants are capable of producing alanylclavam (Figure III.2). As well, *orfC* and *orfD* mutants did not produce alanylclavam; instead, they produced a compound that eluted approximately 15 seconds earlier than alanylclavam (Figure III.2). Since this new compound was derivatized with imidazole (if it were not, it would have appeared in the underivatized sample's chromatogram and would not have shown up on these subtracted chromatograms), it is likely that it represents a different clavam. Indeed, it is possible that this novel clavam represents an intermediate in normal alanylclavam

biosynthesis. Preliminary LC-MS analyses of the *orfC* and *orfD* culture filtrates did not detect the presence of the novel clavam, perhaps reflecting its lack of affinity for the column under the tested conditions used (data not shown). The phenotypes of the mutant strains were observed once again upon growing the strains on a solid soy-based medium (data not shown).

III.1.3 Complementation of an *orfC* mutant

It was of particular interest that both *orfC* and *orfD* mutants produce the novel clavam. This may indicate that both encoded proteins are needed for a single reaction in alanylclavam synthesis, that one protein needs the other to be functional (perhaps the proteins only catalyze reactions when complexed with the other), or that the genes are transcribed in an operon. In the latter case, it is likely that transposon insertion in the upstream *orfC* would have polar effects on the expression of *orfD*. It seemed likely that *orfC* and *orfD* are transcribed together since the *orfC* stop codon overlaps with the *orfD* start codon (Figure I.6).

To test whether or not these two genes are transcribed in an operon, an *orfC* mutant was complemented. *orfC* was inserted into pSET-AT, which confers thiostrepton resistance and integrates into the chromosome at the ΦC31 phage attachment site (*attB*; Bierman et al. 1992). *orfC* mutant strains carrying pSET-AT-*orfC* (Table II.4) were fermented and culture filtrates were analyzed. Chromatograms revealed that the complemented mutants still produced the novel clavam but also produced trace amounts of alanylclavam (Figure III.3).

III.1.4 Generation of *orfF*, *orfG*, and *orfH* mutants

Although the upstream region of the paralogue gene cluster had been sequenced and most genes had already been subjected to mutagenesis, three genes remained unexamined. *orfF*, *orfG*, and *orfH* had not yet been mutated despite earlier attempts to introduce Tn*5062* insertions into them (Cai 2003). The protein encoded by *orfF* resembles an amino acid permease, *orfG*'s encoded protein may be a protein kinase, and the protein encoded by *orfH* shows similarity to a multidrug resistance transporter (Table III.1). To obtain a complete set of mutants in the paralogue gene cluster, these three genes were replaced with an antibiotic resistance cassette via the PCR-targeting REDIRECT© protocol (Gust et al. 2003). Using the orfF-FWD and orfF-REV primer pair, the orfG-FWD and orfG-REV pair, and the orfH-FWD and orfH-REV pair, the REDIRECT© cassette [carrying the apramycin resistance gene, *aac3(IV)*, and a RK2/RP4 origin of transfer, *oriT*] was amplified by PCR. These amplified cassettes, flanked by regions normally flanking either *orfF*, *orfG*, or *orfH*, were introduced into *E. coli* expressing the λ Red functions promoting recombination between the linear PCR product DNA and cosmid 14E10 DNA (Datsenko and Wanner 2000). Recombination between 14E10 and the PCR-products allowed *orfF*, *orfG* and *orfH* to be replaced by the REDIRECT© cassette. Conjugation of each mutant cosmid into wild type *S. clavuligerus*, followed by homologous recombination between mutant cosmid DNA and chromosomal DNA, enabled *orfF*, *orfG*, and *orfH* mutant strains to be obtained. Mutations were confirmed through mycelial PCR (Figures III.4, III.5, III.6).

For *orfF* mutants, it appears that at least Δ*orfF::apra* 1, 2, and 4 are true mutants; the other strains (Δ*orfF::apra* 3 and 5) may contain copies of the wild type chromosome

(Figure III.4). This can be seen by the appearance of a ~1026 bp band in the mycelial PCR reactions for Δ*orfF::apra* 3 and 5 (Figure III.4). The band in the wild type reaction using the 2213F and RED-DN primers is smaller than the 737 bp band produced in *orfF* mutant reactions and, consequently, is of no concern (Figure III.4).

The *orfG* mutant mycelial PCR reactions show that each Δ*orfG::apra* mutant is probably a true mutant. Although two bands do appear in the Δ*orfG::apra* reactions with the 4243F and 5226R primers, one is larger and one is smaller than the 983 bp band produced in the wild type reaction (Figure III.5).

orfH mutant mycelial PCR reactions reveal that Δ*orfH::apra* 2 and 3 may contain copies of the wild type chromosome. However, the bands that appear in the Δ*orfH::apra* 2 and 3 reactions with the 4243F and 5226R primers look as though they may be smaller than the expected wild type product of 983 bp (Figure III.6). Thus, at least Δ*orfH::apra* 1 is a true mutant.

III.1.4.1 Analysis of mutant clavam production profiles

Mutant strains were fermented and culture filtrates were analyzed to determine the levels of clavams produced by each mutant strain. Although one Δ*orfF::apra* strain and one Δ*orfH::apra* strain did not produce any 5*S* clavams, the remaining mutant strains were capable of producing all clavams (Table III.2). Besides these two mutant strains, some trends were observed. Δ*orfF::apra* mutants typically overproduced clavam-2-carboxylate and produced less alanylclavam. Δ*orfG::apra* mutants produced lower amounts of clavam-2-carboxylate (except Δ*orfG::apra* 4; this strain overproduced clavam-2-carboxylate) and overproduced alanylclavam. On the other hand, Δ*orfH::apra*

mutants produced more clavam-2-carboxylate and alanylclavam than wild type cultures. Two trends were consistent for almost all mutants: the overproduction of 2-hydroxymethylclavam and the low production of clavulanic acid. These trends must be viewed with caution because each mutant strain was fermented only once; however, it is apparent that most mutants in any one of the genes are capable of producing clavams.

III.2 **Analysis of the downstream portion of the paralogue gene cluster**

III.2.1 Sequencing of the region downstream of the paralogue gene cluster

In the analysis of the upstream region of the paralogue gene cluster, Cai (2003) constructed a transposon mutant library of the cosmid, 14E10 (Figure III.7). In vitro transposon mutagenesis using 14E10 combined with Tn*5062* (Figure III.8) generated 960 separate mutant cosmid clones within *E. coli* deposited in 96-well plates (Bishop et al. 2004; Cai 2003). Determination of the sequence downstream of the paralogue gene cluster began with restriction digest analysis of 14E10 transposon mutant clones.

Because Tn*5062* contains two *Eco*RI sites, the area in which Tn*5062* inserted could be determined by digesting mutant cosmids with *Eco*RI. The non-mutant form of 14E10 has six *Eco*RI sites; cleavage with *Eco*RI yields 0.1 kb, 2.7 kb, 6.1 kb, 8.1 kb, 11.4 kb, and 19.2 kb fragments (Figure III.9). However, if Tn*5062* is inserted into 14E10, an *Eco*RI digest causes one of the normally-expected bands to disappear and be replaced with three new bands: an invariant 0.8 kb band that is part of Tn*5062* and two new bands whose sizes depend upon the transposon insertion site (Figure III.9). Since the region downstream of the paralogue gene cluster is located in the 19 kb *Eco*RI fragment, mutant cosmids carrying insertions in this area were collected.

Although some of the 19 kb region was made up of unknown sequence, most of it (approximately 11 kb) had already been sequenced. Thus, more precise localization of transposon insertions was desirable before sequencing began. To accomplish this, a restriction site separating the sequenced region of the paralogue gene cluster from the unsequenced downstream region was used. An *Eco*721 restriction site was identified 1264 bp into the known sequence adjacent to the unknown sequence. By cleaving cosmids with both *Eco*721 and *Eco*RI, mutant cosmids more likely to be carrying *Tn*5062 within the unsequenced region were identified.

After identifying 24 mutant cosmids carrying Tn*5062* in the region downstream of the paralogue gene cluster, they were used in sequencing reactions. Tn*5062* carries flanking primer binding sites; sequencing outwards from each transposon insertion allowed most of the downstream region to be sequenced. One of the mutant cosmids, I-B7, never yielded sequencing results even after several attempts at using it as template DNA. Gaps in the sequence were joined by using custom primers and double-stranded sequence was obtained for almost the entire downstream area. However, an area (from nucleotide 3862 to 4276; see Appendix) could not be entirely sequenced on both strands. Even though a small region (41 bp) of double-stranded sequence overlap was obtained so that a single contiguous segment of sequence could be assembled, sequencing from either side of the area to gain completely double-stranded sequence information was not possible. In an attempt to reduce possible secondary structure in the region to facilitate DNA polymerase read-through, an amplicon encompassing part of the problem area was amplified by PCR. Although the primers used could amplify the region, the PCR product could not be sequenced entirely from both strands using the same primers. The area was

sequenced multiple times from a single strand and this sequence is thus deemed to be reliable.

III.2.2 Sequence analysis of the genes downstream of the paralogue gene cluster

Sequencing of the downstream region of the paralogue gene cluster revealed that it is 7690 bp in length and, as determined via FramePlot analysis, comprises six ORFs (Figure III.10 and III.11). The six ORFs were analyzed and sequence comparisons were drawn (Table III.3).

Although it was predicted that a gene encoding a response regulator is located downstream of *snk*, the presence of two response regulator genes was unexpected (Section I.5). The response regulator immediately downstream of *snk*, *res1* (83.1% third-letter codon G+C content), is short when compared to typical response regulators (Table III.3). While most response regulators are made up of two domains, a phosphate-binding pocket and a DNA-binding domain, Res1 is only made up of a phosphate-binding pocket. Downstream, and transcribed in the opposite direction, of *res1* is *res2* (90.2% third-letter codon G+C content), a gene encoding another response regulator. Res2 resembles a canonical response regulator in that it carries both a phosphate-binding domain and a DNA-binding domain (Table III.3). The 3' end of *res2* overlaps the 3' end of *res1* by 23 bp.

Two other genes in the downstream region shared significant similarities. The nucleotide sequences of *orf2/3* (thus named because it was originally misidentified as two separate ORFs) and *orf4* (97 and 95.8% third-letter codon G+C contents, respectively) show 66.7% identity, while their amino acid sequences display 49.1% identity. Indeed,

their encoded proteins show the most similarity to the same protein in the NCBI database (Table III.3). Both proteins have unique N-terminal regions and most amino acid sequence similarity can be found in the C-terminal regions. Examination of the C-termini of both proteins reveals that they contain a repeated sequence varying from 61 to 67 amino acids in length (Figure III.12). Interestingly, these repeat sequences show similarity to FG-GAP domains (found in eukaryotic proteins called integrins) implicated in Ca^{2+} and ligand binding (Springer 1996).

At the N-terminus, Orf2/3 carries a cysteine, histidine-dependent amidohydrolase/peptidases (CHAP) domain (Bateman and Rawlings 2003). CHAP domains are often found in amidases involved in peptidoglycan metabolism (Bateman and Rawlings 2003). The N-terminus of Orf4 carries a transglycosylase domain, which are also frequently involved in the breakdown and biosynthesis of carbohydrates. Both Orf2/3 and Orf4 show evidence of signal peptides, suggesting that they are secreted proteins.

Between *res2* and *orf4* resides *orf5* (89.2% third-letter codon G+C content). Orf5 shows a high degree of similarity to short-chain dehydrogenases, which carry out a variety of redox reactions (Table III.3).

orf1 (94% third-letter codon G+C content) is the ORF furthest downstream of the paralogue gene cluster. Orf1 displays similarity to only a few known proteins (Table III.3). Among *Streptomyces* genomes, Orf1 shows most similarity to SCO3065 (40% identity over 42 amino acids), a hypothetical protein from *S. coelicolor*, and SAV2254 (27% identity over 107 amino acids), an ATP/GTP-binding protein from *S. avermitilis*.

Unfortunately, Orf1 does not appear to contain any known protein domains; however, it does have a likely N-terminal signal peptide, suggesting that it is a secreted protein.

III.2.3 Generation of mutants in *res1*, *res2*, *orf5*, *orf4*, *orf2/3*, and *orf1*

Originally, the Tn*5062* mutant cosmid library was going to be used to introduce transposon insertions in all of the genes found downstream of the paralogue gene cluster. Tn*5062* carries an apramycin resistance gene as well as *oriT* (Figure III.8). These features would allow the introduction of the mutant cosmid into *S. clavuligerus* and the subsequent antibiotic selection of insertional mutants. Thus, mutant cosmids carrying Tn*5062* inserted in the discovered ORFs were conjugated into *S. clavuligerus*. The following mutant cosmids were chosen for insertional mutagenesis: I-D8 (for an *orf1* mutant), II-G10 (for an *orf2/3* mutant), II-E8 (for an *orf4* mutant), II-C1 (for another *orf4* mutant), II-H5 (for an *orf5* mutant), and I-A7 (for a *res2* mutant)(Figure III.11). There were no Tn*5062* mutant cosmids carrying the transposon in *res1* due to the ORF's small size.

Unfortunately, only the conjugation of II-G10 into *S. clavuligerus* yielded exconjugant colonies. Although they were each attempted several times, conjugations involving the other mutant cosmids did not produce colonies. The reason behind the lack of success of producing Tn*5062* insertional mutants in *S. clavuligerus* is unknown but has been observed in the past (Cai 2003).

Since a Tn*5062* mutant could only be obtained for *orf2/3*, the REDIRECT© protocol (described earlier) was used to produce deletion mutant strains in *res1*, *res2*,

orf5, *orf4* and *orf1*. Mycelia of each type of mutant were used in mycelial PCR to confirm mutations (Figures III.13, III.14, III.15, III.16, III.17 and III.18).

All of the above mutant strains appear to be true mutants. Even though the PCR reactions of *orf4* mutants did produce two bands in the wild type-specific reactions (with the NZE3 and NZE18 primers), neither band was the correct size (1484 bp) expected from the wild type chromosome (Figure III.16). Likewise, the band produced with *orf1* mutants using NZE9 and NZE17 was too large to be the 648 bp band produced in wild type reactions (Figure III.18).

III.2.3.1 Analysis of mutant clavam production profiles

The soy culture filtrates of mutants in *res1*, *res2*, *orf5*, *orf4*, *orf2/3*, and *orf1* were subjected to HPLC and bioassay analysis. However, it proved difficult to obtain consistent results. Multiple mutant strains were obtained for each gene; oftentimes, mutant strains bearing the same mutation did not produce comparable levels of clavams. Despite producing significantly varying levels of clavams, it was found that most mutant strains in *orf5*, *orf2/3*, and *orf1* could produce all of the clavams (Figure III.19). Although certain mutant strains in each of these three genes never produced alanylclavam in liquid soy cultures, all mutants produced all of the clavams when grown on solid soy medium (data not shown). Mutants in *res1*, *res2*, and *orf4* all generated consistent patterns of clavam biosynthesis: *res1* mutants overproduced 2-hydroxymethylclavam and alanylclavam while producing less than wild type levels of clavam-2-carboxylate, *res2* mutants were incapable of producing any 5*S* clavams on liquid or solid soy media, and

orf4 mutants overproduced all of the 5*S* clavams (Figure III.19; Table III.4). The lack of 5*S* clavams in *res2* mutant culture filtrates was confirmed via LC-MS (data not shown).

III.2.4 Complementation of *res2* and *snk* mutants

Mutants in both *snk* and *res2* cannot produce 5*S* clavams yet can still produce clavulanic acid (Tahlan 2005). The similarity of these phenotypes was of interest considering the proximity of these two genes whose protein products could interact as is seen in typical sensor kinase/response regulator systems. To confirm the role of these genes in 5*S* clavam biosynthesis, *res2* and *snk* mutants were complemented with single copies of their missing gene. *res2* and its upstream region was cloned as a 1.2 kb *Eco*RI/*Sac*II fragment into the integrative *Streptomyces* vector pSET-AT. However, since this insertion inadvertently replaced a portion of *tsr* (the thiostrepton resistance gene), this new plasmid was named pSET-AT*-*res2* (Table II.4). A complete copy of *tsr* was reintroduced to generate pSET-AT-*res2* (Table II.4). *snk* and its upstream region was cloned into pSET-AT as a 2 kb *Nhe*I fragment to create pSET-AT-*snk* (Table II.4). These plasmids carrying *res2* and *snk* were conjugated into *res2* and *snk* mutants, respectively, to produce the complemented mutants.

The phenotypes of both types of complemented mutants were of interest, but the complemented *snk* mutant was of particular significance. Since *snk* is 70 bp upstream of *res1* and both genes have the same orientation, it was conceivable that the genes are transcribed together in an operon (Figure III.11). If *snk* and *res1* are in an operon, then a complemented *snk* mutant would be expected to produce similar levels of clavams as a

res1 mutant. If the genes are not expressed as single transcript, then the complemented *snk* mutant would exhibit wild type levels of clavam production.

Complemented *snk* and *res2* mutants were grown in liquid soy medium and their culture broths were analyzed via HPLC to determine clavam production profiles. The complemented *snk* mutant clearly overproduced all of the 5*S* clavams, which is similar to the phenotype of the *res1* mutant (Figure III.20). Although this indicates that *snk* and *res1* may be transcribed in an operon, the same phenotype was observed in the complemented *res2* mutant (Figure III.20). Since complementation of either *snk* or *res2* mutants produced the same *res1* mutant-like phenotype, these experiments do not shed light on the transcriptional status of *snk* and *res1*. However, the importance of *snk* and *res2* in 5*S* clavam production is confirmed by these complementation tests.

III.2.5 Overexpression of *res1* and *res2* in *S. clavuligerus*

Because the phenotypes of *res1* and *res2* mutants were so pronounced, it was of interest to observe whether or not the opposite phenotypes are exhibited upon overexpression of either ORF. To this end *res1* was amplified by PCR with res1-pfwd and res1-prev and *res2* was amplified with res2-pfwd and res2-prev using a high fidelity PCR system; the amplicons were then TA-cloned into pCR®2.1-TOPO® (Table II.4; Table II.5). The forward primers used for PCR each had an engineered 5' *Nde*I site and the reverse primers contained a 5' *Bam*HI site for ease of insertion into expression vectors. The ORFs were sequenced from the resulting vectors (TOPO-*res1* and TOPO-*res2*; Table II.4) to ensure that mutations were not introduced during amplification. *res1* and *res2* were then excised and inserted into pHM8a to produce pHM8a-*res1* and

pHM8a-*res2*, respectively (Table II.4; Motamedi et al. 1995). pHM8a carries the *ermE** promoter (P_E) that allows constitutive expression of genes in *Streptomyces* (Motamedi et al. 1995; Weber et al. 1989). The resulting cassettes, carrying either *res1* or *res2* behind P_E and followed by a transcriptional terminator and a hygromycin resistance gene (*hyg*), were then inserted into pSET152 to produce pSET-P_E*res1* and pSET-P_E*res2* (Table II.4). Both of these integrative plasmids were introduced into wild type *S. clavuligerus* and strains carrying the plasmids were isolated and fermented.

It was noticed that during the seed culture growth, strains carrying either plasmid tended to begin lysing before growing for 40 hours. To allow adequate growth in the production medium, soy culture supernatants were analyzed after 120 hours of growth rather than the normal 96 hours after inoculating with 40 hour seed cultures. Before 120 hours of incubated shaking, most of the cultures did not exhibit growth likely because the inoculum had fewer than usual viable cells. Although the pSET-P_E*res2*-containing strains showed fairly consistent patterns of clavam production (Figure III.21), the pSET-P_E*res1*-containing strains did not (data not shown). Only two pSET-P_E*res1* cultures out of five showed growth in soy medium after 120 hours, and while one culture produced low levels of 5*S* clavams, the other overproduced them (data not shown). However, the *res2* overexpression strain did generally overproduce 2-hydroxymethylclavam as was expected based on the *res2* mutant phenotype (Figure III.21).

III.2.6 Generation and analysis of *res1*/*res2* double mutants

Both *res1* and *res2*, based on sequence similarities, likely encode response regulators. Because they are immediately adjacent to each other and because mutants in

either gene display defects in 5*S* clavam biosynthesis, it is quite possible that they participate in the same regulatory pathway. Even if they are not involved in the same regulatory mechanism, it could be possible to determine which of them acts further downstream in regulating 5*S* clavam production by knocking out both genes in a single mutant.

Additionally, each gene carries a single TTA codon. This indicates that they may both be regulated by *bldA*, the rare leucyl tRNA that is active during the later stages of the life cycle (Leskiw et al. 1991a). Interestingly, an *S. clavuligerus bldA* mutant cannot produce alanylclavam (S. E. Jensen, personal communication). If *bldA* is required for the expression of both *res1* and *res2*, then a *res1/res2* double mutant may be deficient in alanylclavam production.

A *res1/res2* double mutant was generated by using the REDIRECT© protocol to replace both *res1* and *res2* with the apramycin resistance gene cassette amplified by PCR (Gust et al. 2003). This was fairly simple because *res1* and *res2* are adjacent; the area encompassing both genes was replaced by a single cassette amplified by PCR with the orf7(RRshort)FWD and res1/2-REV primer pair (Table II.5). Δ*res1/res2::apra* mutants were isolated and then fermented in liquid soy medium.

HPLC analysis of culture filtrates revealed that the double mutant is incapable of producing 5*S* clavams but can still produce clavulanic acid (Figure III.22). Since the phenotype of this mutant is quite similar to that of a *res2* single mutant (Figure III.19), it is likely that *res2* acts later than *res1* in a regulatory pathway governing 5*S* clavam biosynthesis. As well, the *res1/res2* double mutant clavam production profile is different

from that of a *bldA* mutant. For this reason, it is unlikely that both *res1* and *res2* expression are under the control of *bldA*.

III.2.7 Biochemical analysis of Snk, Res1 and Res2

The putative atypical two-component system comprised of Snk, Res1 and Res2 represented a novel way in which clavam biosynthesis is governed in *S. clavuligerus*. In most two-component systems, a sensor kinase autophosphorylates in response to some stimulus and then passes the phosphate molecule to a response regulator. The phosphorylated response regulator is then capable of binding DNA and thereby influencing gene expression. However, the Snk/Res1/Res2 system is of interest not only because it regulates 5*S* clavam production but because it is made up of three components. To gain some insight into how the proteins of this three-component system interact, the purification of the proteins and determination of their phosphorylation activities was undertaken.

III.2.7.1 Overexpression and purification of Snk-His, Res1-His, and Res2-His

To overexpress Snk, Res1 and Res2, the ORF encoding each protein was amplified by PCR with a high fidelity DNA polymerase and then TA-cloned into pCR®2.1-TOPO®. This procedure was described earlier for *res1* and *res2* (Section III.2.5); *snk* was amplified using the snk-pfwd and snk-prev 2 primers and then TA-cloned to produce TOPO-*snk* (Table II.4; Table II.5). As previously described, the forward primers used for PCR each had an engineered 5' *Nde*I site and the reverse primers contained a 5' *Bam*HI site for ease of insertion into expression vectors. Each

ORF was then inserted into the T7 RNA polymerase-controlled pT7-7His expression vector, which confers a C-terminal 7x His-tag to resulting proteins. The resulting expression vectors (pT7-7His-*snk*, pT7-7His-*res1* and pT7-7 His-*res2*) were then introduced into *E. coli* BL21(DE3) for IPTG-induced high-level expression.

Overnight BL21(DE3) cultures were subcultured to produce larger expression cultures. Once the OD_{600} of a culture reached 0.6, protein expression was induced using 0.4 mM IPTG. After induction, cultures were grown for an additional five to six hours and then harvested via centrifugation. Cell pellets were frozen overnight, cells were lysed with lysozyme and were then sonicated. The resulting lysates were centrifuged to obtain clarified lysates.

Clarified lysates were combined with Ni-NTA resin and incubated for an hour at 4°C. Since each expressed protein carries a 7x His-tag, they should bind to Ni-NTA resin. Proteins were eluted by using a stepwise gradient of imidazole. This purification procedure allows the isolation of soluble protein under native conditions. After obtaining relatively pure samples of Snk-His and Res2-His, the samples were concentrated using centrifugation concentrators. Samples were collected at different stages of the purification process and then visualized using SDS-PAGE (Figures III.23 and III.24). Despite the appearance of two protein bands corresponding to the position of Res2-His (Figure III.24), SDS-PAGE using fresh running buffer revealed that only one band was truly present (data not shown).

For an unknown reason, Res1-His could not be isolated from pT7-7His-*res1* carrying cultures. Even purification of non-soluble protein material under denaturing conditions did not reveal the presence of any Res1-His protein. Thus, *res1* was excised

from TOPO-*res1* and inserted into pET-19b as an *Nde*I/*Bam*HI fragment to produce pET-19b-*res1* (Table II.4). This construct confers an N-terminal 10x His-tag to Res1. Once this expression plasmid was obtained, it was transformed into *E. coli* BL21(DE3) for high-level expression upon IPTG induction.

Cultures carrying pET-19b-*res1* were grown overnight and then inoculated into larger expression cultures as was done earlier. After the OD_{600} reached 0.6 and cultures were induced with 1 mM IPTG, cultures were initially shaken at 37°C for five to six hours. This protocol only allowed Res1-His to be produced in an insoluble form and to be purified under denaturing conditions. However, shaking the expression cultures at room temperature (approximately 22°C) for seven hours resulted in the purification of soluble Res1-His under native conditions. The same purification method used for isolating Snk-His and Res2-His under native conditions was used with some modifications to purify Res1-His (Figure III.25). Samples were once again concentrated with centrifugation concentrators. All proteins were resuspended in protein storage buffer, divided into smaller aliquots, and stored at -80°C.

III.2.7.2 In vitro phosphorylation assays

Once relatively pure samples of Snk-His, Res1-His and Res2-His had been isolated, it was possible to test the phosphorylation dynamics between the proteins. Normally, sensor kinases are capable of obtaining the γ-phosphate from ATP and covalently attaching it to an internal histidine residue. The phosphate molecule is then passed onto an aspartate residue on a cognate response regulator. Since there is one sensor kinase but two response regulators, the path of a phosphate molecule through this

system was uncertain. It was possible that Snk phosphorylates only Res1, only Res2, both, or neither. However, because mutagenesis of *snk*, *res1* or *res2* affects 5*S* clavam biosynthesis significantly, it is likely that the encoded proteins do interact on some level with phosphorylation being the most obvious interaction.

Different combinations of Snk-His, Res1-His and Res2-His were incubated together with [γ-^{32}P]ATP at 30°C. After 30 minutes, each reaction was subjected to SDS-PAGE to separate the proteins and the dried gel was exposed to a phosphor screen.

The results from such assays revealed that Snk-His autophosphorylates (Figure III.26). As well, Snk-His is capable of phosphorylating both Res1-His and Res2-His, even when all three proteins are combined (Figure III.26). Despite the difference in phosphorylation levels of Snk-His, Res1-His, and Res2-His in different assays, the same variations were not observed upon repeating the assays (data not shown).

III.3 Development of an in vivo, FLP-mediated technique to produce in-frame, unmarked gene deletions

The original REDIRECT© procedure describes a method to produce in-frame deletion mutants that do not carry any selectable markers (Figure I.8; Gust et al. 2003). First, a preliminary mutant with a cassette amplified by PCR replacing the gene of interest is generated. This cassette carries an antibiotic resistance gene, *oriT*, and is flanked by FRT sites. The cassette replaces the gene of interest on a cosmid in *E. coli* expressing the λ Red recombination system, which promotes recombination when using linear DNA (Murphy 1998). Cosmid DNA that has the gene replaced by the cassette is

transferred to the organism of interest by intergeneric conjugation and homologous recombination allows the mutation to be introduced onto the chromosome.

In the second stage of the process, the mutant cosmid is transformed into a strain of *E. coli* that carries a plasmid that expresses FLP recombinase. FLP recombinase recognizes the flanking FRT sites on the cassette and catalyzes site-specific recombination between the sites to excise intervening DNA (Cherepanov and Wackernagel 1995). This results in the formation of an 81 bp "scar" that replaces the cassette, but retains the original gene's start and stop codons. Because the number of nucleotides in the "scar" is a multiple of three, the new mutation should be in-frame and thus avoid any issues of polarity for downstream genes in the same operon.

This new mutant cosmid that carries the 81 bp "scar" in place of the REDIRECT© cassette is then transformed into the original mutant carrying the cassette. The cosmid must be transformed rather than conjugated because it now lacks *oriT*. The in-frame unmarked mutation replaces the cassette through homologous recombination and antibiotic sensitive transformants will carry the new mutation.

Unfortunately, this method is not suitable for all *Streptomyces* spp. In the case of *S. clavuligerus*, large molecules of DNA (such as cosmids, for example) cannot be introduced via transformation (S. E. Jensen, personal communication). Instead, they must be transferred into the cells through conjugation. Although this does not affect the production of the original mutant carrying the cassette amplified by PCR in its chromosome, it does hamper the second step, the production of the in-frame unmarked mutant. Two ways to bypass this problem were investigated: FLP-mediated excision of the cassette on a conjugative vector within *E .coli* with subsequent conjugation into the

original mutant, and excision of the cassette directly from the chromosome of the original mutant. Both methods were used successfully to produce in-frame markerless *Streptomyces* mutants.

III.3.1 Production of an in-frame unmarked Δ*snk* mutant

In testing different methods of producing in-frame unmarked deletions in *S. clavuligerus*, it was decided that *snk*, among other genes, is a suitable target for such analyses. Since *res1* is only 70 bp downstream of *snk*, it is conceivable that *snk* and *res1* are transcribed in an operon. If they are transcribed as a single unit, then all *snk* mutants tested so far actually lack expression of both *snk* and *res1* and, thus, may not be accurate representations of cells that lack *snk* by itself. Therefore, production of an in-frame Δ*snk* mutant would serve not only to illustrate a method of producing in-frame deletions but would also provide confirmation of the original Δ*snk::apra* mutant phenotype.

In the first method for producing a Δ*snk* mutant, a fragment of DNA was isolated from 14E10-SKN (the mutant cosmid carrying the REDIRECT© cassette in place of *snk*) encompassing the REDIRECT© cassette and flanking regions (Table II.4). A 10.4 kb *Kpn*I/*Nde*I fragment including the region carrying the REDIRECT© cassette in place of *snk* was inserted into pSL1180 to form pSL-Δ*snk::apra*. From there, a 10.4 kb *Eco*RI/*Hind*III Δ*snk::apra* fragment was then cloned into pUWL-KS to produce pUWL-Δ*snk::apra* (Table II.4). This plasmid was transformed into *E. coli* carrying BT340, a plasmid that expresses FLP recombinase and is lost at 42°C (Gust et al. 2003; Table II.3). Two transformant colonies were streaked out onto a non-selective medium for single colonies and then grown at 42°C to induce FLP expression and BT340 loss. Thirty of the

resulting colonies were patched onto media containing either apramycin or ampicillin. Ampicillin resistant (conferred by the pUWL-based plasmid) and apramycin sensitive (due to the loss of the REDIRECT© cassette) colonies were isolated and loss of the REDIRECT© cassette from the new plasmid, pUWL-Δ*snk*, was confirmed via restriction digestion and sequencing (Table II.4).

Attempts to transform pUWL-Δ*snk* into the Δ*snk::apra* mutant were fraught with contamination. Plates spread with transformed *S. clavuligerus* protoplasts usually became overgrown with fungal colonies in a short period of time. For this reason, the 9 kb *Eco*RI/*Hind*III fragment carrying the Δ*snk* mutation and flanking regions was inserted into pUWL-*oriT* to produce pUWL-*oriT*- Δ*snk* (Table II.4). This plasmid was then conjugated into the original Δ*snk::apra* mutant and replica-plating allowed ampicillin and apramycin sensitive exconjugants to be isolated. The overall procedure used in the isolation of Δ*snk* mutants is outlined in Figure III.27. Mycelial PCR of cultures confirmed the identity of the strains (Figure III.28). Although the wild type reaction produced bands in the *snk* mutant-specific reaction with the RED-UP and ANW40 primers, none of the bands appears to be precisely the same size as the 1220 bp band found in the snk mutant reactions (Figure III.28). As well, the bands that appeared in the Δ*snk* 2-1, 2-2, and 2-4 reactions with the ANW23 and ANW32 primers were not 1676 bp as observed in wild type reactions (Figure III.28).

The resulting Δ*snk* strains were fermented in soy medium. HPLC analysis revealed that the culture filtrates, akin to those produced by Δ*snk::apra* mutants, did not contain any 5*S* clavams (Figure III.29).

III.3.2 Production of plasmids for FLP expression in *Streptomcyes*

Although the above method works for producing in-frame deletions, a faster and easier method was desirable. One such protocol involves the excision of the REDIRECT© cassette directly from the chromosome through FLP expression in the original mutant. To accomplish this, *flp* must first be cloned into suitable *Streptomyces* expression vectors that are easily lost from cells. *flp* was first isolated from pEAW115 as a 1.5kb *Xho*I/*Xba*I fragment and inserted into pCR®2.1-TOPO® to generate TOPO-*flp* (Table II.4). *flp* was excised from this plasmid as an *Nde*I/*Bam*HI fragment through a partial digest (*flp* contains an *Nde*I site within its start codon and another further downstream within the gene) and then cloned into pHM8a to produce pHM8a-*flp* (Table II.4). pHM8a-*flp* carries *flp* under the control of the *ermE** promoter (P_E) and upstream of a transcriptional terminator and *hyg* (Motamedi et al. 1995). The entire cassette carrying P_E, *flp*, the transcriptional terminator (*t*), and *hyg* was inserted as a 5.1 kb *Bgl*II fragment into the *Bam*HI sites of pHJ401 and pUWL-*oriT* to produce pHJ401-*flp* and pUWL-*oriT*-*flp*, respectively (Table II.4). These two plasmids can be conjugated into *Streptomyces* cells to enable constitutive expression of *flp*. pHJ401 is a segregationally unstable *E. coli*/*Streptomyces* shuttle vector. pUWL-*oriT*, which carries the pIJ101 origin of replication, is easily lost from *S. clavuligerus* cells. The instability of such plasmids will facilitate their loss to produce mutant strains lacking the plasmids.

III.3.3 Generation of an in-frame unmarked Δ*c7p* mutant

In a past study, a Δ*c7p::apra* mutant was generated (Tahlan et al. 2007). This mutant was unable to produce 5*S* clavams. However, another gene downstream of *c7p* and transcribed in the same direction, *c6p*, is also essential for 5*S* clavam production (Tahlan et al. 2007). If *c7p* and *c6p* are transcribed in an operon, then it is possible that *c7p* is not actually required in clavam biosynthesis. To determine if *c7p* is in fact necessary, it was chosen as the target for the generation of an in-frame deletion.

To produce the Δ*c7p* mutation, pUWL-*oriT-flp* was conjugated into the original Δ*c7p::apra* mutant. Thiostrepton resistant exconjugants were obtained and then sporulated twice without antibiotic selection to allow for the loss of the FLP-expressing plasmid. Replica-plating was used to isolate apramycin sensitive (indicating loss of the REDIRECT© cassette) and thiostrepton sensitive (showing loss of pUWL-*oriT-flp*) colonies. These colonies were used to generate spore stocks of Δ*c7p* mutant strains. The Δ*c7p* mutation was confirmed by mycelial PCR followed by sequencing of generated PCR fragments carrying the 81 bp "scar" in place of *c7p* or the REDIRECT© cassette (Figure III.30). Fermented mutants were unable to produce 5*S* clavams, confirming the importance of *c7p* in clavam biosynthesis (Figure III.31). The general procedure used in this protocol is summarized in Figure III.32.

III.3.4 The *red* gene cluster in *S. coelicolor*

It is clear that the in vivo expression of FLP for the excision of REDIRECT© cassettes directly from the chromosome (Section III.3.3) is faster and less laborious than subcloning a fragment of DNA encompassing the original mutation into a conjugative

vector and then carrying out FLP-mediated excision in *E. coli* (Section III.3.2). The in vivo excision of cassettes would also facilitate the production of in-frame deletions in *Streptomyces* spp. that allow large pieces of DNA to be introduced via transformation. The simple introduction and subsequent loss of a plasmid expressing FLP into the original REDIRECT© mutant carrying the cassette would be simpler than carrying out cassette excision in *E. coli* and then transforming the resulting cosmid into the mutant. Therefore, the utility of in vivo excision was demonstrated in the *Streptomyces* genetic model, *S. coelicolor*.

The *red* gene cluster in *S. coelicolor* was deemed a suitable target to test the effectiveness of the in vivo expression of FLP. The *red* cluster is responsible for producing the distinctive red-pigmented antibiotic, undecylprodigiosin (Rudd and Hopwood 1980). This gene cluster contains four putative transcriptional units carrying genes involved in the biosynthesis of undecylprodigiosin (Cerdeno et al. 2001; Figure III.33). One of the operons contains a gene, *redJ*, that has unknown importance to undecylprodigiosin production (Cerdeno et al. 2001). Within the putative operon, *redJ* is upstream of a number of genes that are required for undecylprodigiosin production. Replacement of *redJ* with a REDIRECT© cassette will likely have polar effects on the downstream genes (*redI*, *redH*, *redG* and *redF*) necessary for undecylprodigiosin biosynthesis. Therefore, the initial REDIRECT© *redJ* mutant, Δ*redJ::apra*, would be no longer able to produce undecylprodigiosin even if RedJ itself was not required. However, replacement of the cassette with the 81 bp "scar" would relieve any downstream polar effects; generation of an in-frame Δ*redJ* mutant will restore

undecylprodigiosin production if RedJ is not necessary for synthesis of the red-pigmented antibiotic.

III.3.4.1 Generation of an in-frame unmarked Δ*redJ* mutant

To test the effectiveness of the in vivo excision of REDIRECT© cassettes in *S. coeliolor*, a Δ*redJ::apra* mutant first had to be constructed. St3f7, a cosmid carrying a region of the *red* cluster encompassing *redJ*, had *redJ* replaced with a REDIRECT© cassette amplified by PCR (produced with the FWD-redJ and REV-redJ primers) to produce St3f7-Δ*redJ::apra* (Table II.4; Table II.5). This mutant cosmid was then conjugated into *S. coelico*lor M145 to generate the initial Δ*redJ::apra* mutant. Apramycin resistant and kanamycin sensitive exconjugant colonies were selected and used to produce spore stocks of Δ*redJ::apra* strains. The phenotypes of these mutants are described below.

Both pHJ401-*flp* and pUWL-*oriT-flp* were conjugated into a Δ*redJ::apra* mutant to allow FLP-mediated excision of the REDIRECT© cassette. Thiostrepton resistant exconjugant colonies were selected and sporulated once without antibiotic selection to allow plasmid loss. Three exconjugant colonies carrying either plasmid were selected and replica-plated to isolate thiostrepton and apramycin sensitive cells. Replica-plating revealed that all of the exconjugants had lost the plasmids. Additionally, out of 51 colonies originally containing pUWL-*oriT-flp*, 33 (65%) colonies appeared to be apramycin sensitive and therefore had lost the REDIRECT© cassette. Out of 89 colonies originally carrying pHJ401-*flp*, four (4.5%) colonies seemed to be apramycin sensitive. Twelve apra^{s}thio^s colonies that had lost pUWL-*oriT-flp* and all four apra^{s}thio^s colonies

that had lost pHJ401-*flp* were patched onto media containing either apramycin or thiostrepton to confirm them as true Δ*redJ* mutants. Three out of 12 colonies that had lost pUWL-*oriT-flp* were truly apra^s, and one out of four colonies that had lost pHJ401-*flp* was indeed apra^s. All tested colonies were thio^s, indicating that the FLP-expressing plasmids had been lost. This suggests that the FLP-expressing plasmids are often lost prior to FLP expression. Genomic DNA was isolated from each Δ*redJ::apra* strain and Δ*redJ* strain. PCR was carried out to confirm the identity of each mutant (Figure III.34). The PCR fragments encompassing the 81 bp "scar" in the Δ*redJ* mutants were sequenced to confirm the replacement of *redJ* and the REDIRECT© cassette with the in-frame deletion.

Although the M145 PCR reaction with the *redJ* mutant-specific primers (redJ-UP and RED-UP) produced a product, it was not the 439 bp product seen with the mutants (Figure III.34). As well, the appearance of bands in *redJ* mutant reactions when using redJ-UP and redJ-INT was not a cause for concern since none of the bands were 658 bp as observed with wild type reactions (Figure III.34).

Each strain was then tested for the production of red pigment. An *S. coelicolor* M145 culture, three Δ*redJ::apra* strains, and the four isolated Δ*redJ* strains were also grown. Cultures were grown in 25 ml liquid R5 medium supplemented with 5% (w/v) PEG 8000 for 120 hours, during which cell growth and undecylprodigiosin production were measured. Undecylprodigiosin was solvent-extracted from mycelia under acidic conditions and quantified by measuring absorbance at 530 nm of the resulting methanol-undecylprodigiosin solutions. Red pigment production was determined relative to the level of cellular growth, measured as OD_{450}, for each culture (Figure III.35).

It is clear that the M145 culture is capable of producing undecylprodigiosin [Figure III.35 (A)]. As expected, the Δ*redJ::apra* cultures did not produce near as much red pigment as did the M145 culture [Figure III.35 (B)]. On the other hand, the Δ*redJ* cultures did not produce much undecylprodigiosin either [Figure III.35 (C)]. Although two Δ*redJ* strains can produce some undecylprodigiosin, even their cultures did not produce appreciable levels of the antibiotic, indicating that *redJ* plays some kind of role in undecylprodigiosin production.

Table III.1. Sequence analysis of genes in the upstream region of the paralogue gene cluster in *S. clavuligerus*.

ORF	Length (aa)[a]	Protein with highest similarity	% Identity/ Similarity[a,b]	Accession Number	Transmembrane domain(s)	Predicted function
orfA	390	Serine hydroxymethyltransferase from *Methanocaldococcus jannaschii*	32/48, over 380 aa	AAB99615	-	Hydroxymethyltransferase
orfB	126	UPF0076 protein PHO854 from *Pyrococcus horkoshii*	49/71, over 119 aa	O58584	-	YjgF/YER057c/UK114 family protein; possible amino acid biosynthetic regulator
orfC	393	Aminotransferase, class I and II from *Solibacter usitatus*	37/54, over 344 aa	YP_824873	-	Aminotransferase
orfD	351	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit from *Acidobacteria bacterium*	39/56, over 314 aa	YP_591159	-	Threonine dehydratase
orfE	102	Hypothetical protein RHA1_o07262 from *Rhodococcus* sp. RHA1	46/62, over 94 aa	YP_707184	+	AzlD type amino acid permease
orfF	253	Possible branched-chain amino acid transporter from *Rhodococcus* sp. RHA1	53/66, over 204 aa	YP_707185	+	AzlC type amino acid permease
orfG	364	Hypothetical protein SAV818 from *Streptomyces avermitilis*	51/65, over 353 aa	NP_821993	-	Protein kinase

Table III.1. (continued).

ORF	Length (aa)[a]	Protein with highest similarity	% Identity/ Similarity[a,b]	Accession Number	Transmembrane domain(s)	Predicted function
orfH	416	Putative transporter from *Streptomyces ambofaciens*	45/60, over 400 aa	CAK50890	+	Multidrug resistance transporter

[a] aa stands for number of amino acids
[b] Identity and similarity of amino acid sequences were reported as percentages over a range of amino acid

Table III.2. Clavam production by *orfF*, *orfG*, and *orfH* mutant strains after 96 hours growth in soy medium[a].

Strain	Clavulanic acid (% of wild type)[b]	2-hydroxymethylclavam (% of wild type)	Clavam-2-carboxylate (% of wild type)	Alanylclavam (% of wild type)
Wild type	100	100	100	100
Δ*orfF::apra* 1	130.5	0	0	0
Δ*orfF::apra* 2	30	218.1	177.9	63.2
Δ*orfF::apra* 3	70.3	107.3	97.5	26
Δ*orfF::apra* 4	47.3	146.8	333.8	73.6
Δ*orfF::apra* 5	57.1	194.6	196.2	124.1
Δ*orfG::apra* 1	4.2	153.5	44	157.7
Δ*orfG::apra* 2	3.9	211.7	39.8	163.3
Δ*orfG::apra* 3	11.7	378.2	77.7	222.4
Δ*orfG::apra* 4	48.9	874	270	249.4
Δ*orfH::apra* 1	18.8	564.2	243.9	184.5
Δ*orfH::apra* 2	14.2	325	123.7	150.3
Δ*orfH::apra* 3	43.8	0	0	0

Table III.2. (continued).

[a] Results from individual fermentations are shown; each mutant strain was fermented once

[b] Clavam production levels were determined by integrating the area beneath HPLC peaks and then comparing these values to wild type peak areas

Table III.3. Sequence analysis of genes in the downstream region of the paralogue gene cluster in *S. clavuligerus.*

ORF	Length (aa)[a]	Protein with highest similarity	% Identity/ Similarity[a,b]	Accession Number	Transmembrane domain	Predicted function
res1	135	Response regulator receiver protein from *Mycobacterium vanbaalenii*	56/65 over 131 aa	YP_952893	-	Response regulator
res2	214	Probable response regulator from *Rhodococcus* sp. RHA1	78/86 over 214 aa	YP_702584	-	Response regulator
orf5	248	Short chain dehydrogenase from *Solibacter usitatus*	47/62 over 241 aa	YP_827794	-	Short chain dehydrogenase
orf4	618	Putative alpha-amylase from *Clostridium beijerinckii*	36/49 over 350 aa	AAB36555	-	Transglycosylase
orf2/3	676	Putative alpha-amylase from *Clostridium beijerinckii*	35/49 over 349 aa	AAB36555	-	Esterase
orf1	200	Type III restriction enzyme from *Syntrophobacter fumaroxidans*	29/44 over 73 aa	YP_847379	-	Unknown

[a] aa stands for number of amino acids
[b] Identity and similarity of amino acid sequences were reported as percentages over a range of amino acid

Table III.4. Clavam production levels expressed as percentages of wild type production by different mutant strains in *res1*, *res2* and *orf4* after 96 hours of growth in liquid soy medium[a].

Strain	Clavulanic acid (% of wild type)[b]	2-hydroxymethlclavam (% of wild type)	Clavam-2-carboxylate (% of wild type)	Alanylclavam (% of wild type)
Wild type	100	100	100	100
Δ*res1::apra* 1	65.3	629.9	72.3	137
Δ*res1::apra* 2	80.6	514.4	29.9	124.6
Δ*res1::apra* 3	107.9	692.1	70.9	152.2
Δ*res1::apra* 4	92.2	586.1	59.1	135.2
Δ*res2::apra* 1	77.9	0	0	0
Δ*res2::apra* 2	24.3	0	0	0
Δ*res2::apra* 3	25.7	0	0	0
Δ*orf4::apra* 1	94.3	226.3	163.3	152.9
Δ*orf4::apra* 2	23	285.9	132	100.3
Δ*orf4::apra* 3	78.8	150	190.2	119.6

[a] Results from individual fermentations are shown; all Δ*res1::apra* mutants were fermented at least three times, and the Δ*res2::apra* and Δ*orf4::apra* mutants were fermented once each

[b] Clavam production levels were determined by integrating the area beneath HPLC peaks and then comparing these values to wild type peak areas

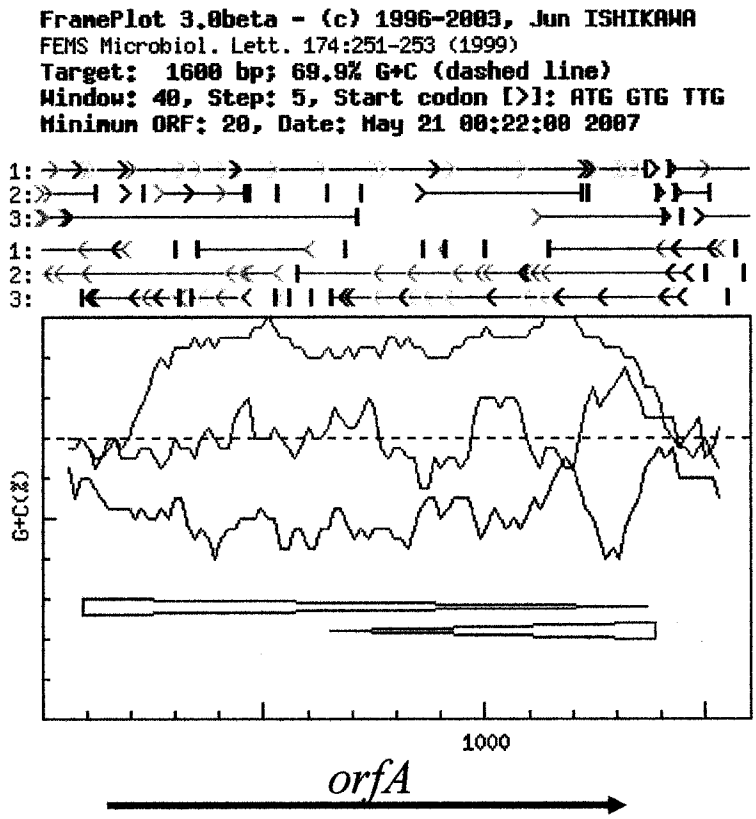


Figure III.1. FramePlot analysis of *orfA*. The complete *orfA* sequence was obtained with sequence information from p2.8-18. Final sequence data were examined using FramePlot, a program used for the prediction of ORFs in high G+C organisms. FramePlot predicts the presence of ORFs based on G+C content in the third-letter position of codons. *Streptomyces* spp. typically have very high G+C levels in the third-letter position of codons. The lines above the graph represent the six possible reading frames, arrowheads represent potential start codons, and vertical bars represent possible stop codons. Triangles at the bottom of the graph represent likely ORFs as predicted by FramePlot.

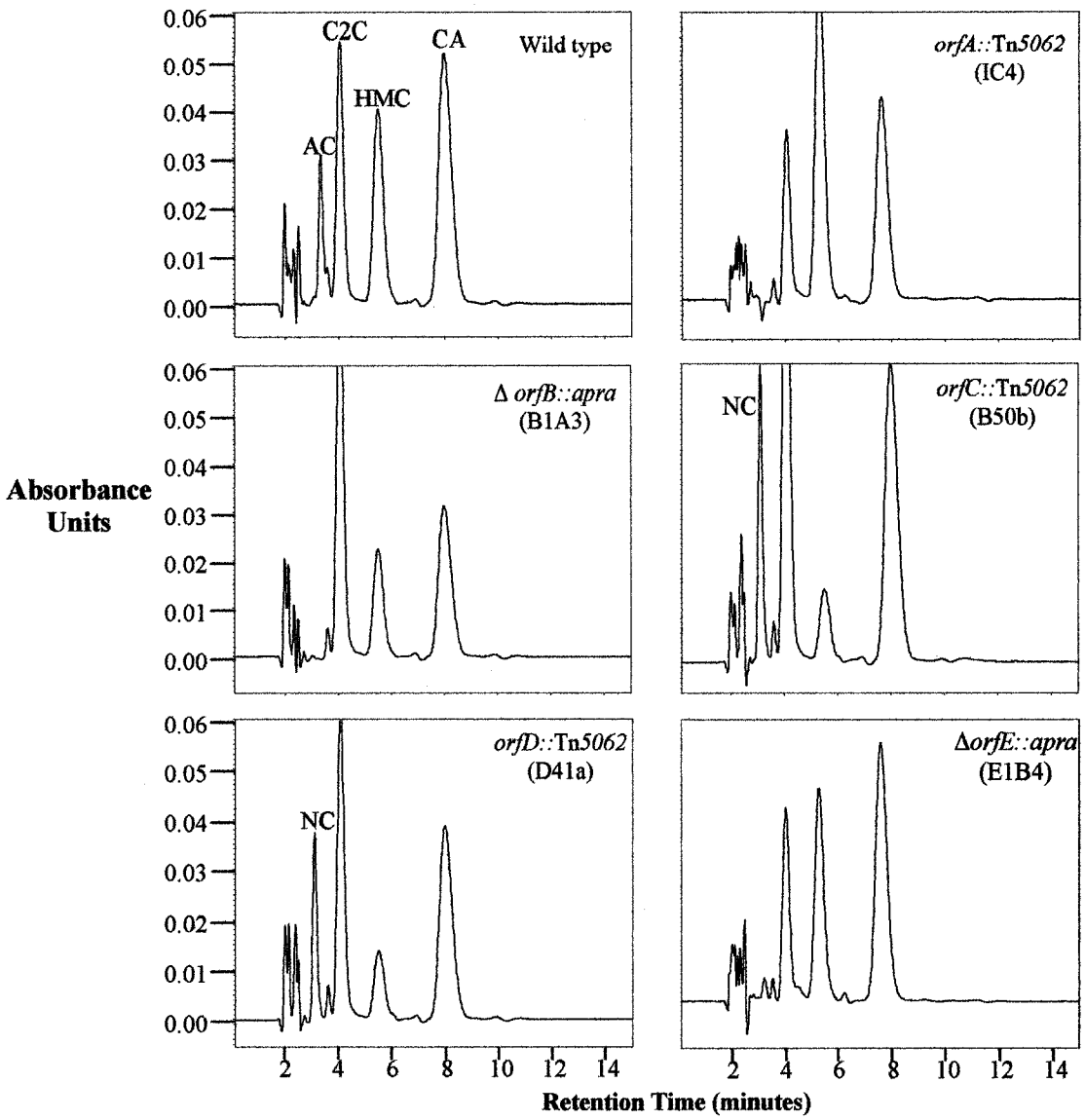


Figure III.2. Clavam production profiles of *orfA*, *orfB*, *orfC*, *orfD*, and *orfE* mutants. Culture filtrates from 96 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting underivatized sample chromatograms from derivatized sample chromatograms. All samples were diluted (1/5) before analysis. NC: novel clavam; AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

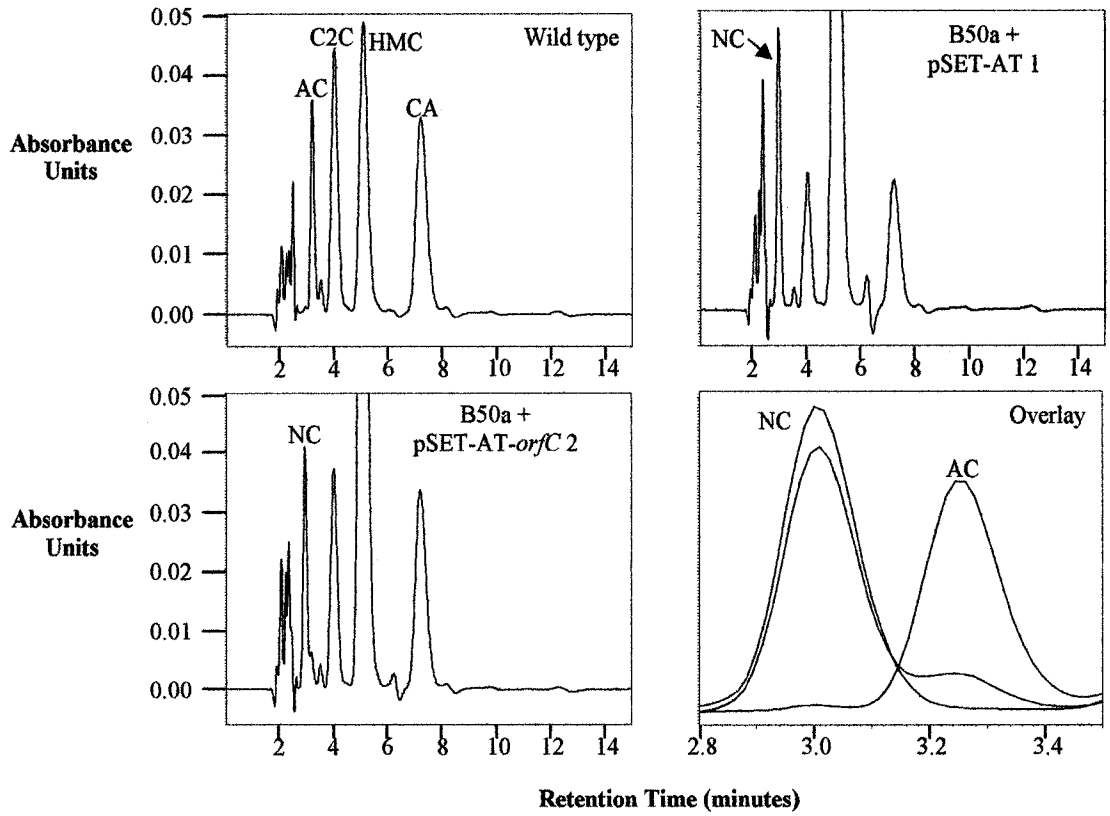


Figure III.3. Clavam production profiles of the B50a+pSET-AT-*orfC* strain. Culture filtrates from 96 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting underivatized samples from derivatized samples. All samples were diluted (1/5) before analysis. The bottom right chromatogram is an overlay of all three other chromatograms shown in the figure from 2.8 to 3.5 minutes. At 3.0 minutes, the top line is the B50a + pSET-AT 1 profile, the middle line is the B501 + pSET-AT-orfC 2 profile, and the bottom line is the wild type profile. NC: novel clavam; AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid. B50a is a mutant with Tn*5062* inserted in *orfC* and B50a+pSET-AT 1 is the B50a mutant carrying pSET-AT integrated in the chromosome (Table II.2; Table II.4).

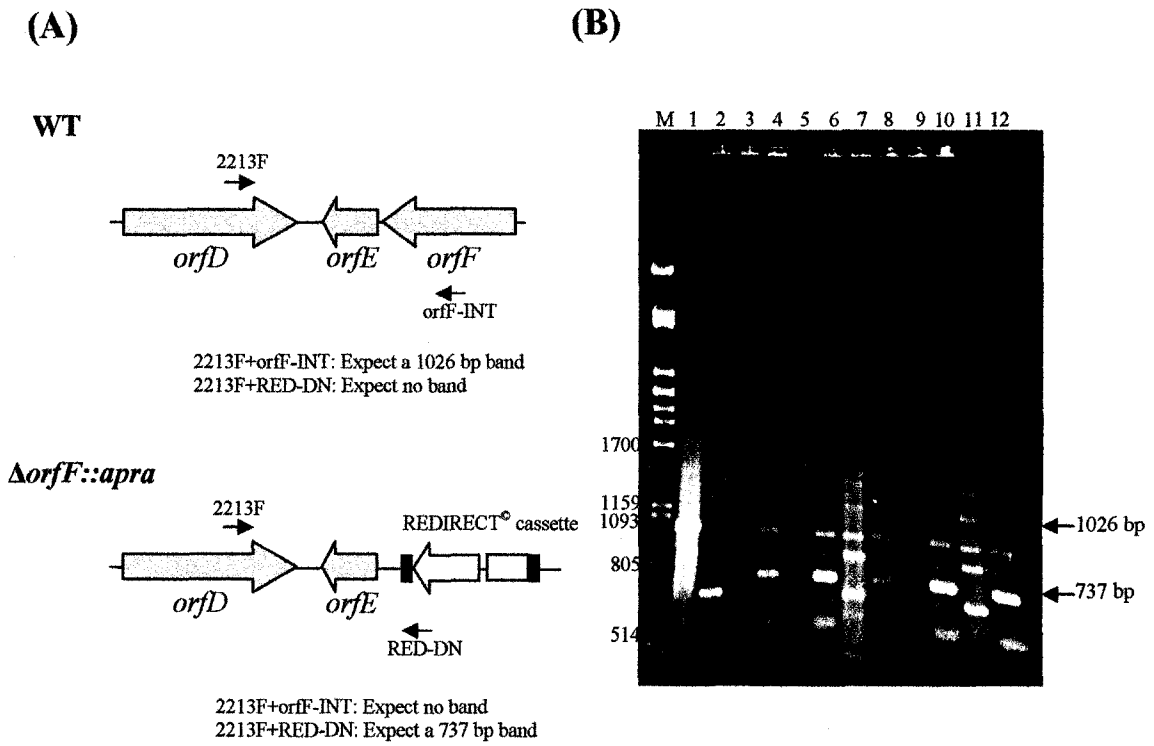


Figure III.4. Expected (A) and observed (B) mycelial PCR products from wild type and *orfF* mutants. Lane M, λ*Pst*I marker. Lanes 1, 3, 5, 7, 9, and 11, reactions containing primers 2213F and orfF-INT. Lanes 2, 4, 6, 8, 10, and 12, reactions containing primers 2213F and RED-DN. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*orfF::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*orfF::apra* 2 mycelium. Lanes 7 and 8, reactions containing Δ*orfF::apra* 3 mycelium. Lanes 9 and 10, reactions containing Δ*orfF::apra* 4 mycelium. Lanes 11 and 12, reactions containing Δ*orfF::apra* 5 mycelium.

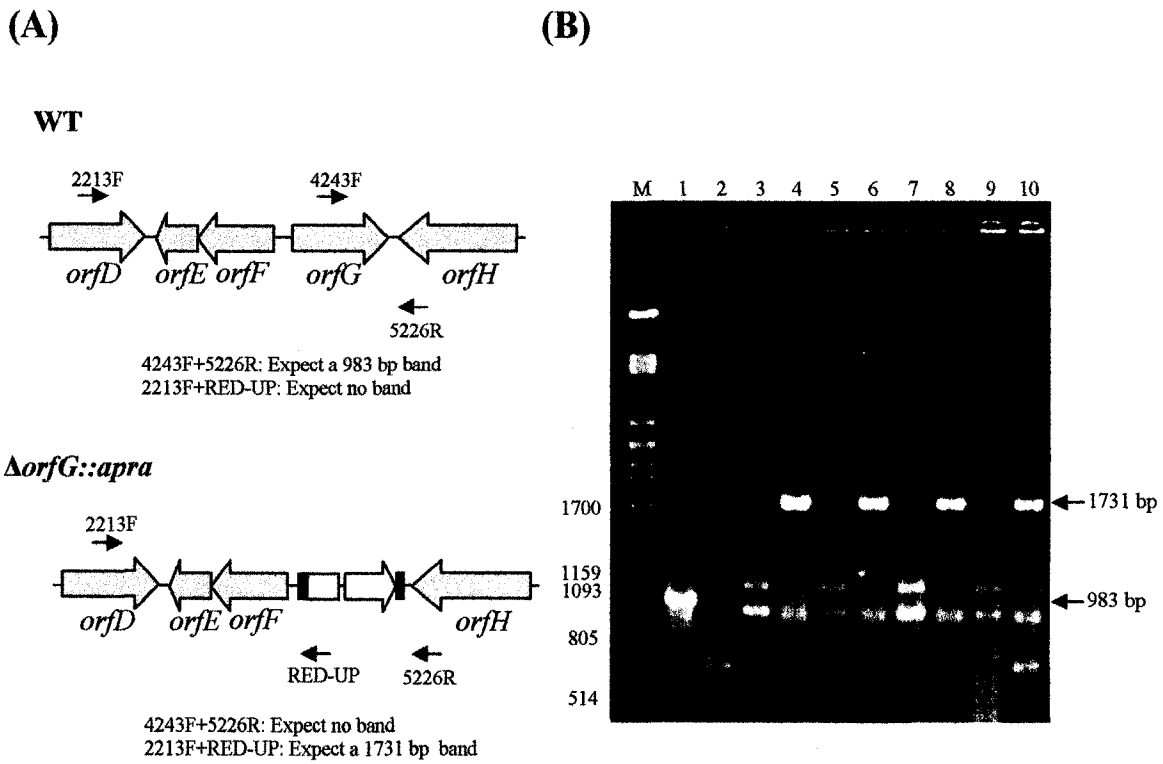


Figure III.5. Expected (A) and observed (B) mycelial PCR products from wild type and *orfG* mutants. Lane M, λ*Pst*I marker. Lanes 1, 3, 5, 7, and 9, reactions containing primers 4243F and 5226R. Lanes 2, 4, 6, 8, and 10, reactions containing primers 2213F and RED-UP. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*orfG::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*orfG::apra* 2 mycelium. Lanes 7 and 8, reactions containing Δ*orfG::apra* 3 mycelium. Lanes 9 and 10, reactions containing Δ*orfG::apra* 4 mycelium.

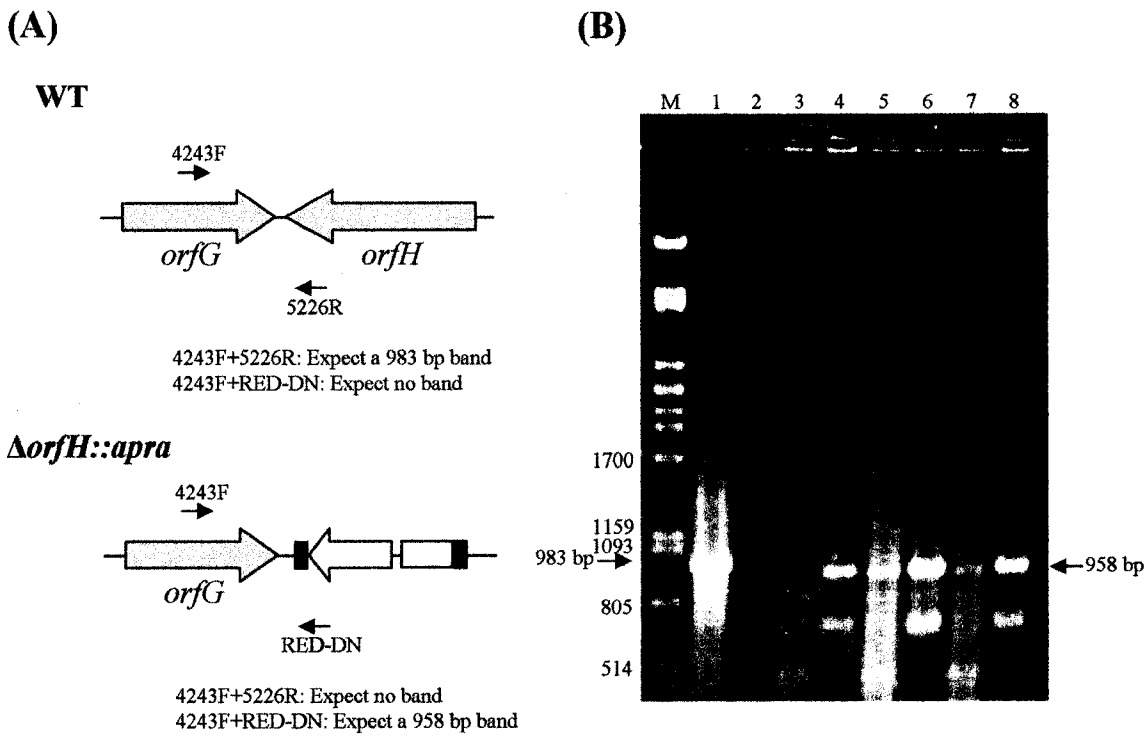


Figure III.6. Expected (A) and observed (B) mycelial PCR products from wild type and *orfH* mutants. Lane M, λ*Pst*I marker. Lanes 1, 3, 5, and 7, reactions containing primers 4243F and 5226R. Lanes 2, 4, 6, and 8, reactions containing primers 4243F and RED-DN. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*orfH::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*orfH::apra* 2 mycelium. Lanes 7 and 8, reactions containing Δ*orfH::apra* 3 mycelium.

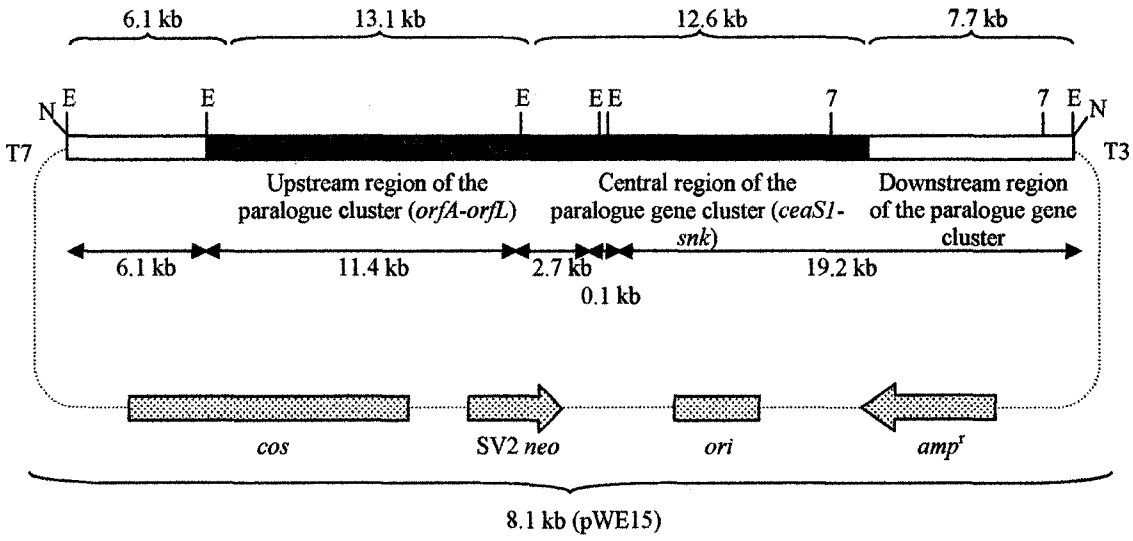


Figure III.7. The cosmid 14E10 carrying *S .clavuligerus* genomic DNA. The chromosomal DNA insert includes the entire known paralogue gene cluster and flanking regions. Boxes indicate regions of the *S. clavuligerus* chromosomal insert. Clear boxes indicate unsequenced regions, the gray box is the sequenced upstream region of the paralogue gene cluster, and the black box is the sequenced central region of the paralogue gene cluster. Dotted lines indicate the vector (pWE15) sequence. *Eco*RI (E), *Not*I (N), and *Eco*721 (7) sites are shown. Indicated above the boxes (with braces) are the sizes of the unsequenced and sequenced regions. Indicated below the boxes (with double-headed arrows) are the sizes of *Eco*RI fragments generated upon 14E10 *Eco*RI digestion.

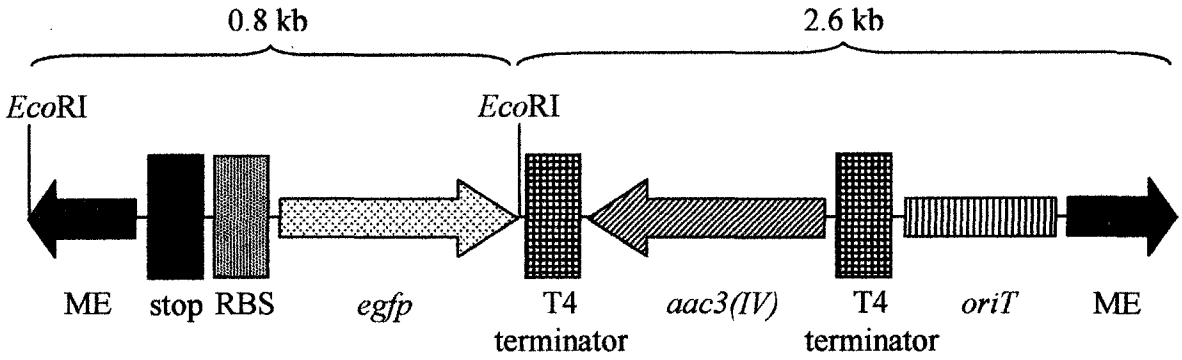
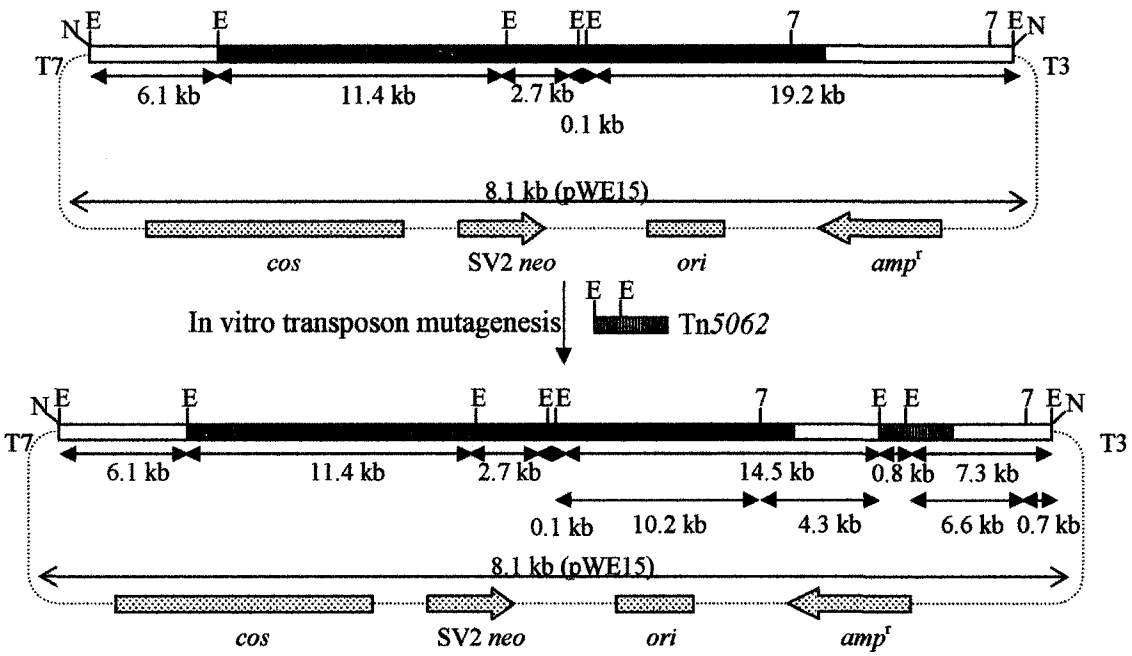
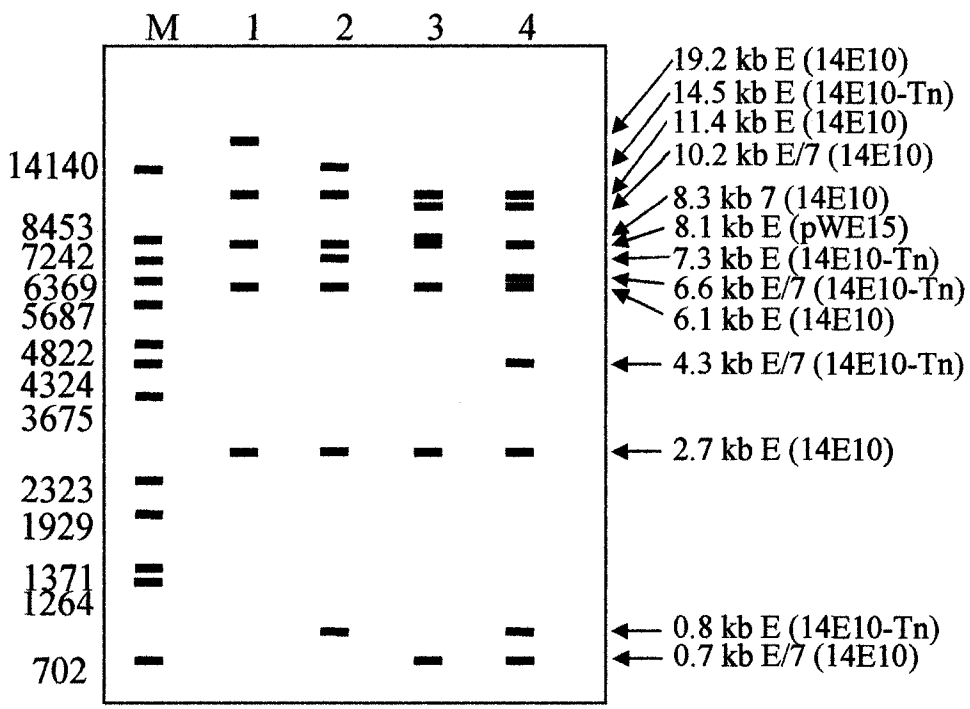


Figure III.8. The Tn*5062* transposon used for in vitro transposon mutagenesis of 14E10. It is composed of two flanking Tn*5* mosaic ends (ME), a three-frame translational stop, a *Streptomyces* ribosome binding site (RBS) preceding a promoterless gene encoding enhanced green fluorescent protein (*egfp*), an apramycin resistance gene [*aac3(IV)*] flanked by transcriptional terminators, and an RK2 *oriT* for inter-generic conjugation.





*The 0.1 kb E fragment (from 14E10) ran off the gel

Figure III.9. Localization of Tn*5062* in mutant 14E10 cosmids. Digestion of mutant cosmids with *Eco*RI followed by agarose gel electrophoresis revealed the location of Tn*5062* insertions. A. Restriction maps of 14E10 and a possible mutant 14E10 cosmid. B. Diagram of an agarose gel depicting the location of bands resulting from *Eco*RI digestion of 14E10 and the hypothetical mutant cosmid. Lane M, *Bst*EII-digested lambda DNA (sizes given in bp); lane 1, *Eco*RI-digested 14E10; lane 2, *Eco*RI-digested mutant 14E10; lane 3, *Eco*1722/*Eco*RI-digested 14E10; lane 4, *Eco*721/*Eco*RI-digested mutant 14E10.

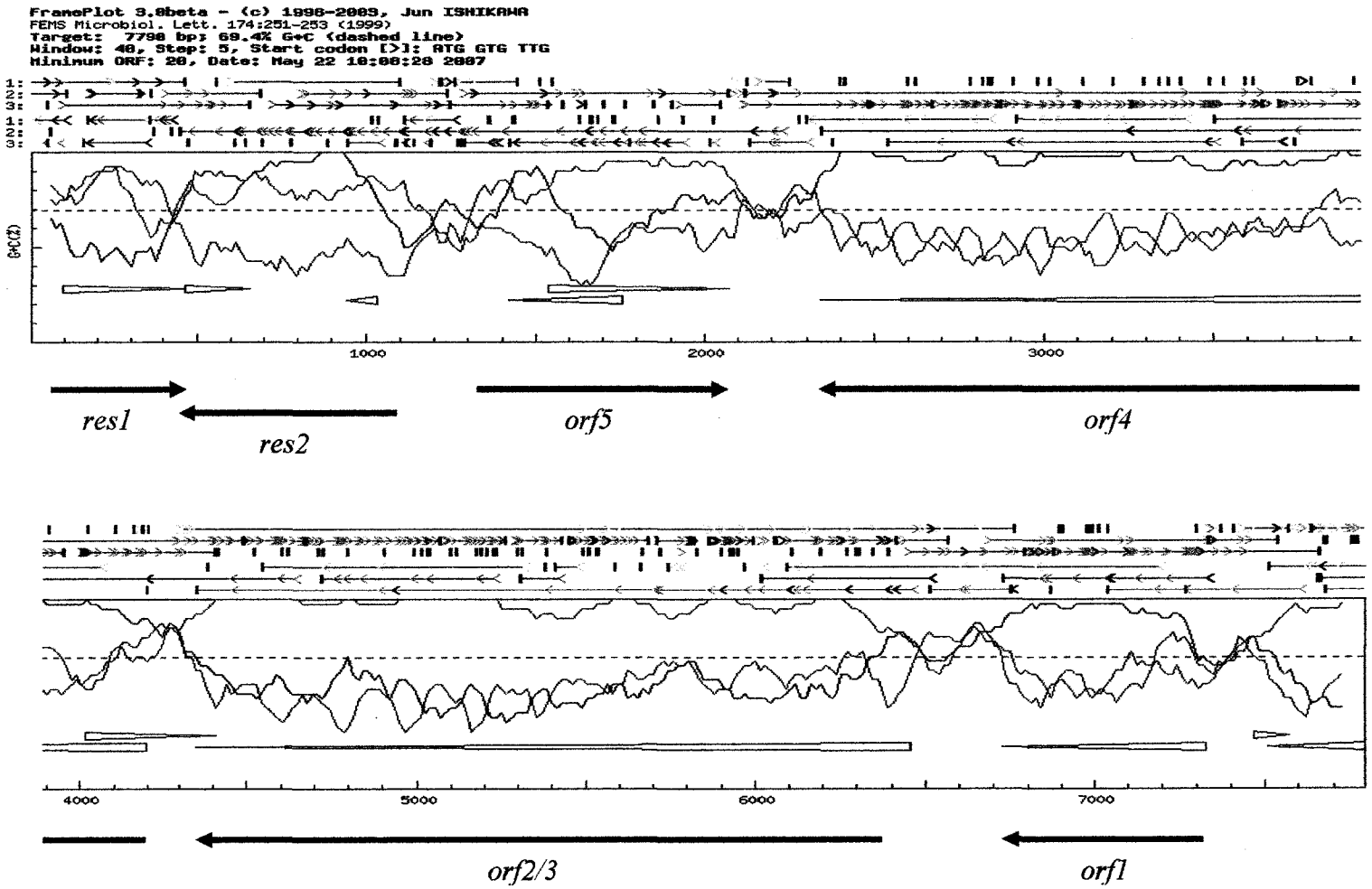


Figure III.10. FramePlot analysis of the region downstream of the paralogue gene cluster. Sequence data were

collected from mutants in 14E10 in which transposons had been inserted into the downstream area. Gaps in the sequence were filled using custom primers and 14E10 as a template.

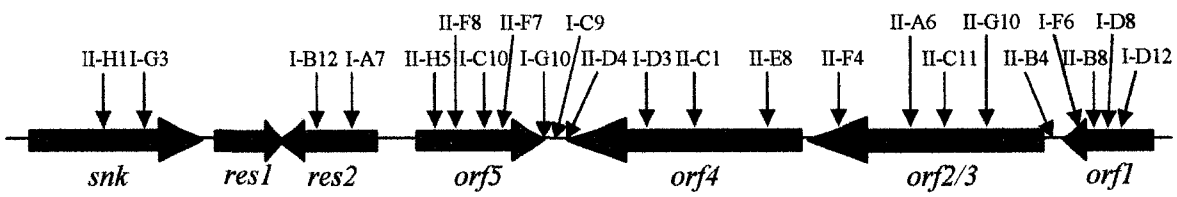
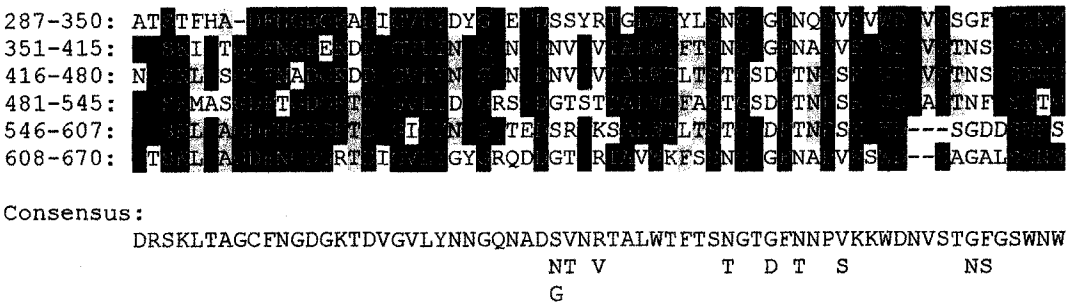


Figure III.11. The ORFs located downstream of the paralogue gene cluster and the location of Tn*5062* insertions.

(A) Orf2/3:



(B) Orf4:

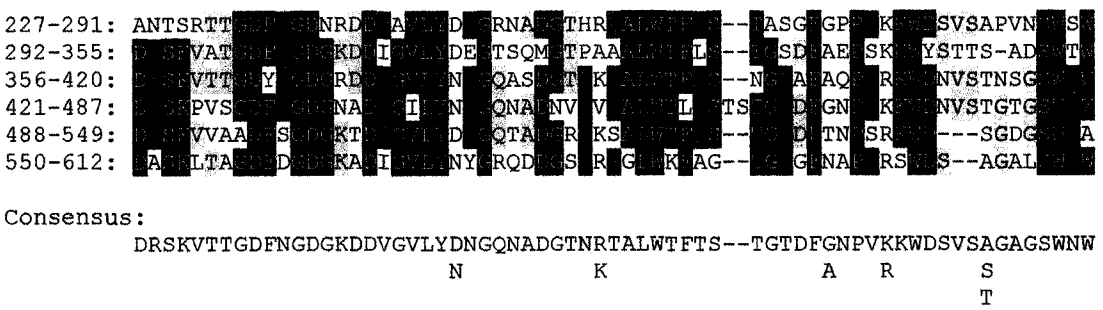


Figure III.12. Repeat amino acid sequences found in Orf2/3 and Orf4. (A) Alignment of the repeated sequence from Orf2/3. Shown below the alignment is the consensus sequence generated from the alignment. (B) Alignment of the repeated sequence from Orf4. Numbers refer to the amino acid positions within either protein. Shown below the alignment is the consensus sequence generated using Jalview (Clamp et al. 2004).

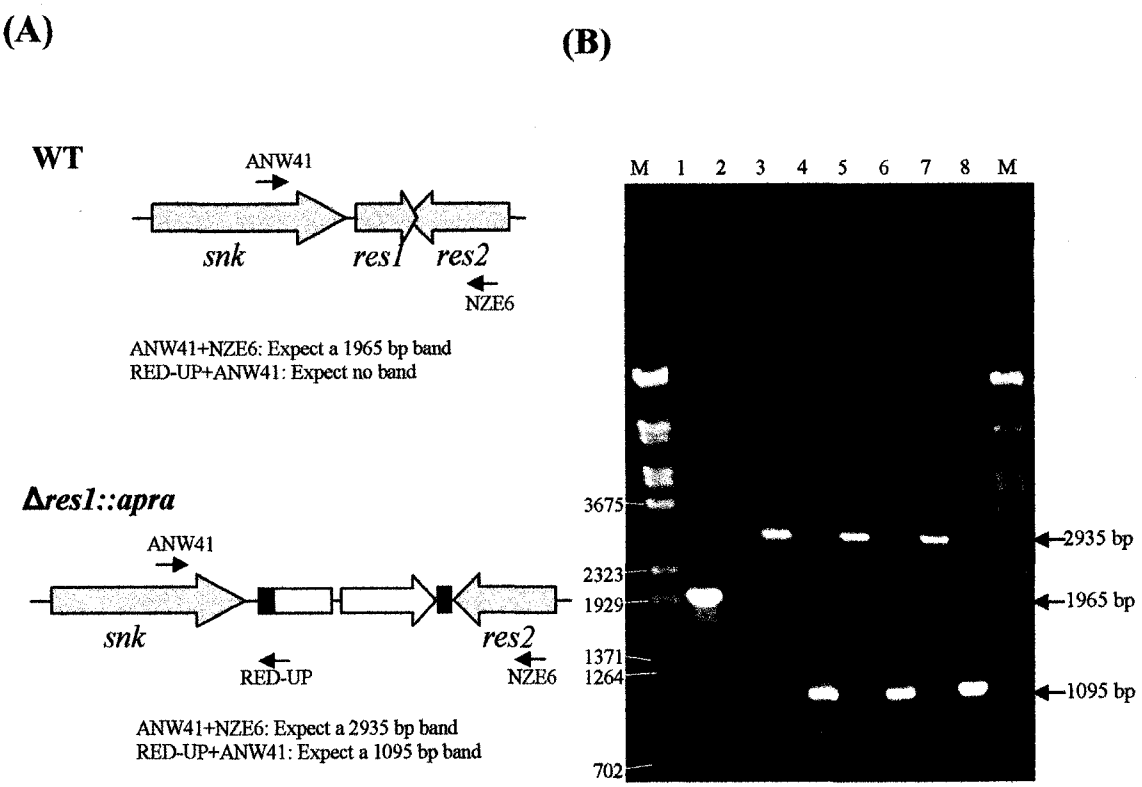


Figure III.13. Expected (A) and observed (B) mycelial PCR products from wild type and *res1* mutants. Lanes M, λ*Bst*EII marker. Lanes 1, 3, 5, and 7, reactions containing primers ANW41 and NZE6. Lanes 2, 4, 6, and 8, reactions containing primers RED-UP and ANW41. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*res1::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*res1::apra* 2 mycelium. Lanes 7 and 8, reactions containing Δ*res1::apra* 3 mycelium.

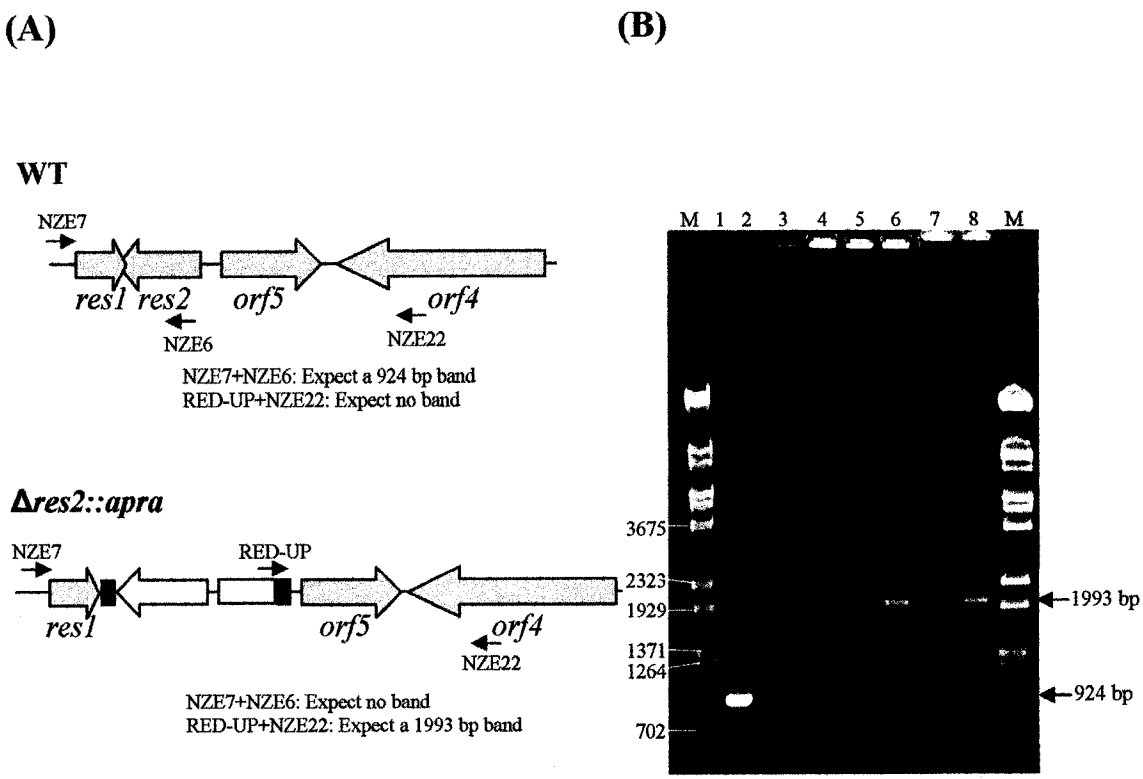


Figure III.14. Expected (A) and observed (B) mycelial PCR products from wild type and *res2* mutants. Lanes M, λ*Bst*EII marker. Lanes 1, 3, 5, and 7, reactions containing primers NZE7 and NZE6. Lanes 2, 4, 6, and 8, reactions containing primers RED-UP and NZE22. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*res2::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*res2::apra* 2 mycelium. Lanes 6 and 7, reactions containing Δ*res2::apra* 3 mycelium.

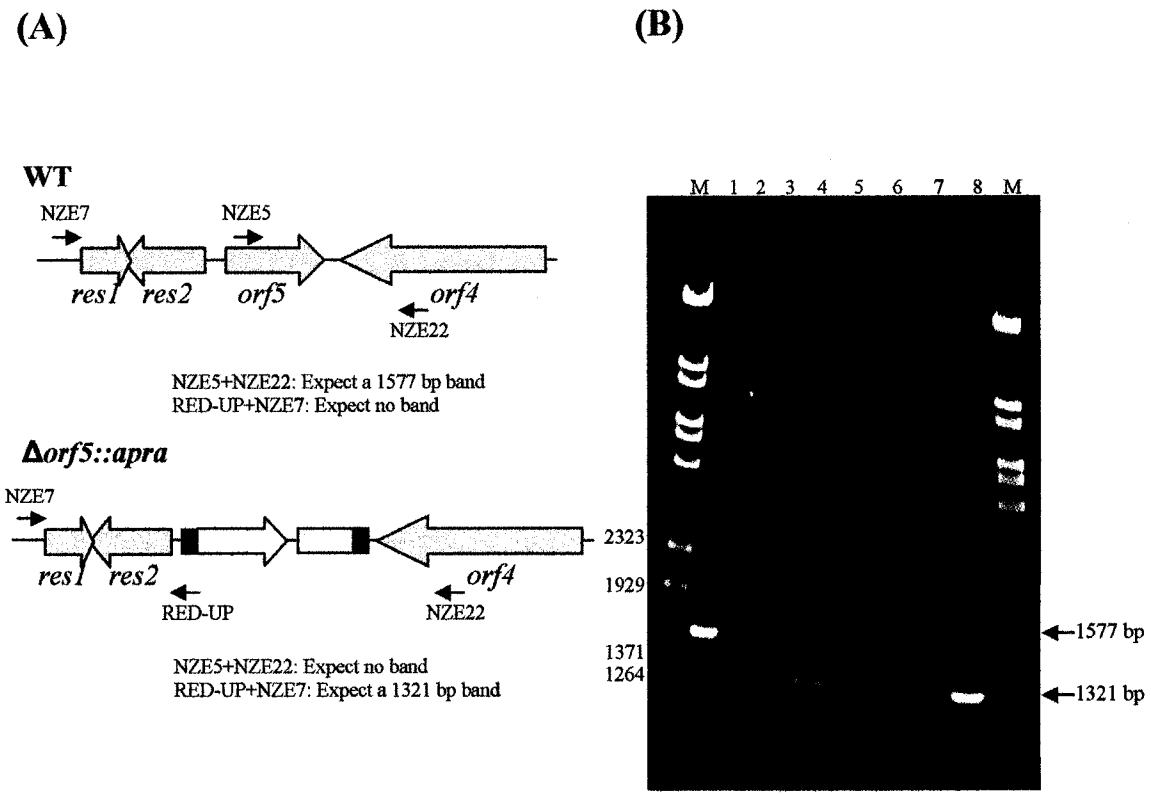


Figure III.15. Expected (A) and observed (B) mycelial PCR products from wild type and *orf5* mutants. Lanes M, λ*Bst*EII marker. Lanes 1, 3, 5, and 7, reactions containing primers NZE5 and NZE22. Lanes 2, 4, 6, and 8, reactions containing primers RED-UP and NZE7. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*orf5::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*orf5::apra* 2 mycelium. Lanes 7 and 8, reactions containing Δ*orf5::apra* 3 mycelium.

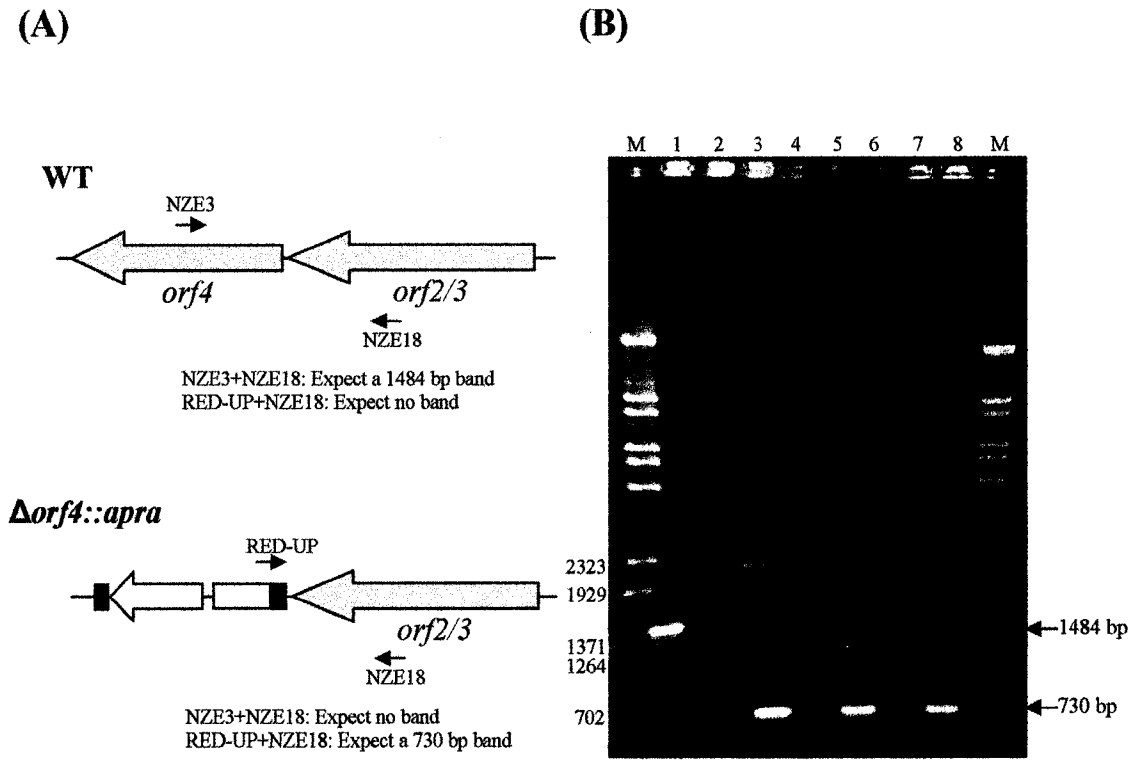


Figure III.16. Expected (A) and observed (B) mycelial PCR products from wild type and *orf4* mutants. Lanes M, λ*Bst*EII marker. Lanes 1, 3, 5, and 7, reactions containing primers NZE3 and NZE18. Lanes 2, 4, 6, and 8, reactions containing primers RED-UP and NZE18. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*orf4::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*orf4::apra* 2 mycelium. Lanes 7 and 8, reactions containing Δ*orf4::apra* 3 mycelium.

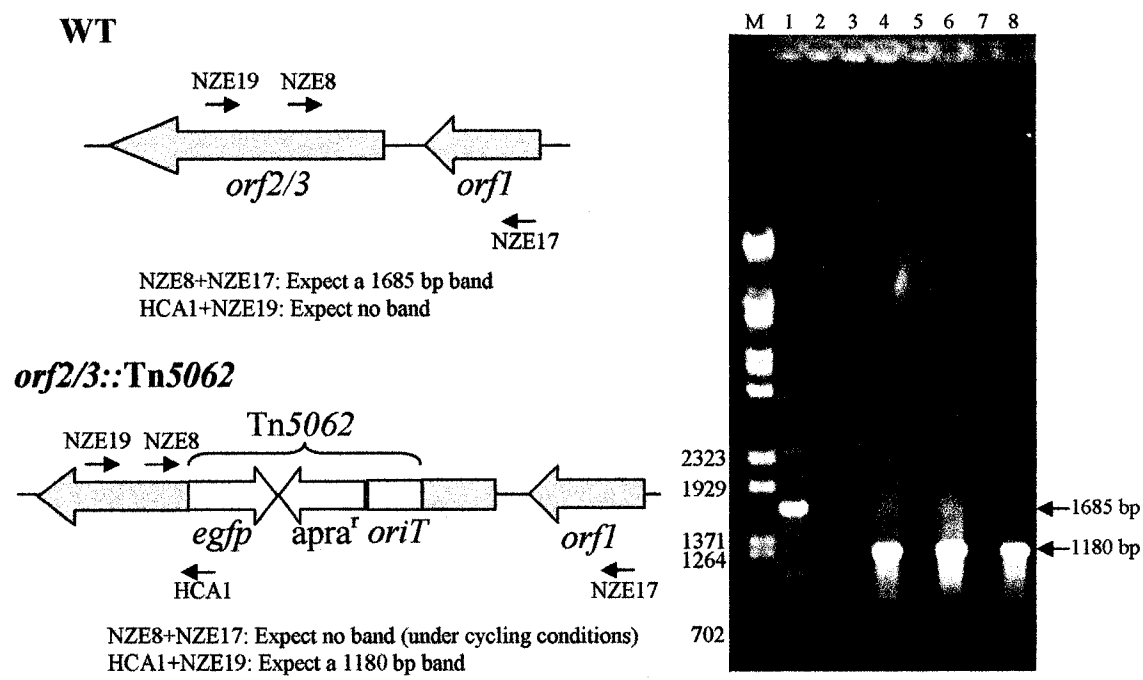


Figure III.17. Expected (A) and observed (B) mycelial PCR products from wild type and *orf2/3* mutants. Lane M, λ*Bst*EII, marker. Lanes 1, 3, 5, and 7, reactions containing primers NZE8 and NZE17. Lanes 2, 4, 6, and 8, reactions containing primers HCA1 and NZE19. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing *orf2/3::*Tn*5062* 1 mycelium. Lanes 5 and 6, reactions containing *orf2/3::*Tn*5062* 2 mycelium. Lanes 7 and 8, reactions containing *orf2/3::*Tn*5062* 3 mycelium.

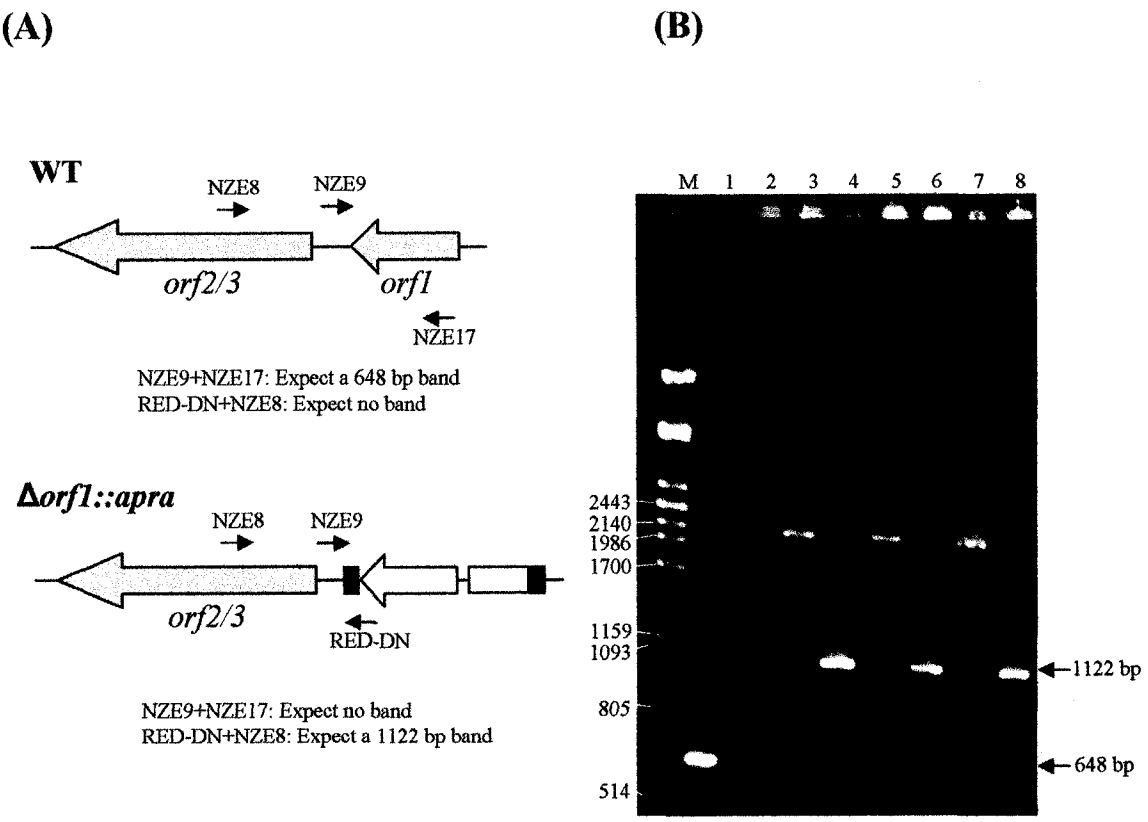


Figure III.18. Expected (A) and observed (B) mycelial PCR products from wild type and *orf1* mutants. Lane M, λ*Bst*EII marker. Lanes 1, 3, 5, and 7, reactions containing primers NZE9 and NZE17. Lanes 2, 4, 6, and 8, reactions containing primers RED-DN and NZE8. Lanes 1 and 2, reactions containing wild type mycelium. Lanes 3 and 4, reactions containing Δ*orf1::apra* 1 mycelium. Lanes 5 and 6, reactions containing Δ*orf1::apra* 2 mycelium. Lanes 7 and 8, reactions containing Δ*orf1::apra* 3 mycelium.

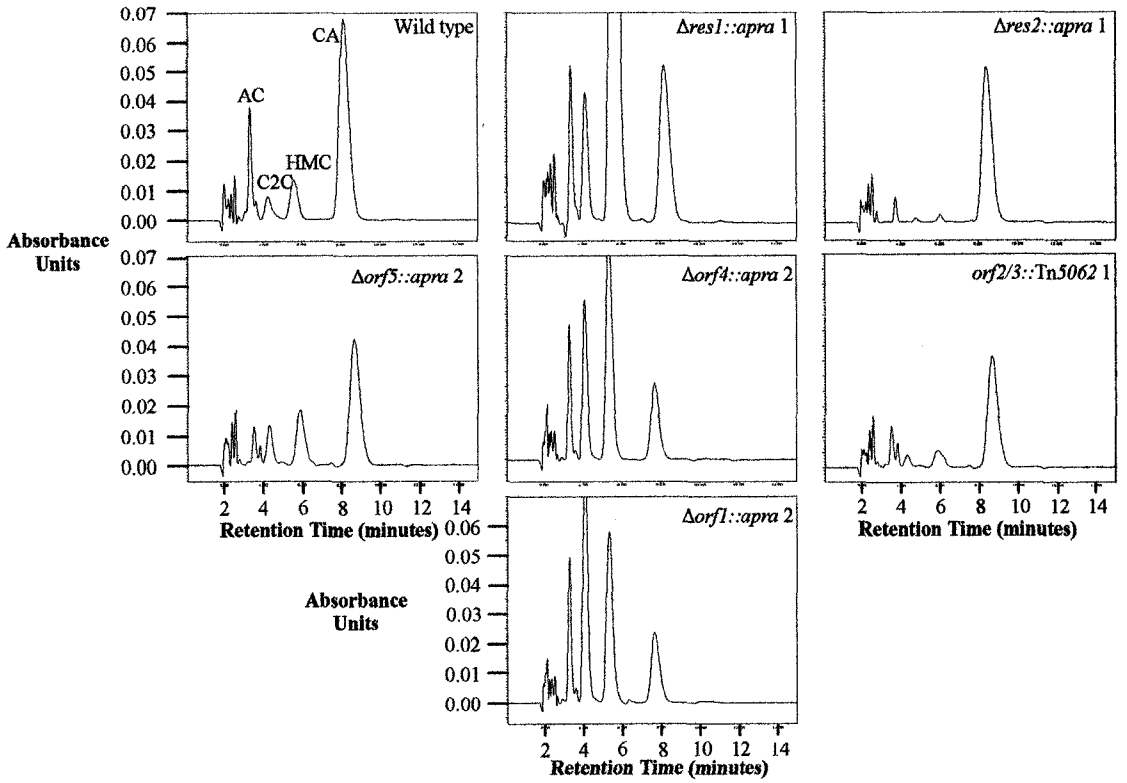


Figure III.19. Clavam production profiles of *orf1*, *orf2/3*, *orf4*, *orf5*, *res2* and *res1* mutants. Culture filtrates from 96 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting underivatized samples from derivatized samples. All samples were diluted (1/5) before analysis. AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

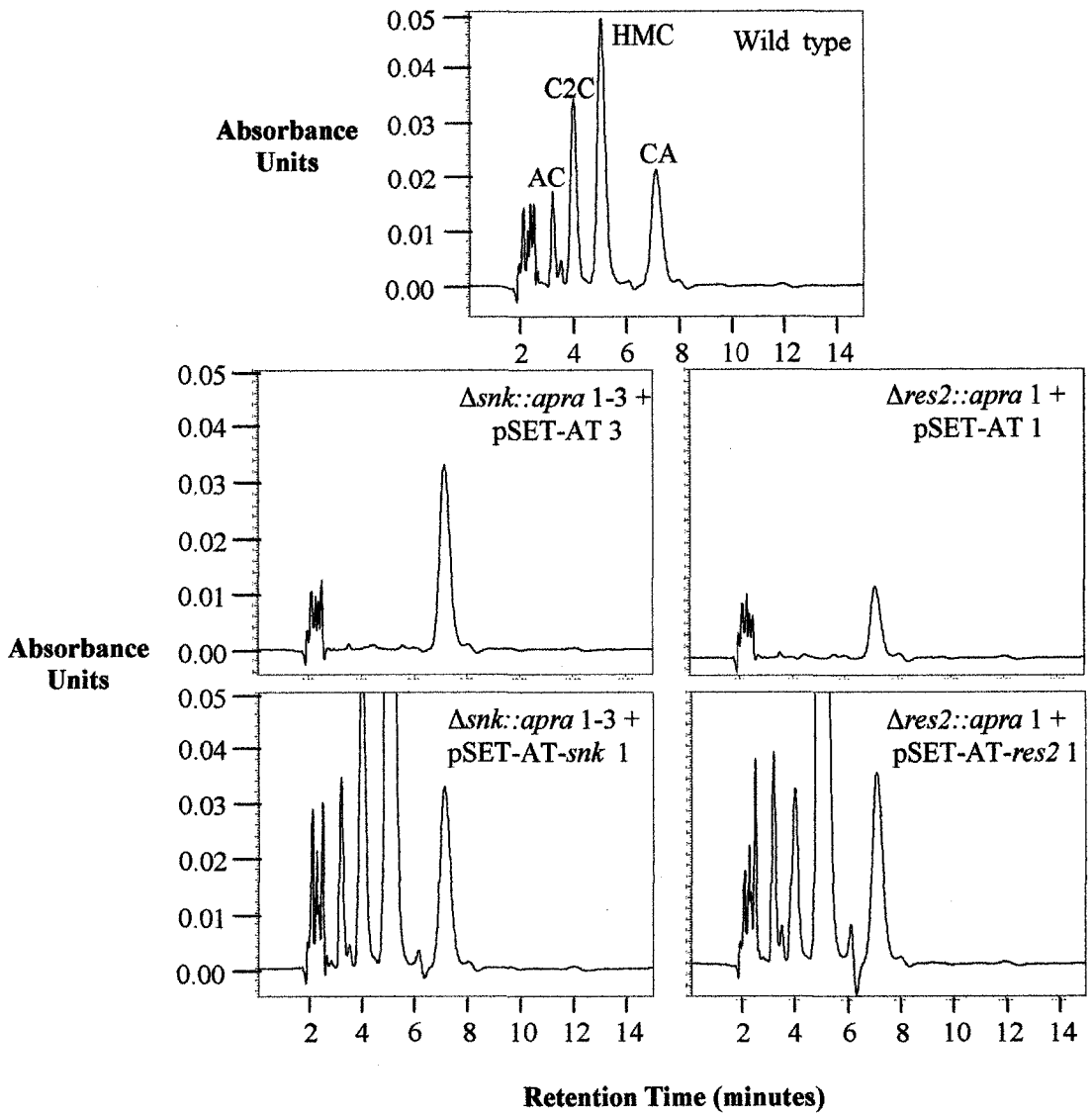


Figure III.20. Clavam production profiles of complemented *snk* and *res2* mutants. Culture filtrates from 96 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting peaks from underivatized samples' chromatograms from derivatized samples' chromatograms. All samples were diluted (1/5) before analysis. AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

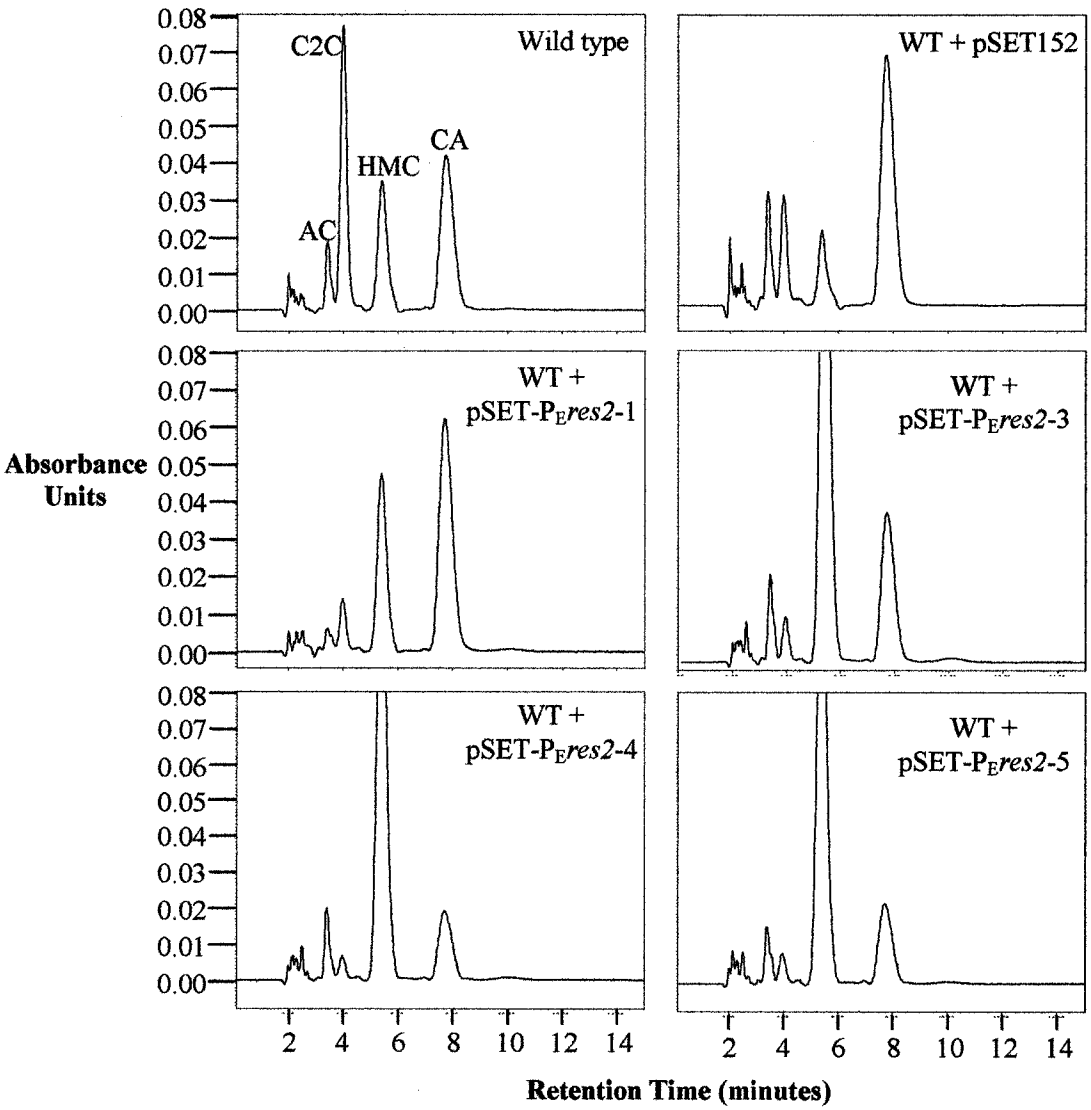


Figure III.21. Clavam production profiles of wild type *S. clavuligerus* carrying pSET-P$_E$*res2*. Culture filtrates from 120 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting underivatized samples from derivatized samples. All samples were diluted (1/5) before analysis. AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

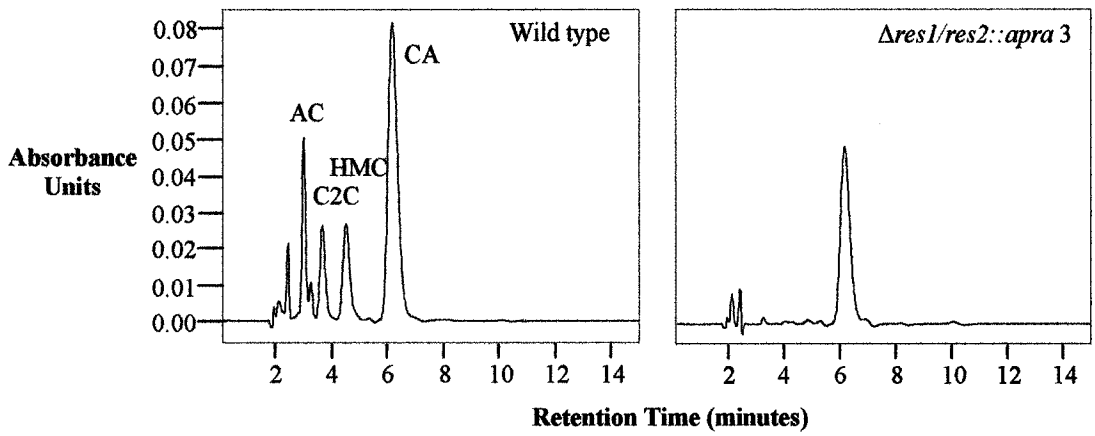


Figure III.22. Clavam production profiles of a *res1/res2* double mutant. Culture filtrates from 96 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting underivatized samples from derivatized samples. All samples were diluted (1/5) before analysis. AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

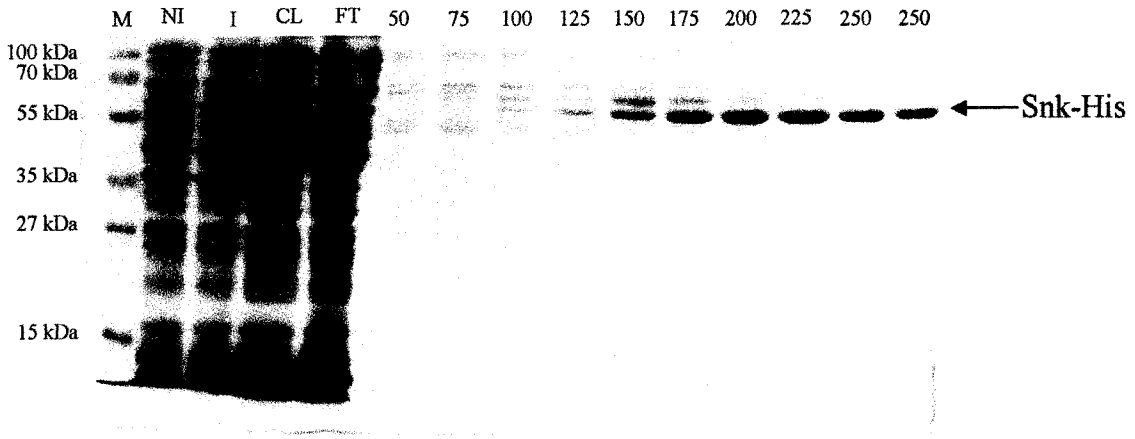


Figure III.23. Production and purification of Snk-His from *E. coli* BL21(DE3) carrying pT7-7His-*snk*. Protein samples were taken during different stages of the overexpression and purification processes and then analyzed via SDS-PAGE. Proteins were visualized by staining the gel with PhastGel Blue R stain. Lane M, PageRuler™ Protein Ladder Plus molecular weight marker. Lane NI, non-induced total protein sample. Lane I, induced total protein sample. Lane CL, cleared lysate. Lane FT, flow-through. Numbers above lanes indicate imidazole concentrations (in mM) in which the protein samples had been eluted.

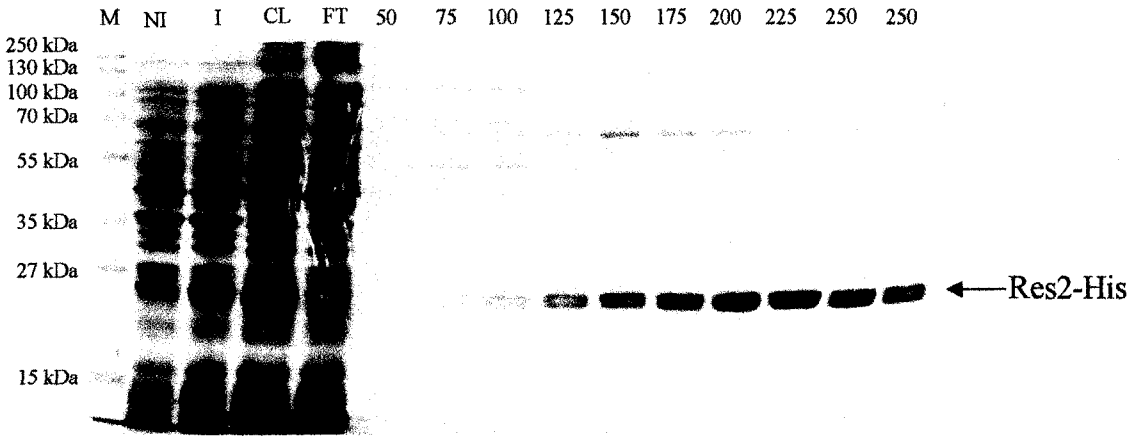


Figure III.24. Production and purification of Res2-His from *E. coli* BL21(DE3) carrying pT7-7His-*res2*. Protein samples were taken during different stages of the overexpression and purification processes and then analyzed via SDS-PAGE. Proteins were visualized by staining the gel with PhastGel Blue R blue. Lane M, PageRuler™ Protein Ladder Plus molecular weight marker. Lane NI, non-induced total protein sample. Lane I, induced total protein sample. Lane CL, cleared lysate. Lane FT, flow-through. Numbers above lanes indicate imidazole concentrations (in mM) in which the protein samples had been eluted.

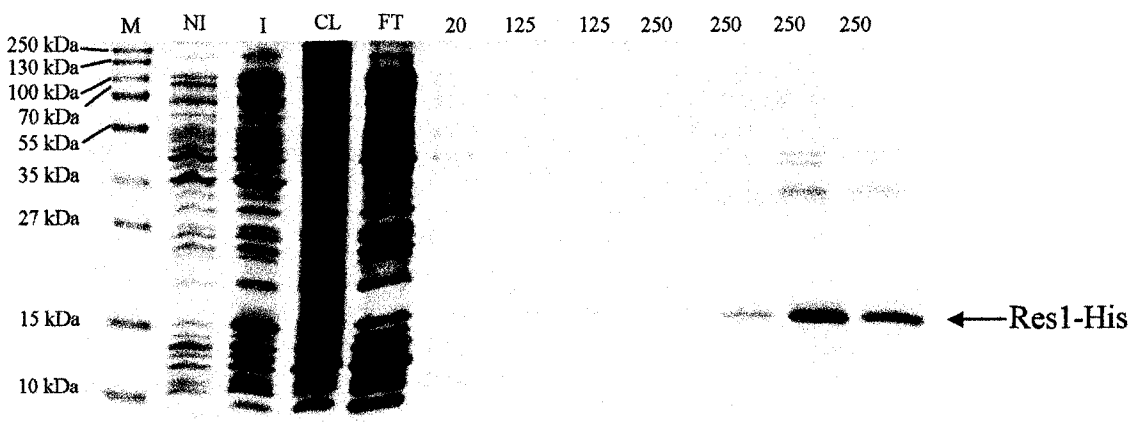


Figure III.25. Production and purification of Res1-His from *E. coli* BL21(DE3) carrying pET-19b-*res1*. Protein samples were taken during different stages of the overexpression and purification processes and then analyzed via SDS-PAGE. Proteins were visualized by staining the gel with PhastGel Blue R blue. Lane M, PageRuler™ Protein Ladder Plus molecular weight marker. Lane NI, non-induced total protein sample. Lane I, induced total protein sample. Lane CL, cleared lysate. Lane FT, flow-through. Numbers above lanes indicate imidazole concentrations (in mM) in which the protein samples had been eluted.

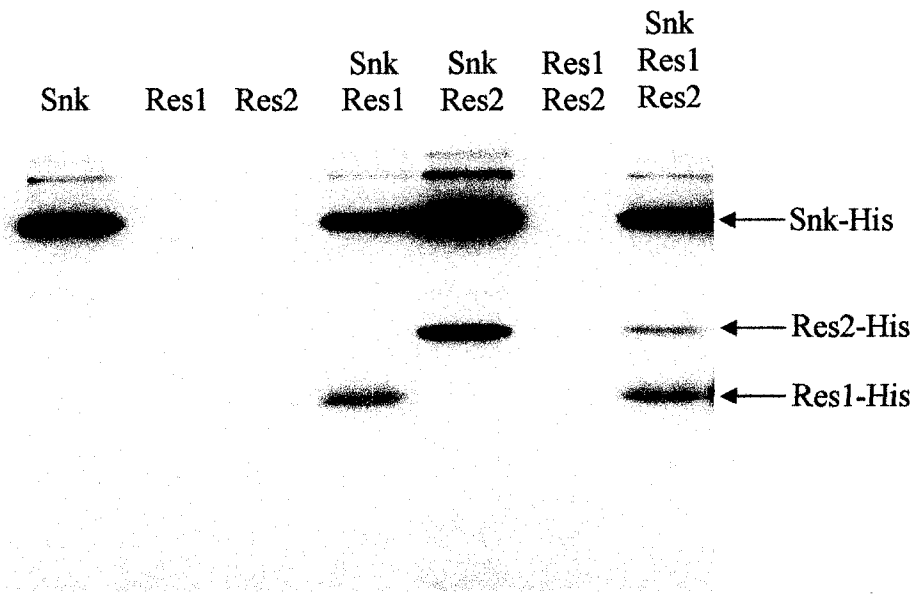


Figure III.26. In vitro phosphorylation of Snk-His, Res1-His and Res2-His. Snk-His, Res1-His and Res2-His were incubated in different combinations with [γ-^{32}P]ATP at 30°C for 30 minutes. In all reactions except that involving all three proteins, 1μM Snk-His was used and 3μM of Res1-His and Res2-His were used. The reaction involving the three proteins contained 1 μM Snk-His and 1.5 μM Res1-His and Res2-His. After 30 minutes, the reactions were cooled on ice, combined with 5x SDS-PAGE loading buffer, and then analyzed via 15% (v/v) SDS-PAGE. The gel was dried and exposed to a phosphor screen overnight.

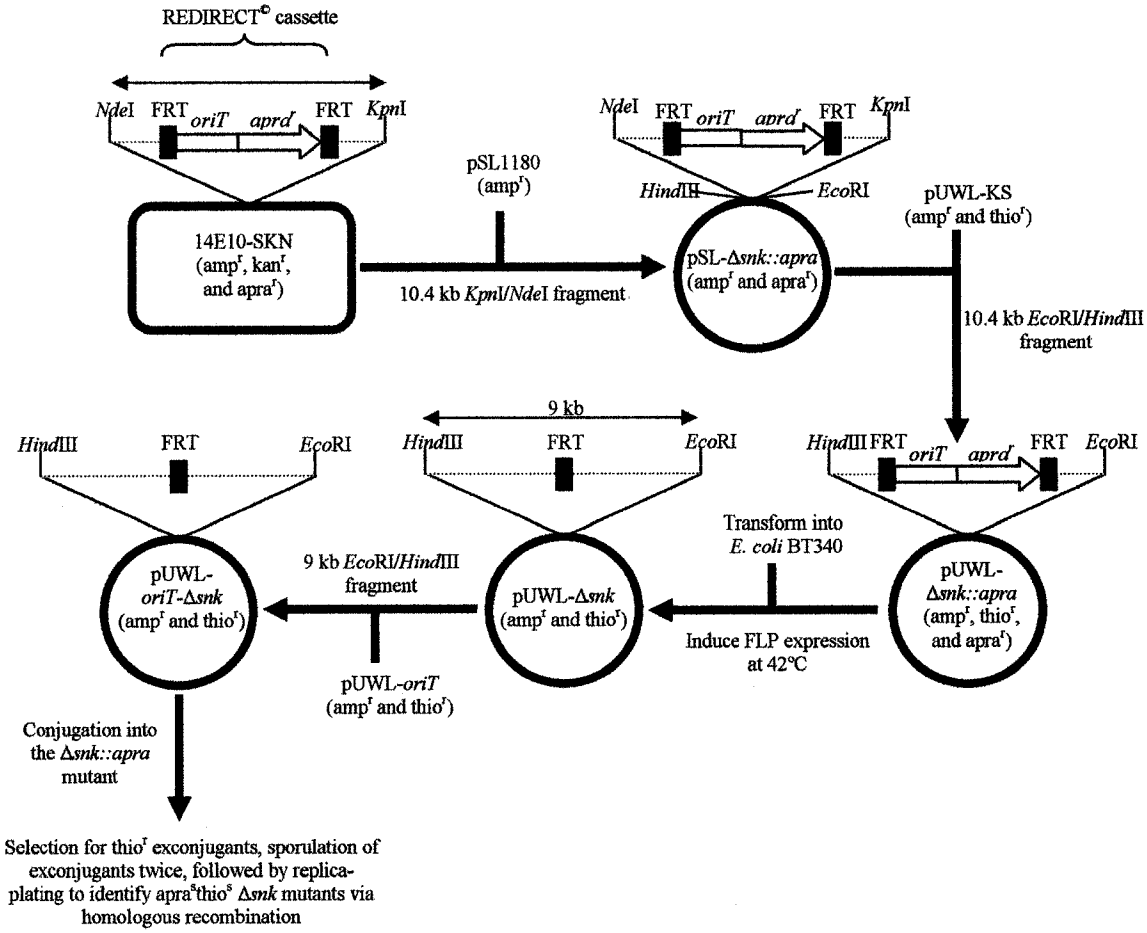


Figure III.27. The procedure used to generate an in-frame, unmarked deletion of *snk*. Refer to the text for more detailed descriptions of all steps.

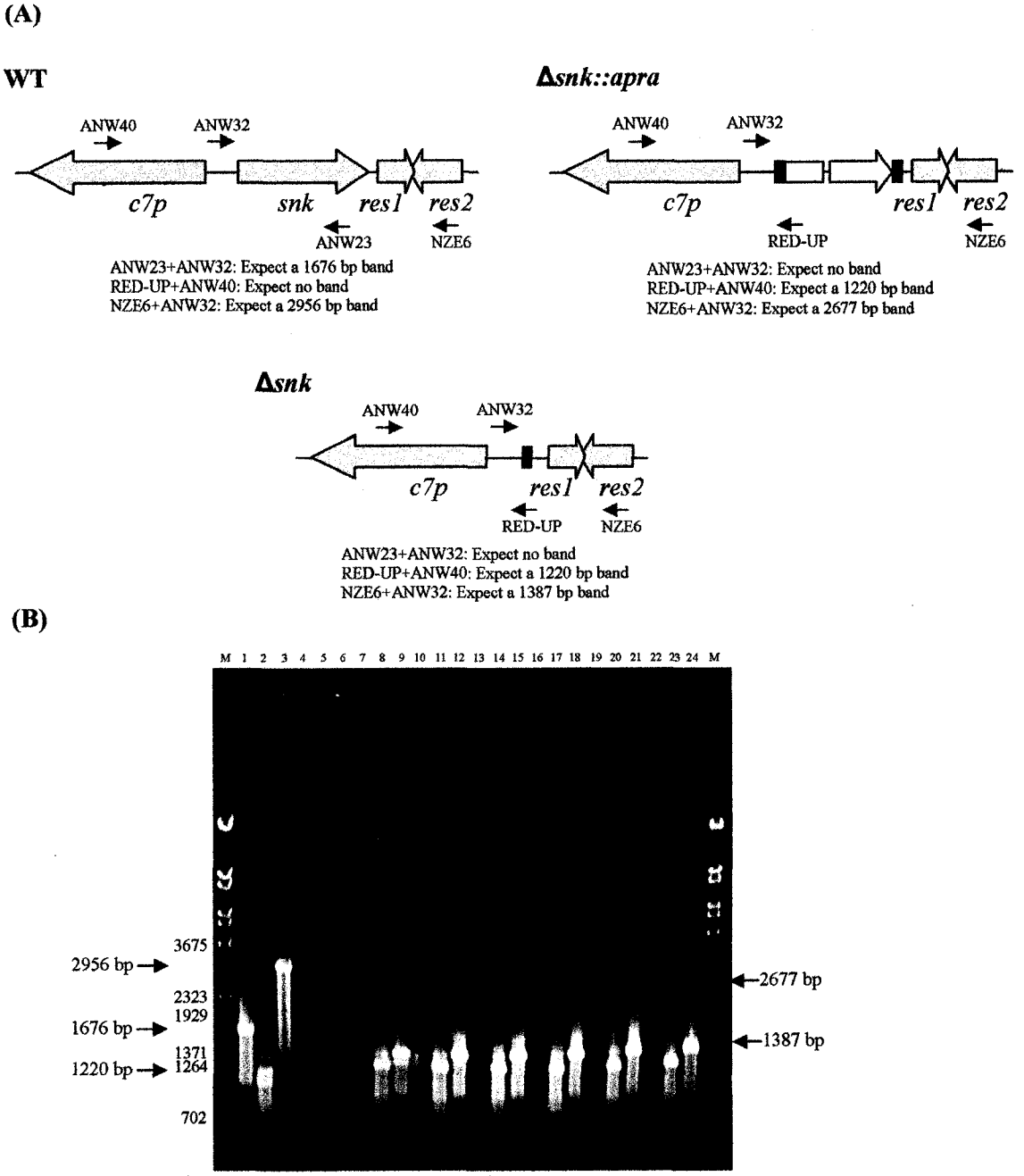


Figure III.28. Expected (A) and observed (B) mycelial PCR products from wild type and Δ*snk* mutants. Lanes M, λ*Bst*EII marker. Lanes 1, 4, 7, 10, 13, 16, 19, and 22, reactions containing primers ANW23 and ANW32. Lanes 2, 5, 8, 11, 14, 17, 20, and 23,

reactions containing primers RED-UP and ANW40. Lanes 3, 6, 9, 12, 15, 18, 21, and 24, reactions containing primers NZE6 and ANW32. Lanes 1, 2, and 3, reactions containing wild type mycelium. Lanes 4, 5, and 6, reactions containing Δ*snk::apra* 1-3 mycelium. Lanes 7, 8, and 9, reactions containing Δ*snk* 2-1 mycelium. Lanes 10, 11, and 12, reactions containing Δ*snk* 2-2 mycelium. Lanes 13, 14, and 15, reactions containing Δ*snk* 2-3 mycelium. Lanes 16, 17, and 18, reactions containing Δ*snk* 2-4 mycelium. Lanes 19, 20, and 21, reactions containing Δ*snk* 3-1 mycelium. Lanes 22, 23, and 24, reactions containing Δ*snk* 3-2 mycelium.

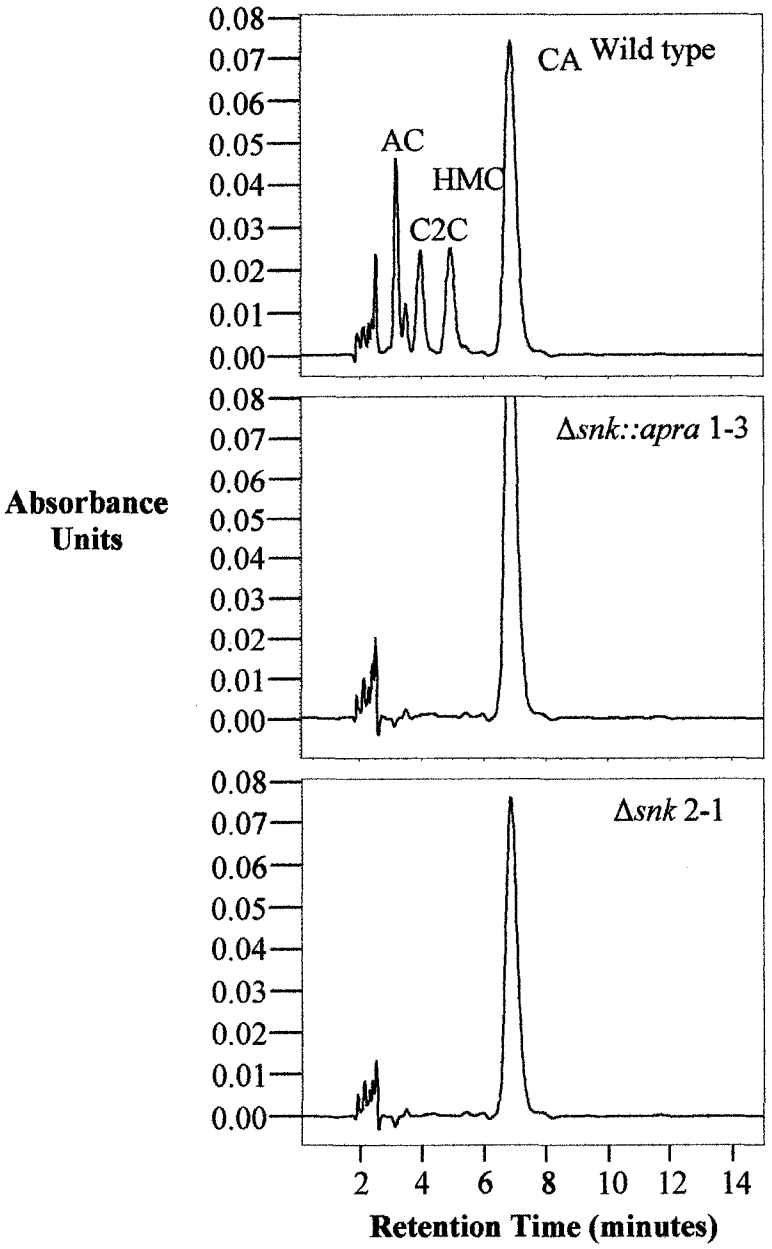


Figure III.29. Clavam production profiles of a Δ*snk* mutant. Culture filtrates from 96 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting underivatized samples from derivatized samples. All samples were diluted (1/5) before analysis. AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

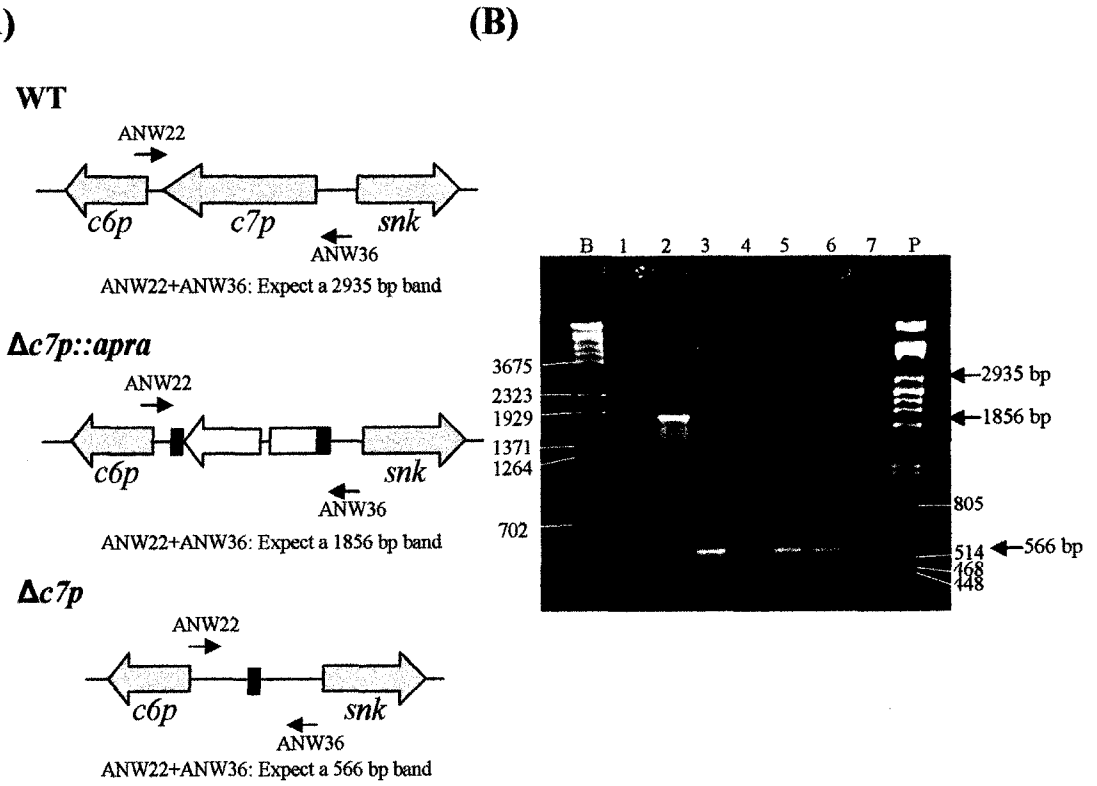


Figure III.30. Expected (A) and observed (B) mycelial PCR products from wild type and Δ*c7p* mutants. Lane B, λ*Bst*EII marker. Lane P, λ*Pst*I marker. Lanes 1, 2, 3, 4, 5, 6, and 7, reactions containing primers ANW22 and ANW36. Lane 1, reaction containing wild type mycelium. Lane 2, reaction containing Δ*c7p::apra* 1-1 mycelium. Lane 3, reaction containing Δ*c7p* 1-1 mycelium. Lane 4, reaction containing Δ*c7p* 1-2 mycelium. Lane 5, reaction containing Δ*c7p* 1-3 mycelium. Lane 6, reaction containing Δ*c7p* 2-1 mycelium. Lane 7, reaction containing Δ*c7p* 2-2 mycelium.

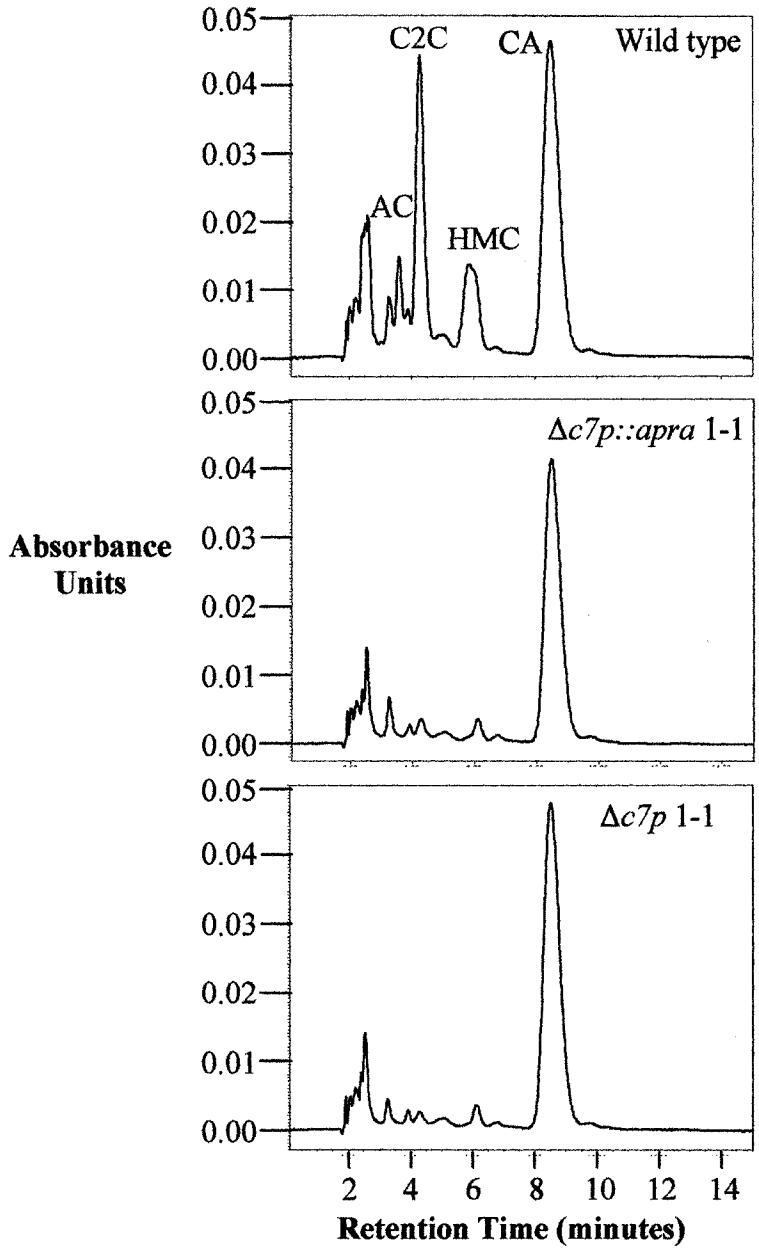


Figure III.31. Clavam production profiles of a Δ*c7p* mutant. Culture filtrates from 96 hour soy cultures were derivatized with imidazole and analyzed via HPLC. Chromatograms were obtained by subtracting underivatized samples from derivatized samples. All samples were diluted (1/5) before analysis. AC: alanylclavam; C2C: clavam-2-carboxylate; HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

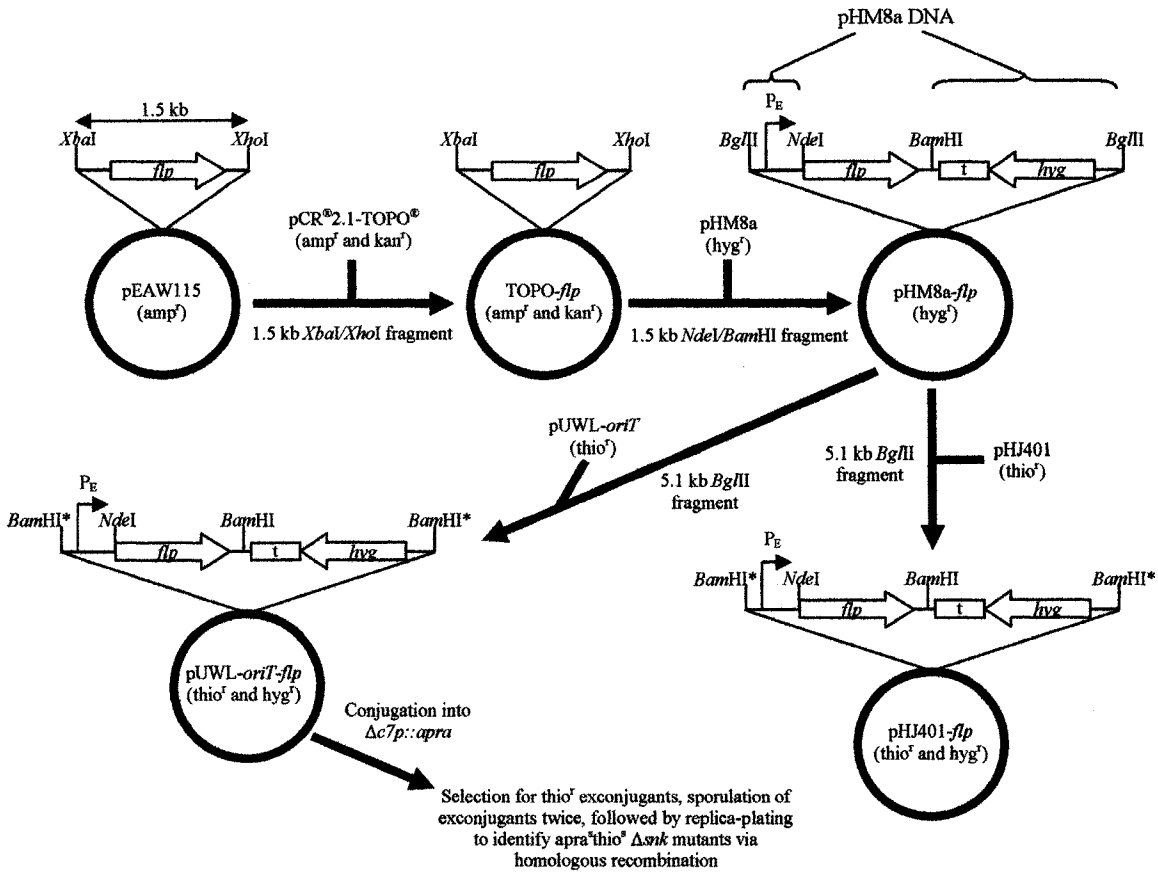


Figure III.32. The procedure used to generate an in-frame, unmarked deletion of *c7p*. Refer to the text for more detailed descriptions of all steps. Restriction sites marked with an asterisk have been altered and are no longer of any use.

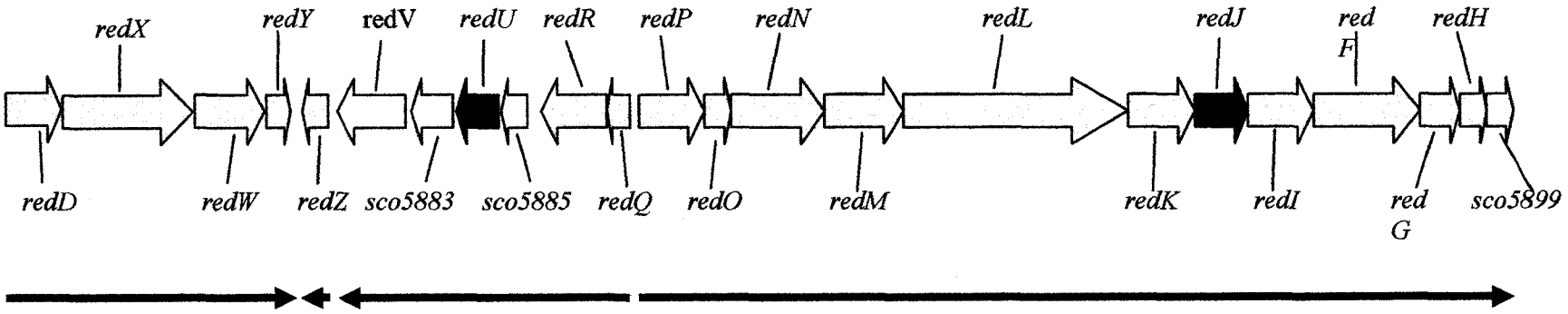


Figure III.33. The *red* gene cluster of *S. coelicolor*. Genes involved in undecylprodigiosin biosynthesis are gray, putative housekeeping genes are black, and genes with unknown functions are white. Arrows below the cluster indicate putative transcripts of the cluster. Adapted from Cerdeno et al. (2001).

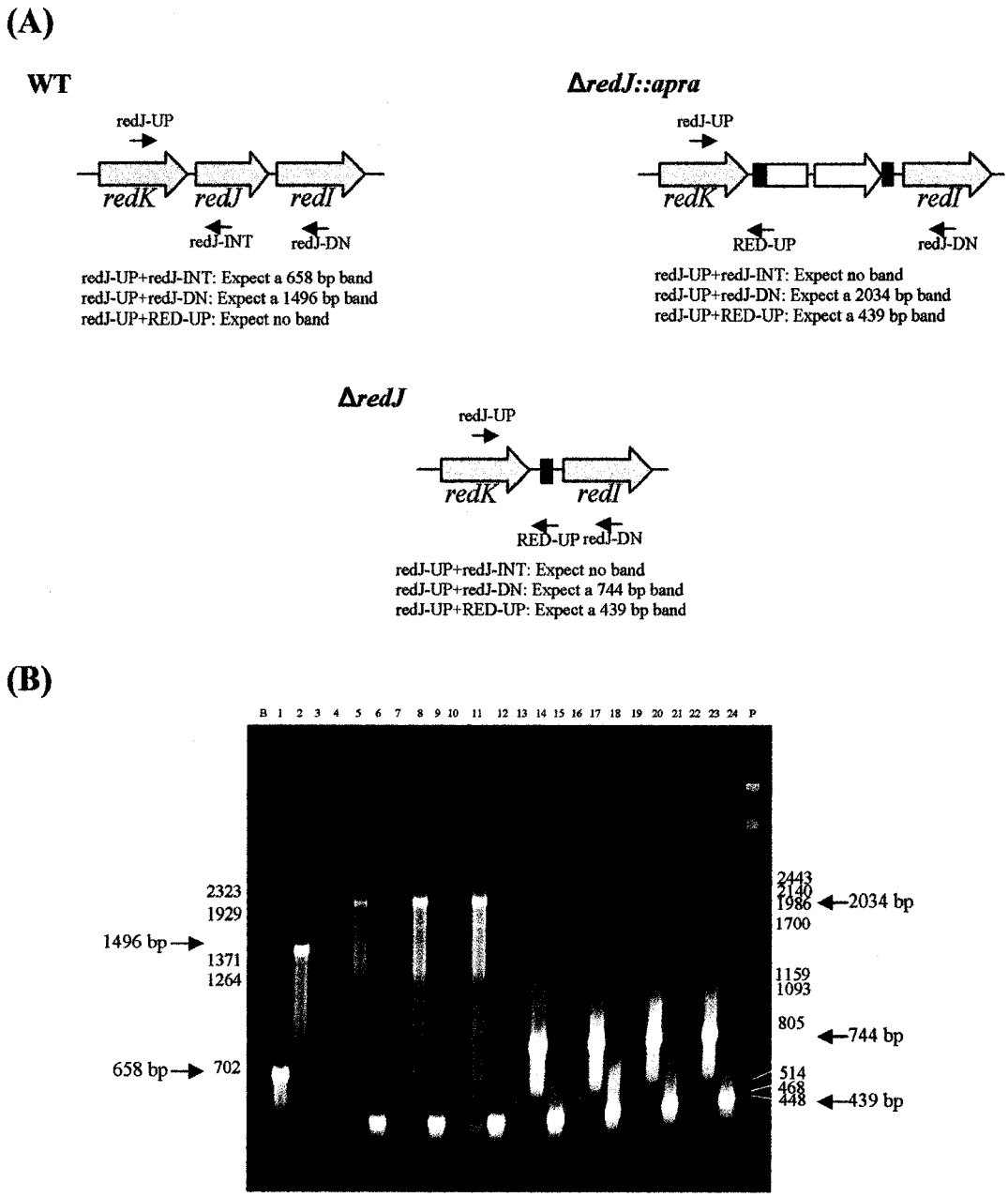


Figure III.34. Expected (A) and observed (B) PCR products using genomic DNA isolated from 48 hour R5 cultures of *S. coelicolor* M145, Δ*redJ::apra* mutants, and Δ*redJ* mutants. Lane B, λ*Bst*EII marker. Lane P, λ*Pst*I marker. Lanes 1, 4, 7, 10, 13, 16, 19, and 22, reactions containing primers redJ-UP and redJ-INT. Lanes 2, 5, 8, 11, 14, 17,

20, and 23, reactions containing primers redJ-UP and redJ-DN. Lanes 3, 6, 9, 12, 15, 18, 21, and 24, reactions containing primers redJ-UP and RED-UP. Lanes 1, 2, and 3, reactions containing wild type mycelium. Lanes 4, 5, and 6, reactions containing Δ*redJ::apra* 2 mycelium. Lanes 7, 8, and 9, reactions containing Δ*redJ::apra* 4 mycelium. Lanes 10, 11, and 12, reactions containing Δ*redJ::apra* 6 mycelium. Lanes 13, 14, and 15, reactions containing Δ*redJ* 1-4 mycelium. Lanes 16, 17, and 18, reactions containing Δ*redJ* 1-5 mycelium. Lanes 19, 20, and 21, reactions containing Δ*redJ* 1-8 mycelium. Lanes 22, 23, and 24, reactions containing Δ*redJ* 2-1 mycelium.

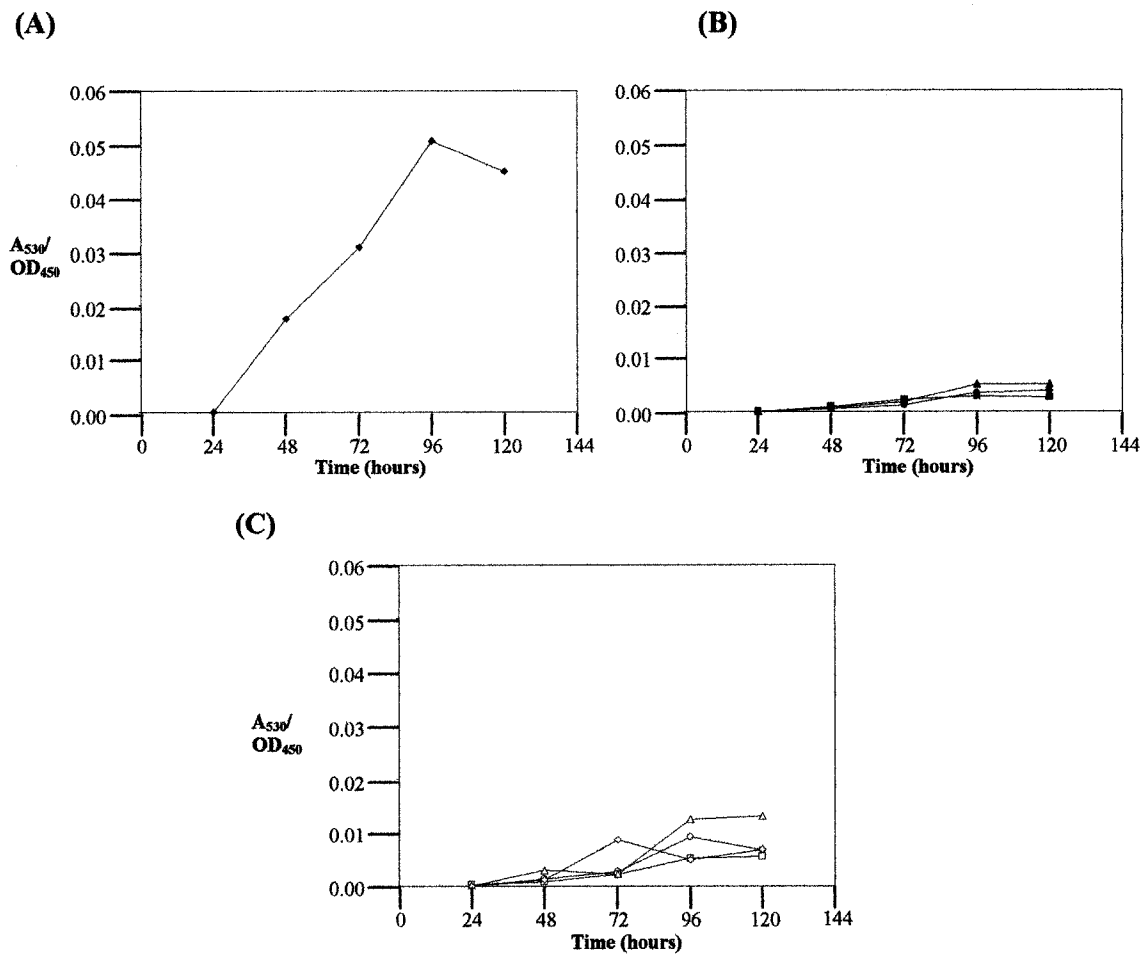


Figure III.35. Undecylprodigiosin production by *S.coelicolor redJ* mutants. (A) Undecylprodigiosin production by a wild type culture (closed diamonds). (B) Undecylprodigiosin production by Δ*redJ::apra* 2 (closed circle),Δ*redJ::apra* 4 (closed square), and Δ*redJ::apra* 6 (closed triangle) mutants. (C) Undecylprodigiosin production by Δ*redJ* 1-4 (open circle), Δ*redJ* 1-5 (open square), Δ*redJ* 1-8 (open triangle), and Δ*redJ* 2-1 (open diamond). All cultures were grown in 25 mL R5 medium supplemented with 5% PEG in spring flasks at 28°C with shaking at 250 rpm.

IV. DISCUSSION

IV.1 The upstream region of the paralogue gene cluster

Twelve genes lie upstream of *ceaS1* in the paralogue gene cluster (Figure I.6). Prior work indicated that *orfA*, *orfB*, *orfC*, and *orfD* may have roles in alanylclavam biosynthesis while the remaining genes were shown to be unimportant for clavam production or remained unexamined (Cai 2003). This study confirmed that *orfA*, *orfB*, *orfC*, and *orfD* are required for alanylclavam biosynthesis and that mutants lacking *orfC* or *orfD* produce an extra clavam metabolite. This novel clavam may represent a biosynthetic intermediate in the pathway leading to alanylclavam. Purification and structural determination of this clavam will shed further light on alanylclavam biosynthesis.

In an earlier study, transposon mutagenesis of *orfF*, *orfG*, and *orfH* was not possible (Cai 2003). Although the genes could be disrupted in cosmids carrying *S. clavuligerus* genomic DNA through in vitro transposon mutagenesis, introduction of the mutant cosmids into *S. clavuligerus* cells did not result in the incorporation of the mutation via homologous recombination. It was originally hypothesized that this indicated that *orfF*, *orfG*, and *orfH* are essential genes, required for basic survival functions (Cai 2003). However, these genes were mutated in the present study, thus demonstrating their non-essential nature. Interestingly, similar problems were observed when trying to introduce mutant cosmids into *S. clavuligerus* carrying the transposon in genes found in the downstream region of the paralogue gene cluster (Section III.2.3). The reason behind this difficulty in introducing Tn*5062* mutations into the *S. clavuligerus* chromosome remains unknown, but demonstrates that while in vitro transposon

mutagenesis offers a simple method to sequence fairly large segments of DNA, it may not be a viable option for producing complete sets of *S. clavuligerus* mutants.

Currently, cosmids carrying *S. coelicolor* genomic DNA are being subjected to in vitro transposon mutagenesis with Tn*5062* (Bishop et al. 2004). This work is being done to eventually generate a library of mutant cosmids that carries a mutation in every *S. coelicolor* gene. Despite the difficulties with introducing Tn*5062* mutations into *S. clavuligerus*, such mutations can be introduced into *S. coelicolor* with apparent ease (Bishop et al. 2004; Hoskisson et al. 2006). A number of studies have used such Tn*5*-based transposons without problems to identify genes involved in morphological differentiation, antibiotic production, and antibiotic resistance in *S. coelicolor* (Gehring et al. 2000; Gehring et al. 2001; Gehring et al. 2004; Sprusansky et al. 2003).

While *orfA*, *orfB*, *orfC*, and *orfD* mutants cannot produce alanylclavam, *orfE*, *orfF*, *orfG*, and *orfH* mutants can still produce all of the clavams. As well, Cai (2003) found that mutants in the remaining genes upstream of the paralogue gene cluster (*orfI*, *orfJ*, *orfK*, and *orfL*) were capable of producing clavams. Thus, a small cluster of four genes immediately upstream of *ceaS1* is involved in alanylclavam biosynthesis. However, *orfF*, *orfG*, and *orfH* mutants showed varying trends in clavam production such that roles in clavam biosynthesis or extracellular transport cannot be ruled out (Table III.2).

IV.1.1 A possible *orfC-orfD* operon

Initially, the production of the novel clavam in both *orfC* and *orfD* mutants was somewhat confounding. Normally, it is expected that the loss of one gene will result in

the accumulation of an intermediate upon which the encoded protein acts. The appearance of the novel clavam in both mutants appears to indicate that the novel clavam is acted upon by both OrfC and OrfD. This is difficult to imagine; both proteins are somehow involved in alanylclavam biosynthesis, so it is not likely that they catalyze essential but different conversions of the exact same intermediate. It is possible that they act in a protein complex that is unstable and nonfunctional if one or the other protein is absent from the complex; in this way, loss of either gene would result in the accumulation of an intermediate acted upon first by OrfC or OrfD in the alanylclavam biosynthetic pathway. However, the existence of an *orfC-orfD* operon seemed to be a far more likely possibility.

The stop codon of *orfC* and the start codon of *orfD* overlap. This is another reason that it seems likely that they are transcribed together in an operon. If the genes are expressed in a single transcript, then an insertional mutation in *orfC* would likely abolish the expression of *orfD*. Whether or not *orfC* is involved in alanylclavam biosynthesis would be masked by the effect of negating *orfD* expression in *orfC* mutants. This means that a polar *orfC* mutant would have the same apparent phenotype as an *orfD* mutant, as is observed. Since the gene downstream of *orfD*, *orfE*, is transcribed in the opposite direction and is not involved in clavam biosynthesis, the phenotype of an *orfD* mutant is probably authentically due to the absence of *orfD* (Figure I.6).

Complementation of the *orfC* mutant with an extra copy of *orfC* led to the partial restoration of alanylclavam biosynthesis with the continued production of the novel clavam (Section III.1.3). While this indicates that *orfC* is indeed required for alanylclavam biosynthesis, it also shows that *orfD* must undergo at least some expression

in the insertional *orfC* mutant. This indicates that *orfD* may have its own promoter. Since Tn*5062* contains two T_4 transcriptional terminators, it does not seem likely that the *orfC* promoter can direct transcription through the transposon to allow *orfD* expression in the *orfC* mutant (Figure III.8). Thus, it is possible that both *orfC* and *orfD* have their own promoters. Looking at the clavam production of the *orfC* mutant compared with that of the complemented mutant, the *orfC* promoter may actually direct both *orfC* and *orfD* transcription, but *orfD* may have its own, weaker promoter. This may explain why so little alanylclavam is produced by the complemented *orfC* mutant; while *orfC* may be provided, *orfD* is being expressed from its own, weaker promoter. With limiting levels of OrfD, alanylclavam will not be produced at wild type concentrations. The locations of promoters in an *orfC-orfD* operon could be determined by carrying out primer extension or S1 nuclease experiments with wild type samples of RNA.

IV.1.2 A putative alanylclavam biosynthetic pathway

It is currently unknown exactly where alanylclavam-specific biosynthesis branches off from the biosynthetic pathway of the rest of the 5*S* clavams (Figure I.5). However, since at least some of the genes required specifically for alanylclavam production have been identified, it is possible to envision how its biosynthetic pathway proceeds.

orfA is most similar to serine and glycine hydroxymethyltransferases (Table III.1). These are pyridoxal phosphate-dependent enzymes that catalyze the reversible interconversion of glycine and 5,10-methylenetetrahydrofolate to serine and tetrahydrofolate (Figure IV.1). This reaction is important in amino acid metabolism,

notably during the breakdown of threonine, glycine, and serine to pyruvate. When comparing the amino acid sequence of OrfA to those of other serine hydroxymethyltransferases, some interesting features become apparent. Of the ten amino acid residues involved in binding pyridoxal phosphate in *E. coli* serine hydroxymethyltransferase, eight are found in OrfA; as well, three of the four amino acids that bind to glycine in *E. coli* serine hydroxymethyltransferase are present in OrfA (Scarsdale et al. 2000; Figure IV.2). This indicates that OrfA may be capable of binding both pyridoxal phosphate and glycine. However, OrfA lacks all but one of the six amino acid residues required for 5,10-methylenetetrahydrofolate binding in *E. coli* serine hydroxymethyltransferase (Scarsdale et al. 2000; Figure IV.2). Such similarities and differences between the proteins suggest that OrfA uses pyridoxal phosphate as a cofactor and catalyzes a reaction between glycine and a compound other than 5,10-methylenetetrahydrofolate.

OrfB shows similarity to proteins of the YjgF/YER057c/UK114 family. While a common function of proteins within this family has not yet been found, a number of possible functions have been proposed for members of the family. One member of the family, isolated from rat liver cells, exhibits ribonuclease activity (Morishta et al. 1999). In *Saccharomyces cerevisiae*, a homologous protein is involved in isoleucine biosynthesis and in maintenance of the mitochondrial genome (Kim et al. 2001; Oxelmark et al. 2000). In *Bacillus subtilis*, the YabJ protein (a member of the YjgF/YER057c/UK114 family) appears to be involved in the repression of purine biosynthesis, possibly by interacting with a DNA-binding repressor protein (Rappu et al. 1999).

Initial work showed that YjgF from *Salmonella typhimurium* is required for isoleucine biosynthesis (Enos-Berlage et al. 1998). Further work demonstrated that *S. typhimurium* YjgF is required for normal IlvB (an aminotransferase directly involved in isoleucine synthesis) activity (Schmitz and Downs 2004). As well, a member of the YjgF/YER057c/UK114 family from *Haemophilus influenzae* is capable of binding to α-keto acids, such as 2-ketobutyrate, the first intermediate of isoleucine biosynthesis (Parsons et al. 2003). This led to the formulation of a model in which YjgF binds to an α-keto acid metabolite, possibly an intermediate in the isoleucine biosynthetic pathway such as 2-ketobutyrate, that normally represses IlvB aminotransferase activity (Schmitz and Downs 2004). By binding to the α-keto acid, YjgF sequesters the metabolite and keeps it from interacting with the aminotransferase; this allows isoleucine production to continue (Schmitz and Downs 2004).

Comparison of the OrfB amino acid sequence to those of other YjgF/YER057c/UK114 family proteins reveals that it contains all of the residues implicated in ligand-binding (Figure IV.3). These amino acids have been identified from crystal structures of the *E. coli* and the human YjgF/YER057c/UK114 family proteins (Misiniene et al. 2005; Volz 1999). Because both OrfB and the *H. influenzae* homologue carry these amino acid residues, it is likely that OrfB is capable of binding to α-keto acids as well (Parsons et al. 2003).

OrfC shows similarity to aminotransferases. Aminotransferases are pyridoxal phosphate-requiring enzymes that catalyze the transfer of amino groups from amino acids to oxo acids such that the ketone oxygen is replaced by an amino group (Figure IV.1). Analysis of the OrfC amino acid sequence reveals that it belongs to family I of the

aminotransferases by virtue of containing ten of the 11 invariant amino acid residues found in the family (Mehta et al. 1993; Jensen and Gu 1996). Family I includes aspartate, alanine, and aromatic amino acid aminotransferases (Mehta et al. 1993). The residue that appears to be missing from OrfC, R386 in the pig aminotransferase sequence, may be present but does not align perfectly with other sequences (Figure IV.4). This residue is found in all aminotransferases and it has been noted that this arginine residue in different aminotransferases does not always align precisely (Jensen and Gu 1996). Thus, R364 in OrfC may be equivalent to R386 found in all aminotransferases (Figure IV.4).

While OrfC contains eight of the 12 residues involved in pyridoxal phosphate binding, it lacks four of the six residues involved in substrate binding (Figure IV.4). The aminotransferases used in the alignment are all family I members; the *E. coli* protein and the pig protein use aspartate as the amino group donor, while the *Paracoccus denitrificans* protein uses aromatic amino acids (such as phenylalanine and tyrosine) as donors (Kamitori et al. 1990; Okamoto et al. 1998; Rhee et al. 1997; Figure IV.4). It appears that OrfC may bind to unusual substrates while still employing pyridoxal phosphate as a cofactor.

OrfD is most similar to threonine and serine dehydratases. Threonine dehydratases (also called threonine deaminases) are enzymes that catalyze the pyridoxal phosphate-dependent dehydration and deamination of threonine or serine to yield 2-ketobutyrate or pyruvate, respectively (Figure IV.1). The amino acid sequence of OrfD reveals that it contains ten of the 11 residues implicated in pyridoxal phosphate binding and all seven of the residues likely involved in substrate (threonine or serine) binding (Figure IV.5). This indicates that OrfD uses pyridoxal phosphate to catalyze a very

similar reaction to threonine dehydratases that involves a substrate that is probably structurally analogous to threonine or serine.

Examination of the types of reactions catalyzed by proteins similar to OrfA, OrfC, and OrfD (Figure IV.1) allows a putative alanylclavam biosynthetic pathway to be imagined. It is possible that alanylclavam-specific biosynthesis may branch off from the rest of the clavam pathway at 2-hydroxymethylclavam (Figure IV.6). From there, OrfA may use glycine as a substrate and attach it to the methylene group bound to –OH (Figure IV.6). This would result in the formation of a clavam metabolite with serine attached to C2 [Intermediate Clavam 1 (IC1)]. OrfD may then catalyze the dehydration/deamination of this intermediate to produce a clavam with pyruvate bound to C2 [Intermediate Clavam 2 (IC2); Figure IV.6]. Alanylclavam would then be finally produced by the transfer of an amino group in place of the keto oxygen through the action of OrfC (Figure IV.6). The role of OrfB in this pathway could be regulatory in nature; OrfB may bind IC2 (an α-keto acid) to keep it from inhibiting the activity of OrfC. As described above, OrfB shows similarity to proteins capable of binding α-keto acids that may decrease aminotransferase activity.

This type of pathway could also help explain the phenotypes observed with *orfC* mutants, *orfD* mutants, and complemented *orfC* mutants. As described earlier, *orfC* and *orfD* may both be transcribed from a strong promoter upstream of *orfC* and *orfD* may also be transcribed from a promoter upstream of *orfD* and within the *orfC* sequence (Section IV.1.1; Figure IV.7A). When *orfC* was insertionally mutated, the transcription through the upstream promoter would have been interrupted such that *orfD* could only be transcribed from its own promoter (Figure IV.7B). If the *orfD*-specific promoter is weak,

there would only be a limiting amount of OrfD; this would result in the accumulation of more IC1 than IC2 (Figure IV.6). Depending on the amount produced and the stability of IC2, it may not even be detectable through normal means, such as HPLC. Introduction of another copy of *orfC* in an *orfC* mutant would allow alanylclavam to be produced again, but, because *orfD* is still transcribed from its own weak promoter, little alanylclavam would be produced while IC1 would still accumulate in abundance (Figure IV.7C). In this case, there are limiting amounts of OrfD such that very little IC1 is converted to IC2, which is further converted into alanylclavam. In this hypothetical scheme, IC1 is the novel clavam observed in *orfC* and *orfD* mutant culture filtrates described earlier (Section III.1.2).

The authenticity of this proposed alanylclavam biosynthetic pathway could be confirmed through labeling and feeding experiments and identification of the novel clavam produced by *orfC* and *orfD* mutants. To confirm that 2-hydroxymethylclavam is produced prior to alanylclavam, mutants in biosynthetic genes unable to produce 2-hydroxymethylclavam (such as *cvm1*, *cvm2*, *cvm5*, or *c6p* mutants) could be fed 2-hydroxymethylclavam and then tested for alanylclavam biosynthesis. Radioactively labeled glycine (possibly using an isotope for one of the carboxylic acid atoms) could be fed to cultures and isolated alanylclavam could then be assayed for radioactivity. Purification and identification of the novel clavam would also provide a significant hint for how the pathway proceeds from reactions catalyzed by OrfC and OrfD.

Once the above experiments are carried out, it will be desirable to purify OrfA, OrfB, OrfC, and OrfD and to use them for in vitro experiments with different clavam

metabolites and other substrates to determine the exact reactions carried out by these proteins.

Another interesting aspect of alanylclavam biosynthesis is that it appears to be regulated by *bldA*. *S. clavuligerus bldA* mutants are capable of producing all of the clavams except alanylclavam (S. E. Jensen, personal communication). This indicates that *bldA* regulates the translation of OrfA, OrfB, OrfC, or OrfD, unless there are other, as of yet unknown, proteins required for alanylclavam production. It is unlikely that *bldA* is involved in either OrfC or OrfD translation, as *bldA* mutants did not accumulate a novel clavam. As well, it is most probable that *bldA* is required for OrfA translation since *orfA* actually contains a single TTA codon. It will be possible to determine whether or not OrfA translation is regulated by *bldA* by using OrfA-specific antibodies in western analysis of *bldA* mutants.

IV.2 The downstream region of the paralogue gene cluster

Transposon mutagenesis of cosmid 14E10 allowed a 7.7 kb region downstream of the paralogue gene cluster to be sequenced, but was not particularly useful in producing mutations in ORFs in the region. Of the six ORFs found in the downstream area, only one (*orf2/3*) could be mutated by homologous recombination between the mutant cosmid and the *S. clavuligerus* chromosome. At least three of the ORFs found did not appear to be involved in clavam biosynthesis. Mutants in *orf5* (generated by PCR targeted mutagenesis), encoding a putative short-chain dehydrogenase, maintained the ability to produce all clavams (Sections III.2.2 and III.2.3.1). Likewise, mutants in *orf2/3* or *orf4* were still able to produce clavams; however, *orf4* mutants overproduced all of the 5*S*

clavams detectable by HPLC (Section III.2.3.1). Although Orf2/3 and Orf4 both contain highly similar amino acid repeats that may be implicated in the binding of Ca^{2+} and an unidentified ligand, the function of these putative proteins is unknown (Section III.2.2). However, the N-terminal portions of these proteins indicate that Orf2/3 and Orf4 are involved in peptidoglycan and carbohydrate metabolism, respectively (Section III.2.2). The exact role of Orf4 is somewhat difficult to imagine. Perhaps the absence of Orf4 affects the availability of different carbohydrates to cells; it is possible that the loss of Orf4 causes cells to be less able to obtain carbohydrates needed for optimal growth. Generally, cells produce secondary metabolites during times of slow growth, usually at the beginning of stationary phase in liquid cultures (Demain 2000). If *orf4* mutants experience an earlier stationary phase, or simply grow slower, than the wild type strain, then it could be expected that secondary metabolites (including antibiotics) will be produced in greater quantities by the mutant compared to wild type. Unfortunately, the growth patterns exhibited by mutants in liquid cultures were not examined during the course of this study. A problem with this hypothesis is that *orf4* mutants did not exhibit increased production levels of clavulanic acid (a secondary metabolite), as would be expected if loss of *orf4* had such global effects as described.

The function of *orf1* is even more difficult to discern; mutagenesis reveals that it is not required for clavam biosynthesis, and it shows limited similarity to a type III restriction enzyme over less than half of its length (Sections III.2.2 and III.2.3.1). Even though *orf5*, *orf2/3* and *orf1* would be interesting to examine, they are not involved in clavam biosynthesis and thus were not analyzed any further.

The two other ORFs found in the downstream region, *res1* and *res2*, were of far more interest. Mutants in *res1* generally overproduced 2-hydroxymethylclavam and alanylclavam and exhibited depressed clavam-2-carboxylate production. *res2* mutants were unable to produce any 5*S* clavams. An earlier study found that the gene upstream of *res1*, *snk*, is required for 5*S* clavam production like *res2* (Tahlan 2005; Figure I.6). Complementation of an *snk* mutant and a *res2* mutant restored their ability to produce 5*S* clavams (Section III.2.4). Because of the likelihood of *snk*, *res1*, and *res2* encoding an atypical two-component system (in this case, three, rather than two, protein components are present) involved in 5*S* clavam biosynthesis, these ORFs and the encoded proteins were subjected to additional scrutiny.

IV.2.1 The Snk/Res1/Res2 three-component system

Snk is a soluble sensor kinase that contains two tandem GAF domains at the N-terminus (Tahlan 2005; Figure IV.8). GAF domains are usually involved in the binding of small molecules such as cyclic nucleotides (cGMP and cAMP), heme, 2-ketoglutarate, formate, and Na^+ (Cann 2007; Ho et al. 2000; Hopper and Bock 1995; Little and Dixon 2003; Sardiwal et al. 2005). These small molecules act as signals. As a GAF domain binds to a molecule, this acts as a regulatory switch that activates other domains in the protein. In the case of a sensor kinase, it is possible that the binding of a molecule to a GAF domain could stimulate autophosphorylation.

Res1 shows similarity to the N-terminus of response regulators; it appears to be composed of only a phosphorylation pocket while lacking the typical C-terminal DNA-binding domain seen in most response regulators (Figure IV.8). On the other hand, Res2

contains both of the normal domains found in response regulators (Figure IV.8). Because of the apparent importance of this putative three-component system to the production of 5*S* clavams, it became important to further characterize the system.

One note of interest is that both *res1* and *res2* contain TTA codons. This indicates that the translation of one, or both, of the response regulators is subject to *bldA* control. It was clear from the *res2* mutant clavam production profile that *res2* alone could not be under *bldA* control. This is because the *res2* mutant clavam production profile does not resemble that of a *bldA* mutant. As mentioned above, a *bldA* mutant can produce all of the clavams except alanylclavam (S. E. Jensen, personal communication). To test if both *res1* and *res2* are regulated by *bldA*, a *res1/res2* double mutant was prepared. Like the *res2* mutant, this mutant could not produce any of the 5*S* clavams, thus indicating that *res1* and *res2* are not both regulated by *bldA* (Section III.2.6). Thus, it appears that res2 is not regulated by *bldA*. Trepanier et al. (2002) suggested that the *bldA* independence exhibited by *ccaR* may be caused by the presence of a G residue 3' of TTA. It was hypothesized that a TTA codon followed by R (G or A) is more likely to be mistranslated in-frame than a TTA codon followed by Y (C or T) (Trepanier et al. 2002). The majority of genes translationally regulated by *bldA* contain at least one TTA codon followed by a C or T (Trepanier et al. 2002). The *res2* TTA codon is followed by a G residue; this may be the reason why *res2* does not appear to be regulated by *bldA*.

It is still possible that *bldA* regulates *res1*. This appears probable since the *res1* TTA codon is followed by a C residue. If *bldA* regulates both *orfA* and *res1*, then *bldA* mutants will still be able to produce all the clavams except for alanylclavam. However, *bldA* mutants should produce elevated levels of 2-hydroxymethylclavam as seen with

res1 mutants (Section III.2.3.1). Unfortunately, quantitative analyses of clavam production by *bldA* mutants have not yet been carried out.

The phenotype of the *res1/res2* double mutant also shows that the *res2* mutation is dominant to the *res1* mutation. This is because the double mutant had the same phenotype as the *res2* mutant rather than that of the *res1* mutant. This indicates that in some way *res2* acts further downstream in the regulatory pathway governing 5*S* clavam biosynthesis.

Overexpression of Res1 and Res2 in *S. clavuligerus* was also attempted (Section III.2.5). The strains bearing the overexpression constructs tended to lyse fairly early in liquid cultures, indicating that the cells were under considerable stress caused by the increased production of Res1 or Res2. This may reflect the toxicity of Res1 and Res2 or simply the burden placed on cells engineered to produce far more protein than normal. Although strains overexpressing Res1 did not reveal consistent trends in clavam production, it appeared that cells overproducing Res2 generally did produce higher than wild type levels of the 5*S* clavams. This lends further evidence of the importance of Res2 in activating 5*S* clavam biosynthesis. However, it should be noted that a complemented *res2* mutant also overproduced the 5*S* clavams, indicating that the simple introduction of *res2* into the chromosome at the ΦC31 attachment site (the site used by the pSET152-based vectors that were employed in the complementation and overexpression experiments) may be responsible for the increased 5*S* clavam production.

The next step in determining how the Snk/Res1/Res2 three-component system actually functions was to characterize the phosphorylation dynamics between the proteins. To do this, the three proteins were purified as His-tagged proteins and then

used for in vitro phosphorylation assays (Section III.2.7). This work revealed that Snk is capable of autophosophorylation in the conditions used and that both Res1 and Res2 can be phosphorylated by Snk, even in the presence of the other response regulator. When this in vitro phosphorylation work is combined with the data obtained from the mutagenesis studies described above, an interesting model for how this system works emerges.

In wild type cells, Snk likely binds to some small molecule through one or both of the GAF domains (Figure IV.9A). This binding causes Snk to autophosphorylate. In the presence of both Res1 and Res2, the phosphate molecule is passed to either. As Res2 becomes phosphorylated, it becomes capable of binding to DNA upstream of genes responsible for 5*S* clavam biosynthesis to activate their expression. Res1 exists simply to funnel phosphate molecules away from Res2. This would allow the level of Res1 expression to modulate the amount of Res2 molecules that receive a phosphate. Such a system provides a further level of control to a typical two-component system.

In an *snk* mutant, Snk could not start off the phosphorylation transfer, so Res2 would never be phosphorylated and 5*S* clavam production would be abolished (Figure IV.9B). A *res1* mutant would not be able to decrease phosphorylation of Res2, so more Res2 molecules will become phosphorylated and the production of 5*S* clavams should increase accordingly (Figure IV.9C). As well, a *res2* mutant will lack the protein necessary for the activation of transcription of the 5*S* clavam biosynthetic genes; thus, a *res2* mutant will not produce 5*S* clavams (Figure IV.9D). From this model, it is clear that a *res1*/*res2* mutant will be unable to produce 5*S* clavams because it lacks *res2*.

Although it is possible that Res1 may actually pass phosphate molecules to Res2, this does not agree with the mutant phenotypes; if Res1 were required for Res2 to become phosphorylated, then a *res1* mutant would have the same phenotype as an *snk* or *res2* mutant. Res1 may form a heterodimer with Res2 and in some way reduce Res2 activity, but testing this possibility would require further analysis to determine whether or not Res1 and Res2 interact.

A number of genes required for 5*S* clavam biosynthesis are known that could be targets for Res2-mediated regulation. This includes *cvm1*, *cvm2*, *cvm5*, *c6p*, and *c7p*. To determine which genes are regulated by the three-component system, reverse transcriptase (RT)-PCR could be carried out. Isolation of RNA from *snk*, *res1*, and *res2* mutants and transcript analysis via RT-PCR using primers specific to genes necessary for 5*S* clavam biosynthesis will likely allow Res2 targets to be identified.

The reasons for having a three-component system that governs antibiotic biosynthesis are not known. It is possible that such a system provides an additional layer of control in Res1 so that 5*S* clavam production can be fine-tuned to meet whatever cellular needs exist for clavams. Res1 may be expressed only at certain times so that 5*S* clavam production is subtly modulated at different periods during the *S. clavuligerus* life-cycle. Likewise, Res1 might be expressed only in certain types of cells. For example, Res1 may be produced only in the substrate mycelium but not in the aerial mycelium or in spores. Once again, the production of Res1 in specific situations would result in a decrease, but not necessarily in a complete loss, of 5*S* clavam production.

Other three-component systems have been identified and studied in other bacteria. One such system in *Pseudomonas syringae*, consisting of CorS (a sensor kinase), CorR (a

typical response regulator), and CorP (a response regulator lacking a DNA-binding domain), is involved in the synthesis of the phytotoxin coronatine (Ullrich et al. 1995). Each gene is necessary for the production of coronatine, the transcription of at least one coronatine biosynthetic gene, and the transcription of *corS* (Ullrich et al. 1995). Interestingly, CorS is capable of phosphorylating CorR but not CorP (Rangaswamy and Bender 2000). The nature of the role of CorP in coronatine biosynthesis remains uncertain.

Another three-component system involved in biofilm formation has been found in *Pseudomonas aeruginosa*. This system consists of RocS1 (a sensor kinase), RocA1 (a typical response regulator), and RocR (a response regulator carrying the phosphorylation pocket yet lacking a DNA-binding domain)(Kuchma et al. 2005; Kulasekara et al. 2005). Mutants in *rocS1*, *rocA1*, or *rocR* produce immature biofilms that lack the macrocolonies and fluid-filled channels formed in wild type biofilms (Kuchma et al. 2005). These genes are involved in the expression of genes needed for type III secretion and genes needed for the production of adhesins; type III secretion and adhesins are important elements needed for normal biofilm formation (Kuchma et al. 2005; Kulasekara et al. 2005). In fact, detailed analyses revealed that RocS1 and RocA1 activate the expression of adhesin biosynthetic genes while RocR somehow represses expression of at least one the adhesin biosynthetic genes transcriptionally activated by RocS1 and RocA1(Kulasekara et al. 2005). Once it was shown that RocS1 interacts with both RocA1 and RocR, it was proposed that RocS1, upon sensing a stimulus, can phosphorylate both RocA1 and RocR (Kulasekara et al. 2005). However, phosphorylation of RocR decreases the expression of

at least one gene that is transcriptionally activated by RocA1, possibly by funneling phosphate molecules away from RocA1 or through some other mechanism.

The CbbRRS system in *Rhodopseudomonas palustris* is also composed of a sensor kinase (CbbSR) and two response regulators (CbbRR1 and CbbRR2)(Romagnoli and Tabita 2006). This three-component system is required for the expression of one of the two groups of genes encoding RubisCO (which is involved in carbon dioxide assimilation). Interestingly, none of these proteins contain a DNA-binding domain; the method by which it regulates the expression of RubisCO is unknown (Romagnoli and Tabita 2006). The full form of CbbSR was capable of phosphorylating both response regulators, but a truncated form of CbbSR lacking the N-terminal transmembrane region could only phosphorylate CbbRR1 (Romagnoli and Tabita 2006). This indicates that specific regions or conformations of CbbSR are responsible for differentiating one response regulator from the other (Romagnoli and Tabita 2006).

It is obvious that three-component systems are quite rare, but they are also clearly involved in a variety of processes. This includes, as described above, coronatine biosynthesis, biofilm formation, carbon dioxide fixation, and, now, clavam production. Continued examination of the Snk/Res1/Res2 system will reveal new intricacies in the regulation of antibiotic production and will expand the knowledge of bacterial regulatory systems in general. As well, understanding the regulatory elements of clavam biosynthesis may lead to increases in clavulanic acid production at an industrial level.

IV.3 The in vivo generation of unmarked in-frame *Streptomyces* mutants

The production of unmarked in-frame *Streptomyces* mutants is usually a long and difficult process. Even in the REDIRECT© procedure, which has greatly facilitated the generation of marked and unmarked *Streptomyces* mutants, excision of the disrupting antibiotic resistance cassette must be carried out in *E. coli* (Gust et al. 2003). The process is made more difficult if the *Streptomyces* species being studied, such as *S. clavuligerus*, cannot be transformed with large, cosmid-sized molecules of DNA. In these cases, a fragment of DNA from the mutant cosmid carrying the antibiotic cassette and flanking regions of *Streptomyces* chromosomal DNA must be cloned into another vector that is smaller (so that the plasmid can be transformed into *Streptomyces* cells) or that contains an *oriT* (so that the plasmid can be conjugated into *Streptomyces* cells). The antibiotic cassette can then be excised in *E. coli* and subsequently introduced into *Streptomyces* cells. This method was used by Tahlan (2005) to generate an unmarked in-frame Δ*ceaS2* mutant and in the current study to produce a Δ*snk* mutant (Section III.3.1). The labor and length of time necessary to carry out this process made it a priority to determine a faster and simpler method to produce unmarked nonpolar mutants.

To facilitate the process, an in vivo approach for excising REDIRECT© antibiotic resistance cassettes was undertaken. This entailed the expression of FLP recombinase in *Streptomyces* mutants carrying the REDIRECT© cassette in the chromosome. By cloning the *flp* gene into two conjugative vectors under the control of the highly active *ermE** promoter to allow high-level FLP expression in *Streptomyces*, the in vivo excision of REDIRECT© cassettes was achieved. However, one of the FLP-expressing plasmids, pHJ401-*flp*, did not yield cassette excision in *S. clavuligerus*. pHJ401 is a

segregationally unstable plasmid (He et al. 2001; Table II.4). It is possible that a toxic level of FLP combined with an already unstable plasmid resulted in the extremely quick loss of the plasmid from cells, such that the cassette could not be excised. The other plasmid, pUWL-oriT-*flp*, appeared to work well in *S. clavuligerus* to produce a Δ*c7p* mutant.

The Δ*c7p* mutant could not produce 5*S* clavams but still produced clavulanic acid, which is the phenotype of the original Δ*c7p::apra* mutant (Tahlan et al. 2007; Section III.3.3). This demonstrates that *c7p* is indeed required for 5*S* clavam biosynthesis and that the original mutant phenotype was not caused by polar effects upon the downstream *c6p* (Figure I.6).

Once it was confirmed that in vivo FLP expression could lead to the excision of REDIRECT© cassettes in *S. clavuligerus*, the demonstration of FLP-mediated in vivo excision of FRT-flanked cassettes in *S. coelicolor* (a genetic model for *Streptomcyes* spp.) was explored. Such a demonstration would show that the FLP-expressing plasmids are functional in multiple *Streptomyces* spp. It was decided that the system could be used in mutating a gene in the middle of an operon encoding proteins required for undecylprodigiosin. The gene chosen for this purpose, *redJ*, is upstream of genes known to be involved in undecylprodigiosin production, while the role of *redJ* itself in biosynthesis was not known (Cerdeno et al. 2001; Figure III.33). After producing a Δ*redJ::apra* mutant via REDIRECT©, the cassette was excised using either pHJ401-*flp* or pUWL-oriT-*flp* (Section III.3.4.1). Both pHJ401-*flp* and pUWL-*oriT-flp* were easily lost from the mutants. This was surprising because pUWL-KS (and all plasmids bearing the pIJ101 origin of replication) is generally quite stable in most *Streptomyces* spp. while

being unstable in *S. clavuligerus* for an unknown reason (S. E. Jensen, personal communication). It is possible that, as mentioned above, high levels of FLP expression may be toxic to *Streptomyces* cells, thus making it beneficial for cells to lose the FLP-expressing plasmid.

As expected, the Δ*redJ::apra* mutants could not produce undecylprodigiosin (Section III.3.4.1). Since *redJ* is upstream of genes necessary for undecylprodigiosin, it made sense that *redJ* disruption would interrupt the expression of downstream genes (Figure III.33). However, it was unexpected that the Δ*redJ* mutants also produced very low levels of undecylprodigiosin (Section III.3.4.1). There are two possibilities for this Δ*redJ* phenotype: *redJ* does play a role in undecylprodigiosin biosynthesis or the 81 bp "scar" left in the Δ*redJ* mutants exerts polar effects on downstream genes. To determine which possibility is correct, RT-PCR could be used to detect the expression of genes downstream of *redJ* in Δ*redJ* mutants and complementation of Δ*redJ* mutants could reveal if *redJ* by itself is necessary for undecylprodigiosin production in the mutants. It is believed that the leftover FRT site in such unmarked in-frame mutants does not generate polar effects (Merlin et al. 2002). It is therefore likely that the former possibility, regarding the essential nature of *redJ* to undecylprodigiosin biosynthesis, is the case.

The RedJ amino acid sequence shows similarity to thioesterases (Cerdeno et al. 2001). In antibiotic biosynthesis, thioesterase domains are encoded in non-ribosomal peptide and polyketide biosynthetic gene clusters and cleave the synthesized antibiotic from the biosynthetic machinery, sometimes causing the cyclization of the antibiotic. Since there are two multienzyme proteins (RedL and RedN) involved in amino acid

addition in undecylprodigiosin biosynthesis, it is possible that RedJ is required for cleaving the growing antibiotic from the multienzymes. Some thioesterases cleave mis-cognate chains from multienzymes to prevent the blockage of antibiotic biosynthesis (Butler et al. 1999). Since Δ*redJ* mutants are able to produce some red pigment (slightly more than Δ*redJ::apra* mutants; Section III.3.4.1), perhaps this editing role of RedJ is more plausible than an essential role in antibiotic chain termination.

Hopefully, this in vivo method of producing unmarked in-frame mutants in *Streptomyces* spp. will be adopted and improved upon by many researchers. With the already wide use of REDIRECT©, it would certainly be simpler for scientists to use the FLP-expressing plasmids in combination with the REDIRECT© method rather than begin using the Cre-*loxP* or the Xis/Int-*attR*/*attL* systems described earlier (Khodakaramian et al. 2006; Raynal et al. 2006; Section I.4). The system described in this thesis could be improved by using a *flp* gene that has been generated for optimum expression in high G+C organisms. This new gene, flp_m, has been used to increase FLP-mediated cassette excision from a frequency of 0% (using the typical *flp* gene) to as high as 59% (using flp_m) in *Mycobacterium bovis* (Song and Niederweis 2007).

Another improvement upon the generation of unmarked in-frame mutants could involve new REDIRECT© cassettes engineered so that they carry an inducible *flp* gene. The initial mutant would carry the FRT-flanked antibiotic resistance cassette including *flp* under the control of a promoter that is induced by the presence of a particular compound, such as thiostrepton or glycerol. The simple addition of the inducing compound to growing cultures of the cassette-bearing mutant would result in FLP-mediated excision of the cassette and the production of unmarked in-frame mutants.

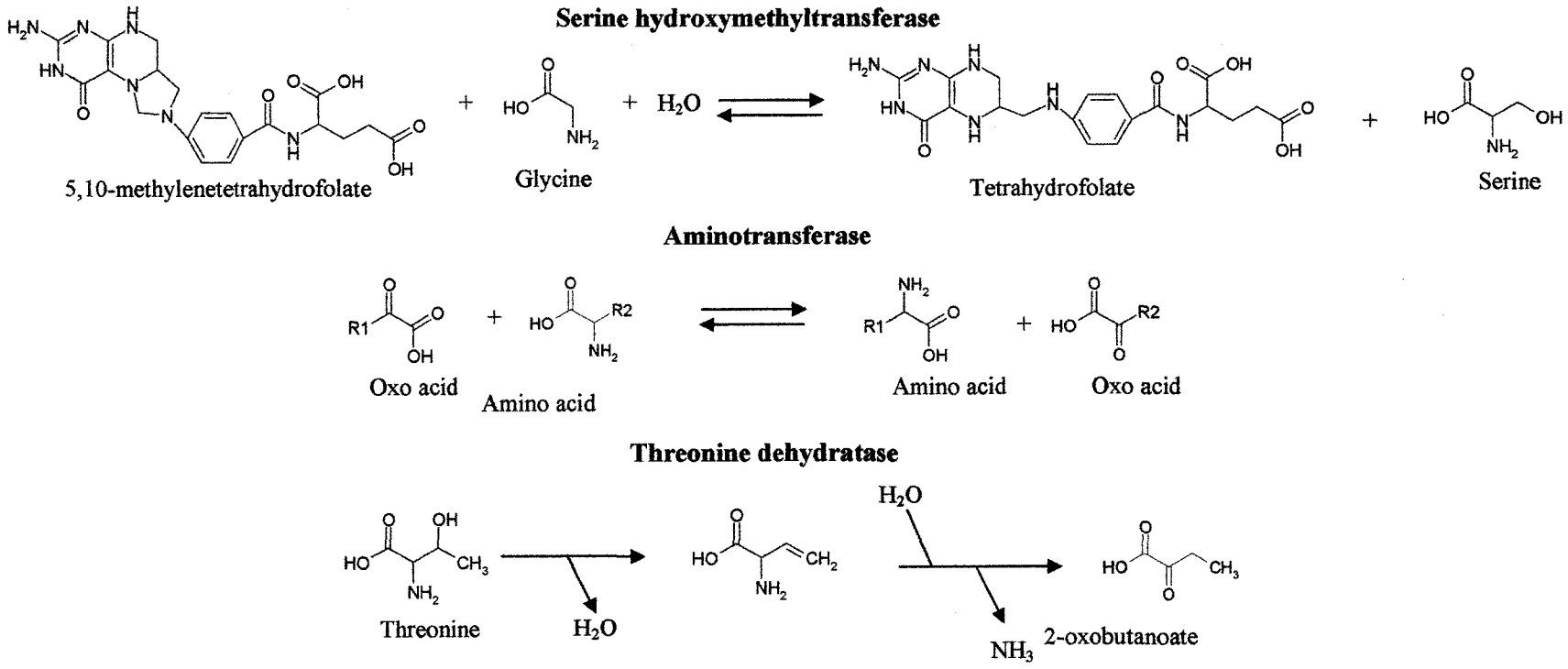


Figure IV.1. Reactions catalyzed by enzymes similar to OrfA, OrfC, and OrfD. OrfA is similar to serine hydroxymethyltransferases, OrfC is similar to aminotransferases, and OrfD is similar to threonine dehydratases.

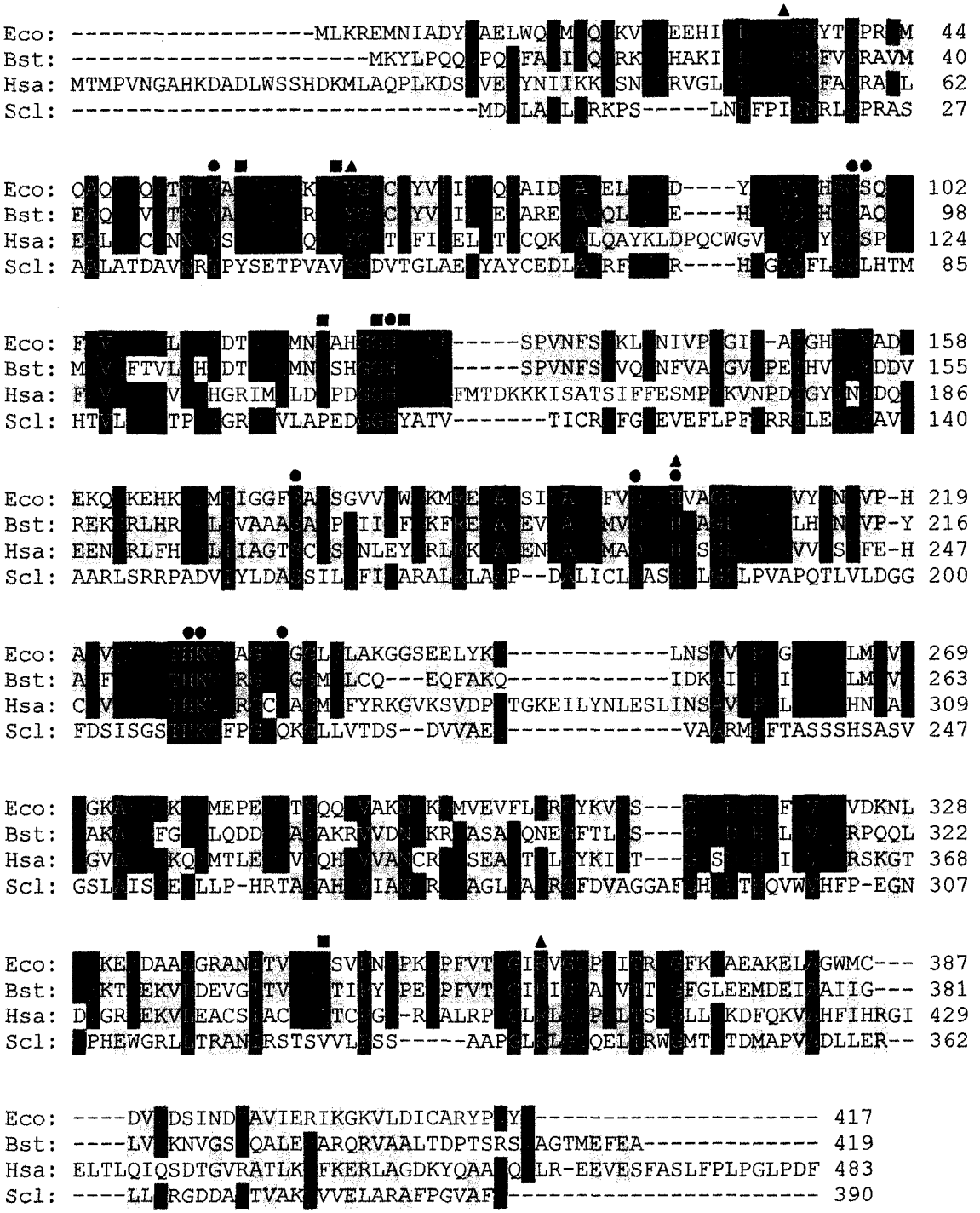


Figure IV.2. The alignment of the amino acid sequence of OrfA with those of serine hydroxymethyltransferases. Eco, the *E. coli* serine hydroxymethyltransferase (P0A825). Bst, the *Bacillus stearothermophilus* serine hydroxymethyltransferase (1KL2B). Hsa, the *Homo sapiens* serine hydroxymethyltransferase 1 (NP_004160.3). Scl, *S. clavuligerus*

OrfA. Black dots indicates residues involved in pyridoxal phosphate binding. Black squares indicate residues involved in 5,10-methylenetetrahydrofolate binding. Black triangles indicate residues involved in glycine binding.

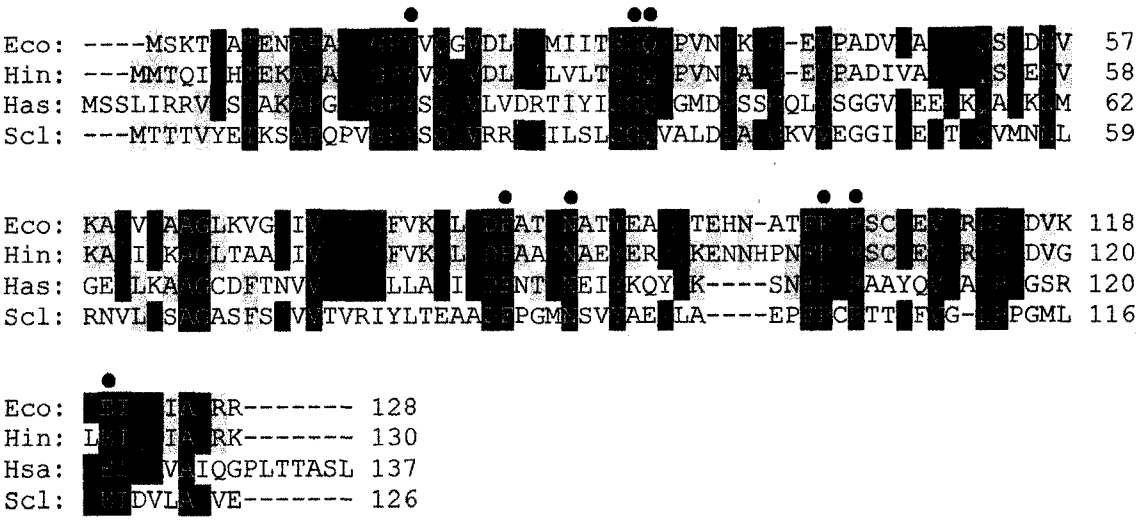


Figure IV.3. The alignment of the amino acid sequence of OrfB with those of YjgF/YER057c/UK114 family proteins. Eco, the *E. coli* YjgF protein (1QU9). Hin, the *H. influenzae* HI0719 protein (1J7H). Hsa, the *Homo sapiens* p14.5 protein (NP_005827.1). Scl, the *S. clavuligerus* OrfB protein. Black circles indicate residues involved in substrate binding.

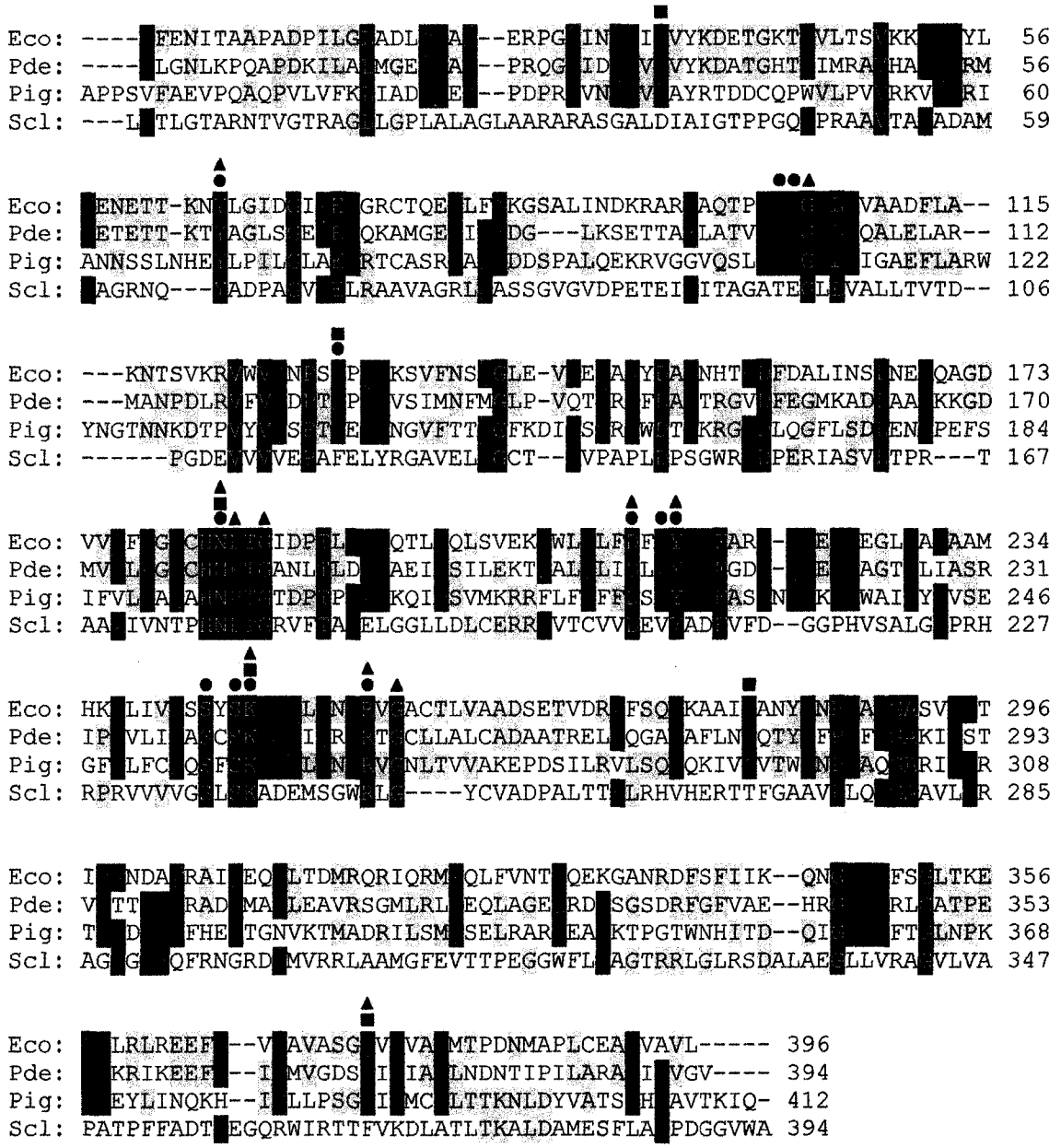


Figure IV. 4. The alignment of the amino acid sequence of OrfC with those of family I aminotransferases. Eco, the *E. coli* AspC protein (1X2A). Pde, the *Paracoccus denitrificans* AroAT protein (P95468). Pig, the pig cytosolic AspAT protein (1AJS). Scl, the *S. clavuligerus* OrfC protein. Black circles indicate residues involved in pyridoxal phosphate binding. Black squares indicate residues involved in substrate

(aspartate or aromatic amino acid) binding. Black triangles indicate the residues found in all family I aminotransferases (Mehta et al. 1993; Jensen and Gu 1996).

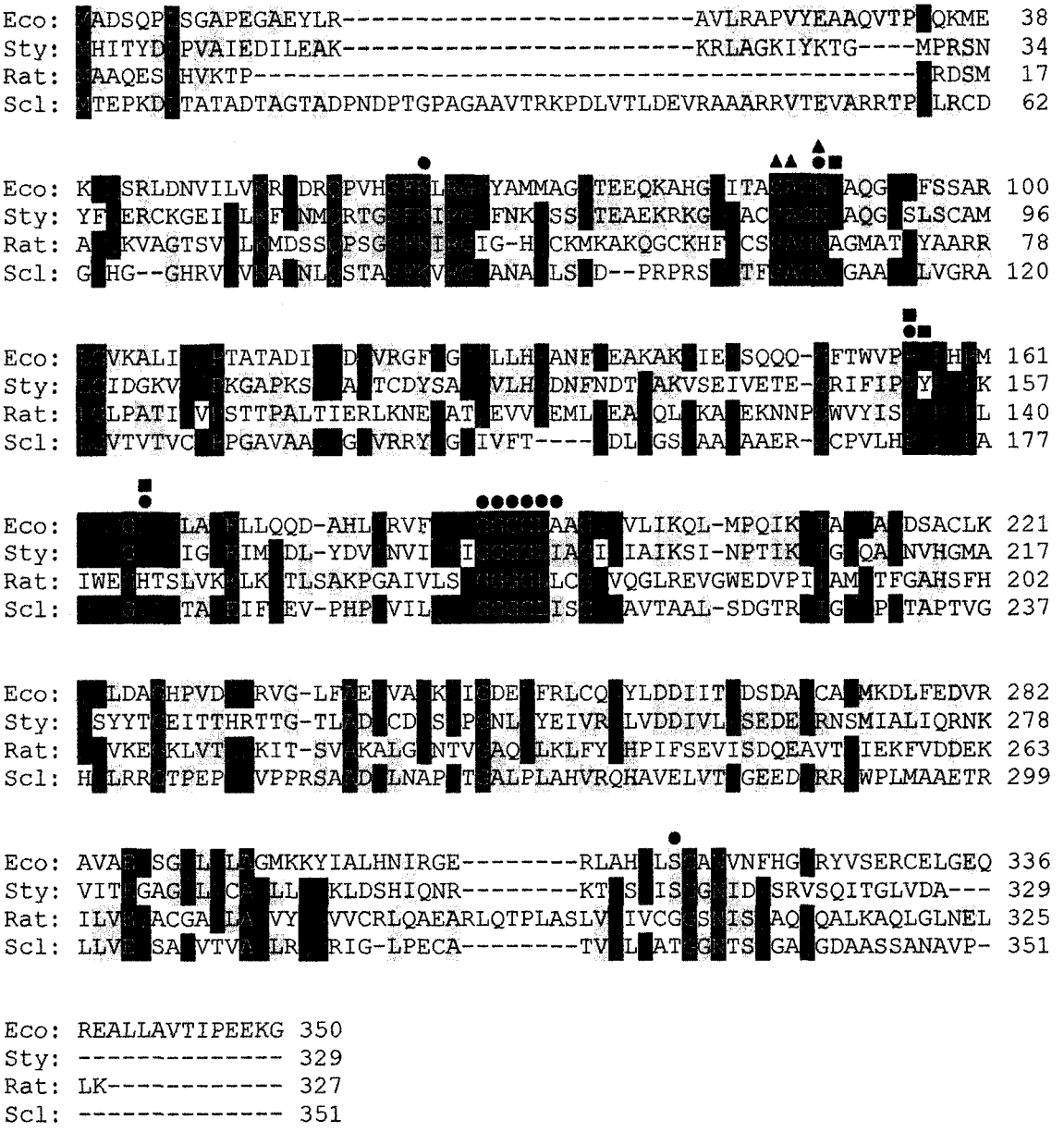


Figure IV. 5. The alignment of the amino acid sequence of OrfD with those of threonine and serine dehydratases. Eco, the first 350 amino acids of the *E. coli* biosynthetic threonine dehydratase (NP_418220.1). Sty, the *Salmonella typhimurium* TcdB protein (P11954). Rat, the rat liver serine dehydratase (1PWH). Scl, the *S. clavuligerus* OrfD. Black circles indicate residues involved in pyridoxal phosphate binding. Black squares

indicate TcdB active site residues. Black triangles indicate residues involved in serine binding in the rat liver serine dehydratase.

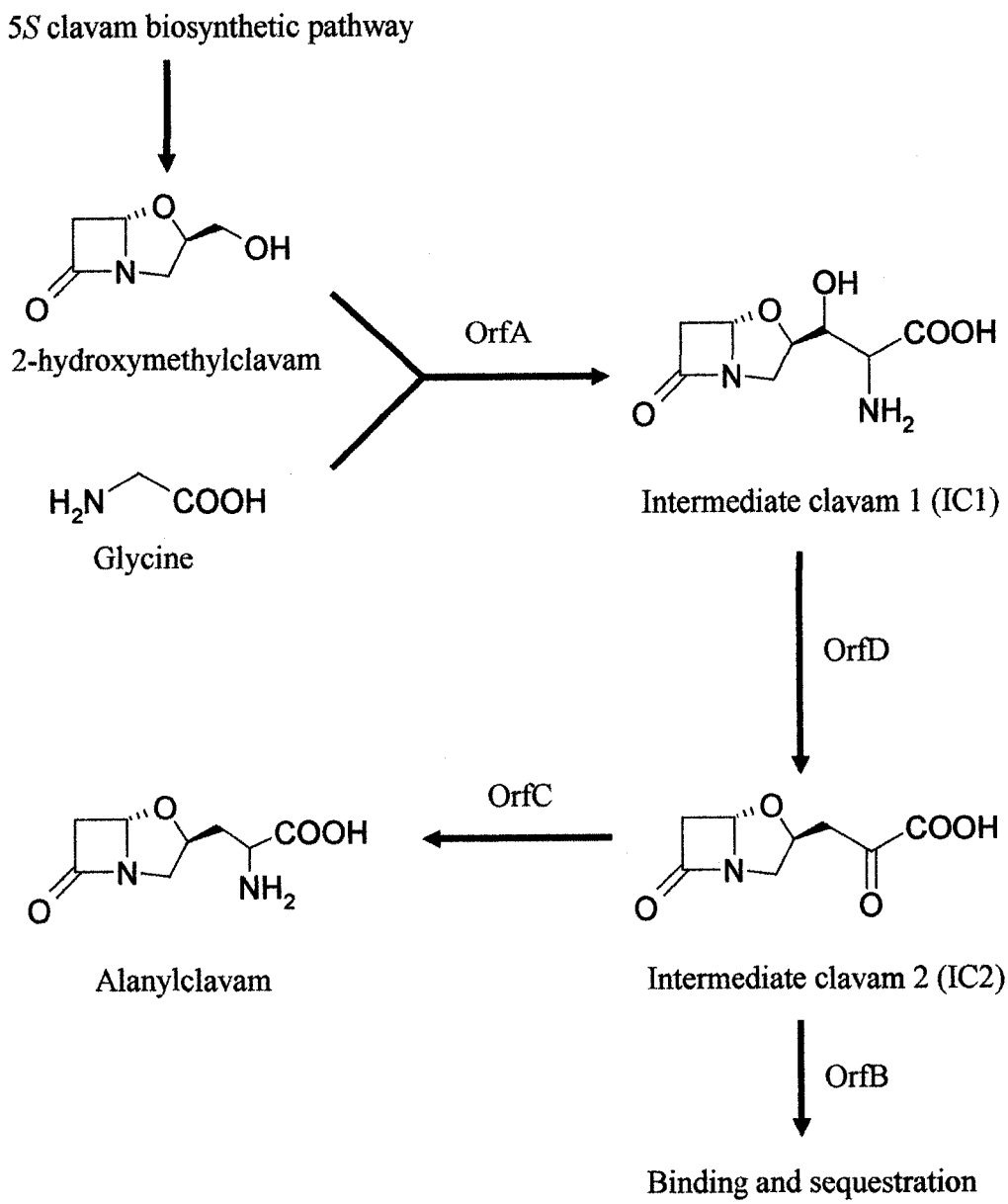


Figure IV.6. The proposed alanylclavam biosynthetic pathway in *S. clavuligerus*. Refer to the text for more detailed descriptions of each step in the pathway.

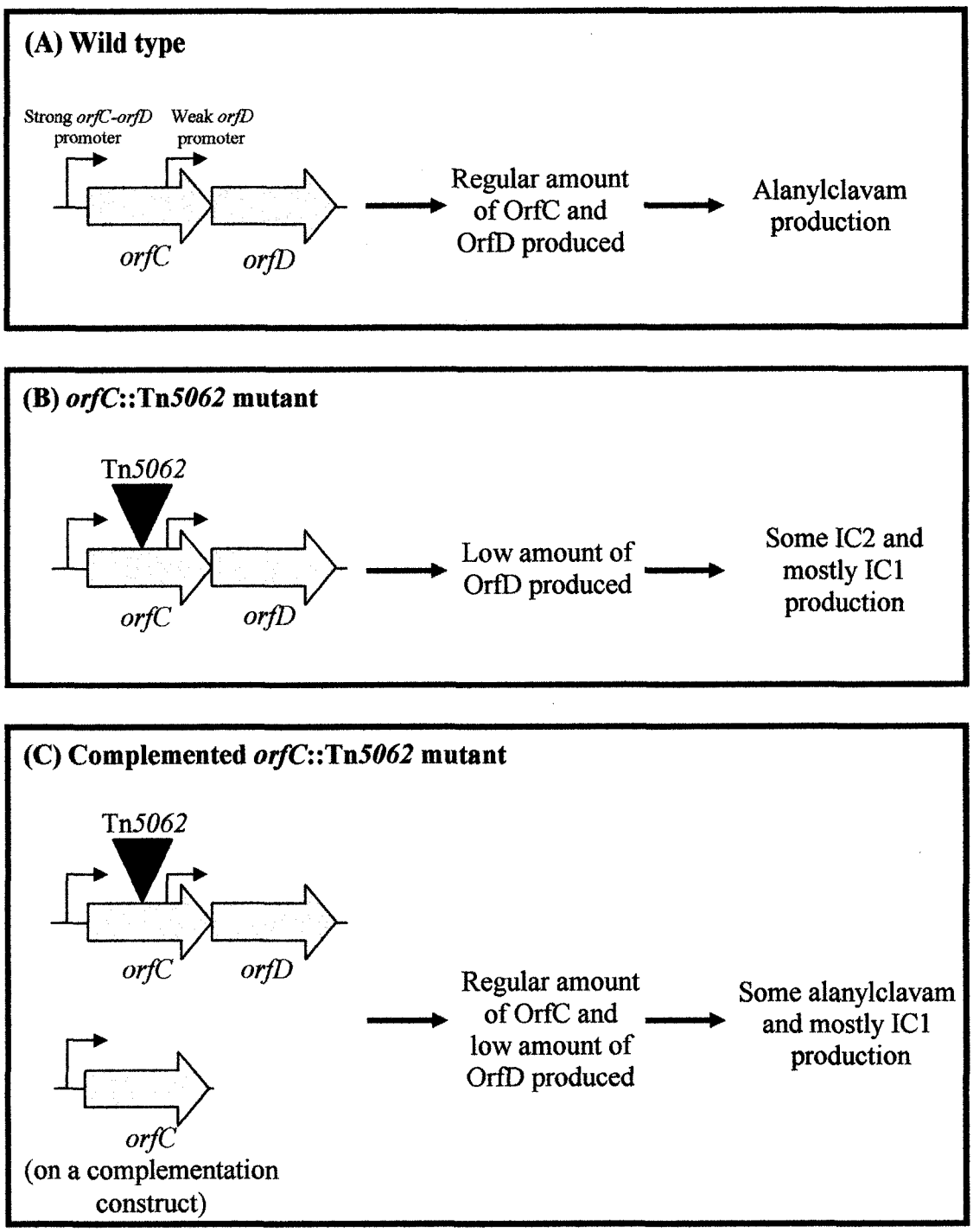


Figure IV.7. The putative *orfC-orfD* operon. Genes are indicated by grey block arrows, promoters are indicated by bent arrows, and the Tn*5062* insertion is indicated by a black triangle above *orfC*. (A), the *orfC-orfD* operon in wild type cells. (B), the *orfC-orfD*

operon in *orfC*::Tn*5062* mutant cells. (C), the *orfC-orfD* operon in complemented *orfC*::Tn*5062* mutant cells. Refer to the text for a more detailed explanation of the operon.

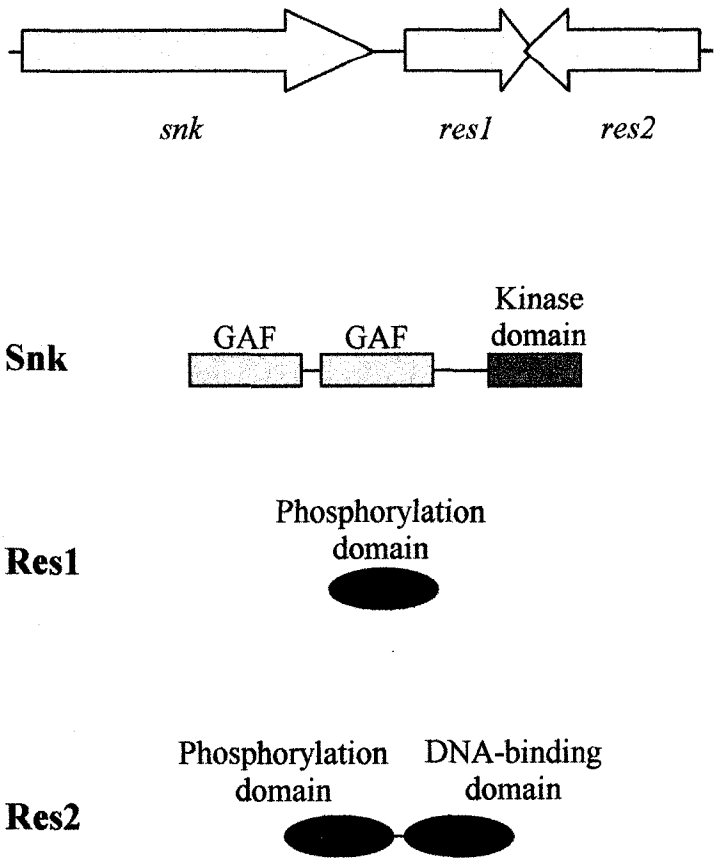


Figure IV.8. The genetic and modular domain organization of the Snk/Res1/Res2 system. Grey arrows indicate genes and rectangles and circles indicate protein domains.

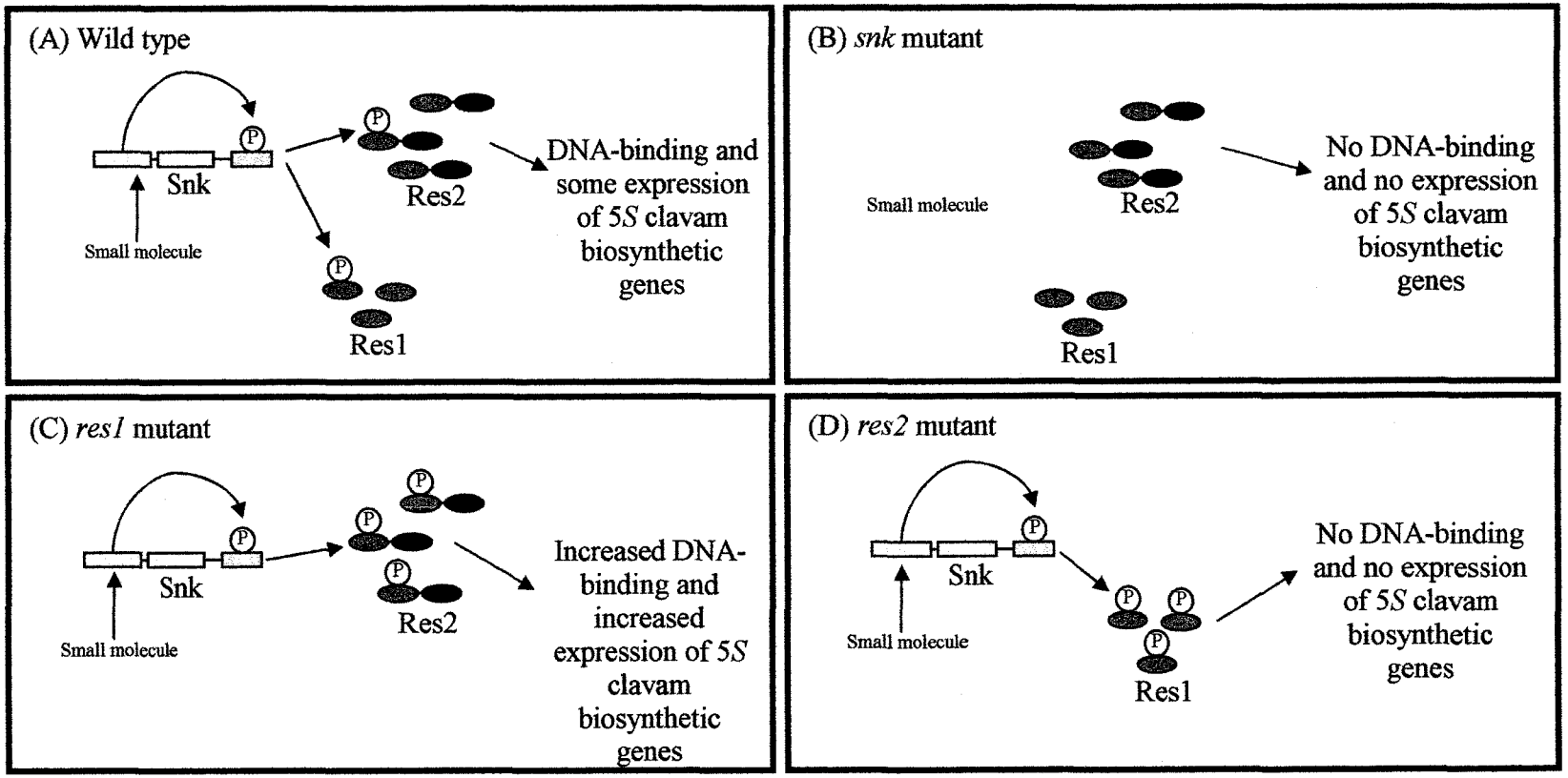


Figure IV.9. The proposed model for the Snk/Res1/Res2 three-component system of *S. clavuligerus*. (A), the system in wild-type cells. (B), the system in *snk* mutants. (C), the system in *res1* mutants. (D), the system in *res2* mutants. Protein domains are as depicted in Figure IV.8. "P" enclosed by a circle represents a phosphate molecule.

V. REFERENCES

Aceti, D.J., and Champness, W.C. 1998. Transcriptional regulation of *Streptomyces coelicolor* pathway-specific antibiotic regulators by the *absA* and *absB* loci. J. Bacteriol. **180**:3100-3106.

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

Aidoo, K.A., Wong, A., Alexander, D.C., Rittammer, R.A., and Jensen, S.E. 1994. Cloning, sequencing and disruption of a gene from *Streptomyces clavuligerus* involved in clavulanic acid biosynthesis. Gene **147**:41-46.

Alexander, D.C. 1998. Regulation of cephamycin biosynthesis in *Streptomyces clavuligerus*. Ph.D. Thesis, University of Alberta, Edmonton, Alta.

Alexander, D.C., and Jensen, S.E. 1998. Investigation of the *Streptomyces clavuligerus* cephamycin C gene cluster and its regulation by the CcaR protein. J. Bacteriol. **180**:4068-4079.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389-3402.

Ambler, R.P. 1980. The structure of β-lactamases. Phil. Trans. R. Sec. Lond. B **289**:321-331.

Ando, N., Matsumori, N., Sakuda, S., Beppu, T., and Horinouchi, S. 1997. Involvement of *afsA* in A-factor biosynthesis as a key enzyme. J. Antibiot. **50**:847-852.

Anton, N., Mendes, M.V., Martin, J.F., and Aparicio, J.F. 2004. Identification of PimR as a positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*. J. Bacteriol. **186**:2567-2575.

Arias, P., Fernandez-Moreno, M.A., and Malpartida, F. 1999. Characterization of the pathway-specific positive transcriptional regulator for actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2) as a DNA-binding protein. J. Bacteriol. **181**:6958-6968.

Arulanantham, H. Kershaw, N.J., Hewitson, K.S., Hughes, C.E., Thirkettle, J.E., Schofield, C.J. 2006. ORF17 from the clavulanic acid biosynthesis gene cluster catalyzes the ATP-dependent formation of *N*-glycyl-clavaminic acid. J. Biol. Chem. **281**:279-287.

Bachmann, B.O., Li, R., and Townsend, C.A. 1998. β-lactam synthetase: a new biosynthetic enzyme. Proc. Natl. Acad. Sci. USA **95**:9082-9086.

Baggaley, K.H., Brown, A.G., and Schofield, C.J. 1997. Chemistry and biosynthesis of clavulanic acid and other clavams. Nat. Prod. Rep. **14**:309-333.

Baldwin, J.E., Lloyd, M.D., Wha-Shon, B., Schofield, C.J., Elson, S.W., Baggaley, K.H., and Nicholson, N.H. 1993. A substrate analogue study on clavaminic acid synthase: possible clues to the biosynthetic origin of proclavaminic acid. J. Chem. Soc. Chem. Commun. 500-502.

Baltz, R.H. 1980. Genetic recombination by protoplast fusion in *Streptomyces*. Dev. Ind. Microbiol. **21**:43-54.

Bao, K., and Cohen, S.N. 2001. Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. Genes Dev. **15**:1518-1527.

Bao, K., and Cohen, S.N. 2003. Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosome by a novel telomere-binding protein essential for linear DNA replication. Genes Dev **17**:774-785.

Bao, K., and Cohen, S.N. 2004. Reverse transcriptase activity innate to DNA polymerase I and DNA topoisomerase I proteins of *Streptomyces* telomere complex. Proc. Natl. Acad. Sci. USA **101**:14361-14366.

Bateman, A., and Rawlings, N.D. 2003. The CHAP domain: a large family of amidases including GSP amidases and peptidoglycan hydrolases. Trends Biochem. Sci. **28**:234-237.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M, Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., and Hopwood, D.A. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417**:141-147.

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

Bignell, D.R., Lau, L.H., Colvin, K.R., and Leskiw, B.K. 2003. The putative anti-anti-sigma factor BldG is post-translationally modified by phosphorylation in *Streptomyces coelicolor*. FEMS Microbiol. Lett. **225**:93-99.

Bignell, D.R.D., Tahlan, K., Colvin, K.R., Jensen, S.E., and Leskiw, B.L. 2005. Expression of *ccaR*, encoding the positive activator of cephamycin C and clavulanic acid production in *Streptomyces clavuligerus*, is dependent on *bldG*. Antimicrob. Agents Chemother. **49**:1529-1541.

Bignell, D.R., Warawa, J.L., Strap, J.L., Chater, K.F., and Leskiw, B.K. 2000. Study of the *bldG* locus suggests that an anti-anti-sigma factor and an anti-sigma factor may be involved in *Streptomyces coelicolor* antibiotic production and sporulation. Microbiology **146**:2161-2173.

Birnboim, H.C., and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **24**:1513-1523.

Bishop, A., Fielding, S., Dyson, P., and Herron, P. 2004. Systematic insertional mutagenesis of a streptomycete genome: a link between osmoadaptation and antibiotic production. Genome Res. **14**:893-900.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.

Brian, P., Riggle, P.J., Santos, R.A., and Champness, W.C. 1996. Global negative regulation of *Streptomyces coelicolor* antibiotic synthesis mediated by an *absA*-encoded putative signal transduction system. J. Bacteriol. **178**:3221-3231.

Broach, J.R., Guarasico, V.R., and Jayaram, M. 1982. Recombination within the yeast plasmid 2mu circle is site-specific. Cell **29**:227-234.

Brown, A.G., Butterworth, D., Cole, M., Hanscomb, G., Hood, J.D., and Reading, C. 1976. Naturally-occurring β-lactamase inhibitors with antibacterial activity. J. Antibiot. **29**:668-669.

Brown, D., Evans, J.R., and Fletton, R.A. 1979. Structures of three novel β-lactams isolated from Streptomyces clavuligerus. J. Chem. Soc. Chem. Commun. 282-283.

Butler, A.R., Bate, N., and Cundliffe, E. 1999. Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*. Chem. Biol. **6**:287-292.

Cai, H. 2003. Transposon mutagenesis and characterization of clavam biosynthetic genes in *Streptomyces clavuligerus*. M.Sc. Thesis, University of Alberta, Edmonton, Alta.

Caines, M.E., Elkins, J.M., Hewitson, K.S., and Schofield, C.J. 2004. Crystal structure and mechanistic implications of N2-(2-carboxyethyl)arginine synthase, the first

enzyme in the clavulanic acid biosynthesis pathway. J. Biol. Chem. **279**:5685-5692.

Cann, M. 2007. A subset of GAF domains are evolutionarily conserved sodium sensors. Mol. Microbiol. **64**:461-472.

Cerdeno, A.M., Bibb, M.J., and Challis, G.L. 2001. Analysis of the prodiginine biosynthesis gene cluster of *Streptomyces coelicolor* A3(2): new mechanisms for chain initiation and termination in modular multienzymes. Chemistry & Biology **8**:817-829.

Chakraburtty, R., and Bibb, M.J. 1997. The ppGpp synthase gene (*relA*) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation. J. Bacteriol. **179**:5854-5861.

Chang, H.C., Chen, M.Y., Shieh, Y.T., Bibb, M.J., and Chen, C.W. 1996. The *cutRS* signal transduction system of *Streptomyces lividans* represses the biosynthesis of the polyketide antibiotic actinorhodin. Mol. Microbiol. **21**:1075-1085.

Chang, P.C., and Cohen, S.N. 1994. Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. Science **265**:952-954.

Chater, K.F. 1989. Aspects of multicellular differentiation in *Streptomyces coelicolor* A3(2). In Genetics and Molecular Biology of Industrial Microorganisms. *Edited by* C.L. Hershberger, S.W. Queener, and G. Hegeman. American Society for Microbiology, Washington D.C. pp. 99-107.

Cherepanov, P.P., and Wackernagel, W. 1995. Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene **158**:9-14.

Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. 2004. The Jalview Java alignment editor. Bioinformatics **20**:426-427.

Cox, M.M. 1983. The FLP protein of the yeast 2-μm plasmid: Expression of a eukaryotic genetic recombination system in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:4223-4227.

Datsenko, K.A., and Wanner, B.L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA **97**:6640-6645.

Demain, A.L. 2000. Small bugs, big business: the economic power of the microbe. Biotechnology Advances **18**:499-514.

Demain, A.L., and Elander, R.P. 1999. The β-lactam antibiotics: past, present, and future. Antonie van Leeuwenhoek **75**:5-19.

Egan, L.A., Busby, R.W., Iwata-Reuyl, D., and Townsend, C.A. 1997. Probable role of clavaminic acid as the terminal intermediate in the common pathway to clavulanic acid and the antipodal clavam metabolites. J. Am. Chem. Soc. **119**:2348-2355.

Elander, R.P. 2003. Industrial production of β-lactam antibiotics. Appl. Microbiol. Biotechnol. **61**:385-392.

Elkins, J.M., Clifton, I.J., Hernandez, H., Doan, L.X., Robinson, C.V., Schofield, C.J., and Hewitson, K.S. 2002. Oligomeric structure of proclavaminic acid amidino hydrolase: evolution of a hydrolytic enzyme in clavulanic acid biosynthesis. Biochem. J. **366**:423-434.

Elson, S.W., Gillett, J., Nicholson, N.H., and Tyler, J.W. 1988. *N*-acyl-derivatives of clavaminic acid produced by a mutant of *Streptomyces clavuligerus*. J. Chem. Soc. Chem. Commun. 979-980.

Enos-Berlage, J.L., Langendorf, M.J., and Downs, D.M. 1998. Complex metabolic phenotypes caused by a mutation in *yjgF*, encoding a member of the highly conserved YER057c/YjgF family of proteins. J. Bacteriol. **180**:6519-6528.

Fernandez-Moreno, M.A., Caballero, J.L., Hopwood, D.A., and Malpartida, F. 1991. The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the bldA tRNA gene of *Streptomyces*. Cell **66**:769-780.

Fischer, G., Holl, A.C., Volff, A.N., Vnadewiele, D., Decaris, B., and Leblond, P. 1998. Replication of the linear chromosomal DNA from the centrally located oriC of *Streptomyces ambofaciens* revealed by PFGE gene dosage analysis. Res. Microbiol. **149**:203-210.

Fleming, A. 1929. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. Brit. J. Exp. Path.**10**:226-236.

Folcher, M., Gaillard, H., Nguyen, L.T., Nguyen, K.T., Lacroix, P., Bamas-Jacques, N., Rinkel, M., and Thompson, C.J. 2001. Pleiotropic functions of a *Streptomcyes pristinaespiralis* autoregulator receptor in development, antibiotic biosynthesis, and expression of a superoxide dismutase. J. Biol. Chem. **276**:44297-44306.

Foulstone, M. and Reading, C. 1982. Assay of amoxicillin and clavulanic acid, the components of Augmentin, in biological fluids with high-performance liquid chromatography. Antimicrob. Agents Chemother. **22**:753-762.

Gehring, A.M., Nodwell, J.R., Beverley, and S.M., Losick, R. 2000. Genomewide insertional mutagenesis in *Streptomyces coelicolor* reveals additional genes involved in morphological differentiation. Proc. Natl. Acad. Sci. USA **97**:9642-9647.

Gehring, A.M., Yoo, N.J., and Losick, R. 2001. RNA polymerase sigma factor that blocks morphological differentiation by *Streptomyces coelicolor*. J. Bacteriol. **183**:5991-5996.

Gehring, A.M., Wang, S.T., Kearns, D.B., Storer, N.Y., and Losick, R. 2004. Novel genes that influence development in *Streptomyces coelicolor*. J. Bacteriol. **186**:3570-3577.

Gomez-Escribano, J.P., Liras, P., Pisaberro, A., and Martin, J.F. 2006. An *rplK*$^{\Delta 29\text{-PALG-}32}$ mutation leads to reduced expression of the regulatory genes *ccaR* and *claR* and very low transcription of the *ceaS2* gene for clavulanic acid biosynthesis in *Streptomyces clavuligerus*. Mol. Microbiol. **61**:758-770.

Gramajo, H.C., Takano, E., and Bibb, M.J. 1993. Stationary-phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. Mol. Microbiol. **7**:837-845.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci. USA **100**:1541-1546.

Hara, O., Beppu, T. 1982. Mutants blocked in streptomycin production in *Streptomyces griseus* – the role of A-factor. J. Antibiot. **35**:349-358.

Hardisson, C., and Manzanal, M.B. 1976. Ultrastructural studies of sporulation in *Streptomyces*. J. Bacteriol. **127**:1443-1454.

Hardisson, C., Manzanal, M.B., Salas, J.A., and Suarez, J.E. 1978. Fine structure, physiology and biochemistry of arthrospore germination in Streptomyces antibioticus. J. Gen. Microbiol. **105**:203-214.

Hashimoto, K., Nihira, T., Yamada, Y. 1992. Distrbution of virginiae butanolides and IM-2 in the genus Streptomyces. J. Ferment. Bioeng. **73**:61-65.

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

He, J., Magarvey, N., Piraee, M., and Vining, L.C. 2001. The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel

shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. Microbiology **147**:2817-2829.

Higgens, C.E., and Kastner, R.E. 1971. *Streptomyces clavuligerus* sp. nov., a β-lactam antibiotic producer. Int. J. Syst. Bacteriol. **21**:326-331.

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., and Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673-4680.

Ho, Y.S.J., Burden, L.M., and Hurley, J.H. 2000. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. EMBO J. **19**:5288-5299.

Hoang, T.T., Karkhoff-Schweizer, R.R., Kutchma, A.J., and Schweizer, H.P. 1998. A broad –host-range Flp-*FRT* recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. Gene **212**:77-86.

Hobbs, G., Frazer, C.M., Gardner, D.C.J., Cullum, J.A., and Oliver, S.G. 1989. Dispersed growth of *Streptomyces* in liquid culture. Appl. Microbiol. Biotechnol. **31**:272-277.

Hodgson, J.E., Fosberry, A.P., Rawlinson, N.S., Ross, H.N.M., Neal, R.J., Arnell, J.C., Earl, A.J., and Lawlor, E.J. 1995. Clavulanic acid biosynthesis in *Streptomyces clavuligerus*: gene cloning and characterization. Gene **166**:49-55.

Hofmann, K., and Stoffel, W. 1993. TMbase – A database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler **374**:166.

Hopper, S., and Bock, A. 1995. Effector-mediated stimulation of ATPase activity by the sigma 54-dependent transcriptional activator FHLA from *Escherichia coli*. J. Bacteriol. **177**:2798-27803.

Horinouchi, S., Suzuki, H., Nishiyama, M., and Beppu, T. Nucleotide sequence and transcriptional analysis of the *Streptomyces griseus* gene (*afsA*) responsible for A-factor biosynthesis. J. Bacteriol. **171**:1206-1210.

Hoskisson, P.A., Rigali, S., Fowler, K., Findlay, K.C., and Buttner, M.J. 2006. DevA, a GntR-like transcriptional regulator required for development in *Streptomyces coelicolor*. J. Bacteriol. **188**:5014-5023.

Howarth, T.T., Brown, A.G., and King, T.J. 1976. Clavulanic acid, a novel β-lactam isolated from *Streptomyces clavuligerus*; X-ray crystal structure analysis. J. Chem. Soc. Chem. Commun. 266-267.

Hutchings, M.I., Hoskisson, P.A., Chandra, G., and Buttner, M.J. 2004. Sensing and responding to diverse extracellular signals? Analysis of the sensor kinases and response regulators of *Streptomyces coelicolor* A3(2). Microbiology **150**:2795-2806.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Omura, S. 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis.* Nat. Biotechnol. **21**:526-531.

Ishikawa, J. and Hotta, K. 1999. FramePlot: a new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. FEMS Microbiol. Lett. **174**:251-253.

Ishizuka, H., Horinouchi, S., Kieser, H.M., Hopwood, D.A., and Beppu, T. 1992. A putative two-component regulatory system involved in secondary metabolism in *Streptomcyes* spp. J. Bacteriol. **23**:7585-7594.

Jensen, S.E. 2004. (University of Alberta.) Personal communication.

Jensen, S.E., Elder, K.J., Aidoo, K.A., and Paradkar, A.S. 2000. Enzymes catalyzing the early steps of clavulanic acid biosynthesis are encoded by two sets of paralogous genes in *Streptomyces clavuligerus.* Antimicrob. Agents Chemother. **44**:720-726.

Jensen, R.A., and Gu, W. 1996. Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. J. Bacteriol. **178**:2161-2171.

Jensen, S.E., Paradkar, A.S., Mosher, R.H., Anders, C., Beatty, P.H., Brumlik, M.J., Griffin, A., and Barton, B. 2004a. Five additional genes are involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus.* Antimicrob. Agents Chemother. **48**:192-202.

Jensen, S.E., Wong, A., Griffin, A., and Barton, B. 2004b. *Streptomyces clavuligerus* has a second copy of the proclavaminate amidinohydrolase gene. Antimicrob. Agents Chemother. **48**:514-520.

Jin, W., Ryu, Y., Kang, S.G., Kim, S.K., Saito, N., Ochi, K., Lee, S.H., and Lee, K.J. 2004. Two *relA/spoT* homologous genes are involved in the morphological and physiological differentiation of *Streptomyces clavuligerus.* Microbiology **150**:1485-1493.

Jones, C., Thompson, A., and England, R. 1997. Guanosine 5'-diphosphate 3'-diphosphate (ppGpp) and clavulanic acid production in *Streptomyces clavuligerus.* World Journal of Microbiology & Biotechnology **13**:633-636.

Kamitori, S., Okamoto, A., Hirotsu, K., Higuchi, T., Kuramitsu, S., Kagamiyama, H., Matsuura, Y., and Katsube, Y. 1990. Three-dimensional structures of aspartate aminotransferase from *Escherichia coli* and its mutant enzyme at 2.5 Å resolution. J. Biochem. **108**:175-184.

Kelly, K.S., Ochi, K., and Jones, G.H. 1991. Pleiotropic effects of a *relC* mutation in *Streptomyces antibioticus*. J. Bacteriol. **173**:2297-2300.

Kenig, M., and Reading, C. 1979. Holomycin and an antibiotic (MM 19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. J. Antibiot. **32**:549-554.

Kershaw, N.J., McNaughton, H.J., Hewitson, K.S., Hernandez, H., Griffin, J., Hughes, C, Greaves, P., Barton, B., Robinson, C.V., and Schofield, C.J. 2002. ORF6 from the clavulanic acid gene cluster of *Streptomyces clavuligerus* has ornithine acetyltransferase activity. Eur. J. Biochem. **269**:2052-2059.

Khaleeli, N., Li, R., and Townsend, C.A. 1999. Origin of the β-lactam carbons in clavulanic acid from an unusual thiamine pyrophosphate-mediated reaction. J. Am. Chem. Soc. **121**:9223-9224.

Khodakaramian, G., Lissenden, S., Gust, B., Moir, L., Hoskisson, P.A., Chater, K.F., and Smith, M.C.M. 2006. Expression of Cre recombinase during transient phage infection permits efficient marker removal in *Streptomyces*. Nucleic Acids Res. **34**:e20.

Khokhlov, A.S., Tovarova, I.L., Borisova, L.N., Pliner, S.A., Schevchenko, I.L., Kornitskaka, E.Y., Ivkina, N.S., and Rapoport, I.A. 1967. A-factor responsible for the production of streptomycin by a mutant strain of *Actinomyces streptomycini*. Dokl. Akad. SSSR **177**:232-235.

Kieser, T., Bibb, M.J., Buttner, M.J., Charter, K.F., and Hopwood, D.A. 2000. Practical *Streptomyces* genetics. John Innes Foundation, Norwich, England.

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

Kim, H.S., Lee, Y.J., Lee, C.K., Choi, S.U., Yeo, S.H., Hwang, Y.I., Yu, T.S., Kinoshita, H., and Nihira, T. 2004. Cloning and characterization of a gene encoding the γ-butyrolactone autoregulator receptor from *Streptomyces clavuligerus*. Arch. Microbiol. **182**:44-50.

Kim, J.M., Yoshikawa, H., and Shirahige, K. 2001. A member of the YER057c/yjgF/Uk114 family links isoleucine biosynthesis and intact mitochondria maintenance in *Saccharomyces cerevisiae*. Genes Cells **6**:507-517.

Kinoshita, H., Tsuji, T., Ipposhi, H., Nihira, T., and Yamada, Y. 1999. Characterization of binding sequences for butyrolactone autoregulator receptors in streptomycetes. J. Bacteriol. **181**:5075-5080.

Kuchma, S.L., Connolly, J.P., and O'Toole, G.A. 2005. A three-component regulatory system regulates biofilm maturation and type III secretion in *Pseudomonas aeruginosa*. J. Bacteriol. **187**:1441-1454.

Kulasekara, H.D., Ventre, I., Kulasekara, B.R., Lazdunski, A., Filloux, A., and Lory, S. 2005. A novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial *cup* genes. Mol. Microbiol. **55**:368-380.

Kyte, J., and Doolittle, R. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105-132.

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

Leblond, P., Fischer, G., Francou, F., Berger, F., Guerineau, M., and Decaris, B. 1996. The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremeties of the linear chromosomal DNA. Mol. Microbiol. **19**:261-271.

Leblond, P., Francou, F.X., Simonet, J.M., and Decaris, B. 1990. Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofciens* strains. FEMS Microbiol. Lett. **72**:79-88.

Leblond, P., Redenbach, M., and Cullum, J. 1993. Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). J. Bacteriol. **175**:3422-3429.

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

Leskiw, B.K., Lawlor, E.J., Fernandez-Abalos, J.M., and Chater, K.F. 1991b. 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., Mah, R., Lawlor, E.J., and Chater, K.F. 1993. Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3(2). J. Bacteriol. **175**:1995-2005.

Lezhava, A., Mizukami, T., Kajitani, T., Kameoka, D., Redenbach, M., Shinkawa, H, Nimi, O., and Kinashi, H. 1995. Physical map of the linear chromosome of *Streptomyces griseus*. J. Bacteriol. **177**:6492-6498.

Li, R., Khaleeli, N., and Townsend, C.A. 2000. Expansion of the clavulanic acid gene cluster: identification and in vivo functional analysis of three new genes required for biosynthesis of clavulanic acid by *Streptomyces clavuligerus*. J. Bacteriol. **182**:4087-4095.

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

Little, R., and Dixon, R. 2003. The amino-terminal GAF domain of *Azotobacter vinelandii* NifA binds 2-oxoglutarate to resist inhibition of NifL under nitrogen-limiting conditions. J. Biol. Chem. **278**:28711-28718.

Liu, G., Tian, Y., Yang, H., and Tan, H. 2005. A pathway-specific transcriptional regulatory gene for nikkomycin biosynthesis in *Streptomyces ansochromogenes* that also influences colony development. Mol. Microbiol. **55**:1855-1866.

Lorenzana, L.M., Perez-Redondo, R., Santamarta, I., Martin, J.F., and Liras, P. 2004. Two oligopeptide-permease-encoding genes in the clavulanic acid cluster of *Streptomyces clavuligerus* are essential for production of the β-lactamase inhibitor. J. Bacteriol. **186**:3431-3438.

Madduri, K., and Hutchinson, C.R. 1995. Functional characterization and transcriptional analysis of the dnrR1 locus, which controls daunorubicin biosynthesis in *Streptomyces peucetius*. J. Bacteriol. **177**:1208-1215.

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

Marsh, E.N., Chang, M.D., and Townsend, C.A. 1992. Two isozymes of clavaminate synthase central to clavulanic acid formation: cloning and sequencing of both genes from *Streptomyces clavuligerus*. Biochemistry **31**:12648-12657.

McKenzie, A.K., Kershaw, N.J., Hernandez, H., Robinson, C.V., Schofield, C.J., and Andersson, I. Clavulanic acid dehydrogenase: structural and biochemical analysis of the final step in the biosynthesis of the β-lactamase inhibitor clavulanic acid. Biochemistry **46**:1523-1533.

McKenzie, N.L., and Nodwell, J.R. 2007. Phosphorylated AbsA2 negatively regulates antibiotic production in *Streptomyces coelicolor* through interactions with pathway-specific regulatory gene promoters. J. Bacteriol. **189**:5284-5292.

McNaughton, H.J., Thirkettle, J.E., Zhang, Z., Schofield, C.J., Jensen, S.E., Barton, B., and Greaves, P. 1998. β-lactam synthetase: implications for β-lactamase evolution. J. Chem. Soc. Chem. Commun. 2325-2326.

Mehta, P.K., Hale, T.I., and Christen, P. 1993. Aminotransferases: demonstration of homology and division into evolutionary groups. Eur. J. Biochem. **214**:549-561.

Mellado, E., Lorenzana, L.M., Rodriguez-Saiz, M., Diez, B., Liras, P., and Barredo, J.L. 2002. The clavulanic acid biosynthetic cluster of *Streptomyces clavuligerus*: genetic organization of the region upstream of the *car* gene. Microbiology **148**:1427-1438.

Mendez, C., Brana, A.F., Manzanal, M.B., and Hardisson, C. 1985. Role of substrate mycelium in colony development in *Streptomyces*. Can. J. Microbiol. **31**:446-450.

Merlin, C., McAteer, S., and Masters, M. 2002. Tools for characterization of *Escherichia coli* genes of unknown function. J. Bacteriol. **184**:4573-4581.

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

Miguelez, E.M., Garcia, M., Hardisson, C., and Manzanal, M.B. 1994. Autoradiographic study of hyphal growth during aerial mycelium development in *Streptomyces antibioticus*. J. Bacteriol. **176**:2105-2107.

Miller, M.T., Bachmann, B.O., Townsend, C.A., and Rosenzweig, A.C. 2001. Structure of β-lactam synthetase reveals how to synthesize antibiotics instead of asparagine. Nat. Struct. Biol. **8**:684-689.

Miller, I.M., Stapley, E.O., and Chaiet, L. 1962. Production of synnematin B by a member of the genus *Streptomyces*. Bacteriol. Proc. **49**:32.

Mistiniene, E., Pozdniakovaite, N., Popendikyte, V., and Naktanis, V. 2005. Structure-based ligand binding sites of protein p14.5, a member of protein family YER057c/YIL051c/YjgF. Int. J. Biol. Macromol. **37**:61-68.

Morishta, R., Kawagoshi, A., Sawasaki, T., Madin, K., Ogasawara, T., Oka, T., and Endo, Y. 1999. Ribonuclease activity of rat liver perchloric acid-soluble protein, a potent inhibitor of protein synthesis. J. Biol. Chem. **274**:20688-20692.

Mosher, R.H., Paradkar, A.S., Anders, C., Barton, B., and Jensen, S.E. 1999. Genes specific for the biosynthesis of clavam metablites antipodal to clavulanic acid are clustered with the gene for clavaminate synthase 1 in *Streptomyces clavuligerus*. Antimicrob. Agents Chemother. **43**:1215-1224.

Motamedi, H., Shafiee, A., and Cai, S.J. 1995. Integrative vectors for heterologous gene expression in *Streptomyces* spp. Gene **160**:25-31.

Murphy, K.C. 1998. Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. J. Bacteriol. **180**:2063-2071.

Musialowski, M.S., Flett, F., Scott, G.B., Hobbs, G. Smith, C.P., and Oliver, S.G. 1994. Functional evidence that the principal DNA replication origin of the *Streptomyces coelicolor* chromosome is close to the *dnaA-gyrB* region. J. Bacteriol. **176**:5123-5125.

Nagarajan, R., Boeck, L.D., Gorman, M., Hamill, R.L., Higgens, C.E., Hoehn, M.M., Stark, W.M., and Whitney, J.G. 1971. β-lactam antibiotics from *Streptomyces*. J. Am. Chem. Soc. **93**:2308-2310.

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

Nicholson, N.H., Baggaley, K.H., Cassels, R., Davison, M., Elson, S.W., Fulston, M., Tyler, J.W., and Woroniecki, S.R. 1994. Evidence that the immediate biosynthetic precursor of clavulanic acid is its *N*-aldehyde analogue. J. Chem. Soc. Chem. Commun. **1994**:1281-1282.

Nielsen, H., Engelbrecht, J., Bunak, S., and von Heijne, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering **10**:1-6.

Ochi, K. 1986. 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.

Ohnishi, Y., Kameyama, S., Onaka, H., Horinouchi, S. 1999. A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus*: identification of a target gene of the A-factor receptor. Mol. Microbiol. **34**:102-111.

Okamoto, A., Nakai, Y., Hayashi, H., Hirotsu, K., and Kagamiyama, H. 1998. Crystal structures of *Paracoccus denitrificans* aromatic amino acid aminotransferase: a substrate recognition site constructed by rearrangement of hydrogen bond network. J. Mol. Biol. **280**:443-461.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M. 2001. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. USA **98**:12215-12220.

Onaka, H., Anido, N., Nihara, T., Yamada, Y., Beppu, T., and Horinouchi, S. 1995. Cloning and characterization of the A-factor receptor gene from *Streptomyces griseus*. J. Bacteriol. **177**:6083-6092.

Onaka, H., and Horinouchi, S. 1997. DNA-binding activity of the A-factor receptor protein and its recognition DNA sequences. Mol. Microbiol. **24**:991-1000.

Oxelmark, E., Marchini, A., Malanchi, I., Magherini, F., Jaquet, L., Hajibagheri, M.A.N., Blight, K.J., Jauniaux, J.C., and Tommasino, M. 2000. Mmfip, a novel yeast mitochondrial protein conserved throughout evolution and involved in maintenance of the mitochondrial genome. Mol. Cell. Biol. **20**:7784-7797.

Paradkar, A.S., Aidoo, K.A., and Jensen, S.E. 1998. A pathway-specific transcriptional activator regulates late steps of clavulanic acid biosynthesis in *Streptomyces clavuligerus*. Mol. Microbiol. **27**:831-843.

Paradkar, A.S., and Jensen, S.E. 1995. Functional analysis of the gene encoding the clavaminate synthase 2 isoenzyme involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus*. J. Bacteriol. **177**:1307-1314.

Parsons, L., Bonander, N., Eisenstein, E., Gilson, M., Kairys, V., and Orban, J. 2003. Solution structure and functional ligand screening of HI0719, a highly conserved protein from bacteria to humans in the YjgF/YER057c/UK114 family. Biochemistry **42**:80-89.

Perez-Llarena, F.J., Liras, P., Rodriguez-Garcia, A., and Martin, J.F. 1997. A regulatory gene (*ccaR*) required for cephamycin and clavulanic acid production in *Streptomyces clavuligerus*: amplification results in overproduction of both β-lactam compounds. J. Bacteriol. **179**:2053-2059.

Perez-Redondo, R., Rodriguez-Garcia, A., Martin, J.F., and Liras, P. 1998. The *claR* gene of *Streptomyces clavuligerus*, encoding a LysR-type regulatory protein controlling clavulanic acid biosynthesis, is linked to the clavulanate-9-aldehyde reductase (*car*) gene. Gene **211**:311-321.

Perez-Redondo, R., Rodriguez-Garcia, A., Martin, J.F., and Liras, P. 1999. Deletion of the *pyc* gene blocks clavulanic acid biosynthesis except in glycerol-containing medium: evidence for two different genes in formation of the C3 unit. J. Bacteriol. **181**:6922-6928.

Pruess, D.L., and Kellett, M. 1983. Ro 22-5417, a new clavam antibiotic from *Streptomyces clavuligerus*. J. Antibiot. **36**:208-212.

Qin, Z., and Cohen, S.N. 1998. Replication at the telomeres of the *Streptomyces* linear plasmid pSLA2. Mol. Microbiol. **28**:893-903.

Rangaswamy, V., and Bender, C.L. 2000. Phosphorylation of CorS and CorR, regulatory proteins that modulate production of the phytotoxin coronatine in *Pseudomonas syringae*. FEMS Microbiol. Lett. **193**:13-18.

Rappu, P., Shin, B.S., Zalkin, H., and Mantsala, P. 1999. A role for a highly conserved protein of unknown function in regulation of *Bacillus subtilis purA* by the purine repressor. J. Bacteriol. **181**:3810-3815.

Raynal, A., Karray, F., Tuphile, K., Darbon-Rongere, E., and Pernodet, J.L. 2006. Excisable cassettes: new tools for functional analysis of *Streptomyces* genomes. Appl. Environ. Microbiol. **72**:4839-4844.

Redenbach, M., Kieser, H.M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., and Hopwood, D.A. 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. Mol. Microbiol. **21**:77-96.

Retzlaff, L., and Distler, J. 1995. The regulator of streptomycin gene expression, StrR, of *Streptomyces griseus* is a DNA binding activator protein with multiple recognition sites. Mol. Microbiol. **18**:151-162.

Rhee, S., Silva, M.M., Hyde, C.C., Rogers, P.H., Metzler, C.M., Metzler, D.E., and Arnone, A. 1997. Refinement and comparisons of the crystal structures of pig cytosolic aspartate aminotransferase and its complex with 2-methylaspartate. J. Biol. Chem. **272**:17293-17302.

Romagnoli, S., and Tabita, F.R. 2006. A novel three-protein two-component system provides a regulatory twist on an established circuit to moculate expression of the cbb_I region *Rhodopseudomonas palustris* CGA010. J. Bacteriol. **188**:2780-2791.

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

Salowe, S.P., Marsh, E.N., and Townsend, C.A. 1990. Purification and characterization of clavaminate synthase from *Streptomyces clavuligerus*: an unusual oxidative enzyme in natural product biosynthesis. Biochemistry **29**:6499-6508.

Salowe, S.P., Krol, W.J., Iwata-Reuyl, D., and Townsend, C.A. 1991. Elucidation of the order of oxidations and identification of an intermediate in the multistep clavaminate synthase reaction. Biochemistry **30**:2281-2292.

Sambrook, J., Fritsch, E.F., and Maniatas, T. 1989. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Santamarta, I., Perez-Redondo, R., Lorenzana, L.M., Martin, J.F., Liras, P. Different proteins bind to the butyrolactone receptor protein ARE sequence located upstream of the regulatory *ccaR* gene of *Streptomyces clavuligerus*. Mol. Microbiol. **56**:824-835.

Santamarta, I., Rodriguez-Garcia, A., Perez-Redondo, R., Martin, J.F., and Liras, P. 2002. CcaR is an autoregulatory protein that binds to the *ccaR* and *cefD-cmcI* promoters of the cephamycin C-clavulanic acid cluster in *Streptomyces clavuligerus*. J. Bacteriol. **184**:3106-3113.

Sardiwal, S., Kendall, S.L., Movahedzadeh, F., Rison, S.C.G., Stoker, N.G., and Djordjevic, S. 2005. A GAF domain in the hypoxia/NO-inducible *Mycobacterium tuberculosis* DosS protein binds haem. J. Mol. Biol. **353**:929-936.

Scarsdale, J.N., Radaev, S., Kazanina, G., Schirch, V., and Wright, H.T. 2000. Crystal structure at 2.4 Å resolution of *E. coli* serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. J. Mol. Biol. **296**:155-168.

Schmitz, G., and Downs, D.M. 2004. Reduced transaminase B (IlvE) activity caused by the lack of yjgF is dependent on the status of threonine deaminase (IlvA) in *Salmonella enterica* serovar *Typhimurium*. J. Bacteriol. **186**:803-810.

Schrempf, H., Bujard, H, Hopwood, D.A., Goebel, W. 1975. Isolation of covalently closed circular deoxyribonucleic acid from *Streptomyces coelicolor* A3(2). J. Bacteriol. **121**:416-421.

Schweizer, H.P. 2003. Applications of the *Saccharomyces cerevisiae* Flp-FRT system in bacterial genetics. J. Mol. Microbiol. Biotechnol. **5**:67-77.

Senecoff, J.F., Bruckner, R.C., and Cox, M.M. 1985. The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. Proc. Natl. Acad. Sci. USA **82**:7270-7274.

Sola-Landa, A., Moura, R.S., and Martin, J.F. 2003. The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*. Proc. Natl. Acad. Sci. USA **100**:6133-6138.

Song, H., and Niederweis, M. 2007. Functional expression of the Flp recombinase in *Mycobacterium bovis* BCG. Gene doi:10.1016/j.gene.2007.05.005.

Springer, T.A. 1996. Folding of the N-terminal, ligand-binding region of integrin α-subunits into a β-propeller domain. Proc. Natl. Acad. Sci. USA **94**:65-72.

Sprusansky, J., Zhou, L., Jordan, S., and Westpheling, J. 2003. Identification of three new genes involved in morphogenesis and antibiotic production in *Streptomyces coelicolor*. J. Bacteriol. **185**: 6147-6157.

Stackebrandt, E., Rainey, F.A., and Ward-Rainey, N.L. 1997. Proposal for a new hierarchic classification system, *Actinobacteria*, classis nov. Int. J. Syst. Bacteriol. **47**:479-491.

Stephen, J., Stemmer, V., and Niederweis, M. 2004. Consecutive gene deletions in *Mycobacterium smegmatis* using the yeast FLP recombinase. Gene **343**:181-190.

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

Stuttard, C. 1982. Temperate phages of *Streptomyces venezuelae*: lysogeny and host specificity shown by phages SV1 and SV2. J. Gen. Microbiol. **128**:115-121.

Stutzman-Engwall, K.J., Otten, S., and Hutchinson, C.R. 1992. Regulation of secdondary metabolism in Streptomyces spp. and overproduction of daunorubicin in *Streptomyces peucetius*. J. Bacteriol. **174**:144-154.

Tahlan, K. 2005. Characterization of the carboxyethylarginine synthase genes involved in clavulanic acid and 5*S* clavam metabolite biosynthesis in *Streptomyces clavuligerus*. Ph.D. Thesis, University of Alberta, Edmonton, Alta.

Tahlan, K., Anders, C., and Jensen, S.E. 2004a. The paralogous pairs of genes involved in clavulanic acid and clavam metabolite biosynthesis are differently regulated in *Streptomyces clavuligerus*. J. Bacteriol. **186**:6286-6297.

Tahlan, K., Anders, C., Wong, A., Mosher, R.H., Beatty, P.H., Brumlik, M.J., Griffin, A., Hughes, C., Griffin, J., Barton, B., and Jensen, S.E. 2007. 5*S* clavam biosynthetic genes are located in both the clavam and paralog gene clusters in *Streptomyces clavuligerus*. Chemistry & Biology **14**:131-142.

Tahlan, K., Park, H.U., Wong, A., Beatty, P.H., and Jensen, S.E. 2004b. Two sets of paralogous genes encode the enzymes involved in the early stages of clavulanic acid and clavam metabolite biosynthesis in *Streptomyces clavuligerus*. Antimicrob. Agents Chemother. **48**:930-939.

Tahlan, K., Park, H.U., and Jensen, S.E. 2004c. Three unlinked gene clusters are involved in clavam metabolite biosynthesis in *Streptomyces clavuligerus*. Can. J. Microbiol. **50**:803-810.

Takano, E., Gramajo, H.C., Strauch, Andres, N., White, J., and Bibb, M.J. 1992. Transcriptional regulation of the *redD* transcriptional activator gene accounts for

growth-phase-dependent production of the antibiotic undecylprodigiosin in *Streptomyces coelicolor* A3(2). Mol. Microbiol. **6**:2797-2804.

Tang, L., Grimm, A., Zhang, Y.X., and Hutchinson, C.R. 1996. Purification and characterization of the DNA-binding protein DnrI, a transcriptional factor of daunorubicin biosynthesis in *Streptomyces peucetius*. Mol. Microbiol. **22**:801-813.

Tipper, D.J., and Strominger, J.L. 1965. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. Proc. Natl. Acad. Sci. USA **54**:1133-1141.

Trepanier, N.K., Jensen, S.E., Alexander, D.C., and Leskiw, B.K. 2002. The positive activator of cephamycin C and clavulanic acid production in *Streptomyces clavuligerus* is mistranslated in a bldA mutant. Microbiology **148**:643-656.

Tseng, H.C., and Chen, C.W. 1991. A cloned *ompR*-like gene of *Streptomyces lividans* 66 suppresses defective defective *melC1*, a putative copper-transfer gene. Mol. Microbiol. **5**:1187-1196.

Ullrich, M., Penaloza-Vazquez, A., Bailey, A.M., and Bender, C.L. 1995. A modified two-component regulatory system is involved in temperature-dependent biosynthesis of the *Pseudomonas syringae* phytotoxin coronatine. J. Bacteriol. **177**:6160-6169.

Valentine, B.P., Bailey, C.R., Doherty, A., Morris, J., Elson, S.W., Baggaley, K.H., and Nicholson, N.H. 1993. Evidence that arginine is a later metabolic intermediate than ornithine in the biosynthesis of clavulanic acid by *Streptomyces clavuligerus*. J. Chem. Soc. Chem. Commun. 1210-1211.

Volz, K. 1999. A test case for structure-based functional assignment: The 1.2 Å crystal structure of the *yjgF* gene product from *Escherichia coli*. Protein Science **8**:2428-2437.

Vujaklija, D., Horinouchi, S., and Beppu, T. 1993. Detection of an A-factor responsive protein that binds to the upstream activation sequence of *strR*, a regulatory gene for streptomycin biosynthesis in *Streptomyces griseus*. J. Bacteriol. **175**:2652-2661.

Waksman, S.A., and Henrici, A.T. 1943. The nomenclature and classification system of the actinomycetes. J. Bacteriol. **46**:337-341.

Ward, J.M., and Hodgson, J.E. 1993. The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster' in three *Streptomyces*. FEMS Microbiol. Lett. **110**:239-242.

Weber, J.M., Schoner, B., and Losick, R. 1989. Identification of a gene required for the terminal step in erythromycin A biosynthesis in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). Gene **75**:235–241.

Wehmeier, U. 1995. New multifunctional *Escherichia coli-Streptomyces* shuttle vectors allowing blue-white screening on XGal plates. Gene **165**:149-150.

White, J., and Bibb, M.J. 1997. *bldA* dependence of undecylprodigiosin production in *Streptomyces coelicolor* A3(2) involves a pathway-specific regulatory cascade. J. Bacteriol. **179**:627-633.

Wietzorrek, A., and Bibb, M. 1997. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. Mol. Microbiol. **25**:1181-1184.

Wildermuth, H. 1970. Development and organization of the aerial mycelium in *Streptomyces coelicolor*. J. Gen. Microbiol. **60**:43-50.

Wise, E.M., and Park, J.T. 1965. Penicillin: its basic site of action as an inhibitor of a peptide cross-linking reaction in cell wall mucopeptide synthesis. Proc. Natl. Acad. Sci. USA **54**:75-81.

Wu,T.K., Busby, R.W., Houston, T.A., McIlwaine, D.B., Egan, L.A., and Townsend, C.A. 1995. Identification, cloning, sequencing, and overexpression of the gene encoding proclavaminate amidino hydrolase and characterization of protein function in clavulanic acid biosynthesis. J. Bacteriol. **177**:3714-3720.

Yamazaki, H., Takano, Y., Ohnishi, Y., and Horinouchi, S. 2003. *amfR*, an essential gene for aerial mycelium formation, is a member of the AdpA regulon in the A-factor regulatory cascade in *Streptomyces griseus*. Mol. Microbiol. **50**:1173-1187.

Yamazaki, H., Tomono, A., Ohnishi, Y., and Horinouchi, S. 2004. DNA-binding specificity of AdpA, a transcriptional activator in the A-factor regulatory cascade in *Streptomyces griseus*. Mol. Microbiol. **53**:555-572.

Zdobnova, E.M., and Apweiler, R. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17**:847-848.

Zhang, Z., Ren, J., Stammers, D.K., Baldwin, J.E., Harlos, K., and Schofield, C.J. 2000. Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. Nat. Struct. Biol. **7**:127-133.

VI. APPENDIX

VI.1 DNA sequences[a,b]

DNA sequence of *orfA*:

```
1     atggatgtcc tggccgcgtt ggagcgcaag cccagcctga atcttttccc    50
51    catcgagaac cggctgtcgc cgcgcgccag tgccgcgctg gccaccgacg   100
101   ccgtcaaccg ctatccgtac tccgagaccc cggtggccgt ctacggcgat   150
151   gtcacggggc tggccgaggt gtacgcgtac tgcgaggacc tggccaagcg   200
201   cttcttcggg gcgcgccacg ccggtgtgca gttcctgtcc ggtctgcaca   250
251   ccatgcacac cgtgctgacc gccctgaccc cgcccggcgg gcgcgtcctg   300
301   gtcctcgcgc cggaggacgg cggccactac gccacggtga cgatctgccg   350
351   gggcttcggc tacgaggtcg agttcttacc cttcgaccgc cggacactgg   400
401   agatcgacta cgcggtcctg gctgcgcgcc tctcccggcg gccggccgac   450
451   gtgatctatc tggacgcgtc cagcatcctg cgcttcatcg acgcgcgggc   500
501   gctgcggctg gccgctcctg acgcgctgat ctgcctggac gcgtcgcaca   550
551   tcctcggtct gctgccggtg gcgccgcaga ccctggtgct cgacggcggc   600
601   ttcgacagca tctccggcag tacgcacaag acgttccccg gcccgcagaa   650
651   ggggctgctc gtcaccgaca gcgatgtcgt cgcggagaag gtcgcggcac   700
701   ggatgccgtt cacggcgtcg agttcgcact cggcgagcgt gggttcgctg   750
751   gcgatctcgc tggaggagct gctgccgcac cggacggcgt acgcgcacca   800
801   ggtcatcgcg aacgcccggg cgctggccgg gctgctcgcc gagcgcggtt   850
851   tcgacgtcgc gggcggggcg ttcggccata cggacaccca ccaggtctgg   900
901   gtgcacttcc ccgaggggaa cacgccgcac gagtgggggc ggctgctcac   950
951   ccgcgcgaac atccgctcga ccagcgtcgt cctgccgagc agcgcggcgc  1000
1001  ccgggctgcg gctcgggacg caggagctga cgcgctgggg gatgacggag  1050
1051  acggacatgg cgccggtggc cgatctgctg gaacggctgc tgctgcgggg  1100
1101  cgacgacgcg gagaccgtgg cgaaggaggt cgtggaactc gcccgggcct  1150
1151  tcccgggcgt ggcgttcgtc tga                               1173
```

DNA sequence of region downstream of the paralogue gene cluster:

```
1     tgcgttgcct gatcgtcgac gacagtccgg tcttcctggc cgccgccagc    50
51    gggttactgc ggcatcaggg catcgtcgtc gtctgtgtcg cggccaatgg   100
101   cgcggaggcg ctgcgggggg ccgagcggga gcggccggac gtggcgctgg   150
151   tggacctcga cctcgggggc gagagcgggc tggatctggc cgagcggctg   200
201   taccgccgat ccgtcccgac gattctgatg tccacccatg ccgagcagga   250
251   ctaccgggac ctgatcaccg ccagtcctgc catcggtttc ctgcccaaga   300
301   tgtcgctgtc ggggtgtgtc atcgtccgcc tgctcgacgg cgccg[redacted]   350
351   [redacted]tc acggcgagta cccgccggtg gtcgtccgcc   400
401   gcctccggca gacccagctt ggccaggagg ctgcggacgt gtttctccac   450
451   ggtgccctcg gtgatccgca gtctgcgggc gatcccggcg ttggaacggc   500
501   cctcggccat cagtgccagc acctcccgct cccgtacgct cagcgcctcc   550
551   agcgggtcct gacggcgccg caccgacacc agttcctgca ctaacgccgg   600
601   gtcgaccacc gagccgcccc gggccacccg ctccagggtg tcgaggaact   650
651   ccgccacgtc ggccacccgg ctcttcagca ggtatccgat gccctgcccg   700
701   ctggccagca gttccatcgc gtgctccacc tcggcgtggg ccgagagcac   750
751   caggatgccg atcccgggca gctcctgccg gatcacccgg gccgccgcca   800
801   gcccctccgt ggtgtgcgtc ggcggcatcc ggatgtccac gatcaccagg   850
851   tcgggccggt gccgccgcac caggggcagc agccccgtcg cgtccccggc   900
901   ctgtcccacc accctatgtc cgcaccgttc aagcaggctg gccaacccct   950
951   cgcgcagcag aacgtcgtcc tcggccagta ctaccctcac tcctgccatg  1000
1001  acgacattat cgactacgca acggacttgt tcggctacgg tttgcgcgtc  1050
1051  tccttgtcct ggacacacat ctccgtacga agacctcatc ccggtgcggc  1100
```

```
1101  gtccgccccg tggctgcgtg atgaccatct gttcgaaccg tgatcggtag  1150
1151  catgcggact tgatcacttc ggtcaactcg cgtgtcacag cgcgtacttg  1200
1201  gcgcaccatc gacgacaggg gtgcagtatg gaacgtttcc gcggcaggac  1250
1251  cgtcctcatc accggcggca ccagcggcat cggtctggcg accgggcacc  1300
1301  gcctggtcgc cgagggtgct caggtcatcg tcacgggccg taccggtgaa  1350
1351  cgcgtcgacg ccgccgtcgg ggaactcggc ccccaggccg gaggtgttgt  1400
1401  cgcggacacc ggtgacctcg gtgccgtcga ggcgctgatg gagaccgtcc  1450
1451  gtgagcggca cggccgtctg gacagcctct tcgtgaacgc gggcacgggc  1500
1501  acgatcgtcc cgttcgagga catcacggag gtggatttcg accatgccgt  1550
1551  gaacgtcaac ctcaagggcg tcttcttcac cgtccaaaga gccctgccgc  1600
1601  tgctgacgaa cggcggctcg atcgtcatca acgcgtcctg gacgatccac  1650
1651  cggggcaaca gcgccctgac gctctactcg gcgaccaagg ccgccgcgca  1700
1701  caatctggcc cgtaccctgg cggcgtccct ggccccccgg ggcatccggg  1750
1751  tgaactccgt cagccccggg tacatcgaca ccccgatgta cccggaggcc  1800
1801  gctctgaccc cggcggaggc cgagaccctc accggccgtg tcgtcgcggg  1850
1851  ccgcttcggc cgtccggagg agatcgcggc ggccgtcgcc ttcctcgcct  1900
1901  cgtccgacgc gtcctacgtc aacggccagg acctggtggt cgacggcggt  1950
1951  ctgatcggcg cgatccccgg ctgacccggc cccgggcgaa cggccgtgtg  2000
2001  ccggctccgg ctccgcgtgc ggtagagtga tcatgtgctc gcgaccggga  2050
2051  cagccggtac gggaggcgcg cgttggtggt ccaaggaaag acaccccact  2100
2101  tcccgtgggg ggatgcaggt gcaaggcctg cccagcgctc cagacgaagg  2150
2151  ccctgatctc cggacgacgg agatcagggc cttcctcgtg ccgctcacac  2200
2201  ccgcggggcc ggtccgccgg agcgaccgga gccgctggct gtcaggcgag  2250
2251  gtcgctggcg taccagttcc agctcagggc cccggcgctg tcccaggacc  2300
2301  tgaccggcgc gttgaagccc gtcccggtcc cggcgaactt ccacaggccg  2350
2351  gtgcggttgc tgccgtcgtc ctggcggccg tagttgtaga ggacgccgat  2400
2401  gtcggccttg ccgtcgccgt cgaagtcccc ggcggtcagc ttgctcgcgt  2450
2451  cggcgttcca gcttccgtcg ccggagtccc acttcctgga cgggttggtg  2500
2501  aagtccgtgc ccgtgctggt gaaggtccac agcgcggact tgttgcgtcc  2550
2551  gtcggcggtc tggccgttgt cgtagaggac accgacgtcc gtcttgccgt  2600
2601  ccccggagaa gtcggcggcg acgaccttgg agcggtccca gttccaggag  2650
2651  ccggtgccgg tcgagacgtt gtcccacttc ttgaccgggt tgccgaagtc  2700
2701  ggtgccggtg ctggtgctgg tgagcgtcca cagggcggtg acgttgacgt  2750
2751  tgtcggcgtt ctggccgttg ttgtagagga tgccgacatc ggcgttgccg  2800
2801  tcgccgttga agtcaccgga taccggcttg gagcggtccc agttccagga  2850
2851  gccggagttg gtggagacgt tgtcccactt cctgaccggc tgggcgaagg  2900
2901  cggtgccgtt gctggtgaac gtccacagcg cggtcttgtt ggtgccgtcg  2950
2951  gaggcttggc cgttgttgta caggacgccg acgtcgtccc ggccgtcgcc  3000
3001  gttgtagtcg cccgtggtga ccttggagcg gtcccaggtc caggagtcgg  3050
3051  ccgaggtggt gctgtaccac ttcttggacg gttcggcgaa gtccgagccc  3100
3101  gtgctgagga aggtccacag cgccgccgcc ggggtgccca tctgcgaggt  3150
3151  gccctcgtcg tagaggacgc cgatgtcgtc cttgccgtcg ccgttgaagt  3200
3201  cgcccgtggc gaccttcgaa cgatcccagc tccaggagtt gacgggcgcc  3250
3251  gagacgctgt cccacttctt gaccggcggg ccgaagcccg aggcggtgga  3300
3301  ggtgaacgtc cacagcgcgg tccggtgggt gccgtcggcg tttctgccgt  3350
3351  tgtcgtacag gacggcgaca tcgtcccggt tgtcgccgtt gaagtcaccc  3400
3401  gtggtgcggg aggtgttggc gcggttggcg gtgaggaaag agaccgcgtc  3450
3451  ggacatcggt cccgtctgcc cggagtcgtt cacggcccgg acggtgaccg  3500
3501  tgtgcgcgcc gtcgctgagc gtcgtcgtgt cgagcgatat ggagtagtgg  3550
3551  gggcccgcgc cggtgtccgt gccgacggcg gcaccgtcga catagaaggt  3600
3601  cgccttggtg ggcttgccga cggggttggt cacctgcgcg ccgagggtga  3650
3651  tggtgcccgt cgcgacggtg cccgatgcca ggccggtgac cttcgagacc  3700
3701  ggcggtacgg ggtcggggta cgggtcggcg ttgtccgcgc ccaggttcgc  3750
3751  ggcggggccg cagtgcggcc aggcgccctg gccctggacc gccaggacct  3800
3801  tctcggcggt gaggatctgc tccctcttgg tggcgaggtc ggcgcggggc  3850
3851  gcgtagtcca gaccgccgaa gcccgcccag gtcggctggg agaactggag  3900
3901  gccgccgtag tagccgttgc cggtgttgat ggaccagttg ccgccggact  3950
```

```
3951  cgcactgggc gaccttgtcc caggtcgcga cggaggcggc tgcggcgggc  4000
4001  ggggcggtga gcagaccggt ggccaggccg gccgagacgg cgagggcggt  4050
4051  ggcgacggtg agcgcgcggg ggctgcggga gggtgcggtg actctcatcg  4100
4101  tggtgactcc aggggacagc ggggcggggg cggggcggga cgcccggccg  4150
4151  gtggaccgga gcgggtccgg tcggccggcc gggcgtctgc gcgggggcgc  4200
4201  gtggtgcgtg cgtgcggggg ctgtgccgcg tgcgtgccgg ggcggtgtca  4250
4251  gccgagcttg ctggcgtacc agttccagct cagggccccg gcgctgtccc  4300
4301  aggacttgac cggcgcgttg aagcccgtcc cgttgctgga gaacttccac  4350
4351  acggcggtgc ggttggtgcc gtcgtcctgg cggccgtagc cgtagaggac  4400
4401  gccgatgtcg gtcctgccgt cgccgttgaa gtccccggcg gtcagcttgc  4450
4451  tggtgtccga gttccagctg tcgtcgccgg agtcccactt cttggagggg  4500
4501  ttggtgaagt cggtgccggt gctggtgagc gtccacagcg cggacttgtt  4550
4551  ccggctgtcc tcggtctggc cgttgttgta gaggatgccg acgtccgtct  4600
4601  tgccgtcgcc gttgaagtca ccggcggtga gcttggagcg gtcccaggtc  4650
4651  cacgagccga agttcgtcga cgcgttgtcc cacttcttgg acgggttggt  4700
4701  gaagtcgctg ccggtgctgg cgaaggtcca cagcgcggtc gtggaggtgc  4750
4751  cgtcggcgga ccggccgttg tcgtacagga cacccacgtc cgtcttgccg  4800
4801  tcgccggtga agtcgccgga ggccatcttg gagcggtccc agttccagga  4850
4851  gccggagttg gtcgagacgt tgtcccactt cttggagggg ttggtgaagt  4900
4901  cgctgccggt gctggtgagc gtccacagcg cggtgacgtt gacgttgtcg  4950
4951  gcgttctggc cgttgttgta gaggacgccg acgtcgtcct tgccgtcggc  5000
5001  gttgaagtcg ccggaggtga gcttcgaccg gttccagttc caggagccgg  5050
5051  agttggtgga gacgttgtcc cacttcttga ccggcgcgtt gaagccggtc  5100
5101  ccgttgctgg tgaaggtcca cagggcggtg acgttgacgt tgtcggcgtt  5150
5151  ctggccgttg ttgtagagga cgccgacgtc gtccttctcg tccccgttga  5200
5201  agtcacctgt ggtgatcttg gagcggtccc agttccagga gccgaagccg  5250
5251  gacgagacgt tgtcccagac cttcacgggc tggttgaaac cggtgccgtt  5300
5301  gctcagatac gtccacagac cggtgcggta gctgctgtcg gcctcctgcc  5350
5351  cgtagtcgta caggacaccg atgtcggcct tgccgtcccc gttgaagtcc  5400
5401  gcgtggaagg tagaggtggc ctcgccgttg gagatgatga tctgggcccg  5450
5451  ggaggcgccc gtgttcccgt ggatgtccgt cacctccgcc gagacggtgt  5500
5501  gcttgccgtc ggggaagcgg ctggtgtcga gggcgaaggt gtacggggac  5550
5551  cgggtcagct tgccgctgac cggctcgccg tcgacgtagt actgcacgga  5600
5601  cgccacgtcg taggtaccgg ccgtgacgga cgcggtcagg ttcgcgacac  5650
5651  cggagaccgt cccggaggca ccgccgaccg agacgcgcgg cggtgtggcc  5700
5701  cgctggagct tgggcgcgat gtagccgctg atgcgctggc cccaggggga  5750
5751  gtcgcccacc cggtaggagg tcgtcgagtg ggtggagacc atgccctcgc  5800
5801  ggttgaccga gtggctctgg ttgccgccgg tgatggtgac gaccccgttg  5850
5851  ctgacgccgg tcacgatggc gacgtgctgg gcccagccgt tggcgtagcc  5900
5901  gtacacgacg gcgtcaccga cgtccggggt gctgtcgagg gtgccgtacc  5950
5951  gctgtccgta gtggtagaag ctggacgcgg cgtcggtcag catgctctgg  6000
6001  ccctgcacat tgaagcgcga ccagacccag cccacgaagt cggcgcacca  6050
6051  ggcgtgggcc tggccgccgt ggccgtcgca gctgttgttc tggttgaggc  6100
6101  cgttgtcgta cccgccgccc gcgcaggggc cgttgccgtt ctggctcttg  6150
6151  gcgagggccg ccatgtccgt gaactcgtcc gcgtgggccg gggcggtggt  6200
6201  gaggccggtg agcgcggcgg cgacgacgag agaggcgagg ccggttctga  6250
6251  gggagcggtg cagggcggga cgggacatgg aggtactcct ggtgagagac  6300
6301  atggaggtgc cctgaggaga ggacgccgag gaggtgcggc agggcgtgga  6350
6351  gcccaaggct ggtgcgcggc ggcggtcgcg gtgttgtccg ccgccgccct  6400
6401  cgtgccatca gcttgcgtcc gctcgatggc ctgcgggaag gagtaccggg  6450
6451  tgtctcgaag gtgcctgaca gcttcgtgcc agggagccgg cgggcacaac  6500
6501  gcgccgcacc ccggcggcgc cctgcgatgt cgtcccgggg agggacggca  6550
6551  gggcaccgcc ggggtgcggg cggcggagcg gtggtgggtc ggtggcggac  6600
6601  cgggcggggc cggtccgtcc cgggtcagga ctggaggggc gagccattca  6650
6651  gctcccaggt gacggggccc tgggtgtcgc cttcgaggtc gtagacgacc  6700
6701  gtcgcgtcgg tggaccacca caggatggtg tccaccgggt tgtccaccgg  6750
6751  gatggcgtgg cagttcaggg tggtggcgtg accccagtgc gcgttgaggt  6800
```

```
6801  tggcggtgaa gtcgcaggtc ctggaggacc gcacggtgtc cgcgccggtc  6850
6851  tggccgggac ggctctccag ccgggtggtg accttgaagg aggtgaagcg  6900
6901  cttgctccgg tcgatgacct ggtccaccag gatcatccag ttgatcccgg  6950
6951  tccggccgga gacccaggcc ggtcccgtcg ccggggcgtg cagctgcccg  7000
7001  cagttctcga tgtacacgtg cgtgtcgggc ttgtccccgg gcaggtccag  7050
7051  atcgcggccg ttggcgccac tcgacgtgct gcacttctcc gggcccgccg  7100
7101  cggccacggc gggcgtaccg gtggtggtcg cggaggccgg ggcggcgaac  7150
7151  gcggtcccac tggtcagaag ggccagcgtg gcggcggccg tggagagctt  7200
7201  gagggacgtg gtgcgggaac ggcgcatgga atctcctgag aaatgtggtg  7250
7251  cgtttgtcgg atgttttccg gtgagcgacc ggtattccgg aaggacccca  7300
7301  ccggcggtga gccgggcggc cggggaggta cgggcgctgt cgtggggtgt  7350
7351  gccgcggttt ccgtggggtg cgccgcgctt tccgcggcgt tcggtccgcc  7400
7401  ccgggtcaga cctgctgcgg cgagccttcc agctcccagg tgatgggccc  7450
7451  ctgggtgtcg ccttcgaggt cgtagacgat cgtcgcgtcg gtggaccacc  7500
7501  agaggatggt gtccaccggg ttgtccaccg gcagggcgta gcagtacagg  7550
7551  ttcatgccgc ttaagtcggc gttaaggttg gcggtgaagt cgcaggtcct  7600
7601  ggaggactgc acggtgtccg tgccggtctg gctggggcgg ctctccagcc  7650
7651  gggtggtgac cttgaaggag gtgaagcgct tgctccgctc             7690
```

[a] ORF sequences are highlighted in light gray and areas where ORFs overlap are highlighted in dark gray

[b] start codons are underlined and stop codons are bold