National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

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

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

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

AVIS

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

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

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

Canada

UNIVERSITY OF ALBERTA

HIGH-LEVEL EXPRESSION IN *ESCHERICHIA COLI* AND FUNCTIONAL CHARACTERISATION OF ISOPENICILLIN N SYNTHASE FROM *STREPTOMYCES CLAVULIGERUS*

BY

MICHEAL DURAIRAJ

A THESIS

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

DEPARTMENT OF MICROBIOLOGY

EDMONTON, ALBERTA

SPRING, 1995

ISBN 0-612-01684-6

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: MICHEAL DURAIRAJ

TITLE OF THESIS: HIGH-LEVEL EXPRESSION IN *ESCHERICHIA COLI* AND FUNCTIONAL CHARACTERISATION OF ISOPENICILLIN N SYNTHASE FROM *STREPTOMYCES CLAVULIGERUS*

DEGREE FOR WHICH THIS WAS PRESENTED: DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED: 1995

.....

(Student's signature)

Student's permanent address:

M. Michealpattanam (P.O)

Bamb[illegible]r (Via), Ramnad (Dt)

Tamilnadu-623 706, INDIA.

Date: 25/1/95

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled HIGH-LEVEL EXPRESSION IN *ESCHERICHIA COLI* AND FUNCTIONAL CHARACTERISATION OF ISOPENICILLIN N SYNTHASE FROM *STREPTOMYCES CLAVULIGERUS* submitted by Micheal Durairaj in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

.....

Dr. S. E. Jensen-Supervisor

.....

Dr. William Paranchych

.....

Dr. M. A. Pickard

.....

Dr. John Bell

.....

Dr. David H. Sherman-External Examiner

Date: January 12, 1995.

DEDICATION

I would like to dedicate this thesis to Dr. Susan E. Jensen, my supervisor as a mark of my appreciation for the concern and care given to me and my family during my Ph.D program.

ABSTRACT

Isopenicillin N synthase (IPNS), which catalyses the oxidative cyclisation of the acyclic tripeptide δ-(L-α- vl)-L-cysteinyl-D-valine into isopenicillin N, is produced at low levels in *Stre ces clavuligerus*. In order to obtain large amounts of IPNS for crystallographic studies, the gene encoding IPNS (*pcbC*) was subcloned into *Escherichia coli* using the pT7 series of plasmid vectors. When the native ribosome binding site of *pcbC* was used to provide translation initiation signals, transcription but no translation was observed. The polymerase chain reaction was used to introduce an *Nde*I restriction endonuclease site into the translation initiation codon of *pcbC*, allowing the gene to be inserted behind an *E. coli*-type ribosome binding site. This construct directed high-level expression of IPNS, yielding protein which accumulated predominantly in an inactive form in inclusion bodies.

Inclusion bodies were solubilised by treatment with high concentrations of urea under reducing conditions. Optimisation of refolding conditions to recover active IPNS revealed that a dialysis procedure carried out at a protein concentration of approximately 1mg/mL at a pH of 8.0 yielded maximal recovery of active IPNS. As a result of these optimisation studies, the folding efficiency increased from an initial value of less than 30% to nearly 70%. Pure solubilised IPNS of more than 95% purity was obtained by passing this solubilised material through a DEAE-Trisacryl ion exchange column. The recovery rate of active, pure solubilised IPNS during purification was nearly 60% of the starting material.

Soluble IPNS was required for comparison with the solubilised IPNS to ensure that the refolded material resembled native IPNS. Expression attempts at different temperatures indicated that IPNS was produced predominantly in a soluble form when expression was conducted at 20°C. The extent of expression was about 20% of the total soluble protein. This high-level production of soluble IPNS facilitated the purification of soluble IPNS to near homogeneity in four steps. Comparison of the solubilised and

soluble IPNS by activity measurements and CD studies revealed that they are very similar and so either form should be suitable for crystallographic investigations

In order to avoid the possibility of oligomer formation during crystallisation, the four cysteine residues present in the IPNS protein were mutated individually into serine residues by PCR and single-strand site-directed mutagenesis. Functional analysis of these single mutants showed that cysteine 104 was very important while the remaining cysteine residues at positions 37, 142 and 251 were not critical for activity. Activity analysis of a solubilised preparation of an IPNS triple mutant (C37S, C142S and C251S) showed that it had lost more than 90% of its activity, presumably due to structural alterations. This suggested that these three cysteine residues are involved in folding. However, a soluble preparation of IPNS from the same triple mutant retained more than 50% of the wild type activity. It is possible that when folding occurs in vivo, the folding defect in the triple mutant is suppressed leading to the formation of more active enzyme. The large amounts of solubilised wild type IPNS, soluble wild type IPNS and triple mutant IPNS available from this study will be useful for crystallisation studies.

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor Dr. S.E. Jensen, for her constant guidance, encouragement and support offered during my graduate studies. I am unable to find proper words to express my thanks for her endless endurance in dealing with me during my studentship.

I would like to thank Dr. Brenda K. Leskiw for providing laboratory space and for her guidance towards my research project. I would also like to express my thanks to my supervisory committee members Dr. William Paranchych and Dr. M.A. P[illegible] for their suggestions and constructive criticism during my graduate studies.

I am thankful to all the staff members and non-acadmic staff of the then, Department of Microbiology for helping me in every possible way with regard to my research work and to K. Oikawa and R.K. Sherbourne for their assistance.

I am very thankful to Dr. Astrid K. Petrich, Annie Wong, Richard Mah and Farzana Damji for offering their assistance on many occasions in a variety of ways. Dr. Ashish Paradkar and Dr. Roy Mosher deserve my special thanks for their friendship which made my life easier.

I would like to thank the Canadian Commonwealth Committee, the Government of Canada as well as the Government of India for the Commonwealth Scholarship provided to pursue my Ph.D program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACV	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine
bis-ACV	ACV in dimer form
CD	Circular Dichroism
DTT	Dithiothreitol
DMSO	Dimethyl sulfoxide
DEAE	Diethylaminoethyl
EDTA	Ethylenediaminetetraaceticacid
HPLC	High Performance Liquid Chromatography
IPNS	Isopenicillin N synthase
IPTG	Isopropyl-β-D-thiogalactopyranoside
LX112	Embedding resin
NEM	N-ethylmaleimide
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pcbC	Gene encoding IPNS
SDS	Sodium dodecyl sulfate
Tris	Tris (hydroxymethyl) aminomethane
TDEG	Tris-DTT-EDTA-Glycerol buffer
TE	Tris-EDTA buffer
TEA	Tris-EDTA-acetate buffer
tfe	Trifluroethanol
tip	translation initiation point
tsp	transcription start point
X-gal	5-bromo-4 chloro-3-indolyl-β-D-galactoside
YT	Yeast extract-Tryptone medium

1. INTRODUCTION

1.1 Biosynthesis and regulation of β-lactam antibiotic production:

β-lactam antibiotics are widely used for the treatment of bacterial infections because of their clinical effectiveness and low toxicity, and they account for more than 50% of the world consumption of antibiotics. Despite the discovery, nearly 4 decades ago, of cephalosporins from *Cephalosporium acremonium* (Newton and Abraham, 1955) and the discovery in 1929 of their precursors, the penicillins produced by *Penicillium notatum* (Fleming, 1929), considerable research is still being devoted towards the isolation and development of new β-lactam antibiotics. This includes potent natural compounds as well as semisynthetic penicillins and cephalosporins, designed to combat penicillinase producing organisms and penicillin-resistant opportunistic pathogens. While the chemical nature and the mode of action of these antibiotics has been known for a long time, rapid progress has only been made during the last 15 years in the elucidation of the biochemical machinery which assembles these antibiotics *in vivo*. This progress became possible with the development of cell-free assay systems for antibiotic biosynthetic enzymes, coupled with the advent of recombinant DNA technology.

The discovery of the production of cephamycins (7-α-methoxy cephalosporins) in 1971 by *Streptomyces* spp. stimulated the pace of research towards the isolation of new β-lactam antibiotics and the characterisation of their metabolic pathways. Understanding of the biosynthetic pathways could give insights into the mechanisms of regulation, making it possible to systematically alter various control elements resulting in augmentation of antibiotic production. Early efforts into improving penicillin and cephalosporin production from fungal species were directed towards increasing the volumetric productivity of cultures by manipulating fermentation conditions and by strain improvement programs. These strain improvement programs used random mutagenesis and screening to select strains with superior capabilities. Studies were also directed

towards the development of semisynthetic penicillins by chemical modification of the naturally produced penicillins and cephalosporins. However, these yield improvement studies failed to contribute much meaningful information towards understanding the details of the biochemical routes and the regulatory aspects of the production of antibiotics. Increased production of antibiotics by improved strains did not directly contribute towards the deciphering of the biosynthetic pathways, because these strains contained increased amounts of enzymes for the various biosynthetic steps which facilitated the study of enzymatic reactions in cell-free systems. This, in turn, made it possible to investigate their reaction chemistry, and finally to purify and design oligonucleotide probes to clone the various biosynthetic genes based on N-terminal amino acid sequence information derived from those enzymes.

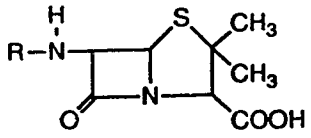
Genetic analysis of one *Penicillium chrysogenum* β-lactam overproducer strain developed by classical strain improvement programs showed that it had undergone amplification and contained 8-10 copies of one of the penicillin biosynthetic genes (Smith *et al.*, 1989) while another strain contained a 14-fold amplification of a DNA segment which contained two biosynthetic genes (Barredo *et al.*, 1989a). This demonstrated the feasibility of improving antibiotic production by introducing additional copies of biosynthetic genes into producer organisms. Strain improvement by genetic engineering has met with mixed results so far, but it is still at an early stage of development. When multiple copies of one of the cephalosporin biosynthetic genes were introduced into *C. acremonium* by *in vitro* genetic manipulation, no increase in the cephalosporin yield was observed. However, when the same process was used to increase the gene dosage of a different gene encoding a later enzyme in the biosynthetic pathway, this led to increased yield of cephalosporin C (Skatrud *et al.*, 1989a). Similarly, kinetic analysis of the rate limiting steps of cephamycin C production in *Streptomyces clavuligerus* followed by subsequent targeted insertion of an additional copy of a gene to overcome the limiting step led to improvement in the production of cephamycin C

(Malmberg *et al.*, 1993). These findings suggest that there is considerable scope for improving antibiotic yields by *in vitro* genetic manipulations, either by introducing additional copies of relevant genes or by removing the bottlenecks imposed by regulatory elements at various levels of antibiotic production. However, successful exploitation of recombinant technology, will require a detailed understanding and characterisation of the biochemical pathways and their mode of regulation.

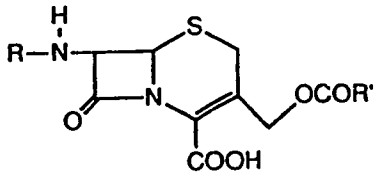
Sensitive screening techniques aimed at identifying new β-lactam producers have shown that these antibiotics are not only produced by fungi but also by both Gram positive and Gram negative bacteria, particularly the actinomycetes. In contrast to the fungi which produce only penicillins and cephalosporins as end products, the actinomycetes produce both classical antibiotics such as penicillins and cephamycins (modified forms of cephalosporins) and non-classical antibiotics such as oxypenams (eg-clavulanic acid), carbapenems (eg-thienamycin) and monobactams (eg-sulfazecin, nocardicin A). The structures of some of these classical and non-classical β-lactam compounds are shown in Figure 1. For example, *Streptomyces cattleya* produces penicillin N, cephamycin C and the non-classical β-lactam antibiotic thienamycin. *Streptomyces clavuligerus* produces penicillin N, cephamycin C and the non-classical β-lactam antibiotic clavulanic acid (Jensen, 1986). Clavulanic acid is a poor antibiotic but a potent inhibitor of β–lactamase enzymes and it represents a major achievement of antibiotic research in the fight against β-lactamase producing pathogenic organisms. Clavulanic acid mixed with amoxycillin under the trade name of Augmentin is used to treat pathogenic organisms producing β-lactamases (Wolfe *et al.*, 1984). Cephamycin C has a greater intrinsic resistance to β-lactamases than penicillins and cephalosporins.

This capability of the bacterial β-lactam producing species to produce such a diverse group of β-lactam antibiotics, together with the continuing demand and economic importance of these antibiotics has focussed much interest towards research on antibiotic production in those species. The similarity in the structure of penicillins, cephalosporins

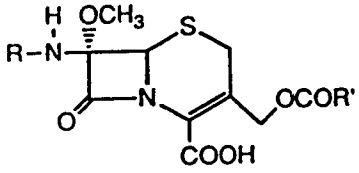
Figure 1: Structures of various classes of β-lact[illegible]ibiotics. Penams (penicillins), cephems (cephalosporins and cephamycins), oxypenams (clavulanic acid), carbapenems (thienamycin), monobactams (sulfazecin, nocardicin A)



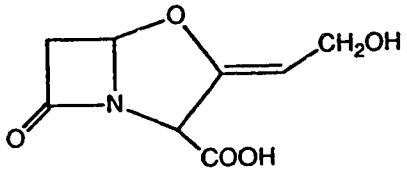
Penicillins



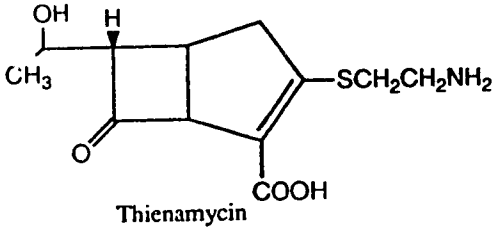
Cephalosporins



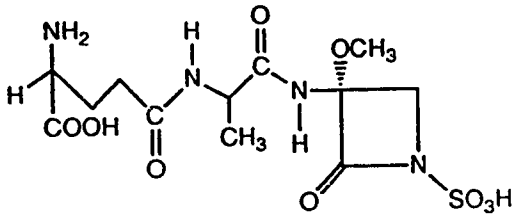
Cephamycins



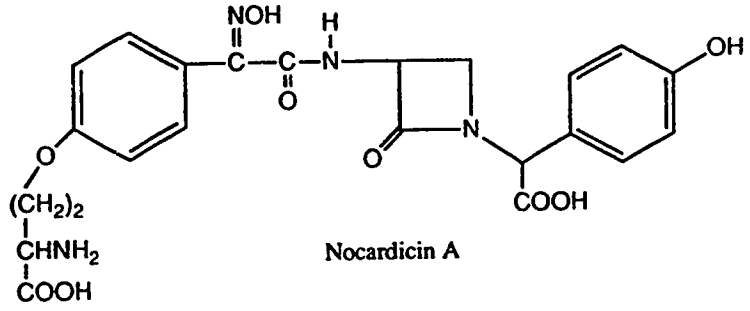
Clavulanic acid



Thienamycin



Sulfazecin



Nocardicin A

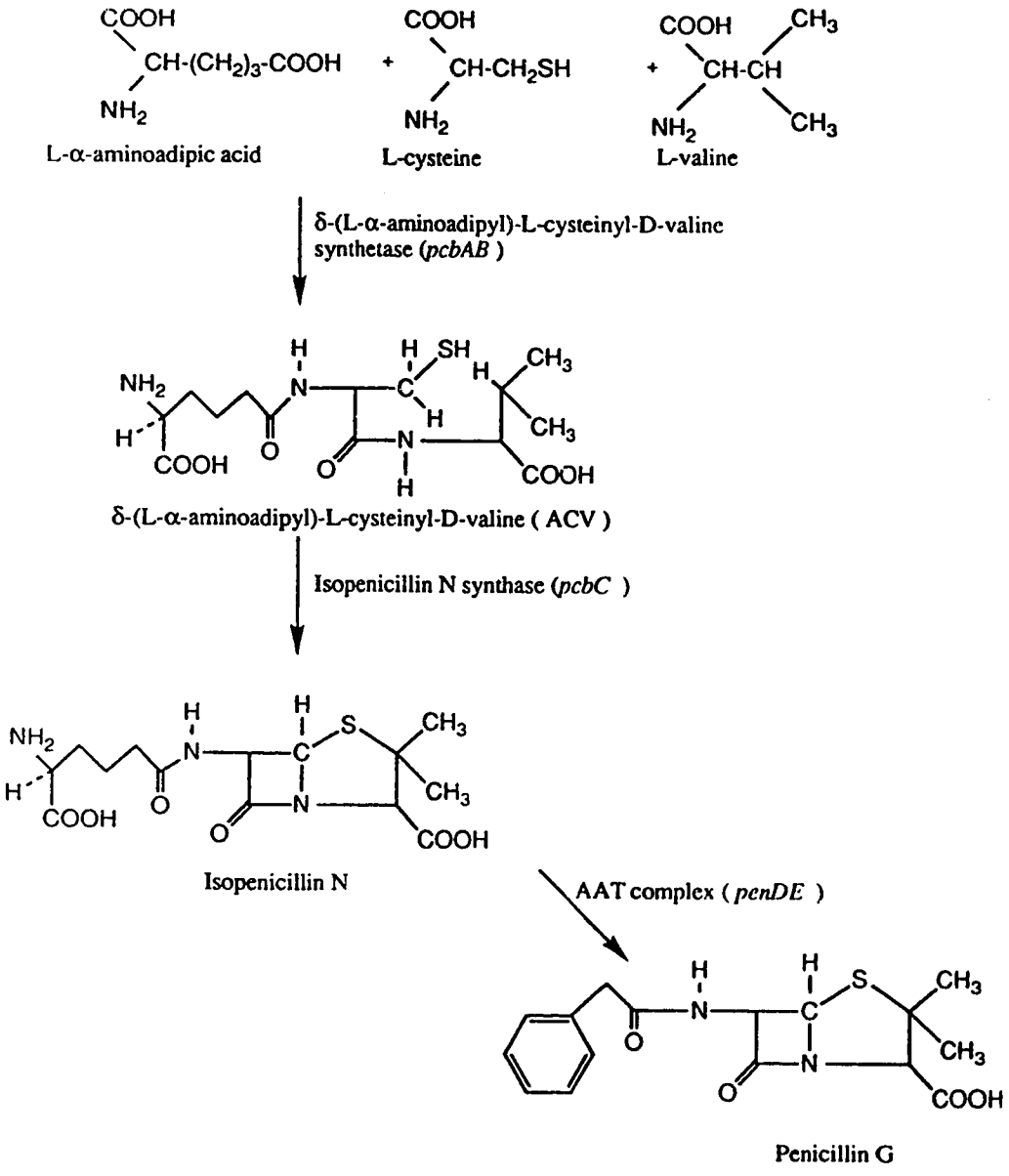
and cephamycins, together with the observation that a single species often produces more than one type of β-lactam antibiotic, suggested that these antibiotics might have common biosynthetic steps. It was also observed that while some penicillin producers do not produce cephalosporins and/or cephamycins, cephalosporin and cephamycin producers invariably produce at least small amounts of penicillin. This implies that penicillins could be intermediates in the biochemical pathway to cephalosporins and /or cephamycins.

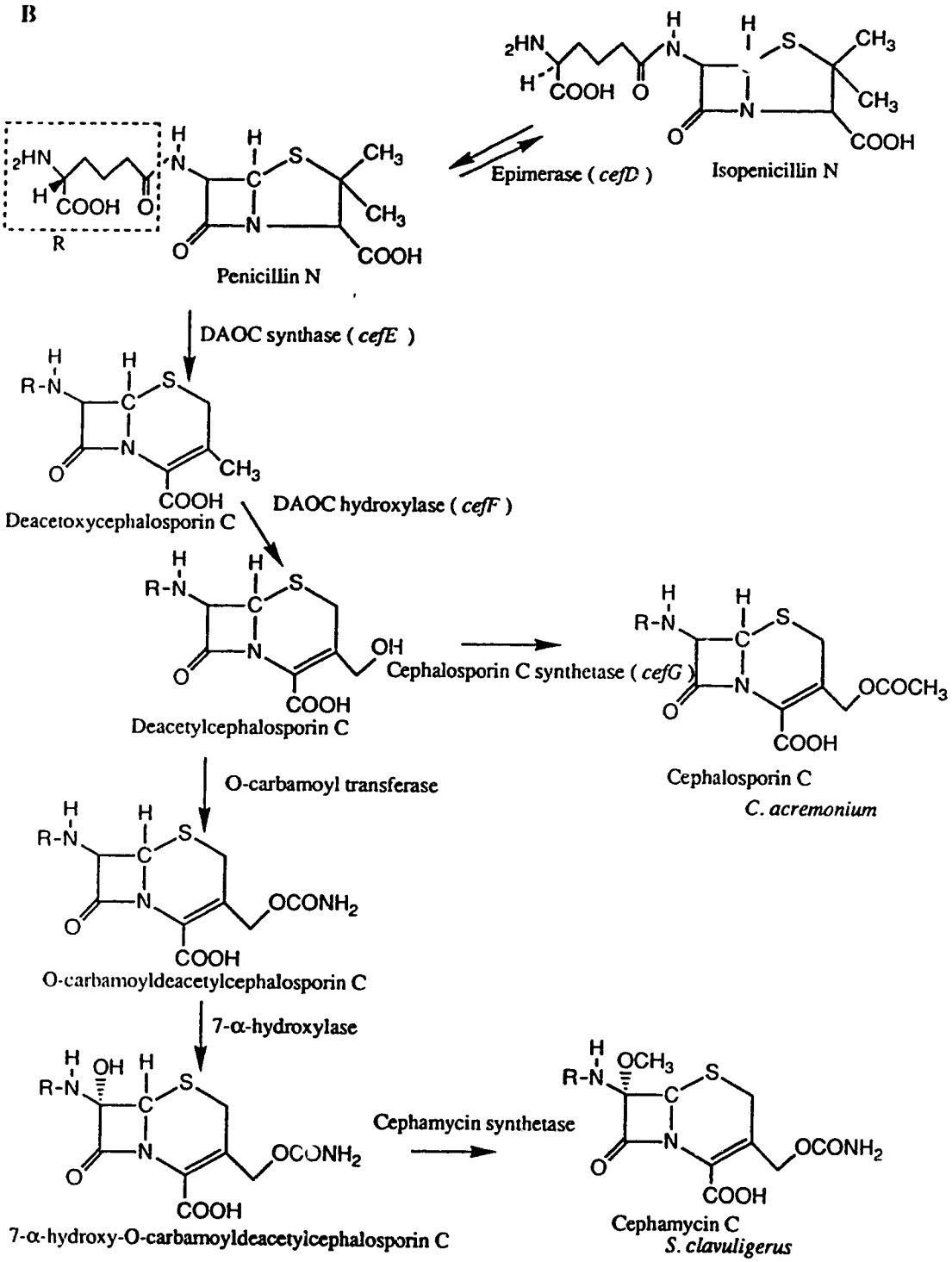
Through the advent of recombinant DNA technology, many of the genes encoding the enzymes involved in the various steps for the synthesis of these antibiotics have now been identified and excellent reviews have appeared in the literature covering the biosynthesis, genetic organisation and regulation of antibiotic synthesis in both fungal and bacterial β-lactam producers (Martin and Liras, 1989; Aharonowitz *et al.*, 1992; Jensen and Demain, 1995). It has been recognised that the biochemical routes terminating with the formation of isopenicillin N are shared by both fungal and bacterial penicillin, cephalosporin and cephamycin producers and the pathways diverge thereafter. This section will briefly narrate the developments that have taken place in the characterisation of the β-lactam biochemical pathway during the last fifteen years in both the fungal and bacterial penicillin, cephalosporin and cephamycin producers.

The biosynthetic pathway leading to the formation of penicillins, cephalosporins and cephamycins begins with the formation of a linear tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) catalysed by the enzyme ACV synthetase. The reaction taking place is a non-ribosomal polymerisation of the three amino acid components of penicillins and cephalosporins/cephamycins, namely L-α-aminoadipic acid, L-cysteine and L-valine (Figure-2). Cysteine and valine are primary metabolites in both bacteria and fungi. In contrast, α-aminoadipic acid is formed as an intermediate during the synthesis of lysine in fungi, but it is formed as a breakdown product of lysine involving additional steps in bacteria. The enzyme catalysing the first step in the breakdown of lysine,

Figure 2: Biochemical pathways occurring in the penicillin, cephalosporin and cephamycin producers. A: Biochemical steps culminating in the formation of penicillin G as end product in fungal penicillin producers. B: Various biosynthetic steps leading to the formation of cephalosporin C in fungal cephalosporin producers and cephamycin C in bacterial cephamycin producers. Enzymes catalysing the respective steps have been indicated followed by their gene designation.

A


L-α-aminoadipic acid
L-cysteine
L-valine
δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase (pcbAB)
δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV)
Isopenicillin N synthase (pcbC)
Isopenicillin N
AAT complex (penDE)
Penicillin G


B
Penicillin N
Epimerase (cefD)
Isopenicillin N
DAOC synthase (cefE)
Deacetoxycephalosporin C
DAOC hydroxylase (cefF)
Deacetylcephalosporin C
Cephalosporin C synthetase (cefG)
Cephalosporin C
C. acremonium
O-carbamoyl transferase
O-carbamoyldeacetylcephalosporin C
7-α-hydroxylase
7-α-hydroxy-O-carbamoyldeacetylcephalosporin C
Cephamycin synthetase
Cephamycin C
S. clavuligerus

referred to as L-lysine-ε-aminotransferase (LAT), has been found to be specifically required for the synthesis of cephamycin, and it is considered as the first enzyme in the synthesis of β-lactam antibiotics in prokaryotes. Preliminary studies from *P. chrysogenum* and *C. acremonium* have also indicated the presence of LAT activity although its significance and relevance in the synthesis of β-lactam antibiotics has not been ascertained yet. The LAT enzyme has been partially characterised and its corresponding gene (*lat*) has been cloned from both *S. clavuligerus* (Madduri *et al.*, 1989; Tobin *et al.*, 1991) and *Nocardia lactamdurans* (Kern *et al.*, 1988; Coque *et al.*, 1991a).

ACV synthetase has been purified from *Aspergillus nidulans* (Van Liempt *et al.*, 1989), *C. acremonium* (Baldwin *et al.*, 1990) and *S. clavuligerus* (Jensen *et al.*, 1990; Zhang and Demain, 1990). The corresponding gene (*pcbAB*) has been cloned and sequenced from *P. chrysogenum* (Diez *et al.*, 1990), *C. acremonium* (Gutierrez *et al.*, 1991), *A. nidulans* (MacCabe *et al.*, 1991) and *N. lactamdurans* (Coque *et al.*, 1991b). In *S. clavuligerus*, the complete sequence of the *pcbAB* gene has not yet been determined, but the gene has been located and determined to be closely linked to other penicillin biosynthetic genes (Tobin *et al.*, 1991; Doran *et al.*, 1990). ACV synthetase is a single multifunctional polypeptide with molecular weight of approximately 400 kDa, and it catalyses the ATP-dependent activation of each of the constituent L-amino acids, epimerises L-valine, binds the activated amino acids as thioesters and finally polymerises the amino acids into ACV (Van Liempt *et al.*, 1989). In every producer strain studied so far, the enzyme, as deduced from the gene sequence, consists of three domains denoted by A, C and V. These domains share considerable sequence similarities at the nucleotide and amino acid level with each other, and with the similar domains from other producer organisms. In addition, these domains show considerable similarity to the domains found in peptide synthetases from other species, namely gramicidin S synthetase I and tyrocidin synthetase I from *Bacillus brevis* (Kleinkauf and von Dohren, 1990). Hence, it is

thought that each domain is involved in the activation of one of the constituent amino acids. Binding studies using the first domain of ACV synthetase from *S. clavuligerus*, produced by recombinant DNA techniques, with L-α aminoadipic acid strengthen the above observation (Kadima, 1993).

The second step in the formation of β-lactam antibiotics is the oxidative cyclisation of ACV into isopenicillin N by isopenicillin N synthase (IPNS). The *pcbC* gene which encodes IPNS has been shown to be located immediately adjacent to *pcbAB* in all species studied to date. The IPNS enzyme requires oxygen, ferrous ion and ascorbate as cofactors for its maximal activity. The reaction involves the removal of four protons and the closure of two ring systems leading to the formation of the penicillin nucleus. The penicillin nucleus consists of a four membered β-lactam ring and a five membered thiazolidine ring. This enzyme has been the most extensively studied among the β-lactam biosynthetic enzymes. As it is also the subject of this thesis, a detailed discussion of this enzyme is therefore deferred to a later section.

Isopenicillin N, formed in the previous step, undergoes two types of processing depending on the type of producer organism. Isopenicillin N can be converted into hydrophobic penicillins in *P. chrysogenum* and *A. nidulans*, species which produce only penicillins as end products, through the action of the isopenicillin N amidohydrolase and acyl-CoA:6APA acyltranferase complex (AAT). This pair of activities is found within a single heterodimeric enzyme consisting of two subunits of 11 kDa and 29 kDa each. The enzyme is encoded by a single gene (*penDE*) which gives rise to a polypeptide of 40 kDa and presumably undergoes posttranslational processing to yield the two subunits. The *penDE* genes from both of these fungal producers are reported to have three introns in their N-terminal halves and the positions of the introns are the same in both species (Barredo *et al.*, 1989b; Tobin *et al.*, 1990).

Isopenicillin N, in cephamycin and cephalosporin producers, undergoes epimerisation of the L-α-aminoadipyl side chain into the D-enantiomer to produce

penicillin N which then undergoes ring expansion leading to cephalosporins and cephamycins. Epimerase activity was detected first in *C. acremonium* (Konomi *et al.*, 1979) and then purified from *S. clavuligerus* (Jensen *et al.*, 1983; Usui and Yu, 1989) and *N. lactamdurans* (Laiz *et al.*, 1990). Epimerase genes (*cefD*) have been cloned and sequenced from *S. clavuligerus* (Kovacevic *et al.*, 1990) and *S. lipmanii* (Skatrud, 1992). The epimerase has a molecular weight of 43 kDa and requires pyridoxal phosphate for activity and catalyses a reversible reaction resulting in formation of approximately equal amounts of isopenicillin N and penicillin N at equilibrium.

Following epimerisation of the side chain, the five membered thiazolidine ring of penicillin N is then converted into a six-membered dihydrothiazine ring in deacetoxycephaloporin C (DAOC), through an oxidative reaction catalysed by DAOC synthetase (DAOCS), also known as expandase. DAOC then gets further hydroxylated to form deacetylcephalosporin C (DAC) by DAOC hydroxylase. DAOCS activity was first demonstrated in cell free extracts of *C. acremonium* (Kohsaka and Demain, 1976) and subsequently in S. *clavuligerus* (Jensen *et al*, 1985). It requires ferrous ion, molecular oxygen and DTT as cofactors, the same as those required for IPNS activity (Kupka *et al.*, 1983), but in addition, it also requires α-ketoglutarate. Attempts to purify DAOCS and DAOC hydroxylase from *C. acremonium* and *S. clavuligerus* demonstrated that there are structural differences in the two enzymes between these two producers. In *C. acremonium*, both of the activities are associated with a single enzyme, whereas in *S. clavuligerus*, two separate enzymes are involved. Expression of the *C. acremonium* DAOCS encoding gene, *cef E*, in *E. coli* confirmed that it possesses both DAOCS and DAOC hydroxylase activities in a single polypeptide with a molecular weight of 36 kDa, and so the gene was renamed *cefEF* (Samson *et al.*, 1987a; Skatrud and Queener, 1989a). In contrast, *S. clavuligerus* has two separate genes encoding the two separate enzymes, each with a molecular weight of 34 kDa. It was also noticed that each of the *S. clavuligerus* DAOCS and DAOC hydroxylase enzymes was capable of carrying out the

other reaction, although at a greatly reduced rate (Baker *et al.*, 1991). The *cefE* and *cefF* genes of *S. clavuligerus* share 71% similarity at the nucleotide sequence level and 59% similarity at the amino acid sequence level, which can partly explain the bifunctional nature of the enzymes.

In the cephalosporin producer, *C. acremonium*, cephalosporin C synthetase catalyses the transfer of an acetyl group from acetyl-CoA to the hydroxyl moiety on the C-3 methyl group of the sulfur containing ring of DAC, leading to the formation of cephalosporin C. However, in *S. clavuligerus*, DAC is converted into O-carbamoyl-DAC (OCDAC) by the enzyme O-carbamoyltransferase which transfers a carbamoyl group from carbamoyl phosphate. The O-carbamoyltransferase enzyme has been partially purified from *S. clavuligerus* and requires Mg^{2+}, Mn^{2+} and ATP (Martin and Liras, 1989). OCDAC is then hydroxylated to give 7-α-hydroxy-OCDAC by a 7-α-hydroxylase which has been purified to near homogeneity (Xiao *et al.*, 1991) followed by O-methylation by cephamycin synthetase to give cephamycin C. Additional enzymes are postulated to explain the formation of cephamycins A and B (Demain and Solomon, 1983) by some *Streptomyces* spp., and a variety of 7-formamido derivatives of cephalosporins by *Flavobacterium* sp 12154 (Singh *et al.*, 1984), but these activities remain to be explored.

Since penicillin and cephamycin formation is frequently accompanied by the formation of non-classical β-lactam antibiotics such as clavulanic acid and thienamycin in *Streptomyces* spp., this suggested that the formation of all of these β-lactam compounds might share the same pathway. However, current evidence from whole cell studies as to the origin of the carbon atoms in clavulanic acid and thienamycin indicates that these compounds arise from very different precursors than those involved in penicillin/cephalosporin/cephamycin biosynthesis (Elson and Oliver, 1978; Elson *et al*, 1982; Albers-Schonberg *et al*, 1976). Information regarding the enzymology and genetics of clavulanic acid biosynthesis is just beginning to emerge, but the biosynthetic

pathway is clearly separate from that leading to the penicillins, cephalosporins and cephamycins.

Analysis of the organisation of the genes involved in the biosynthesis of β-lactam antibiotics has revealed that they are clustered and closely linked in most of the organisms studied (Aharonowitz *et al.*, 1992). It has been found that the *lat*, *pcbAB* and *pcbC* genes are not only closely linked but also cotranscribed in *S. clavuligerus* (Petrich *et al.*, 1994). Coordinate transcriptional regulation appears to be an important feature of β-lactam gene expression in some but not all producer strains. The appearance of *pcbC*-derived mRNA in *P. chrysogenum*, *A. nidulans* and *S. clavuligerus* is always accompanied by the production of antibiotics, although there is variation from species to species in the pattern of appearance of mRNA. In *A. nidulans*, under conditions in which penicillin synthesis was repressed, no transcripts of biosynthetic genes could be detected, suggesting common regulation of these genes at the transcriptional level (MacCabe *et al.*, 1990). Exogenous addition of methionine has been shown to stimulate the production of cephalosporin in *C. acremonium*, and cells grown in the presence of methionine were found to contain increased levels of mRNA of *pcbAB*, *pcbC*, *cefEF* genes (Velasco *et al.*, 1994). Coordinated synthesis of various intermediates of β-lactam antibiotics would be beneficial in terms of efficient production of β-lactams, and clustering of various biosynthetic genes provides a simple mechanism for achieving this coordinate control.

1.2 Isopenicillin N Synthase (*pcbC*):

IPNS catalyses the oxidative cyclisation of the linear tripeptide ACV leading to the formation of the bicyclic molecule isopenicillin N, which possesses weak but significant antibiotic properties. Studies using cell free preparations from *S. clavuligerus* and *C. acremonium* have shown that the enzyme requires molecular oxygen, ferrous ion and ascorbate, similar to other oxygenases, for its maximal activity (Jensen *et al.*, 1982a: Abraham *et al.*, 1981; Sawada *et al.*, 1980). Dithiothreitol is required for the protection

of the enzyme from inactivation by oxygen and for the reduction of the substrate ACV which is provided in the form of a dimer in *in vitro* studies. *S. clavuligerus* and *P. chrysogenum* possess a disulphide reductase that recognises bis-ACV as a substrate and might be required for the maximal activity of the enzyme *in vivo* (Aharonowitz *et al.*, 1993; Cohen *et al.*, 1994). Unlike other oxygenases, IPNS is unusual in its reaction chemistry in that oxygen is not incorporated into the product, isopenicillin N.

The enzyme has been purified from *C. acremonium* (Pang *et al.*, 1984, Hollander *et al.*, 1984), *P. chrysogenum* (Ramos *et al.*, 1985), *S. clavuligerus* (Jensen *et al.*, 1986a) and *N. lactamdurans* (Castro *et al.*, 1988). Its reported molecular weight ranges from 26.5 kDa in *N. lactamdurans* to 40 kDa in *C. acremonium*. IPNS from *N. lactamdurans* displays a Km for ACV of 0.18 mM, similar to that of *P. chrysogenum* (0.13 mM) and lower than the reported value for *C. acremonium* (0.3 mM) and *S. clavuligerus* (0.32 mM). The optimal temperature for the activity of IPNS from *P. chrysogenum* and *N. lactamdurans* is 25°C while that of *S. clavuligerus* is 20°C. Optimum pH for activity is in the range of 7-8. The pI of the IPNS of *N. lactamdurans* is 6.55 whereas the IPNS of *C. acremonium* has a pI of 5.05 and that of *P. chrysogenum* is 5.5. IPNS from *N. lactamdurans* is apparently a monomer, since the molecular weight of the non-denatured form as calculated by gel filtration, is identical to that of the SDS-denatured protein, and this monomer form is likely to be the same for IPNS from other producer organisms.

The *pcbC* gene was the first β-lactam synthetase encoding gene to be cloned, and that occurred in 1985, from the eukaryotic filamentous fungus *C. acremonium* (Samson *et al.*, 1985). In cloning the *pcbC* gene from *C. acremonium*, two pools of 17-mer oligodeoxynucleotides were synthesised, both containing 32 different oligomers, such that they comprised all possible nucleotide sequences encoded by two short peptides located at the amino-terminal portion of the IPNS protein. Screening the genomic library with the DNA probes led to the identification of an open reading frame (ORF) encoding the IPNS protein. Upon expression in *Escherichia coli*, IPNS activity could be detected.

Using the *pcbC* gene of *C. acremonium* as a heterologous probe, corresponding fungal *pcbC* genes from *P. chrysogenum* (Carr *et al.*, 1986) and *A. nidulans* (Ramon *et al.*, 1987) were cloned. Similarly, using synthetic probes of 48-mer and 60-mer length, each designed based on the N-terminal amino acid sequence of the purified IPNS enzyme, *pcbC* genes from *S. clavuligerus* (Leskiw *et al.*, 1988) and *S. lipmanii* (Weigel *et al.*, 1988) respectively, were cloned and sequenced. These *pcbC* genes from *S. clavuligerus* and *S. lipmanii* were used as heterologous probes to isolate *pcbC* genes from *N. lactamdurans* (Coque *et al.*, 1991b), *Streptomyces jumonjinensis* (Shiffman *et al.*, 1988), *Flavobacterium sp.* (Shiffman *et al.*, 1988), *Streptomyces griseus* (Garcia-Dominuguez *et al.*, 1991), and all of these genes have now been cloned and sequenced.

IPNS genes from the fungal, Gram positive and Gram negative organisms form a closely related family of genes sharing significant homology at both the level of amino acids and nucleotides. There is at least 72% homology between members of each group while between the groups, a minimum of 57.1% homology has been observed. Further comparison of deduced amino acid sequences and alignment of the nine *pcbC* genes that have been characterised so far to display maximum similarity, has shown the presence of identical amino acids at the corresponding sites in all nine sequences and approximately 40% of the amino acid changes at numerous additional sites are conservative (Figure 3). Secondary structure prediction of all the IPNS proteins by computer analysis further emphasizes the structural similarity of these proteins (Weigel *et al.*, 1988).

While there are extensive similarities within the nine *pcbC* genes, the sequence identity is scattered throughout the genes and no recognisable sequence motifs could be detected. In the absence of a structural model of IPNS or other characterised proteins with related sequence features, progress has been limited in identifying the functional domains of the IPNS proteins. However, sequence comparison within the *pcbC* genes has identified two highly conserved domains surrounding two cysteine residues

Figure 3: Alignment and comparison of the amino acid sequences of IPNS enzymes. The deduced amino acid sequences of the nine IPNS enzymes were compared using Multalin program version 3.0 (Corpet, 1988). Gaps have been introduced to display maximum similarity. The position of cysteine residues have been highlighted using bold letters.

```
1        10         20         30         40         50         60
MGSVPVPVANVPRIDVSPLFGDDKEKKLEVARAIDAASRDTGFFYAVNHGVDLPWLSRET
MGSVSK..ANVPKIDVSPLFGDDQAAKMRVAQQIDAASRDTGFFYAVNHGINVQRLSQKT
MASTPK..ANVPKIDVSPLFGDNMEEKMKVARAIDAASRDTGFFYAVNHGVDVKRLSNKT
    MNRHADVPVIDISGLSGNDMDVKKDIAARIDRACRGSGFFYAANHGVDLAALQKFT
   MKMPSAEVPTIDVSPLFGDDAQEKVRVGQEINKACRGSGFFYAANHGVDVQRLQDVV
 MPVLMPSAHVPTIDISPLFGTDAAAKKRVAEEIHGACRGSGFFYATNHGVDVQQLQDVV
 MPILMPSAEVPTIDISPLSGDDAKAKQRVAQEINKAARGSGFFYASNHGVDVQLLQDVV
 MPIPMLPAHVPTIDISPLSGGDADDKKRVAQEINKACRESGFFYASHHGIDVQLLKDVV
 MPVLMPSADVPTIDISPLFGTDPDAKAHVARQINEACRGSGFFYASHHGIDVRRLQDVV

        70         80         90        100        110        120
NKFHMSITDEEKWQLAIRAYNKEHESQIRAGYYLPIPGKKAVESFCYLNPSFSPDHPRIK
KEFHMSITPEEKWDLAIRAYNKEHQDQVRAGYYLSIPGKKAVESFCYLNPNFTPDHPRIQ
REFHFSITDEEKWDLAIRAYNKEHQDQIRAGYYLSIPEKKAVESFCYLNPNFKPDHPLIQ
TDWHMAMSAEEKWELAIRAYNPAN.PRNRNGYYMAVEGKKANESFCYLNPSFDADHATIK
NEFHRTMSPQEKYDLAIHAYNKNN.SHVRNGYYMAIEGKKAVESFCYLNPSFSEDHPEIK
NEFHGAMTDQEKHDLAIHAYNPDN.PHVRNGYYKAVPGRKAVESFCYLNPDFGEDHPMIA
NEFHRNMSDQEKHDLAINAYNKDN.PHVRNGYYKAIKGKKAVESFCYLNPSFSDDHPMIK
NEFHRTMTDEEKYDLAINAYNKNN.PRTRNGYYMAVKGKKAVESWCYLNPSFSEDHPQIR
NEFHRTMTDQEKHDLAIHAYNENN.SHVRNGYYMARPGRKTVESWCYLNPSFGEDHPMIK

       130        140        150        160        170        180
EPTPMHEVNVWPDEAKHPGFRAFAEKYYWDVFGLSSAVL.RGYALALGRDEDFFTRHSRR
AKTPTHEVNVWPDETKHPGFQDFAEQYYWDVFGLSSALL.KGYALALGKEENFFARHFKP
SKTPTHEVNVWPDEKKHPGFREFAEQYYWDVFGLSSALL.RGYALALGKEEDFFSRHFKK
AGLPSHEVNIWPDEARHPGMRRFYEAYFSDVFDVAAVIL.RGFAIALGREESFFERHFSM
AGTPMHEVNSWPDEEKHPSFRPFCEEYYWTMHRLSKVLM.RGFALALGKDERFFEPELKE
AGTPMHEVNLWPDEERHPRFRPFCEGYYRQMLKLSTVLM.RGLALALGRPEHFFDAALAE
SETPMHEVNLWPDEEKHPRFRPFCEDYYRQLLRLSTVIM.RGYALALGRREDFFDEALAE
SGTPMHEGNIWPDEKRHQRFRPFCEQYYRDVFSLSKVLM.RGFALALGKPEDFFDASLSL
AGTPMHEVNVWPDEERHPDFRSFGEQYYREVFRLSKVLLLRGFALALGKPEEFFENEVTE

         190        200        210        220        230        240
DTTLSSVVL.IRYPYLDPYPEPAIKTADDGTKLSFEWHEDVSLITVLYQSDVQNLQVKTP
DDTLASVVL.IRYPYLDPYPEAAIKTAADGTKLSFEWHEDVSLITVLYQSNVQNLQVETA
EDALSSVVL.IRYPYLNPIPPAAIKTAEDGTKLSFEWHEDVSLITVLYQSDVANLQVEMP
DDTLSAVSL.IRYPFLENYP..PLKLGPDGEKLSFEHHQDVSLITVLYQTAIPNLQVETA
ADTLSSVSL.IRYPYLEDYP..PVKTGPDGEKLSFEDHFDVSMITVLYQTQVQNLQVETV
QDSLSSVSL.IRYPYLEEYP..PVKTGPDGQLLSFEDHLDVSMITVLFQTQVQNLQVETV
ADTLSSVSL.IRYPYLEEYP..PVKTGADGTKLSFEDHLDVSMITVLYQTEVQNLQVETV
ADTLSAVTL.IHYPYLEDYP..PVKTGPDGTKLSFEDHLDVSMITVLFQTEVQNLQVETA
EDTLSCRSLMIRYPYLDPYPEAAIKTGPDGTRLSFEDHLDVSMITVLFQTEVQNLQVETV

         250        260        270        280        290        300
QGWQDIQADDTGFLINCGSYMAHITDDYYPAPIHRVKWVNEERQSLPFFVNLGWEDTIQP
AGYQDIEADDTGYLINCGSYMAHLTNNYYKAPIHRVKWVNAERQSLPFFVNLGYDSVIDP
QGYLDIEADDNAYLVNCGSYMAHITNNYYPAPIHRVKWVNEERQSLPFFVNLGFNDTVQP
EGYLDIPVSDEHFLVNCGTYMAHITNGYYPAPVHRVKYINAERLSIPFFANLSHASAIDP
DGWRDLPTSDTDFLVNAGTYLGHLTNDYFPSPLHRVKFVNAERLSLPFFFHAGQHTLIEP
DGWRDIPTSENDFLVNCGTYMAHVTNDYFPAPNHRVKFVNAERLSLPFFLNGGHEAVIEP
DGWQDIPRSDEDFLVNCGTYMGHITHDYFPAPNHRVKFINAERLSLPFFLNAGHNSVIEP
DGWQDLPTSGENFLVNCGTYMGYLTNDYFPAPNHRVKFINAERLSLPFFLHAGHTTVMEP
DGWQSLPTSGENFLINCGTYLGYLTNDYFPAPYHRVKYVNAERLSLPFFLHAGQNSVMKP

         310        320        330      338
WDPATAKDGAKDAAKDKPAISYGEYLQGGLRGLINKNGQT      C. acremonium
FDPREP.....NGKSDREPLSYGDYLQNGLVSLINKNGQT      A. nidulans
WDPSKE.....DGKTDQRPISYGDYLQNGLVSLINKNGQT      P. chrysogenum
FAPPPYA.....PPGGNPTVSYGDYLQHGLLDLIRANGQT      Flavobacterium
FFPDGAP....EGKQGNEAVRYGDYLNHGLHSLIVKNGQT      N. lactamdurans
FVPEGAS.....EEVRNEALSYGDYLQHGLRALIVKNGQT      S. clavuligerus
FVPEGAA.....GTVKNPTTSYGEYLQHGLRALIVKNGQTN     S. jumonjinensis
FSPEDTR.....GKELNPPVRYGDYLQQASNALIAKNGQT      S. griseus
FHPEDTG.....DRKLNPAVTYGEYLQEGFHALIAKNVQT      S. lipmanii
```

implicated in the IPNS catalytic reaction. Several observations suggest that free cysteine thiols are required for activity. The *C. acremonium* IPNS enzyme is inactivated when oxidised, and its activity is reversibly restored if it is treated with a disulphide reductant such as DTT. In *C. acremonium*, the IPNS enzyme contains only two cysteine residues, whereas IPNS enzymes from other species contain a variable number of cysteine residues, from two to four. Chemical blocking of the cysteine thiol groups results in loss of activity (Jensen *et al.*, 1986a; Perry *et al.*, 1988). Site-directed *in vitro* mutagenesis of the cysteine groups present in the *pcbC* gene from *C. acremonium* such that either, or both, of the two cysteine residues in the IPNS enzyme was replaced by a serine residue, resulted in a C106S mutant which was found to have lost 97% of its original activity and a C255S mutant which had lost 50% of its activity. When both cysteine residues were changed to serine, the specific activity of the double mutant was only marginally lower than that of the C106S mutant (Samson *et al.*, 1987b). The structural comparisons of the wild type recombinant IPNS and the double mutant of *C. acremonium* by circular dichroism studies (CD) gave identical spectra, which suggested that the loss of activity which occurred was not due to structural alterations (Kriauciunas *et al.*, 1991). Cysteine thiols therefore seem to be important, but not absolutely required for activity, and so presumably are not involved in the active site of the enzyme. Evaluation of kinetic parameters of the C106S mutant showed a 5 fold increase in K_m while the C255S mutant did not differ from the wild type enzyme, which led to speculation that the cysteine 106 residue might be involved in substrate binding. The deduced amino acid sequence of the IPNS from *N. lactamdurans* does not have a cysteine 249 (equivalent to cysteine 255 in *C. acremonium*) residue and this position is instead occupied by an alanine residue. Hence, it is likely that cysteine 255 is not involved in the function of IPNS and possibly has a role in the structural maintenance. Spectroscopic analysis of the wild type and serine mutant IPNS enzymes of *C. acremonium* in the presence of ACV and nitric oxide (NO) identified the ligands that the ferrous ion was coordinated with as three histidine

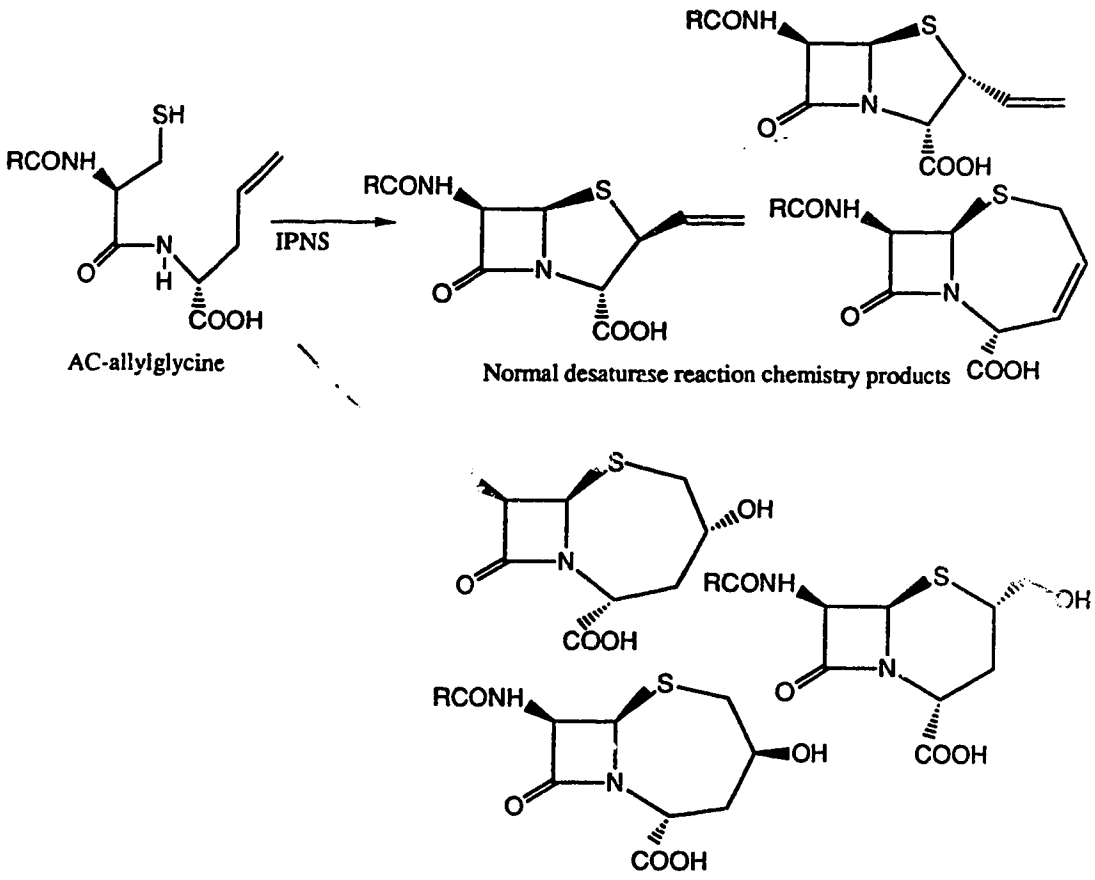
residues, one aspartate residue, NO and ACV (substrate) suggesting, an octahedral symmetry. No involvement of the endogenous cysteine residues in binding to iron was noted (Ming *et al.*, 1991; Jiang *et al.*, 1991; Orville *et al.*, 1992; Scott *et al.*, 1992; Randall *et al.*, 1993). Although simultaneous thiolate ligation between the metal ion and an endogenous cysteine residue has not been disproved so far, available evidence suggests that the cysteine residues are neither involved in the active site catalysis nor in direct substrate binding.

Alhough IPNS proteins, in general, are conserved between fungal and bacterial β-lactam producers, more homology was detected within the members of the fungal and bacterial groups than between the groups, as mentioned earlier. IPNS protein from *S. clavuligerus* contains four cysteine residues, two residues at positions analogous to those in fungal IPNS and the remaining two residues at positions 37 and 142 (equivalent to 38 and 144 in *C. acremonium*). Unlike the cysteine residue at position 104, the cysteine residues at positions 37 and 142 are not highly conserved among various IPNS. Nevertheless, when additional cysteine residues are found in IPNS, they are restricted to only these two positions, and they are only found in the bacterial IPNS proteins suggesting that they might have some specific function unique to bacterial producers. Hence, the IPNS protein from *S. clavuligerus* can be considered as a representative bacterial IPNS protein, useful for comparison with the well studied *C. acremonium* IPNS which is a representative fungal IPNS protein.

Early studies on the substrate specificity of IPNS indicated that in some respects it does not tolerate large changes in the structure of the substrate, ACV (Jensen *et al.*, 1982a). For example, any shortening of the aminoadipate residue even by one carbon atom, abolishes all enzyme activity. However overall, the enzyme has a relatively broad substrate specificity and this property has been exploited in the synthesis of novel, unnatural penicillins from ACV analogues (Baldwin and Abraham, 1988; Wolfe *et al*, 1984). In addition to showing a fairly broad substrate specificity, some substrate

analogues, such as tripeptides in which valine has been replaced by either allylglycine or by α-aminobutyrate, give more than one product upon incubation with the enzyme and cofactors (Baldwin *et al.*, 1991) (Figure 4). Purified IPNS enzymes from *C. acremonium*, *P. chrysogenum*, *N. lactamdurans* and *S. clavuligerus* have also been shown to convert phenylacetyl-L-cysteinyl-D-valine, an ACV analogue containing a phenylacetyl group in place of L-α-aminoadipate, directly into penicillin G without the need for the acyl-CoA:6APA acyltranferase activity which is required for the *in vivo* synthesis of penicillin G (Martin and Liras, 1989). In view of these potential commercial applications, the central importance of IPNS activity to the biosynthetic pathway, and the scientific curiosity of its unusual oxidative chemistry, IPNS has become the most extensively studied enzyme to date among the β-lactam biosynthetic enzymes. Despite extensive genetic, biochemical and biophysical investigations of the *C. acremonium* enzyme, very little progress has been made in unraveling the functional domains of the enzyme. Though IPNS has a relaxed substrate specificity, the efficiency of conversion of unnatural substrate analogues is usually very low and often yields more than one type of product. Therefore, successful exploitation of IPNS for the synthesis of unnatural penicillins with greater intrinsic activity or resistance to β-lactamases would likely require site-specific alteration of residues located in the active site of the enzyme in order to enhance the activity of the enzyme against the unnatural substrates. However, in the absence of a full understanding of the structure of the enzyme in general and the active site in particular, any attempts at altering enzyme activity or specificity by site directed mutagenesis are likely to be futile. Three-dimensional structural characterisation of IPNS by X-ray crystallographic studies would provide the structural information needed to undertake such studies. However, crystallisation requires large amounts of IPNS protein. Purified preparations of IPNS could be obtained from *S. clavuligerus* after an elaborate purification procedure involving a large number of steps, and then only in very low yield.

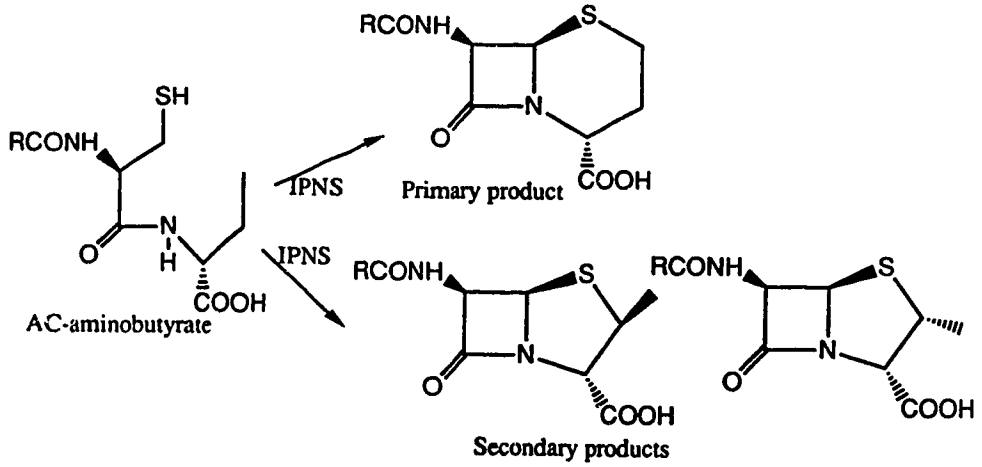
Figure 4: Processing of analogues of L-α-aminoadipyl-L-cysteinyl-D-valine (ACV) by IPNS isolated from *C. acremonium* to indicate its relaxed substrate specificity, normal reaction chemistry as a desaturase and unusual reaction chemistry as a monooxygenase.



AC-allylglycine

Normal desaturase reaction chemistry products

Unusual desaturase/monooxygenase reaction chemistry products



AC-aminobutyrate

Primary product

Secondary products

Hence, to further these studies it has become necessary to over-produce IPNS from *S. clavuligerus* in a heterologous host, and that was the central goal of this thesis research.

1.3 Isopenicillin N Synthase Expression:

E. coli has served as a host for the high-level expression of numerous heterologous genes including *pcbC* genes, and excellent reviews have appeared (see Goeddel, 1990). Heterologous gene expression in *E. coli* is a very widely used technique, but in some cases expression may be limited by regulatory controls which are incompatible between the host organism and the gene to be expressed. The kinds of problems encountered can be categorised at three levels, namely, at the level of transcription, at the level of translation initiation and elongation, and at the level of post-translational processing. Transcriptional controls are exerted primarily through promoters, and promoter strength determines the efficiency of transcription initiation. Since the high-level expression of heterologous proteins can deprive the host cell of nutritional resources required for the normal growth processes, promoters with regulatable control signals are desirable. In this way, the growth phase of the host organism can be separated from the expression phase to reduce the metabolic load imposed on the organism. Promoters with regulatable traits become very important particularly when the protein to be expressed is toxic to the cell or interferes with the host organisms' functions. Researchers have developed a variety of strong promoters with strict regulatory control signals. Expression of the recombinant protein thus becomes inducible and, in general, induction can be delayed until after sufficient cell density is obtained in order to circumvent the drawbacks mentioned above.

There are several reports indicating successful exploitation of different strong promoters to achieve high-level expression of heterologous genes, and in general transcriptional control is no longer a limiting parameter in the level of expression of heterologous proteins. The bacteriophage lambda-derived P_L promoter has been used to control expression of human gamma interferon to the level of 25% of total cellular

protein (Simons *et al.*, 1984). Transcription from P_L promoter is tightly regulated by the lambda repressor, and by combining the P_L promoter with a temperature sensitive version of the lambda repressor (cI857), transcription from P_L becomes thermally inducible by heating to 42°C. Similarly the *trp* promoter has been used in the expression of human insulin-like growth factor-II (Misoka *et al.*, 1989). The *trp* promoter is regulated by the *trp* repressor-tryptophan complex and induction is achieved by the addition of 3β-indoleacrylic acid. *Lac* , *lac* UV5 and *tac* , a hybrid promoter derived from the *lac* UV5 and *trp* (De Boer *et al.*, 1983; Amann *et al.*, 1983) promoters, are all negatively regulated by the *lac* repressor and have been used widely for the purpose of high-level expression of genes. Typically, expression is induced by the addition of isopropyl-β-D-thiogalactoside (IPTG), a gratuitous inducer (Glick and Whitney, 1987).

Recently, a series of T7 bacteriophage promoter-based expression systems, which give very high expression levels, has been developed. A major advantage of these expression systems is that the T7 RNA polymerase is insensitive to rifampicin while *E. coli* RNA polymerase is sensitive. Addition of rifampicin after induction of T7 RNA polymerase leads to selective channelling of the protein synthesising machinery towards the expression of heterologous genes. In one such system, T7 RNA polymerase specific for the T7 promoter is provided on a helper plasmid and is regulated by the temperature sensitive lambda repressor (cI857) making heterologous gene expression heat inducible (Tabor and Richardson, 1985). In addition to the powerful T7 promoter, this pT7 series of vectors offers different features including translation initiation signals optimised for *E. coli*. This series of plasmid expression vectors was chosen to express the *pcbC* gene from *S. clavuligerus* in *E. coli* in this study. Rosenberg *et al.* (1987) have developed a similar expression system in which the gene for T7 RNA polymerase is integrated into the *E. coli* chromosome under the regulation of the *lac* UV5 promoter and so heterologous gene expression is inducible by IPTG. Landman *et al.* (1991) successfully used the system

developed by Rosenberg *et al.* (1987) to express the *pcbC* gene from *S. jumonjinensis* up to a level of 51% of the total cellular protein.

The primary impediment to high level expression of heterologous proteins has often been determined to be at the translation level, more particularly at the level of translation initiation (Olins and Rangwala, 1990). Efficiency of translation initiation is influenced by the Shine-Dalgarno sequence or ribosome binding site (RBS) which typically has the consensus nucleotide sequence AAGGAG and is located just upstream of the coding region of genes (Shine and Dalgarno, 1974), and the spacing between the RBS sequence and the translation initiation codon (Hui and de Boer, 1987; Jacob *et al.*, 1987). Analysis of the optimal translation initiation control signals in *E. coli* has revealed that the spacing between the RBS sequence and the start codon should be 7-9 bases and there should be an absence of potential secondary structure in the region upstream of the start codon that could sequester the RBS sequence. Although secondary structure involving the RBS sequence can negatively affect translation initiation, the spacer sequence and the start codon themselves can be part of secondary structure without affecting the initiation process (Schottel *et al.*, 1984). There are a few well documented "exceptions" [e.g. c*I* mRNA from *E. coli* (Ptashne *et al.*, 1976), *ermE* mRNA from *Streptomyces erythraeus* now called *Saccharopolyspora erythrea*(Bibb *et al.*, 1986)], which differ from the above mentioned dogma for efficient translation initiation. These mRNAs lack a leader sequence and the sequence responsible for ribosome binding (analogous to the SD sequence) is suspected to lie within the coding region of the gene. High-level expression of these "exceptions" may demand different approaches, or it might be sufficient if these genes are placed downstream from a strong promoter.

Even if the translation initiation process is activated efficiently and successfully, the elongation process, the decay of mRNA and the degradation of expressed proteins can also influence the level of expression of heterologous genes. It has been reported that there is a strong correlation between codon usage in genes and the abundance of

corresponding tRNA species. There is a preferential usage of codons corresponding to major tRNA species in highly expressed genes, while poorly expressed genes use codons corresponding to minor tRNA species (Ikemura, 1981; Andersson and Kurland, 1990; Kurland, 1991). It also has been demonstrated that the codon usage can influence the translation rate (Sorensen *et al.*, 1989). Heterologous genes derived from both prokaryotes and eukaryotes with biased codon usage by virtue of differential usage of nucleotide bases (G+ C content) in their genomes, often contain codons corresponding to rare tRNA species. The high levels of mRNA generated by strong promoters could then experience a shortage of tRNA molecules resulting in a reduced translation rate and lower level of expression. Despite this prediction, there are several examples of high-level expression of genes with a high content of rare codons. However, the above prediction has been justified in the case of the tetanus toxin fragment C by the demonstration of an improvement in the expression level after the removal of rare codons (Makoff *et al.*, 1989). Similarly, removal of rare codons at the 5' end of the genes encoding the components of polyketide synthase of *Streptomyces glaucescens* led to considerable improvement in the expression level in *E. coli* (Gramajo *et al.*, 1991).

A direct correlation has also been noted between the rate of mRNA synthesis and translation, and the rate of mRNA decay and the degradation of proteins (Hargrove and Schmidt, 1989). Messenger RNA molecules are rapidly degraded by RNases in prokaryotes since they lack both the protective 5' capping and 3' polyadenylation that is seen in eukaryotic mRNA. Although RNase negative host strains can be used to slow down the rate of mRNA degradation, it is usually sufficient to overcome mRNA degradation through an increase in the number of transcripts by using very strong promoters and high-copy-number plasmids.

E. coli, by unidentified mechanisms, senses " non-self " proteins and activates the production of a series of proteases, including Lon and other heat shock response proteases, which can lead to the degradation of expressed proteins (Goff and Goldberg,

1985; Allen *et al.*, 1992; Chung, 1993). Proteolytic cleavage is disadvantageous, not only because it reduces the level of expression, but also because it creates a non-homogeneous mixture of expressed protein species. This can seriously limit the utility of the expressed proteins, and removing the degraded protein fragments from intact proteins is very difficult. Dramatic improvements have been achieved in some cases by expressing the cloned genes in *lon*⁻ host strains, deficient in the production of Lon protease. Other strategies aimed at controlling proteolysis include the use of *htpR* mutant host strains lacking a regulator of heat shock genes (Baker *et al.*, 1984), or host strains containing the *pin* proteolysis inhibition gene from T4 bacteriophage (Gottesman, 1990). Somatomedin-C (IGF-1) expression was increased three fold in a *lon* mutant, 15 fold in an *htpR* mutant and 33 fold in *lon*, *htpR* double mutant, when compared to the wild type *E. coli* 12 (Buell *et al.*, 1985). Similarly, expression of fibroblast interferon was increased from 300 units/mL in the wild type *E. coli* 12, to 2400 units/mL in an *E. coli* strain containing the *pin* gene (Simon *et al*, 1983).

1.4 Inclusion Body Formation (IB):

To date, many of the heterologous proteins expressed in *E. coli* at high-level have been reported to result in the formation of insoluble, inactive refractile material referred to as inclusion bodies (IB). These IB can be visualised by phase contrast microscopy within *E. coli* cells that are expressing foreign proteins at high levels (Kane and Hartley, 1988). The IB consist of the expressed proteins, present in the form of incorrectly folded aggregates (predominantly), along with lesser amounts of the four subunits of RNA polymerase (α, σ, β and β') as well as a combination of ompC, ompF and ompA proteins. These components of IB are independent of the nature of the cloned gene, promoter, plasmid vector, fermentation regimen or medium composition. However, the presence of other protein components does seem to be influenced by changes in vector construction and growth conditions. In addition, plasmid encoded proteins seem to be selectively included in IB depending on the presence and orientation of promoters of the various

genes on the plasmid. In some cases, plasmid-encoded proteins other than the heterologous gene product constitute as much as 40% of total IB protein (Rinas and Bailey, 1993). In addition to protein components, IB also contain 16S and 23S rRNA as well as both covalently closed circular and nicked forms of plasmid DNA (Hartley and Kane, 1988).

A protein molecule adopts its unique three dimensional equilibrium structure spontaneously under physiological conditions (Anfinsen, 1973). However in many cases, over-expressed prokaryotic and eukaryotic protein molecules reach incorrectly folded, non-native aggregates defined as IB (Kane and Hartley, 1988). Several reasons have been proposed to explain the accumulation of incorrectly folded protein. These include high local concentrations of the protein in the cytoplasm, lack of cellular compartmentation causing the protein to be produced in a reducing environment which prevents formation of S-S bonds necessary for proper folding, lack of mammalian post-translational modifying enzymes, lack of proper accessory folding proteins (called chaperonins) during production, limited flexibility of the polypeptide chain and instability of the native tertiary structure of some proteins (Schein, 1989). In addition, properties of the host cell, growth temperature and media composition (Kane and Hartley, 1988) or a combination of these and other still uncharacterised factors may all play a role. In order to solve the problem of inactive IB formation, research has been directed towards an understanding of the mechanisms of the protein folding pathway; a long standing complicated problem of modern research. The transition of a polypeptide chain from the disordered, unfolded state present at the time of synthesis, to the ordered, native state is called protein folding. It is well established, based on *in vitro* studies that the information necessary to drive this transition is present in the amino acid sequence itself. However, several lines of evidence based on *in vitro* and *in vivo* studies indicate that chaperonins may be necessary to guide the proper folding of a given polypeptide *in vivo*. Investigations aimed at elucidating the mechanisms of protein folding based on the

Salmonella phage tailspike protein P22 and numerous other examples, have revealed that protein molecules are folded in a series of defined stages identified by characteristic intermediates (Creighton, 1990; Haase-Pettingell and King, 1988). These intermediates contain compact, native-like backbone secondary structures but lack specific tertiary strucure, and have been referred to as " molten globule intermediates " (Kuwajima *et al.*, 1989; Martin *et al*, 1991). Certain amino acid residues in the primary amino acid sequence of a protein seem to be involved in directing a protein molecule to the final functional molecule with proper three dimensional structure. These important residues were identified for the tailspike protein by creating temperature sensitive mutants which can fold properly and form mature, functional protein molecules at the permissive temperature but not at the non-permissive temperature. It was also found that these temperature sensitive folding mutations could be suppressed by chaperonins at non-permissive temperatures. Hence, it is postulated that while the specific amino acid residues which make up the primary sequence of the protein direct the folding pathway of the molecule through series of intermediates, additional chaperone molecules are necessary to facilitate the constructive folding process by suppressing undesirable inter and intra molecular interactions between partially folded intermediates. Under non-physiological conditions, these partially folded intermediates can become trapped and lead to aggregate formation. There are a number of studies that support the need for chaperone molecules for the correct folding of polypeptides. Several refolding studies using members of the *E. coli* heat shock protein family belonging to the hsp60 (GroEL/GroES) and hsp70 (DnaK/ DnaJ/GrpE) class of proteins, demonstrated that these heat shock proteins bind only non-native and unfolded polypeptides and are required for the ATP-dependent folding of proteins (Gragerov *et al.*, 1992; Buchner *et al.*, 1991). The increased folding efficiency observed in the presence of chaperone molecules is the result of suppression of aggregate formation. DnaK can reactivate RNA polymerase from aggregates and prevent aggregate formation (Skowyra *et al.*, 1990). Similarly, the

GroEL/ GroES complex has been shown to bind unprocessed (unfolded) pre-β-lactamase and guide its folding and transport in *E. coli* (Laminet *et al.*, 1990). *In vitro* folding of proteins at high concentration also encourages the aggregation process, and reduction in the protein concentration leads to increased folding. The folding efficiency observed in the presence of chaperone molecules is higher than in the absence of chaperone molecules for any given protein concentration, although overall efficiency of folding is reduced as protein concentration increases.

Inclusion bodies are found to contain 16S rRNA, 23S rRNA, RNA polymerase and plasmid template DNA. This suggests that transcription of the plasmid, translation of the mRNA and aggregation of the newly synthesised protein must all occur in close proximity to one another. Based on these observations the following hypothesis has been developed to explain IB formation. When polypeptide chains are synthesised in a normal coupled transcription/translation system, the newly synthesized polypeptide chains are sequestered by chaperone molecules, and folding is guided by the primary amino acid sequence. However, when the rate of transcription, as dictated by the strength of the promoter and polymerase binding, exceeds a certain level, there may be insufficient chaperone molecules to deal with the high concentration of nascent polypeptides. This, in turn, could trigger precipitation of the polypeptides along with the transcription and translation machinery of the cell leading to IB formation. Although a simplified picture of aggregate formation has been presented above, the detailed mechanisms by which proteins fold and IB form are still obscure.

1.5 Solubilisation of Inclusion Bodies:

Expression of proteins at high-level is carried out to obtain proteins for a variety of uses, ranging from structure-function studies to the commercial utilisation of the expressed proteins. Many of the eukaryotic genes being expressed produce proteins of commercial utility and formation of aggregates of the expressed proteins can be a fortuitous advantage if it simplifies the purification of the expressed proteins. The

purification of recombinant proteins from IB involves the simple disruption of cells and isolation of particulate material to yield preparations which often exceed 50% purity (Kane and Hartley, 1988). After solubilization of the IB and refolding of the solubilized protein, a single chromatographic step can yield material which is more than 90% pure. An additional benefit associated with the production of recombinant proteins sequestered within IB is that the proteins are more resistant to proteolysis than the soluble protein (Marston, 1986).

In vitro refolding of IB material to get active proteins involves the addition of concentrated denaturant, either urea or guanidine-HCl, to the protein material followed by the removal of the denaturant to facilitate the renaturing process. Various strategies have been employed for the removal of denaturant to allow for the refolding of proteins. Depending on the nature of the proteins, some of those more commonly used include dialysis (London *et al.*, 1974; Goldberg, 1985), gel filtration chromatography (Durairaj *et al.*, 1992; Ghatge *et al.*, 1994), rapid dilution (Rinas *et al.*, 1992), antibody-assisted folding (Carlson and Yarmush, 1992), co-solvent assisted folding (Cleland and Wang, 1990), chaperonin-assisted folding (Brown *et al.*, 1992) and ion-exchange chromatography (Creighton, 1986; Hoess *et al.*, 1988). Despite the advantages associated with purification of recombinant proteins produced as insoluble IB, the difficulties encountered in reactivating the insoluble protein can more than offset the advantages gained in purification. Even in cases where quantitative yields of refolded material are attained, these yields can often only be achieved when refolding is carried out at very low protein concentrations (5 μg/ mL) resulting in large volumes of very dilute protein solution which are difficult to work with (London *et al.*, 1974). Furthermore, it is possible that the conformation of the solubilised protein molecules might not be identical to the native, soluble proteins. As a result of these drawbacks, in some cases, it may be more advantageous to develop a system supporting soluble expression of the recombinant proteins, despite the greater difficulty of purification of soluble proteins.

1.6 Soluble Expression of IPNS:

While the form in which a recombinant protein is expressed, soluble or insoluble, is largely influenced by promoter strength, other factors can also play a role. It has been observed that secretion of recombinant proteins, alteration of amino acid residues in the primary sequences of proteins to encourage correct folding, manipulation of growth conditions, and coexpression of the molecular chaperonins GroEL and DnaK to facilitate the folding of proteins can all lead to soluble expression of heterologous genes. Human growth hormone was expressed in the soluble form by suppressing IB formation by coexpressing the DnaK protein (Blum *et al.*, 1992). Thioredoxin which can be produced up to a level of 40% of the total soluble protein has been fused to a variety of mammalian cytokines and growth factors to express them in soluble, biologically active form while the same cytokines resulted in the IB when expressed using other systems (LaVallie *et al.*, 1993). Growth at lower temperature has been used to express human epidermal growth factor and P22 tailspike protein in the soluble form (Chalmers *et al.*, 1990; Haase-Pettingell and King, 1988). However, it appears that none of these methods is universally applicable and each has to be explored on a case by case basis. In order to obtain recombinant IPNS protein of *S. clavuligerus* in the soluble form, growth at lower temperature was investigated in this thesis and detailed discussion is deferred to later sections.

1.7 Functional Analysis of Isopenicillin N Synthase:

IPNS from *S. clavuligerus* contains four cysteine residues and they have been implicated in the function of IPNS, although the mechanistic aspects are still relatively obscure. Functional analysis of the role of the two cysteine residues in the IPNS of *C. acremonium* was carried out by Samson *et al.* (1987b) by substituting serine residues in the place of cysteine residues. Results of their analysis indicated that the C106S mutant loses 97% activity while the C255S mutant loses approximately 50% of activity.

However, Kriauciunas *et al.* (1991), based on their investigations using the same mutants, concluded that C106S and C255S mutants are much more active. Although both of these groups of investigators came to the same conclusions as to the role of these residues in the function of IPNS, they differed significantly in their interpretation of the importance of these residues for the function of IPNS. IPNS from *S. clavuligerus* contains two cysteine residues at positions analogous to those in the fungal enzyme but it also contains two additional cysteine residues. Site-directed mutagenesis of cysteine residues into serine residues in IPNS from *S. clavuligerus* and analysis of the activity of these mutants may help to clarify the contradictory findings with the *C. acremonium* mutants.

Free thiol groups present in cysteine residues have been observed to promote oligomer formation via intra and inter molecular disulphide linkages in proteins with multiple cysteine residues. During the process of crystallisation, which involves high concentrations of proteins, oligomer formation is likely to occur, especially for proteins with multiple cysteine residues such as the IPNS from *S. clavuligerus* (M. N. James, personal communication; Mark *et al.*, 1984). Oligomer formation can block single crystal formation and therefore prevent crystallographic analysis. Substitution of serine residues for cysteine residues by site-directed mutagenesis should eliminate the possibility of oligomer formation while maintaining the structural integrity of IPNS.

1.8 Mutation of *pcbC*:

Oligonucleotide directed mutagenesis is a technique widely used to introduce controlled changes into genes, making it possible to probe structure-function relationships in the proteins they encode. The usual strategy is to anneal a mutagenic oligonucleotide primer to a single-stranded DNA template, extend the primer by DNA polymerisation to form a genome length complementary strand, ligate the ends and transform competent bacterial cells to obtain progeny containing the desired mutation (Zoller and Smith, 1987). Although a high frequency of mutants should be detected after transformation, efficiency of mutagenesis is reduced by various factors including the nature of the primer,

kind of DNA polymerase used for elongation, nature of the single-strand template, annealing and elongation conditions and competence of the bacterial cells used.

In the last few years, novel approaches have been developed to enhance the efficiency of mutagenesis by overcoming some of the problems inherent in the mutagenesis protocol. By various means, a strand asymmetry is created between the template and the *in vitro* synthesised strands of DNA and this asymmetry is used to achieve preferential elimination of the non-mutant strand. A few examples of selection systems based on this asymmetry are, mismatch repair systems which distinguish methylated from unmethylated DNA, nonsense mutations differentiating wild type from amber codons, restriction modification systems differentiating *Eco* K from *Eco* B type restriction, inability of restriction enzymes to digest α-S-dNTP/ dNTP substituted templates, and uracil N glycosylase-based repair systems differentiating thymine from uracil-substituted templates.

The use of hemimethylated DNA is based on the observation that wild type *E. coli* cells repair heteroduplex DNA containing mismatches in favour of the methylated rather than the unmethylated strand. Hence, a hemimethylated gapped duplex is constructed by using a methylated, less-than-genome length linear DNA strand and an unmethylated single-stranded circular DNA in such a way that the gap in the duplex contains the region to be mutagenised. The mutagenic oligonucleotide is then hybridised within the gap, polymerised and ligated. After transfection, the methylated, mutation-containing strand will be preferentially produced (Kramer *et al.*, 1982).

In the second strategy, mutagenic oligonucleotide and a second oligonucleotide to correct the amber mutation appearing in the newly synthesised strand, are annealed to a full length circular recombinant single-stranded DNA template containing an amber mutation in an essential vector gene, elongated and ligated. After, transfection into a wild type host (not an amber suppressor), the mutagenic strand will be preferentially produced (Kramer *et al.*, 1984; Carter *et al.*, 1985).

In the restriction modification approach, the template strand contains a recognition sequence for one of the two restriction modification systems, *Eco* K or *Eco* B. Mutagenic oligonucleotide and a second oligonucleotide containing the sequence for the alternate restriction enzyme not present in the non-mutant strand are annealed, polymerised and ligated. After transformation into an appropriate host, the mutant strand is selectively enriched because it is resistant to restriction (Carter *et al.*, 1985).

The fourth strategy is based on the observation that a DNA strand containing phosphorothioate-modified internucleotide bonds is not cleaved by certain restriction enzymes. An *in vitro* polymerisation reaction is used to extend a mutagenic oligonucleotide in reaction mixtures containing three normal dNTPs and one α-phosphorothioate dNTP analogue. Restriction digestion of the resulting duplex DNA nicks the normal strand only. The nick is converted by exonuclease III digestion, to a gap extending across the region to be mutagenised, and the gap is then repolymerised using normal dNTPs. After transfection, the mutant strand is isolated (Taylor *et al.*, 1985a &b). This strategy is limited by the need to have an appropriate restriction enzyme site flanking the mutagenic region, which is uncommon.

In the fifth strategy, mutagenic oligonucleotide is annealed to the template isolated from a *dut⁻ ung⁻* strain, polymerised and ligated. After transfection into a wild type host, mutant strand is preferentially expected. This strategy was designed based on the error free repair mechanism operating in *E. coli* cells. In *E. coli dut⁻* cells which lack a functional dUTP diphosphohydrolase, dUTP accumulates and gets incorporated into DNA. *E. coli ung⁻* strains, defective in uracil N-glycosylase which normally removes uracil residues from DNA, are unable to remove the uracil residues. After *in vitro* polymerisation of the mutagenic strand using normal dNTPs and transfection into a *dut⁺ ung⁺* host, the non-mutant strand containing uracil residues is recognised by the uracil N-glycosylase and the repair process is initiated. During the repair process, multiple nicks

introduced by the AP endonuclease lead to degradation of the non-mutant template resulting in selective production of the mutant strand (Kunkel, 1985).

All of the strategies described above involve similar procedures which theoretically should generate 50:50 mixtures of mutant and wild type genes. The strategies differ only in the mechanisms used to select the desired mutants. However, an alternative approach using the polymerase chain reaction (PCR) has the ability to generate essentially homogeneous preparations of the mutant gene only. This procedure involves the denaturation of the template, annealing of mutagenic and accessory oligonucleotides followed by elongation by thermostable DNA polymerases. Repetition of the denaturation, annealing and elongation processes for a number of cycles leads to the introduction of the intended mutation and selective amplification of the region flanked by the oligonucleotides. Cloning of the PCR product eliminates the need for identification of mutants and can greatly reduce the time required in comparison to single-strand site-directed mutagenesis procedures. However, this strategy requires appropriately located restriction enzyme sites for further genetic manipulation and reconstruction of intact genes when the desired location for the mutation is within the open reading frame (ORF) of a gene rather than at one end or the other (Innis *et al.*, 1990).

1.8 Rationale and Objectives:

The possibility of using IPNS enzymes from different organisms to obtain penicillin G or other unnatural antibiotics by direct cyclisation of ACV analogues has been demonstrated by several groups. Even though the cyclisation of ACV analogues can be achieved using purified preparations of IPNS, the efficiency of conversion is low in comparison to the natural substrate, ACV. Furthermore, *in vitro* synthesis of unnatural β-lactam antibiotics often yields more than one product, depending on the nature of the substrate. It is not known yet whether this is a reflection of the catalytic activity of the enzyme, or simply a manifestation of non-enzymatic chemical reactivity of a single

enzymatically produced product. Altering the specificity and/or the affinity of IPNS for different substrates by site-directed *in vitro* mutagenesis of the cloned *pcbC* gene is one approach to overcoming this barrier. However, in the absence of a three dimensional structural model for IPNS and with only limited data on the functional domains and active site, it is difficult to predict which site-directed mutations might alter the active-site without destroying the enzyme activity altogether. Any attempt to obtain structural information on the IPNS from *S. clavuligerus* by x-ray crystallography will require milligram amounts of protein. However, the production level of IPNS is extremely low in *S. clavuligerus,* and after an elaborate purification procedure, only 0.2 mg of purified protein could be obtained from two litres of culture (Jensen *et al.*, 1986a).

E. coli expression systems have been widely used for the high-level expression of *pcbC* genes of both prokaryotic and eukaryotic origin, including the *pcbC* genes from *C. acremonium*, *S. jumonjinensis*, *S. lipmanii* and *Flavobacterium* sp. Since the behaviour of IPNS proteins from different producer organisms, when expressed in *E. coli*, seems to be indistinguishable from the native IPNS proteins in their biochemical properties, an attempt has been made in this thesis to produce high levels of IPNS from *S. clavuligerus* in *E. coli*. In previous studies, expression of *pcbC* from *S. clavuligerus* was carried out in *E. coli* using a two-cistron expression system, but analysis of the expression level achieved revealed that the yield of IPNS protein was no greater than that found in the native host (Doran *et al.*, 1990). Furthermore, to overcome possible oligomer formation during crystallisation, IPNS protein with cysteine residues replaced by serine residues, was required. The particular cysteine residues to be replaced were to be chosen based on activity analyses of the four individual cysteine to serine IPNS mutants. This thesis reports on the progress made in obtaining large quantities of IPNS from *E. coli*, activity analyses, functional characterisation, and obtaining a triple mutant IPNS of *Streptomyces clavuligerus*.

2. MATERIALS AND METHODS

2.1 Materials:

Purified bis-ACV was generously provided by Saul Wolfe, Simon Fraser University, Burnaby, B.C, Canada. Oligonucleotides required for sequencing and mutagenesis were obtained from the DNA Synthesis Laboratory, Department of Biological Sciences, University of Alberta. N-terminal amino acid sequence information was obtained from the Alberta Peptide Institute, University of Alberta. Circular dichroism (CD) studies were performed by K. Oikawa, Department of Biochemistry, University of Alberta. Electron microscopic analysis was performed by R. K. Sherburne, Department of Medical Microbiology and Infectious Diseases, University of Alberta. [α^{32}P]dATP and [γ^{32}P]ATP were obtained from ICN Biochemicals. Hybond-N hybridisation transfer membranes were purchased from Amersham. NAP-5 Sephadex G-25 DNA grade columns were purchased from Pharmacia. Nuc Trap Push columns were purchased from Stratagene. Molecular weight marker proteins were obtained as a Mol-ranger Kit, Pierce Chemical Co. DEAE-Trisacryl chromatography resin was from Reactifs IBF, Villeneuve-la-Garenne, France. Sephadex G-25 resin and the Superose 12 HR 10/30 and Mono Q HR 5/5 prepacked high-performance liquid chromatography (FPLC) columns were purchased from Pharmacia. All other chemicals were of reagent grade.

2.2 Enzymes:

Sequenase version 2 was purchased from United States Biochemical Corporation. *Taq* DNA polymerase was purchased from either Boehringer Mannheim or Promega Corporation. Deep vent DNA polymerase was purchased from New England Biolabs. Restriction enzymes, T4 polynucleotide kinase and T4 DNA ligase were purchased from Boehringer Mannheim.

2.3 Bacterial Strains and Plasmids:

E. coli INV1αF' For was used as the host strain and pCR™ 1000 and pCR™II (Invitrogen Corp., CA) were used as cloning vectors for the cloning of PCR products. *E. coli* MV1193 (Vieira and Messing, 1987), *E. coli* XL-1 Blue (Invitrogen) and *E. coli* GM48 (Yanisch-Perron *et al.*, 1985) were used as host strains for the construction of pT7-4, pT7-5, pT7-6 and pT7-7 vector derivatives as well as for the construction of pUC119 derivatives. Overnight cultures of strains carrying the vectors and their derivatives grown in 2xYT supplemented with ampicillin (100 μg/mL) or ampicillin (100 μg/mL) and kanamycin (75 μg/mL) were stored as 10% glycerol stocks at -80°C. 2xYT medium consists of Bacto tryptone (1.6%), Bacto yeast extract (1%), and NaCl (0.5%). Competent cells of strains were prepared essentially as recommended either by Hanahan (1983) or Chung *et al.* (1989).

pUC119 and M13K07 helper phage were gifts from J. Vieira, Waksman Institute of Microbiology, Rutgers University. The plasmid pIPS-1 consists of a 1.7 kb fragment of *S. clavuligerus* DNA containing the *pcbC* gene together with 335 bp upstream sequence and 385 bp of downstream sequence, cloned into pUC119, and was available from a previous study (Doran *et al.*, 1990). pIJ2925 is a derivative of the *E. coli* plasmid pUC18 with an altered multiple cloning site flanked by *Bgl*II sites (Leskiw *et al.*, 1991). It was kindly provided by M.J. Bibb (John Innes Institute, Norwich, England).

E. coli DH5α containing the pT7 series of plasmids, helper plasmid pGP1-2, which carries the T7 RNA polymerase gene under the control of λP_L promoter (Tabor and Richardson, 1985) and *E. coli* K38 were kindly provided by S. Tabor (Harvard Medical School). *E. coli* K38 (Russel and Model, 1984) served as the host for the expression of *pcbC*.

2.4 Computer Analysis of *pcbC* Sequence:

DNA sequence analysis for the purpose of designing primers and for restriction site analysis was carried out on an Apple Macintosh computer using the software

program DNA Strider version 1. 2 (purchased from C. Marck, Gif-Sur-Yvette, France). Automated DNA sequence analysis was performed at the DNA sequencing laboratory, Department of Microbiology, University of Alberta.

2.5 Plasmid Isolation and Genetic Manipulation:

Plasmid and phagemid DNA was isolated by the modified Birnboim and Doly method described in Sambrook *et al.* (1989). When phenotypic selection of transformants was not possible, recombinant constructs were screened by a gel cracking procedure described by Ahmed (1987). Briefly, colonies resulting from overnight incubation were suspended in 50 µL of Klar cracking buffer [0.05 M Tris-HCl (pH 7.0), 1% SDS, 0.002 M EDTA, 0.4 M sucrose and 0.01% bromphenol blue] using a tooth-pick and incubated at 37°C for 10 minutes. Cell suspensions were then centrifuged for 15 minutes at 13,000 xg and the supernatants were electrophoresed in 1% agarose gels. Recombinant plasmids were identified based on size differences.

2.6 Transformation of *E. coli* Strains MV1193. -1 Blue and GM-48:

One hundred microlitre amounts of *E. coli* competent cells, prepared either according to Hanahan (1983) or Chung *et al.* (1989) were combined in 1.7 mL microfuge tubes with appropriate amounts of DNA, and incubated for 40 minutes in an ice bath. Samples were then heat-shocked for 90 seconds at 42°C and placed on ice for 2 minutes. Nine hundred microlitre amounts of 2xYT, prewarmed at 37°C, were then added to the transformation mixtures and 100 µL aliquots were plated out on 2xYT agar plates supplemented with 100 µg/mL of ampicillin. In the case of transformation of *E. coli* MV1193 and XL-1 Blue, X-gal/IPTG were also included. Transformants were screened after overnight growth at 37°C.

2.7 Transformation of *E. coli* K38:

The transformation procedure mentioned earlier was repeated except that the transformation mixture was not subjected to heat-shock, and the 2xYT medium was prewarmed to 30°C, added to the transformation mixture and incubated overnight at

30°C, before transformants were selected on 2xYT agar supplemented with 100 μg/mL ampicillin and 75 μg/mL kanamycin.

2.8 Radiolabelling of the Oligonucleotide Probe OCS-1:

One microlitre (9 U) of T4 polynucleotide kinase was added to a labelling mixture containing 1 μL (1.5 μg/μL) of oligonucleotide OCS-1, 5 μL (50 μCi) of [γ^{32}P]ATP, 1 μL of 10x T4 polynucleotide kinase buffer and 2 μL distilled H_2O. The mixture was incubated for 1 hour at 37°C. Labelled oligonucleotides were separated from unincorporated [γ^{32}P]ATP using Nuc Trap Push columns (Stratagene, CA) as recommended by the supplier.

2.9 Phosphorylation of Oligonucleotides:

One microlitre (9 U) of T4 polynucleotide kinase and 5 μL (1.5 μg/μL) of mutagenic oligonucleotides OCS-1 and OCS-3 were added to separate phosphorylation mixture tubes containing 0.5 mM ATP, 100 mM Tris-HCl (pH 8.0), 20 mM $MgCl_2$ and 50 mM DTT in a total reaction volume of 50 μL. The mixtures were incubated at 37°C for 45 minutes followed by a 10 minute incubation at 65°C.

2.10 Site-directed Mutagenesis by PCR:

2.10.1 Introduction of an *Nde*I site in front of *pcbC* using *Taq* DNA polymerase:

Introduction of an *Nde*I site at the translation initiation codon of the *pcbC* gene required that three nucleotides, TTC, immediately preceeding the wild type *pcbC* gene, be changed to CAT (shown underlined below). This alteration was achieved by the polymerase chain reaction [PCR (Innis *et al.*, 1990)] using

OND-1: 5' TAGGAATTCCATATGCCAGTTCTGATGC 3' and

OFU-1: 5' CGCCAGGGTTTTCCCAGTCACGAC 3' (Forward Universal Primer)

as the oligonucleotide primers. PCR reactions contained 50 mM Tris-HCl (pH 8.8), 2 mM $MgCl_2$, 10 μg of bovine serum albumin, 100 μM of each deoxynucleoside triphosphate, 0.05% Tween 20, 5% dimethylsulfoxide, 5 units of *Taq* DNA polymerase, 50 pmol of each primer, and 0.1 pmol of pIPS-1 plasmid template (linearised by digestion

with *Pvu*II restriction endonuclease and alkali denatured), in a final reaction volume of 100 µL. The reaction mixture was incubated for 3 minutes at 95°C and then 25 cycles of PCR were conducted at 95°C, 65°C and 72°C for 0.5, 1.0 and 5.0 minutes, respectively, followed by a 10 minute incubation at 72°C. This reaction should give rise to a 1.457 kb PCR product consisting of the intact *pcbC* gene preceeded by an *Nde*I site and followed by 338 bp of downstream sequence.

2.10.2 Construction of pCR1014:

The pCR1000 vector (Invitrogen, San Diego, CA) supplied in the linearised form with terminal overhanging T nucleosides at the 3' ends, was designed specifically for cloning PCR products based on the observation that PCR products resulting from the polymerase chain reaction using *Taq* DNA polymerase have overhanging A nucleotides at their 3' ends (Mead *et al.*, 1991). The PCR reaction mixture described above was analysed by agarose gel electrophoresis and the 1.45 kb product was isolated from the gel and purified using a GeneClean kit (Bio 101, La Jolla, CA) according to the manufacturer's instructions. The purified fragment was cloned into the plasmid vector pCR1000 as recommended by the supplier with the following minor modifications.

Ligation: 1 µL of T4 DNA ligase was added to a ligation mixture containing 2 µL (50 ng) of the PCR product, 2 µL (50 ng) of the pCR1000 vector, 1 µL of 10x ligation buffer in a total volume of 10 µL and incubated at 12°C overnight.

Transformation: 2 µL of 0.5 M β-mercaptoethanol was added to a tube containing 50 µL of *E. coli* INVαF' For competent cells and gently tapped to mix. Five microlitres of the ligation mixture was added to the cells and incubated at 4°C for 30 minutes followed by 60 seconds incubation at 42°C. After chilling for 2 minutes at 4°C, 450 µL of SOC medium [Bacto tryptone (2%), Bacto yeast extract (0.5%), NaCl (10 mM), KCl (2.5 mM), $MgCl_2$ (10 mM), $MgSO_4$ (10 mM) and glucose (20 mM)] pre-warmed to room temperature, was added. The cell suspension was shaken for 1 hour at 37°C and 100 µL amounts were plated out on 2xYT agar plates supplemented with 100

μg/mL ampicillin and 1 mg of X-gal. The plates were incubated overnight at 37°C and then held for an additional 12 hours at 4°C for the development of blue color. White colonies were selected, cultivated in 2xYT medium overnight and used to isolate plasmid DNA. Plasmid DNA preparations were analyzed by restriction endonuclease digestion and electrophoresis, and plasmids carrying the desired 1.457 kb inserts were designated pCR1014.

2.10.3 Introduction of an *Nde*I site in front of *pcbC* using vent DNA polymerase:

Introduction of an *Nde*I site at the translation initiation codon of the *pcbC* gene by PCR was repeated using deep vent DNA polymerase and

OND-1 : 5' TAGGAATTCCATATGCCAGTTCTGATGC 3' and

OCS-4: 5' ATGTAGGTACCGGAGTTGACCAGGAACTCGTTCT 3'

as the primers. PCR reactions contained 0.1 pmol of *Pvu*II linearised, alkali denatured pIPS-1 plasmid template, 50 pmol of each primer, 5% dimethylsulfoxide, 1x NEBuffer [10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$ and 0.1% Triton X-100], 200 μM of each of the four deoxynucleotides and 4 U of Deep Vent DNA polymerase in a final reaction volume of 100 μl with a 100 μl mineral oil overlay. After incubating at 95°C for 3 minutes, 25 cycles of PCR were carried out with each cycle at 95°C, 60°C and 72°C for 0.5, 1.0 and 5 minutes respectively in a Techne model-PHC-2 thermal cycler followed by an additional 10 minute incubation at 72°C at the termination of the 25 cycles. Amplified PCR DNA was extracted once with phenol/chloroform (1:1), precipitated using 98% ethanol and resuspended in TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA] using standard procedures as described in Sambrook *et al.* (1989). The PCR product DNA was digested with *Sal*I and cloned into the plasmid vector pUC119, which had been digested with *Sma*1/ *Sal*1 restriction enzymes. *E. coli* XL-1 Blue competent cells were transformed with the pUC119 recombinant plasmid and transformants with recombinant constructs were identified based on blue/white color selection using X-gal/IPTG (Sambrook *et al.*, 1989). The resulting plasmid pUC07VW,

contains an insert consisting of a 0.7 kb fragment from the 5' end of the *pcbC* gene. pUC07VW was sequenced to verify the introduced mutations and fidelity in the *pcbC gene* sequence by automated DNA sequencing using forward and reverse universal primers.

2.10.4 Creation of C104S and C251S mutations:

Conversion of the cysteine residues at amino acid positions 104 and 251 of the IPNS protein encoded by the *pcbC* gene, into serine residues was carried out simultaneously by PCR using

OCS-2 : 5' TCCCGGGCAGGAAGGCCGTCGAGTCCTTCTCTTACCTC 3' and

OCS-4 : 5' ATGTAGGTACCGGAGTTGACCAGGAAGTCGTTCT 3'

oligonucleotides as primers and pIPS-1 plasmid as the template. The C and G nucleotides underlined in OCS-2 and OCS-4 respectively, replace the G and C nucleotides in the wild type *pcbC* gene and result in replacement of cysteine residues by serine residues. PCR reaction conditions were essentially the same as described earlier for the introduction of an *Nde*1 site by PCR using *Taq* DNA polymerase. DNA fragments resulting from the PCR reaction were purified from 5% polyacrylamide gels as recommended by Sambrook *et al.* (1989) and cloned into the pCR™II plasmid vector (Invitrogen Corp., San Diego, CA) as mentioned earlier for cloning of PCR products using pCR1000. The resulting plasmid named pCRII24 was sequenced to verify the introduced mutations and fidelity in the *pcbC* gene sequence by automated DNA sequencing using forward and reverse universal primers.

2.11 Site-directed Mutagenesis Using a Single-stranded Template:

2.11.1 Creation of C37S and C142S mutations:

These two mutations were introduced into the *pcbC* gene by single-strand oligonucleotide-directed mutagenesis based on the mutagenesis principle described by Kunkel *et al.* (1985). OCS-1 and OCS-3 oligonucleotides

OCS-1: 5' ACGGGGCCTCCCGCGGCTCG 3'

OCS-3: 5' GGCCGTTCTCCGAGGGCTAC 3'

with new cytosine residues (shown underlined) in place of guanine were designed to introduce these two mutations. Annealing mixtures containing 1.0 pmol of single-stranded pUC07VW plasmid DNA template described in Section 2.10.3, 40 pmol of one of the phosphorylated mutagenic oligonucleotides and 3 µL of 5x Sequenase buffer [200 mM Tris-HCL (pH 7.5), 100 mM $MgCl_2$, 250 mM NaCl] in a total volume of 15 µL were warmed to 80°C and then cooled slowly to room temperature to facilitate annealing of the oligonucleotides to the template.

Five microlitres of the annealing mixture was then added to elongation mixtures consisting of 0.4 mM of ATP, 5 mM DTT, 4 µL of 5x Sequenase buffer, 300 µM of each dNTPs, 3 units of Sequenase version 2 and 1 U of T4 DNA ligase in a total reaction volume of 20 µL. After incubation for 4 hours at 37°C, 10 µL of the mixture was used to transform *E. coli* MV1193.

2.11.2 Screening for C37S and C142S Mutants:

2.11.2.1 Colony hybridisation:

E. coli MV1193 colonies arising from C37S mutagenesis were screened by colony hybridisation. The hybridisation procedure developed by Amersham was adopted with the following modifications. A Hybond-N nylon hybridisation membrane (8.2 cm diam. circle) was placed on the surface of a 2xYT agar plate supplemented with 100 µg/mL ampicillin. Fresh transformants grown overnight were spotted on the membrane surface and incubated at 37°C for 4 hours. The membrane was carefully lifted from the agar plate and placed, colony side up on an absorbant filter paper soaked in 10% SDS for 10 minutes. The membrane was then transferred to a filter paper soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 7 minutes, again colony side up. The membrane was then placed on a filter paper soaked in neutralising solution [1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 0.001 M EDTA] for 3 minutes, colony side up. After repeating the

neutralisation step for another 3 minutes, the membrane was washed twice in 2x SSPE [20x SSPE= 3.6 M NaCl, 0.2 M sodium phosphate, 0.02 M EDTA (pH 8.0)] for 15 minutes each time and then air dried by placing on a dry filter paper for 30 minutes. The DNA in the sample was fixed by baking in a vacuum oven at 80°C for 2 hours.

The membrane was placed in a hybridisation cylinder (3 x 15 cm) containing 10 mL of 5x SSPE, 5x Denhardt's solution (100x Denhardt's solution = 2% bovine serum albumin, 2% Ficoll™, 2% polyvinylpyrrolidone), 0.5% SDS and 300 µg of denatured salmon sperm DNA and allowed to prehybridise for 1 hour at 55°C. Following prehybridisation, labelled oligonucleotide probe (OCS-1) was added to the prehybridisation solution and incubated for 15 hours at 55°C. Then the membrane was washed at 55°C with 20 mL of wash solution containing 2x SSPE, 0.1% SDS for 30 minutes. The wash solution was replaced with 20 mL of 1x SSPE, 0.1% SDS and washed for an additional 30 minutes at 55°C. The membrane was then wrapped in SaranWrap and exposed to Kodak X-OMAT-AR film to form an autoradiogram. The exposure of the X-ray film was limited to a maximum of one hour.

2.11.2.2 DNA sequence analysis:

Screening of *E. coli* MV1193 transformants to identify those C142S mutants which had been formed by site directed mutagenesis was carried out through DNA sequencing by the chain termination method of Sanger *et al.* (1977). Single-stranded templates were isolated from individual clones as described by Vieira and Messing (1987) using M13KO7 as the helper phage. Oligonucleotide (OSP-1) 5' ACGAGGTGAACCTCTGGCC 3' was employed as the sequencing primer in the sequencing reactions and [α^{32}P] dATP was included to incorporate radioactive label into the newly synthesized strands. Since the mutagenesis involved the substitution of cytosine in the place of guanine, for initial screening purposes, only sequencing reactions containing dideoxycytosine were employed. The promising isolates were then subjected

to dideoxy sequencing using all four bases to confirm the DNA sequence surrounding the site of mutagenesis. Once a clone containing the desired C142S mutation was found (plasmid designated pUC07V3), the entire sequence of the mutant *pcbC* gene in pUC07V3 was verified by automated sequencing.

The dideoxy sequencing procedure used for screening of clones was carried out as follows. Annealing mixtures containing 3.5 µL of single-strand DNA template isolated from different clones, 0.5 µL (0.1 µg/µL) of appropriate sequencing primer and 1 µL of 5x Sequenase buffer were heated to 70°C and then slowly cooled to room tempertaure. Two microlitres of a labelling mixture containing 20 mM DTT, 0.6 µM deaza GTP, 0.6 µM dCTP, 0.6 µM dTTP, 25 µCi of [α^{32}P]dATP and 26 units of Sequenase version 2 in a total volume of 50 µLwas then added to each annealing mixture and incubated at room temperature for 5 minutes. Two and a half microlitres of termination mixture containing 80 µM deaza GTP, 80 µM dATP, 80 µM dTTP, 80 µM dCTP, 8 µM ddCTP and 50 mM NaCl was then added to each sample tube and incubated at 37°C for an additional 5 minutes. Two microlitres of stop dye containing 1 mg of xylene cyanol FF, 1 mg of bromphenol blue and 20 µL of 0.5 M EDTA in a total volume of 1 mL of deionised formamide was added to each tube, boiled to reduce the volume to 3 µL and loaded on to a denaturing 6% polyacrylamide sequencing gel (37.5:1, acrylamide: N,N'methylene bisacrylamide) prewarmed to 50°C. Following electrophoresis, the gel was exposed to Kodak X-OMAT-AR film at -70°C.

2.12 Expression of *pcbC* to give Isopenicillin N synthase:

2.12.1 Insoluble form:

pcbC expression was carried out essentially as described by Tabor and Richardson (1985). *E. coli* K38 cells carrying expression vector derivatives together with the helper plasmid pGP1-2 were inoculated into expression medium [2% Bacto-Tryptone (Difco), 1% yeast extract (Difco), 0.5 % NaCl, 0.2% glycerol, 50 mM potassium phosphate buffer (pH 7.2), ampicillin (100 µg per mL) and kanamycin (75 µg per mL)] to an initial OD_{600}

of 0.05. The cultures were incubated at 30°C until an OD_{600} of 1.5 was reached, and then transferred to 42°C for 30 min to induce production of T7 RNA polymerase. Cultures were then incubated at 37°C for a further 2 hours.

2.12.2 Effect of temperature on *pcbC* gene expression:

E. coli K38 cells carrying the pMD0P7 construct were cultivated in 4 x 400 mL of expression medium at 30°C in four separate flasks at 280 rpm until the OD_{600} reached 1.5 from an initial optical density of 0.05. The flasks were then transferred to 42°C for the induction of T7 RNA polymerase. After 30 minutes of heat induction, each of the four flasks was incubated at a different temperature, 20, 25, 30 or 37°C for two hours. Crude extracts and pellet extracts were prepared from each culture as mentioned in Section 2.12.5 and 2.12.6 respectively. The solubilisation and refolding of inclusion bodies, protein content and IPNS activity from each culture was carried out as discussed in Section 2.13.1.1, 2.19 and 2.20 respectively. *E. coli* K38 cells carrying pMD0R7 were cultivated as described above with the final incubation at 37°C, to serve as a negative control for inclusion body formation. Culture samples from each temperature were subjected to electron microscopic analysis along with the negative control.

2.12.3 Electron Microscopic Analysis: (conducted by R. K. Sherburne)

Ten millilitre of 2% glutaraldehyde in 0.1M cacodylate buffer (pH 7.4) was added to 10 mL of culture samples at room temperature. After incubating for 75 minutes, each suspension was centrifuged at 4000 xg for 15 minutes. The pellet was treated with 2 mL of 2% glutaraldehyde in 0.1 M cacodylate buffer again for 30 minutes. The pellet was then washed repeatedly with 0.1 M cacodylate buffer for 10 minutes, 20 minutes and 30 minutes and transferred to 1% osmium tetroxide in 0.1 M cacodylate buffer. The suspension was centrifuged for 5 minutes at 4000 xg and the sedimented cells were transferred to fresh 1% osmium tetroxide in 0.1 M cacodylate buffer for 16 hours at 4°C. The pellet was then washed with 0.1 M cacodylate buffer for 3 x 20 minutes followed by a final wash for 10 minutes in distilled water. The pellet was then dehydrated in a graded

series of alcohol in distilled water at 25% for 15 minutes, 50% for 15 minutes, 75% for 15 minutes, 90% for 2 x 15 minutes, 100% for 2 x 15 minutes followed by treatment with propylene oxide for 2 x 15 minutes. The pellet was then treated with propylene oxide: LX 112 (1:1) for 16 hours, uncapped. After treating the pellet with fresh LX112 in propylene oxide for 2 hours, the pellet was placed in a mould and cured for 24 hours at 60°C. Specimens prepared from cultures grown at various temperatures were sectioned and analysed by transmission electron microscopy using a Philips Model 410 electron microscope.

2.12.4 Soluble form:

When soluble IPNS was required, conditions for expression of *pcbC* were the same as those described for expression of insoluble IPNS except that, after incubation at 42°C for 30 minutes to induce T7 RNA polymerase, the culture was transferred to 20°C instead of 37°C.

2.12.5 Preparation of crude extract:

Expression cultures were harvested by centrifugation for 10 min at 12000 xg, and cell pellets were washed with one half of the original culture volume of washing buffer [50 mM Tris HCl (pH 7.5), 0.1 mM dithiothreitol, 0.01 mM EDTA, 10% glycerol] and resuspended in 1/50 of the original culture volume of IPNS assay buffer (washing buffer containing 1.0 mM dithiothreitol). The cells were disrupted by sonication for 2 x 15 s at intensity setting 7 (Sonifier Cell Disruptor 350, 0.75 in. diameter probe, Branson sonic Power Co.) and centrifuged at 17000 x*g* for 30 minutes. The resulting supernatant (crude extract) contained soluble IPNS protein; pellet material contained insoluble IPNS protein in inclusion bodies. Crude extracts from cultures used to express *pcbC* to give soluble IPNS were used as the starting material for purification of soluble IPNS.

2.12.6 Preparation of pellet extract:

Pellet material from cultures expressing *pcbC* to give insoluble IPNS was resuspended in IPNS assay buffer and the ultrasonic treatment and centrifugation was

repeated twice. The final pellet was resuspended in 2 mL of IPNS assay buffer for every 400 mL of original culture volume and this suspension of inclusion bodies was used for refolding studies.

2.13 Refolding studies:

2.13.1 Effect of desalting method on renaturation:

2.13.1.1 Gel filtration method:

Inclusion body suspension (2 mL) was added to 8 mL of denaturing solution [6.25 M urea, 62.5 mM dithiothreitol, 1.25 mM EDTA, and 50 mM Tris-HCl (pH 7.5)] to give a final concentration of 5M urea, 50 mM DTT, 1 mM EDTA and a final protein concentration which varied depending on the concentration of inclusion bodies in the starting suspension. After two hours at 21°C, the solubilized protein extract was desalted on a Sephadex G-25 gel filtration column which was equilibrated and eluted using IPNS assay buffer at 4°C.

2.13.1.2 Dialysis method:

Inclusion body suspension (0.2 mL) was added to 0.8 mL of denaturing solution, incubated for 2 hours at room temperature and then dialysed against a 250 fold excess of IPNS assay buffer for 8 hours at 4°C, and then against a 250 fold excess of fresh IPNS assay buffer for 16 hours. The protein concentration and IPNS activity of the extract were determined.

2.13.2 Effect of pH on renaturation:

Inclusion body suspension (2 mL) was added to 8 mL of denaturing solution containing Tris-HCl prepared with pH values of 7.22, 7.46 and 8.0. The volumes of the solutions were then adjusted to give protein concentrations of approximately 1 mg/mL using appropriately diluted denaturing solution to keep the final concentration of components in the denaturing solution constant. After incubation for 2 hours at room temperature, the protein extracts were renatured by dialysis as described in Section

2.13.1.2, against IPNS assay buffer containing Tris-HCl with pH values corresponding to those used during denaturation, and the IPNS activities were determined.

2.14 Purification of solubilised IPNS:

Ion exchange (DEAE-Trisacryl) chromatography:

Isopenicillin N synthase resulting from refolding of insoluble material was loaded on to a DEAE-Trisacryl column (1.6 x 30 cm) equilibrated with starting buffer [200 mM Tris-HCl (pH 8.0), 0.01 mM EDTA, 1 mM DTT and 10% glycerol] and eluted with a linear Tris-HCl gradient comprising 250 ml each of starting buffer versus limiting buffer [300 mM Tris-HCl (pH 8.0), 0.01 mM EDTA, 1 mM DTT and 10% glycerol]. IPNS elution from the column was monitored using a LKB 2238 UVICORD SII detector at 280 nm and a LKB 2210 1-channel recorder. Five millilitre (100 drop) fractions were collected and assayed for IPNS activity and characterised by SDS-PAGE analysis.

2.15 Purification of soluble IPNS:

2.15.1 Fractionation by ammonium sulfate precipitation:

Solid ammonium sulphate was added gradually to crude extract at 4°C with constant stirring and the material precipitating between 50 and 70% saturation was collected by centrifugation at 17000 xg for 30 minutes. The pellet was resuspended in 8 mL of TDEG buffer [50 mM Tris-HCl (pH 8.0), 0.01 mM EDTA, 1 mM DTT and 10% glycerol] and dialysed overnight against a 100 fold excess of TDEG buffer.

2.15.2 Ion-exchange (DEAE-Trisacryl) chromatography:

Dialysed material resulting from the ammonium sulfate fractionation was loaded on to a DEAE-Trisacryl column (1.6 x 30 cm) and eluted as described earlier for solubilised IPNS. Fractions displaying IPNS activity were pooled and concentrated by ultrafiltration using an Amicon PM 10 membrane to 1/20 of the original volume.

2.15.3 Gel filtration (Superose 12 HR 10/30) chromatography:

Concentrated IPNS containing material from the DEAE-Trisacryl column was applied repeatedly in 0.25 mL aliquots to a Superose 12 column equilibrated with TDEK

buffer [100 mM Tris-HCl (pH 7.5), 1 mM DTT and 0.01 mM EDTA and 100 mM KCl], and eluted at a flow rate of 0.5 mL/min. The column was attached to a Pharmacia FPLC system with the following components; 2 P-500 pumps, a P-1 peristaltic pump, a UV-M monitor, a LCC-500 liquid chromatography controller and a Frac-100 fraction collector. Two minute (1.0 mL) fractions were collected and assayed for IPNS activity.

2.15.4 Ion-exchange (Mono Q HR 5/5) chromatography:

Fractions containing IPNS from the Superose purification step were pooled, concentrated by ultrafiltration and applied repeatedly in 0.3 ml aliquots to a Mono Q HR 5/5 column equilibrated with TDEK buffer. After each sample application, the column was washed for 5 minutes using TDEK buffer and then IPNS was eluted at a flow rate of 0.5 mL/min with a linear KCl gradient spanning 30 minutes with TDEK containing 200 mM KCl as the limiting buffer. IPNS containing fractions were pooled, concentrated by ultrafiltration using a disposable filtration cell (Filtron Corporation) with a molecular weight cut off of 10,000 daltons, and dialysed against a 100 fold excess of TDE buffer (50 mM Tris-HCl (pH 7.5), 0.01 mM EDTA and 1 mM DTT) before being stored at -80°C.

2.16 Circular Dichroism: (conducted by K. Oikawa)

Circular dichroism (CD) measurements were done on a Jasco J-720 spectropolarimeter (Jasco Inc., Easton, MD) interfaced to an Epson Equity 386/25 computer and controlled by Jasco software. The thermostatted cell holder was maintained at 25°C with a Lauda RMS circulating water bath (Lauda., Westbury, NY). The instrument was routinely calibrated with ammonium d-(+)-10 camphor sulfonate at 290.5 nm and 192 nm, and with d-(-)-pantoyllactone at 219 nm. Each sample was scanned 16 times and noise reduction applied to remove the high frequency before calculating molar ellipicities. The voltage to photomultiplier was kept below 500 V to prevent distortion of the CD spectrum. Cell used was 0.05 cm and 1 cm microcell for near UV (250-320 nm) (calibrated for pathlength). Protein concentrations were 0.02 to

0.4 mg/mL range for all the samples. The mean residue weight for the protein was 111.535. CD spectra for the protein were analysed by the Contin program version 1.0 of Provencher and Glockner.

2.17 N-ethyl maleimide (NEM) treatment of IPNS:

Solubilised, purified IPNS preparations from the wild type and the different mutants were treated with 1 mM N-ethylmaleimide for 30 minutes at room temperature. Excess N-ethyl maleimide was separated from the inhibitor-treated IPNS by gel filtration chromatography using a NAP-5 Sephadex G-25 DNA grade column (Pharmacia) which was equilibrated and eluted with TG buffer [Tris-HCl (pH 7.2) and 10% glycerol] as recommended by the supplier.

2.18 SDS-PAGE analysis:

Protein samples from various stages of the isolation and purification procedure were analysed by electrophoresis on 10% polyacrylamide gels containing 0.1% sodium dodecyl sulfate (SDS) as described by Blackshear (1984). Chymotrypsinogen A (25,000), ovalbumin (45,000) and bovine serum albumin (67,000) were used as molecular weight markers.

2.19 Protein Assay:

Protein concentrations were estimated by the dye binding method of Bradford (1976) as recommended by BioRad, Richmond, CA, USA using bovine γ globulin as standard.

2.20 Isopenicillin N synthase assay:

IPNS activity was determined by a reverse-phase high performance liquid chromatography procedure as described previously (Jensen *et al.*, 1982b). In brief, reaction mixtures containing 0.29 mM bis-ACV, 4 mM dithiothreitol, 2.8 mM sodium ascorbate, 45 μM $FeSO_4$, 10% glycerol, 50 mM Tris-HCl (pH 7.2) and enzyme in a total reaction volume of 80 μL were incubated from 10-60 minutes at 20°C. The reactions were terminated by the addition of 80 μL of methanol, and centrifuged for 5 minutes at

13000 xg to sediment precipitated material. Supernatant samples (20 µL) were analysed by HPLC under isocratic conditions using methanol (5%)- 50 mM potassium phosphate buffer (pH 4.0, 95%) as the mobile phase at a rate of 2 mL/min. Column effluent was monitored for UV absorbing material at 214 nm. The HPLC system consisted of a µBondapak C18 column, M-6000 pump, WISP-model 712 sample injector and Lambda-Max model 480 LC spectrophotometer linked through a System Interface Module to a NEC Powermate SX/16 Computer equipped with BASELINE 810 software. All components were from Millipore Waters.

Crude expression extracts contained β-lactamase activity arising from the *bla* gene on the vectors. This β-lactamase activity destroyed the isopenicillin N product of IPNS. Therefore IPNS activity was estimated from the disappearance of the substrate, ACV rather than the appearance of product. However, purified IPNS preparations were free of β-lactamase activity and so the appearance of isopenicillin N as well as the disappearance of ACV could both be used as measures of IPNS activity. One unit of isopenicillin N synthase activity is defined as that amount of enzyme which produces 1 µmol of isopenicillin N / min.

3. RESULTS

Isopenicillin N synthase is produced at very low levels in *S. clavuligerus*, and after a multi-step purification procedure, two litres of culture yielded only 0.2 mg of purified IPNS protein. Structural characterisation by X-ray crystallography requires milligram amounts of proteins, and obtaining those amounts of IPNS protein from *S. clavuligerus* would be tedious and impractical. Therefore, attempts were made to express the *pcbC* gene at high-level in *S. lividans*, a related *Streptomyces* species. Isopenicillin N synthase expression was achieved but the level of expression was no greater than the level seen in *S. clavuligerus* (Jensen *et al.*, 1989). Therefore, attempts were initiated to express the *pcbC* gene in *E. coli*. *E. coli* is the best characterised species biochemically and genetically, and offers the advantage of having served as a host for expression of numerous heterologous genes, including *pcbC* genes from other fungal and bacterial species. Early expression studies of the *pcbC* gene from *S. clavuligerus* in *E. coli* in our laboratory used a two-cistron expression system under the transcriptional control of the *lac* promoter, but once again the level of expression achieved was the same as that seen in *S. clavuligerus* (Doran *et al.*, 1990). Although the reasons for the low level of expression were not clear, the feasibility of expression in *E. coli* was demonstrated, which served as a starting point for the current high-level expression studies. Even though the *lac* promoter is a strong promoter and it has been used for high-level expression of many heterologous genes, the present study focussed on a T7 promoter-based dual plasmid expression system. This plasmid expression system provides several advantages over *lac* promoter-based expression systems. The pT7-4, pT7-5, pT7-6 and pT7-7 vectors contain the bacteriophage T7 gene 10 promoter, which is specifically recognised by T7 RNA polymerase, located upstream of a multiple cloning site, and they also encode an ampicillin resistance gene (*bla*). In pT7-4, *bla* is oriented in the same direction as the T7 promoter and therefore expression of *bla* is influenced by the T7 promoter as well as by

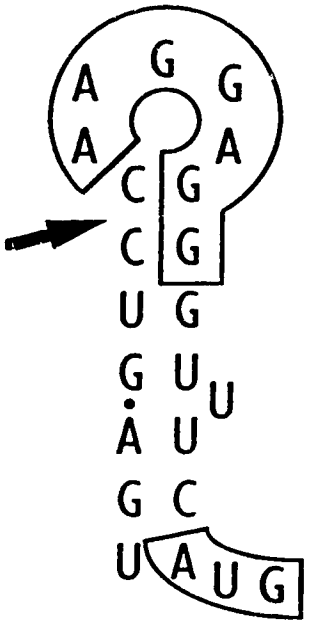
its own promoter. Hence, in pT7-4-based expression constructs, *bla* would serve as a reporter gene for the overall level of transcriptional activity in the plasmid, including the gene of interest. In pT7-5 and pT7-6, *bla* is oriented opposite to the T7 promoter and therefore expression is controlled solely by its own promoter. pT7-7 is similar to pT7-5 except that it contains the ribosome binding and translation initiation sites of the T7 gene 10 located between the gene 10 promoter and the multiple cloning site. These additional features of pT7-7 facilitate the placement of the gene to be expressed under translation initiation signals supporting optimal expression in *E. coli*.

3.1 High-level expression of the *pcbC* gene:

In order to increase the level of expression of the *pcbC* gene, a 1.7 Kb DNA fragment containing the *pcbC* gene, flanked by 335 bp of upstream sequence and 383 bp of downstream sequence, was excised from the plasmid pIPS-1 (see Section 2.3) by *Sma*1 digestion and inserted into the similarly digested T7-based expression vector pT7-4. However, the resulting plasmid (pJD004) gave no detectable production of IPNS protein upon induction of T7 RNA polymerase (data not shown). These early studies were conducted by a previous postdoctoral fellow, J.L. Doran. Secondary structure analysis of the nucleotide sequence upstream of the *pcbC* initiation codon in pJD004 showed that the putative transcript may form an imperfect stem-loop structure incorporating the ribosome binding site (Figure 5). Formation of this secondary structure could interfere with ribosome binding, thereby limiting translation initiation.

To eliminate this possible impediment to expression, pIPS-1 was digested with *Xba*1 and *Sty*1 to remove a DNA fragment containing most of the sequence upstream of *pcbC* except for 13 bp. The truncated plasmid (pIPS-1A) was recircularised by ligation after the " sticky ends " were filled in with the Klenow fragment of DNA polymerase I. A 1.4 Kb DNA fragment containing the *pcbC* gene, 13 bp of upstream sequence and 383 bp of downstream sequence, was excised from pIPS-1A by digestion with *Eco*RI and *Hin*dIII, and cloned into pT7-4 to generate pMD004 (Figure 6). In pMD004, *pcbC*

Figure 5: Schematic representation of the potential secondary structure in the RNA encoded by the DNA sequence immediately upstream of *pcbC*. DNA sequences encoding the translation initiation codon (AUG) and a Shine-Dalgarno motif (AAGGAGG) are boxed. An arrow indicates the cleavage site used in subsequent studies to remove most upstream sequence. The theoretical free energy of formation (ΔG) of the stem-loop structure, calculated according to Zuker and Steigler. (1981), is indicated.

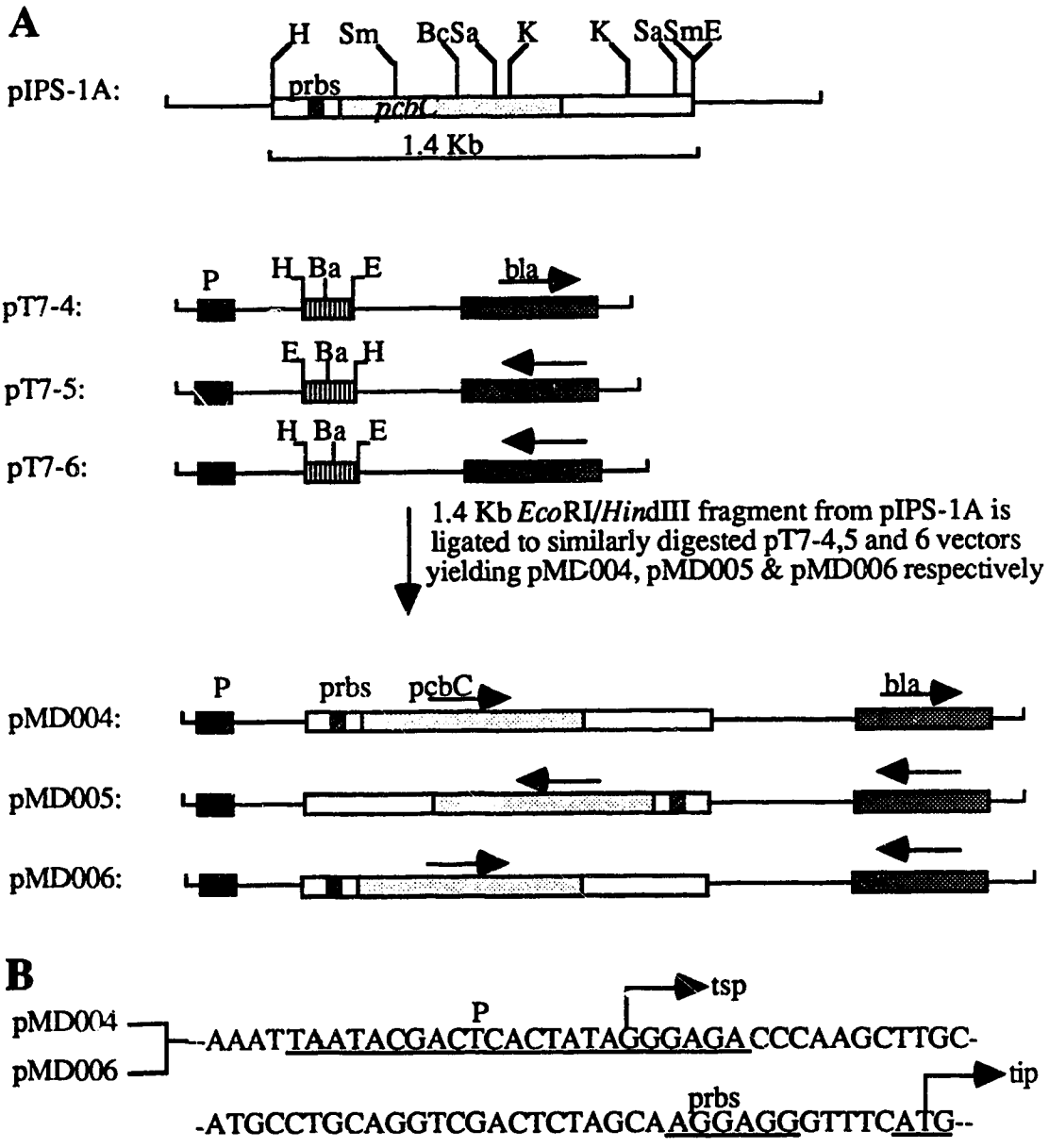


ΔG = -12.1 Kcal/mol

Figure 6: Construction of pMD004, pMD005 and pMD006 expression vectors.

A: A 1.4 Kb DNA fragment containing the 990 bp open reading frame of the *pcbC* gene, 13 bp of upstream sequence including the ribosome binding site (prbs) and 383 bp of downstream sequence carried in pIPS-1A was released by digestion with *Eco*RI and *Hin*dIII and was cloned into similarly digested pT7-4, pT7-5 and pT7-7 vectors to yield pMD004, pMD005 and pMD006 vectors, respectively. P denotes the T7 gene 10 promoter; *bla* denotes the β-lactamase gene; Bc, E, H, K, Sa and Sm symbols designate sites for *Bcl*1, *Eco*R1, *Hin*d111, *Kpn*1, *Sal*1 and *Sma*1 restriction enzymes. The β-lactamase gene is shown as a dark crosshatched box while the *pcbC* gene is shown as a light shaded box. The multiple cloning site in the pT7 vectors is shown as a vertically striped box.

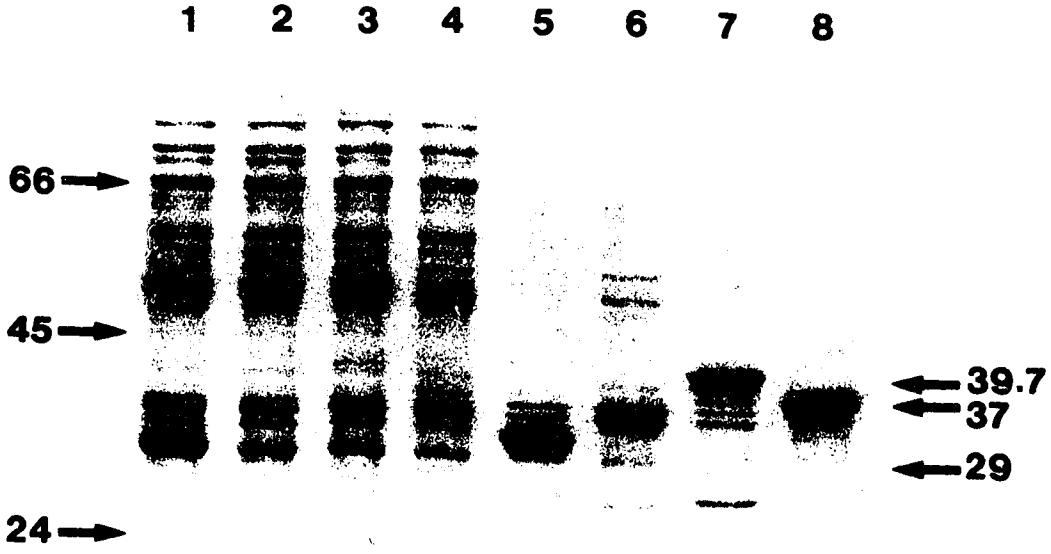
B: Close up representation of the DNA sequence from the T7 gene 10 promoter to the *pcbC* start codon region of pMD004 and pMD006 recombinant plasmids. tsp denotes the transcription start point; tip indicates the translation initiation point; the 23 bp conserved DNA sequence recognised by the T7 RNA polymerase (T7 gene 10 promoter), the ribosome binding site (prbs) and the translation start codon (ATG) of the *pcbC* gene have been underlined.


A
pIPS-1A:
H Sm BcSa K K SaSmE
prbs
pcbC
1.4 Kb
pT7-4:
P H Ba E
bla
pT7-5:
E Ba H
pT7-6:
H Ba E
1.4 Kb EcoRI/HindIII fragment from pIPS-1A is ligated to similarly digested pT7-4,5 and 6 vectors yielding pMD004, pMD005 & pMD006 respectively
pMD004:
P
prbs
pcbC
bla
pMD005:
pMD006:
B
pMD004
pMD006
P
tsp
--AAATTAATACGACTCACTATAGGGAGACCCAAGCTTGC-
tip
prbs
-ATGCCTGCAGGTCGACTCTAGCAAGGAGGGTTTCATG--

retains its native RBS but the upstream arm of the potential stem-loop structure is removed at the position indicated by the arrow in Figure 5. The DNA sequence upstream of the start codon of the *pcbC* gene in the pMD004 construct is given in Figure 6B. When *pcbC* expression was tested using this construction, no major protein band of the size expected for IPNS (molecular weight 37, 000 daltons) was detected by SDS-PAGE, in either the soluble or particulate fraction of the cell extract (Figure 7). This inability of pMD004 to direct expression of *pcbC* suggested that some factor other than occlusion of the RBS by secondary structure was preventing expression of *pcbC*. Examination of SDS-PAGE profiles demonstrated the presence of a prominent protein band of molecular weight 29,000 in soluble and insoluble fractions, which presumably represented the β-lactamase encoded by *bla*. *bla* is located downstream from *pcbC* in pMD004, and so the production of large amounts of β-lactamase suggested that transcription of both *pcbC* and *bla* was being directed by the T7 promoter, but that only the *bla* region of the transcript was being translated efficiently.

The occurrence of rare codons is no more frequent in *pcbC* than in other heterologous genes which are expressed at high-levels in *E. coli* (Makoff *et al.*, 1989) nor are these rare codons crowded near the 5' end of the gene (Leskiw *et al.*, 1988), which could reduce ribosomal loading and hence decrease translation initiation efficiency (Liljenstrom and Von Heijne, 1987). Furthermore, examination of the 13 bp of sequence upstream of *pcbC* in pMD004 revealed that the native RBS of *pcbC* matches the *E. coli* consensus Shine Dalgarno sequence or RBS (AGGAGG). However, the spacing between the RBS and the initiator codon is only 5 bp, much less than the optimal 7-10 bp in *E. coli*, and the C residue immediately upstream of the initiation codon may have an additional detrimental effect on translation initiation (Makoff and Smallwood, 1988; Hwang *et al.*, 1990). Possibly, the translational control signals for *bla* are recognised much more efficiently than those for the *pcbC*, and in the presence of high levels of *bla* transcript, translation of *pcbC* is not favored.

Figure 7: SDS-PAGE analysis of cell extracts from *E. coli* carrying pMD004, pMD006, pMD0N7, and pMD0P7. pMD0N7 and pMD0P7 constructs are described on page 65 and 68, respectively. Lanes 1, 2, 3 and 4 contained 30 µg of protein from the soluble fraction of cell extracts derived from cultures carrying pMD004, pMD006, pMD0N7, and pMD0P7 respectively. Lanes 5, 6, 7, and 8 contained 15 µg of protein from the particulate fraction (after solubilisation) of cell extracts derived from cultures carrying pMD004, pMD006, pMD0N7, and pMD0P7, respectively. The positions of molecular weight marker proteins bovine serum albumin (66,000), ovalbumin (45,000) and trypsinogen (24,000) are indicated with arrows. The arrow labelled 29K indicates the position of β-lactamase band.


1
2
3
4
5
6
7
8
66
45
24
39.7
37
29

To examine this possibility, the 1.4 Kb *pcbC* gene-containing fragment was moved into pT7-6 by *Eco*RI/*Hin*dIII digestion to give pMD006 (Figure 6). In this construction, the *bla* gene is in the opposite orientation to the T7 promoter, transcribed only from its own promoter, and should not compete with the *pcbC* for expression. When expression extracts from cells carrying pMD006 were examined by SDS-PAGE, again no IPNS protein was evident (Figure 7). The SDS-PAGE profile of the pMD006 extract was not appreciably different from that of pMD005, a negative control in which the *pcbC* gene itself was in the opposite orientation relative to the T7 promoter (not shown). The failure to detect expression of IPNS in cells carrying pMD006 indicated that high-level expression of *bla* was not a factor restricting *pcbC* expression, but rather that the sequence upstream of *pcbC* was probably not capable of supporting efficient translation initiation in *E. coli*.

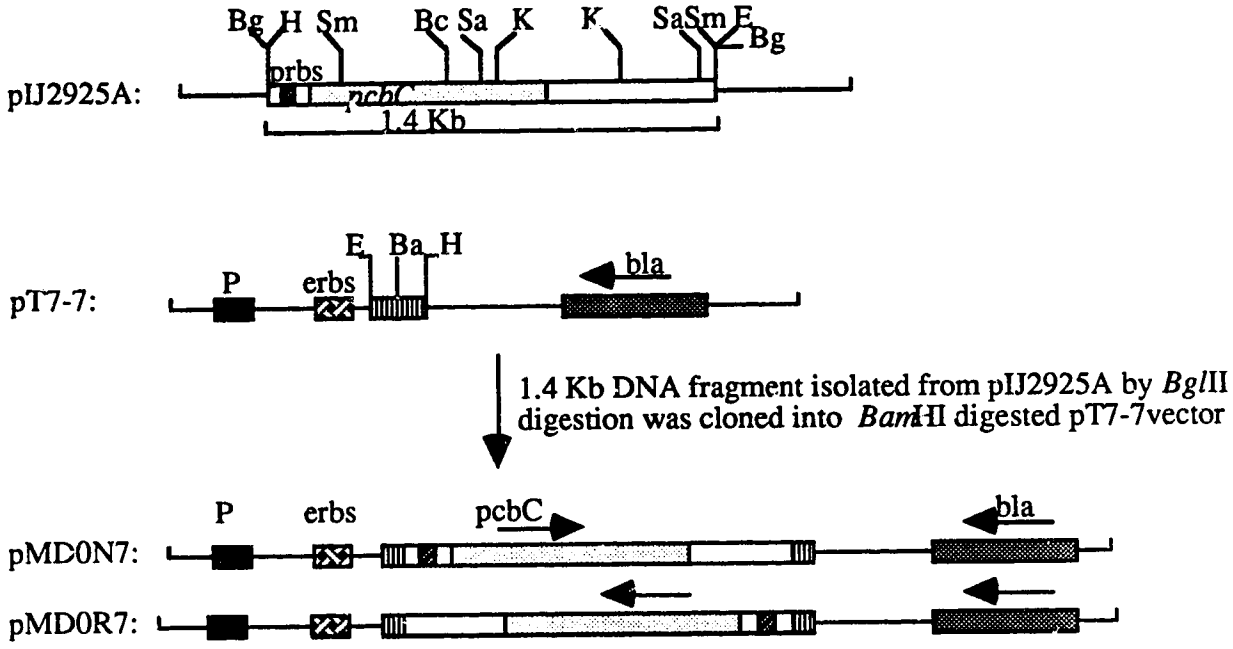
In order to examine the ability of the native *pcbC* upstream sequence to initiate translation, the 1.4 Kb *pcbC* gene-containing fragment from pIPS-1A was cloned into pT7-7 resulting in the formation of pMD0N7 (Figure 8). This was accomplished by first transferring the 1.4 Kb *pcbC* gene-containing fragment from pIPS-1A into the intermediate vector pIJ2925, so that the fragment would be flanked by *Bgl*II sites. The fragment was then removed by digestion with *Bgl*II and cloned into pT7-7 which had been digested with *Bam*HI. This construction results in the *pcbC* gene being fused in frame to a fragment from the 5' end of the T7 gene 10. An *E. coli*-type RBS precedes the fused gene, and within the fused gene, the native *S. clavuligerus* RBS precedes the *pcbC* gene. Therefore, transcription driven by the T7 promoter could give rise to either a fusion protein, authentic IPNS protein or a mixture of the two, depending on which of the two translation initiation sites was preferred (Figure 8B). Examination of expression extracts by SDS-PAGE showed that high-levels of the 39,700 dalton fusion protein were produced, but no 37,000 dalton IPNS protein (Figure 7). The fusion protein was found almost exclusively in the insoluble fraction of the expression extract. These results

Figure 8: Construction of pMD0N7 and pMD0R7 expression vectors

A: A 1.4 Kb DNA fragment containing the 990 bp open reading frame of the *pcbC* , 13 bp of upstream sequence including the ribosome binding site (prbs) and 383 bp of downstream sequence carried in pIJ2925A was released by digestion with *Bgl*11 and cloned into pT7-7 digested with *Bam*H1 yielding pMD0N7 and pMD0R7. The two recombinant plasmids pMD0N7 and pMD0R7 represent the two possible orientations of the *pcbC* gene. The β-lactamase gene has been shown as a dark crosshatched box while the *pcbC* gene has been indicated by the light shaded box. P-stands for the T7 gene 10 promoter. erbs-stands for the ribosome binding site optimised for *E. coli*. *bla* denotes the β-lactamase gene. The multiple cloning site in pT7-7 is shown as a vertically striped box; Bc, Bg, E, H, K, Sa and Sm designate sites for *Bcl*1, *Bgl*11, *Eco*R1, *Hin*d111, *Kpn*1, *Sal*1 and *Sma*1 restriction enzymes.

B: Close up representation of the DNA sequence from the T7 gene 10 promoter to the *pcbC* start codon region of pMD0N7. tsp denotes the transcription start point; tip-1 and tip-2 denote the two possible translation initiation points; the T7 gene 10 promoter sequence (P), *E. coli* ribosome binding site (erbs), ribosome binding site of the *pcbC* gene (prbs) and the translation start codon have been underlined.

A



B

P tsp

pMD0N7: - - -AAATTAATACGACTCACTATAGGGAGACCACAACGGTTT

erbs

CCCTCTGAAATAATTTTGTTTAACTTTAAGAAGGAGATATAC

tip-1

AT ATG GCT AGA ATT CGC GCC CGG GGA TCT GCA GCC AAG

prbs

CTT GCA TGC CTG CAG GTC GAC TCT AGC AAG GAG GGT

tip-2

TTC ATG ---

confirm that there was a strong preference for the *E. coli* translation initiation signals over *pcbC* translation initiation signals. They also further demonstrated that the translation initiation signals of the *pcbC* gene could not support efficient translation initiation in *E. coli*. pMD0R7, a construct in which the *pcbC* gene containing-fragment was fused in reverse orientation relative to the 5' end of the T7 gene 10, did not direct the synthesis of either fusion protein or IPNS, as expected.

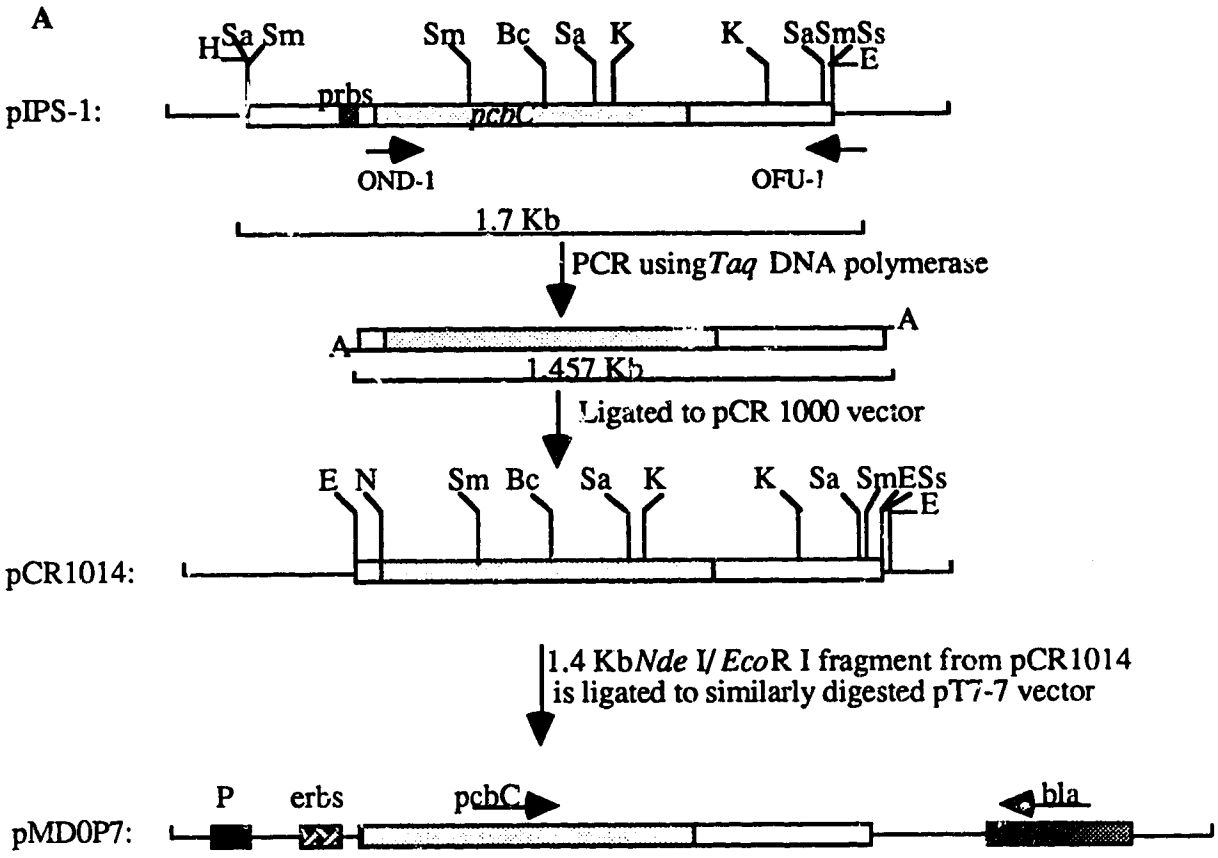
The pMD0N7 construct directed high level expression of fusion protein, but for physicochemical analysis, authentic IPNS is required. Analysis of the initiator codon region of the *pcbC* gene to try to find a way to remove the sequence upstream from the translation initiator codon failed to identify any appropriate restriction enzyme sites. The translation initiation codon of the T7 gene 10 fragment within pT7-7 contains an unique *Nde*I site as part of the multiple cloning site, and creation of a *Nde*I site in the translation initiation codon of the *pcbC* gene should allow expression of authentic IPNS under the control signals optimised for *E. coli*. Introduction of a *Nde*I site required an alteration of three nucleotides upstream of the initiator codon of *pcbC*, from TTC into CAT. OND-1 (5' TAGGAATTCCATATGCCAGTTCTGATGC 3') mutagenic primer was designed to incorporate these three nucleotides as well as to introduce an *Eco*RI site upstream of the *Nde*I site to facilitate further manipulations of the resulting PCR product. Introduction of a *Nde*I site into the translation initiation codon of *pcbC* gene was achieved by PCR using OND-1 as upstream primer, OFU-1 (Forward Universal primer) as downstream primer and pIPS-1 plasmid template as described in Section 2.10.1. The resulting mutant *pcbC* gene-containing fragment was purified by GeneClean and cloned into the pCR1000 vector (Invitrogen, CA) to give pCR1014.

The 1.4 Kb *pcbC* gene-containing fragment, excised by *Nde*I and *Eco*RI digestion from pCR1014, was cloned into similarly digested pT7-7 vector to give pMD0P7 (Figure 9). *E. coli* strain K38 carrying the pMD0P7 construct produced authentic IPNS at high-level upon induction. The presence of IPNS was evident from the appearance of a protein

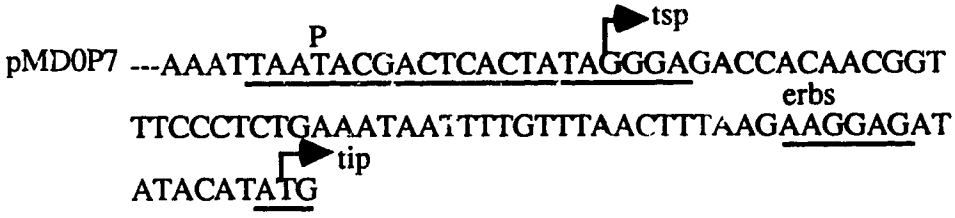
Figure 9: Construction of the pMD0P7 expression vector.

A. A pIPS-1 phagemid template containing the 990 bp open reading frame of the *pcbC* gene, 335 bp of upstream sequence and 385 bp of downstream sequence was subjected to PCR amplification by *Taq* DNA polymerase using the oligonucleotide primers 5' TAGGAATTCCATATGCCAGTTCTGATGC 3' (OND-1) and 5' CGCCAGGGTTTTCCCAGTCACGAC 3' (OFU-1). The nucleotide triplet within OND-1 which results in formation of an *Nde*I site is shown underlined. The resulting 1.457 Kb PCR product was cloned into the pCR1000 vector yielding pCR1014. A 1.4 Kb DNA fragment containing the 990 bp open reading frame of the *pcbC* gene, 385 bp of downstream sequence was released from pCR1014 by *Nde*I and *Eco*RI digestion and ligated to similarly digested pT7-7 vector to generate pMD0P7. The open reading frame of the *pcbC* gene has been shown as a light shaded box while the β-lactamase gene has been shown as a dark crosshatched box. prbs and erbs represent the *pcbC* gene ribosome binding site and optimised *E. coli* ribosome binding site respectively. H, Sa, Sm, Bc, K, E, Ss and N designate restriction enzyme sites for *Hind*111, *Sal*1, *Sma*1, *Bcl*1, *Kpn*1, *Eco*R1, *Sst*1 and *Nde*1 respectively.

B: Close up representation of the DNA sequence from the T7 gene 10 promoter to the *pcbC* start codon in pMD0P7. The T7 gene 10 promoter (P), optimised *E. coli* ribosome binding site (erbs) and the start codon (ATG) have been shown underlined. " tsp " denotes the transcription start point and " tip " indicates the translation initiation point.


A
H Sa Sm
Sm Bc Sa K
K Sa Sm Ss
E
pIPS-1:
prbs
pchC
OND-1
OFU-1
1.7 Kb
PCR using Taq DNA polymerase
A
A
1.457 Kb
Ligated to pCR 1000 vector
E N
Sm Bc Sa K
K Sa Sm E Ss
E
pCR1014:
1.4 Kb Nde I/ Eco R I fragment from pCR1014
is ligated to similarly digested pT7-7 vector
P
erbs
pchC
bla
pMD0P7:

B


P
tsp
pMD0P7 ---AAATTAATACGACTCACTATAGGGAGACCACAACGGT
erbs
TTCCCTCTGAAATAATTTTGTTTAACTTTAAGAAGGAGAT
tip
ATACATATG

band with a molecular weight of 37,000 Da in SDS-PAGE gels (Figure 7). TheIPNS protein was associated predominantly with the insoluble fraction of the cell extracts.

When cell extracts from the various pT7 constructs were assayed for IPNS activity, only low levels of activity were observed in the soluble fraction of expression extracts from pMD0N7 and pMD0P7, and no activity was seen in any of the particulate fractions (Table 1). In contrast, SDS-PAGE of expression extracts had shown that the expressed protein was associated predominantly with the particulate fraction, and the intensity of the stained protein bands suggested that much larger amounts of IPNS protein were present than the activity measurements would indicate. From this it was concluded that the IPNS protein (or fusion protein) expressed from constructs pMD0N7 and pMD0P7 was predominantly in an inactive form in inclusion bodies.

3.1.1 Nucleotide sequence analysis of the intact mutant *pcbC* gene in pMD0P7:

The PCR procedure employed to introduce an *Nde*I site at the translation initiation codon of the *pcbC* gene was carried out using *Taq* DNA polymerase, a thermostable enzyme isolated from *Thermus aquaticus*. It has been documented that DNA synthesis catalyzed by *Taq* polymerase is capable of introducing unintended nontemplate-encoded mutations as well as introducing nontemplate-encoded terminal adenylation. In order to verify the fidelity of the *pcbC* gene nucleotide sequence after the PCR procedure, the 1.4 Kb PCR product resulting from pCR1014 digestion with *Eco*RI and *Sst*I was cloned into similarly digested pUC119 vector generating pUC11914. The insert in pUC11914 contains the 990 bp open reading frame of the *pcbC* gene and 385 bp downstream sequence. pUC11914 was then digested with *Sal*I to release a 675 bp fragment which included 290 bp from the 3' end of the *pcbC* gene and 385 bp of downstream sequence, and the plasmid was recircularised yielding pUC07. pUC07, containing 0.7 Kb of *pcbC* gene sequence from the 5' end, was sequenced using the forward and reverse universal primers by automated sequencing. pUC11914 was used as the template and OSP-11

Table 1: IPNS activity in cell extracts from *E.coli* cultures carrying *pcbC* expression constructs

Expression construct	IPNS activity (units x 10^3 per mg protein)	
	Soluble	Particulate
pMD004	ND[a]	ND
pMD006	ND	ND
pMD0N7	0.048	ND
pMD0P7	0.718	ND

a ND means not detected

(5'GGTGCAGAACCTCCAGGTGG3') was used as the primer for sequencing the 290 bp 3' segment of the *pcbC* gene. The deduced amino acid sequence encoded by the wild type *pcbC* gene and the sequence obtained by analysis of the PCR-generated *pcbC* gene are shown in Figure 10. Scanning of the entire *pcbC* gene sequence for mutations revealed that three transversion and one transition mutations have been introduced during the PCR procedure, although one of the transversion mutation is silent. In order to ensure that the mutations were actually introduced by the PCR procedure, the wild type *pcbC* gene from *S. clavuligerus* was resequenced in these regions using subclones derived from the pIPS-1* (Petrich *et al.*, 1994) plasmid as templates. No discrepancies were found in the original *pcbC* gene sequence, and therefore, the mutations were concluded to have arisen from the PCR procedure. The distribution of the mutations over the entire sequence, and the nature of the base changes indicated that the mutations were introduced randomly and that there was no apparent specificity towards any particular base substitutions.

3.1.2 Construction of pMD0P7VW, an expression vector encoding the wild type IPNS:

Activity analysis of the pMD0P7 construct carrying the above mentioned mutations indicated that it was active. However, it is desirable to produce wild type IPNS without any unintended mutations for crystallographic studies, and so the mutagenesis was repeated using a modified procedure. A derivative of the pT7-7 vector containing the authentic *pcbC* gene without any unintended muations (pMD0P7VW) was generated in a procedure consisting of three steps. In the first step, an *Nde*I site was introduced by PCR at the translation initiation codon of the *pcbC* gene present in the pIPS-1 plasmid using OND-1 (5' TAGGAATTCCATATGCCAGTTCTGATGC 3'), OCS-4 (5' ATGTAGGTACCGGAGTTGACCAGGAAGTCGTTCT 3') primers and deep vent DNA polymerase as discussed in Section 2.10.3. The 0.773 Kb product resulting from the PCR procedure was purified, digested with *Sal*I and the 0.7 Kb DNA

A:

```
        10         20         30         40         50
MPVLMPSAHV PTIDISPLFG TDAAAKKRVA EEIHGACRGS GFFYATNHGV
        60         70         80         90        100
DVQQLQDVVN EFHGAMTDQE KHDLAIHAYN PDNPHVRNGY YKAVPGRKAV
       110        120        130        140        150
ESFCYLNPDF GEDHPMIAAG TPMHEVNLWP DEERHPRFRP FCEGYYRQML
       160        170        180        190        200
KLSTVLMRGL ALALGRPEHF FDAALAEQDS LSSVSLIRYP YLEEYPPVKT
       210        220        230        240        250
GPDGQLLSFE DHLDVSMITV LFQTQVQNLQ VETVDGWRDI PTSENDFLVN
       260        270        280        290        300
CGTYMAHVTN DYFPAPNHRV KFVNAERLSL PFFLNGGHEA VIEPFVPECA
       310        320        329
SEEVRNEALS YGDYLQHGLR ALIVKNGQT
```

B: **IPNS in pMD0P7 construct:**

Codon Position	Codon change	Aminoacid change
20	GGA to CGA	Glycine to Arginine
117	ATC to ACC	Isoleucine to Threonine
150	CTG to CAG	Leucine to Glutamine
275	GCG to GCT	Alanine to Alanine

Figure 10: Deduced amino acid sequence of wild type IPNS and the locations of mutations in the IPNS encoded by pMD0P7.

(A): Wild type amino acid sequence of the IPNS protein.

(B): Mutations observed in the IPNS protein encoded by the pMD0P7 construct. The bases introducing the mutations have been highlighted by bold letters.

fragment containing the mutant *pcbC* gene fragment with the *Nde*I site was ligated into pUC119 which had been digested with *Sma*I and *Sal*I. *E. coli* XL-1 Blue transformants containing the inserts were identified by colour selection using X-gal/IPTG. Two isolates containing the proper inserts were verified by restriction mapping analysis and subjected to automated DNA sequencing to scan the entire 0.7 Kb sequence for the presence of unintended mutations. Both isolates were free of undesirable mutations and so the recombinant plasmid was termed pUC07VW. It contained 0.7 Kb of the *pcbC* gene sequence starting from the 5' ATG start codon, and ending at an internal *Sal*I site. In the second step, in order to regenerate the intact *pcbC* gene, a 0.675 Kb DNA fragment containing 290 bp of *pcbC* gene sequence from the 3' end of the gene and 385 bp of downstream sequence was isolated from pIPS-1 by *Sal*I digestion and ligated to pUC07VW, also digested with *Sal*I, yielding pUC14VW. In the third step, a 1.4 Kb DNA fragment containing the intact *pcbC* gene was isolated from pUC14VW by *Nde*I/*Hin*dIII digestion and ligated to similarly digested pT7-7 vector generating pMD0P7VW (Figure 11).

3.2 Recovery of active IPNS from inclusion bodies:

Active IPNS protein was recovered from the particulate fraction of expression extracts by disaggregation and solubilisation of the insoluble inclusion bodies (IB) with urea under reducing conditions, followed by gradual removal of the urea to facilitate the renaturation of IPNS. Several methods which have been developed for the recovery of active recombinant proteins after denaturation were discussed earlier in Introduction. The choice of size exclusion chromatography for recovery of active IPNS was based on the observation that this procedure worked well for renaturation of IPNS from *S. jumonjinensis* (Y. Aharonowitz, personal communication) as well as the high concentration of protein material that could be obtained in a single recovery step.

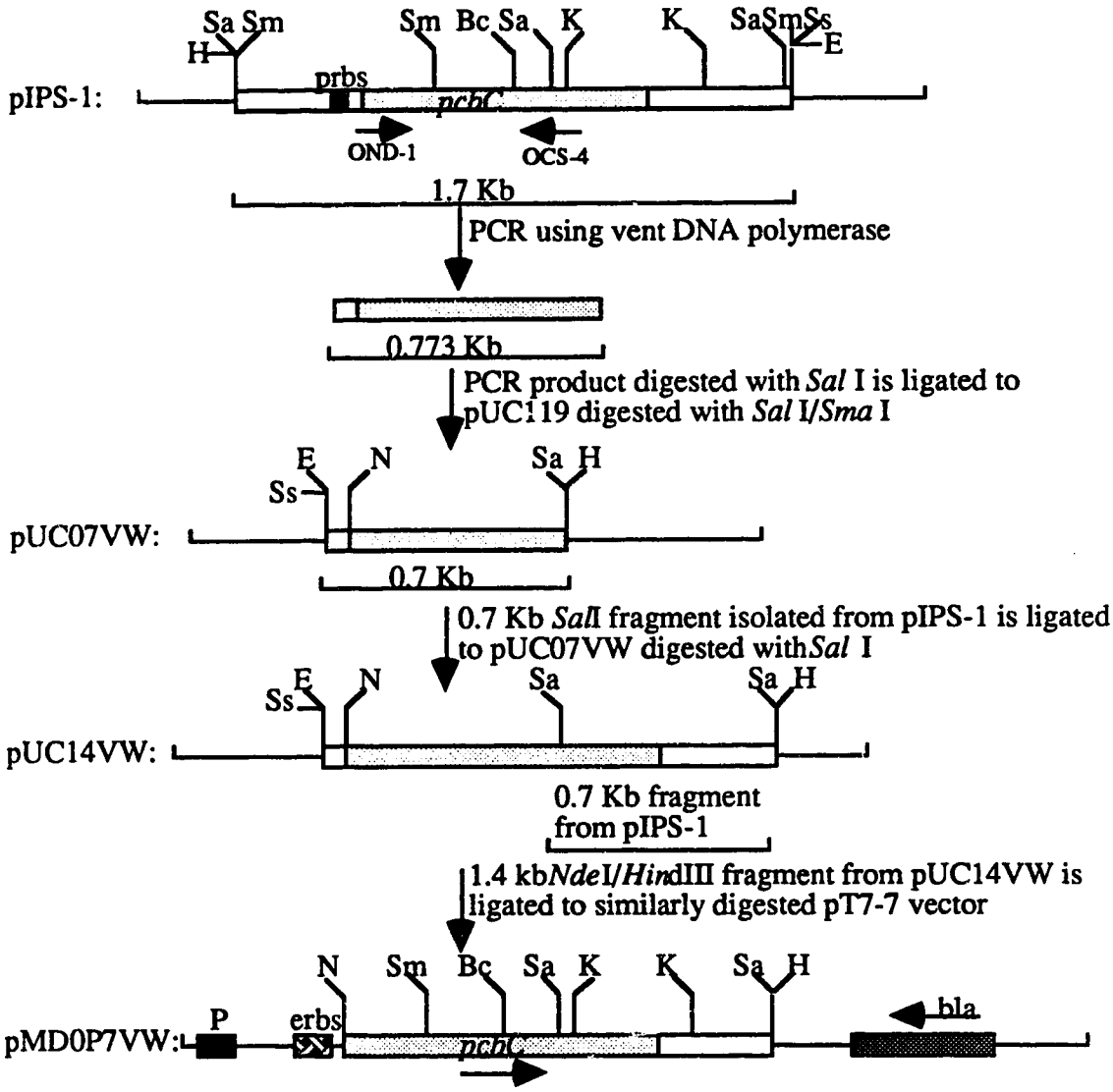
Refolding studies involved the addition of 8 mL of denaturing solution to

Figure 11: Construction of pMD0P7VW, an expression vector encoding the wild type IPNS.

A pIPS-1 phagemid template containing the intact *pcbC* gene, 335 bp of upstream sequence and 383 bp of downstream sequence was subjected to PCR amplification to introduce an *Nde* I site at the initiator codon of the *pcbC* gene using vent DNA polymerase with, 5' TAGGAATTCCATATGCCAGTTCTGATGC 3' (OND-1) and 5' ATGTAGGTACCGGAGTTGACCAGGAAGTCGTTCT 3' (OCS-4) as oligonucleotide primers. The underlined three nucleotides, when incorporated into the PCR product will create an *Nde* I site at the start codon of the *pcbC* gene. The resulting 0.773 Kb PCR product was purified from 5% polyacrylamide gels, digested with *Sal*1 and cloned into *Sal*1 / *Sma*1 digested pUC119 to give pUC07VW.

In the second step, a 0.7 Kb *Sal* I fragment isolated from pIPS-1 was ligated to pUC07VW digested with*SalI* to generate the intact *pcbC* gene yielding pUC14VW.

In the third step, a 1.4 Kb DNA fragment containing the intact *pcbC* gene was isolated from pUC14VW by *Nde*I and *Hin*dIII digestion and cloned into similarly digested pT7-7 yielding pMD0P7VW. pMD0P7VW is identical to pMD0P7 (see Figure 9) except that pMD0P7VW is free of the undesirable mutations present in pMD0P7. prbs denotes the ribosome binding site of the *pcbC* gene; erbs indicates the ribosome binding site optimised for *E. coli*. The β-lactamase gene (*bla*) is shown as a crosshatched box while the *pcbC* gene is shown as a light shaded box. P, denotes the T7 gene 10 promoter. Bc, E, H, K, N, Sa, Sm and Ss designate *Bcl*1, *Eco*R1, *Hin*d111, *Kpn*1, *Nde*1, *Sal*1, *Sma*1 and *Sst*1 restriction enzyme sites.


Sa Sm
Sm Bc Sa K
K
Sa Sm Ss
H
E
pIPS-1:
prbs
pchC
OND-1
OCS-4
1.7 Kb
PCR using vent DNA polymerase
0.773 Kb
PCR product digested with Sal I is ligated to pUC119 digested with Sal I/Sma I
E N
Sa H
Ss
pUC07VW:
0.7 Kb
0.7 Kb SalI fragment isolated from pIPS-1 is ligated to pUC07VW digested with Sal I
E N
Sa
Sa H
Ss
pUC14VW:
0.7 Kb fragment from pIPS-1
1.4 kb NdeI/HindIII fragment from pUC14VW is ligated to similarly digested pT7-7 vector
N Sm Bc Sa K
K
Sa H
bla
P
erbs
pMD0P7VW:
pchC

2 mL of IB suspension resulting in a denaturant concentration of 5M urea, 50 mM DTT, 1 mM EDTA and 50 mM Tris-HCl (pH 7.5) and an initial protein concentration of 10 mg/mL. After incubation at room temperature for 2 hours, the denatured protein material was desalted by gel filtration on a Sephadex G-25 column equilibrated and eluted with IPNS assay buffer (50 mM Tris-HCl (pH 7.5), 0.01 mM EDTA, 1 mM DTT and 10% glycerol) at 4°C. The specific activity of the refolded material containing the authentic IPNS protein was 80 x 10^{-3} units per mg protein. Subsequent purification of this refolded material by ion exchange chromatography on a DEAE-Trisacryl column yielded pure IPNS protein as judged from SDS-PAGE gels stained with Coomassie blue. However, the specific activity of the purified material was only about 140 x 10^{-3} units per mg of protein, which was less than the activity values reported for native IPNS protein purified from *S. clavuligerus* (204. 1 x 10^{-3} units/mg protein). This specific activity value for purified IPNS from *S. clavuligerus* was considered to be a minimum value since IPNS is a labile enzyme and some loss of activity was likely during the 8 step purification process. The even lower specific activity of pure recombinant IPNS suggested that the refolding was not efficient. Studies on refolding of various recombinant, over-expressed proteins have indicated that the efficiency of folding is seriously affected by high concentration of proteins, and when proteins are refolded at high protein concentrations they tend to form aggregates. Quantitative recovery of activity is often observed only when the protein concentration is of the order of a few micrograms per millilitre. Therefore, studies were conducted to attempt to standardise the recovery conditions and methods used to obtain active IPNS material. These studies concentrated on the gel filtration and dialysis methods of renaturation, in view of their success in preliminary experiments, and their widespread applicability for the refolding of a variety of proteins.

Various amounts of IB suspension were treated with denaturing solution to yield protein solutions with different protein concentrations. After denaturation, each solution

was then desalted by gel filtration or by the dialysis method as discussed in Materials and Methods and the results are shown in Table 2 &3. Based on the activity results from both of the methods, it was concluded that a protein concentration of approximately 1 mg/mL gives maximal recovery of IPNS activity. However, the recovery of activity by the dialysis method, with a specific activity of 130 x 10^{-3} units of IPNS activity per mg of protein, was higher than the recovery obtained by the gel filtration method, with a specific activity of 117 x 10^{-3} units of IPNS activity per mg of protein.

3.2.1 Effect of pH on refolding of IPNS:

Among the factors affecting the folding of a protein, hydrophobic interactions play a dominant role (Dill, 1990). However, electrostatic interactions can also be significant in the folding and stabilisation or destabilisation processes, depending on the isoelectric point (pI) of the protein being folded and the pH of the buffer system. Theoretical calculation of the isoelectric point of *S. clavuligerus* IPNS based on the deduced amino acid composition of the protein gave a value of 5.16, similar to the experimentally determined isoelectric points of IPNS proteins from *P. chrysogenum* and *C. acremonium*. In order to examine the effect of pH and determine the optimum pH for denaturation and refolding of IPNS, IB suspensions were treated with denaturant solution at three different pH values to yield preparations with protein concentrations of 1 mg/mL. After denaturation, the preparations were desalted by dialysis, and assayed for IPNS activity and protein concentration (Table 4). It was apparent that pH played a significant role in folding efficiency since there was a more than 50% improvement in the recovery of active IPNS when pH was increased from 7.22 to 8.0. Denaturation and refolding of IPNS at pH values higher than 8.0 was not attempted in view of the presence of thiol groups in the IPNS protein which must be maintained in the reduced form for activity, and the reduced stability of thiol protective reagents such as DTT at elevated pH.

Table 2: Renaturation of solubilised IPNS by gel filtration chromatography

Initial protein[a] Concentration (mg / mL)	Final protein concentration (mg / mL)	IPNS specific activity (units x 10^3 per mg protein)
10.0	3.60	72
5.0	1.80	100
2.5	0.95	117
1.8	0.65	116
1.2	0.45	112
0.6	0.23	108

a estimated protein concentration based on a trial solubilisation and renaturation by gel filtration chromatography

Table 3: Renaturation of solubilised IPNS by dialysis

Initial protein[a] Concentration (mg / mL)	Final protein concentration (mg / mL)	IPNS specific activity (units x 10^3 per mg protein)
5.0	5.8	97
2.5	3.0	102
1.2	1.2	130
0.6	0.7	134

a estimated protein concentration based on a trial solubilisation and renaturation by gel filtration chromatography

Table 4: Effect of pH on renaturation of solubilised IPNS

pH	IPNS specific activity (units x 10^3 per mg protein)
7.22	117
7.46	145
8.00	184

3.3 Soluble Expression of IPNS:

In order to compare the secondary structure of active IPNS obtained by refolding of inactive material from IB suspensions with IPNS protein produced in a soluble form, it was necessary to obtain large amounts of soluble IPNS protein. Native IPNS protein isolated from *S. clavuligerus* was not available in the amounts required, and the difficulties associated with the purification of milligram amounts of IPNS from *S. clavuligerus* was one of the reasons for undertaking this study on high-level expression. As an alternative, purification of soluble recombinant IPNS material from *E. coli* was considered. However, when *E. coli* expression cultures are cultivated at 37°C, very low amounts of soluble IPNS protein are obtained (see Table 1), and purification to homogeneity would take as many purification steps as from *S. clavuligerus*. Therefore, alternative expression conditions were explored. Reducing the growth temperature during expression is a commonly used procedure for increasing the production of soluble recombinant protein, and so this technique was used, despite the induction of gene expression in our dual plasmid expression system being under thermal control. The high processive ability of the T7 RNA polymerase to transcribe the genes under its control suggested that once T7 RNA polymerase expression had been induced by heat shock, expression cultures could be transferred to low temperature for subsequent expression of the *pcbC* gene. Moreover, production of IPNS in both soluble and insoluble form using the same expression system and in identical genetic backgrounds would be the optimum system for comparison of results.

In order to check the temperature sensitive production of soluble IPNS protein, *E. coli* K38 cells carrying the pMD0P7 recombinant plasmid containing the *pcbC* gene were cultivated in 400 mL volumes of culture at 20, 25, 30 and 37°C for 2 hours after the initial 30 minute induction at 42°C. The conditions of growth prior to induction were identical to the conditions for *pcbC* expression in the insoluble form. *E. coli* K38 cells, after the 2 hour expression period, were harvested, washed and resuspended in 8 mL of

IPNS assay buffer. Cells were then disrupted by sonication and centrifuged to sedim particulate material. The pellets were resuspended in 2 mL of IPNS assay buffer and denatured and refolded by the gel filtration method as described in Section 2.13.1.1. The protein concentrations and IPNS activity of the various fractions were determined and tabulated in Table 5. Activity analysis of the crude extracts showed that the production of soluble functional IPNS is temperature sensitive. Increasing amounts of soluble activity in the crude extracts, seen as the temperature was lowered, were accompanied by a corresponding decrease in the amount of IPNS in the insoluble material. Samples of crude extract (30 μg protein) and refolded IB extract (15 μg protein) were analysed by SDS-PAGE and the results are displayed in Figure 12. SDS-PAGE analysis showed the same trends as were evident from activity measurements. In particular, there was a dramatic increase in the soluble activity as the temperature was lowered from 25 to 20°C, which is in agreement with the SDS-PAGE gel results. An estimated 80% of the expressed IPNS was present in the soluble form at 20°C.

In order to use a consistent terminology, from here on in this thesis, IPNS produced in cultures of *S. clavuligerus* will be referred to as native IPNS; recombinant IPNS expressed in a soluble form in *E. coli* will be referred to as soluble IPNS and recombinant IPNS expressed in an insoluble form in *E. coli*, denatured and refolded will be referred to as solubilised IPNS.

3.3.1 Electron Microscopic analysis:

In order to visualise the structure of the IB of IPNS protein formed upon expression at various temperatures, *E. coli* K38 cells containing the pMD0P7 construct were subjected to *pcbC* expression at 20, 25, 30 and 37°C. In order to obtain a suitable negative control, *E. coli* K38 cells containing the pMD0R7 construct, which does not express *pcbC*, were subjected to growth under expression conditions at 37°C. Specimens prepared from the five culture samples were analysed by transmission electron microscopy as discussed in Section 2.12.3 and the results are shown in Figure 13.

Table 5: Effect of temperature on the expression of IPNS

Temperature of expression (^{o}C)	Total soluble activity (units x 10^3)	Total solubilised activity (units x 10^3)	Total activity (units x 10^3)
37	941	8336	9277
30	1396	5952	7348
25	2193	5338	7531
20	4913	890	5803

Figure 12: SDS-PAGE analysis of cell extracts from *E. coli* carrying pMD0P7 after expression at different temperatures. Lanes 1 through 4 contained 30 μg of protein from the soluble fraction of cell extracts derived from 20, 25, 30, and 37°C cultures. Lanes 5 through 8 contained 15 μg of protein from the particulate fraction (after solubilisation) from 20, 25, 30, and 37°C cultures. The positions of molecular weight marker proteins bovine serum albumin (66,000), ovalbumin (45,000) and trypsinogen (24,000) are indicated with arrows.

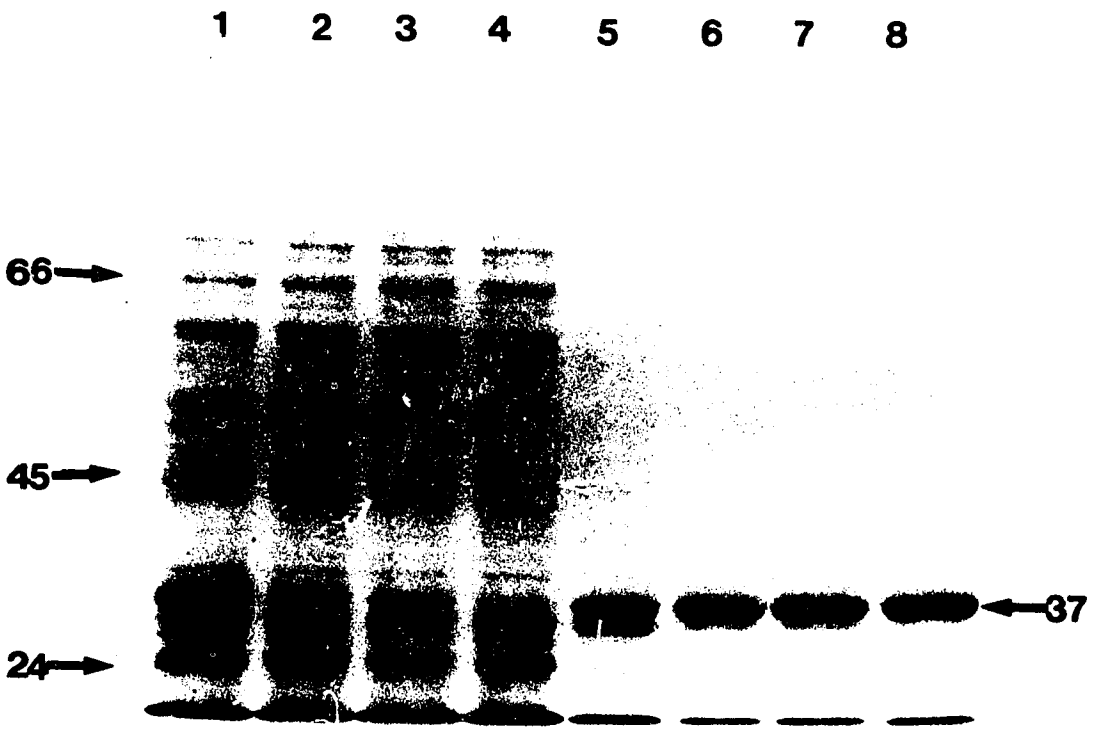

1
2
3
4
5
6
7
8
66
45
24
37

Figure 13: Electron Microscopic analysis of *E. coli* cells after ... ssion at different temperatures.

A: *E. coli* cells ... rying pMD0R7 construct from 37°C culture.

B, C, D, and E: *E. coli* cells carry ng pMD0P7 expression construct from 37, 30, 25, and 20°C ... ures respectively. Inclusion bodies have been ... icated with arrows.

Magnification: 65,000 x

A

B


C

D


E

Careful observation of more than 300 cells revealed the presence of darkly stained irregular shaped inclusion bodies (IB) at the poles of essentially all of the cells expressing *pcbC* at 37°C (Kane and Hartley, 1988). While IB formation was predominant in cultures expressing *pcbC* at 37°C, it was much reduced at 20°C. In addition, the size of the IB was also found to decrease as the temperature dropped from 37 to 20°C. There was a dramatic decrease in the size of the IB as the temperature dropped from 25 to 20°C, in accordance with the IPNS activity measurements and SDS-PAGE results. As expected, the negative control sample did not contain any identifiable IB.

3.4 Purification of IPNS:

3.4.1 Purification of Solubilised IPNS:

IPNS has been shown to display maximal activity in Tris-HCl buffer and this buffer has been used for the purification of IPNS from its native host (Jensen *et al.*, 1986a). One of the more effective purification steps involved chromatography on DEAE Trisacryl. Since solubilised IPNS was expected to show similar chromatography characteristics to the native IPNS from *S. clavuligerus*, ion exchange chromatography on DEAE-Trisacryl resin and Tris-HCl buffer was used for the purification. When solubilised IPNS was eluted using a 0.05 to 0.5 M linear Tris-HCl (pH 8.0) gradient, a relatively high recovery of IPNS protein was obtained (> 80%; data not shown). However, the specific activity of the IPNS obtained was relatively low. Analysis of different fractions across the IPNS peak by SDS-PAGE with Coomassie blue staining revealed the presence of contaminating proteins. Their distribution pattern indicated that most contaminating proteins were found in the ascending portion of the peak and their concentration decreased gradually to undetectable levels towards the centre of the peak. Fractions in the descending portion of the peak were nearly pure. These results suggested that the gradient was likely too steep for optimum purification. In order to achieve high recovery while removing the contaminating proteins eluting in the ascending portion of the IPNS peak, the gradient was made shallower by increasing the volume of the starting

and limiting buffers and narrowing the range of the gradient by increasing the concentration of the starting buffer and lowering the concentration of the limiting buffer. Optimal recovery with maximum purity was obtained using a gradient from 0.2 to 0. 3 M Tris-HCl (8. 0) with 250 mL each of starting and limiting buffer, and a column with dimensions of 1.6 x 30 cm. Under these elution conditions, 55% of the loaded IPNS material could be recovered with a specific activity of 241 x 10^{-3} units /mg protein (Table 6). Based on the recovery of pure, active IPNS protein from the above purification protocol, the yield of solubilised IPNS was estimated to be around 250 mg of IPNS protein per litre of culture when expressed at 37^{o}C, which is much higher than the earlier reported estimate (Durairaj *et al.*, 1992). Although significant improvements were achieved in the folding efficiency of IPNS (as discussed earlier), a quantitative comparison of the distribution of contaminating proteins compared to IPNS in the SDS-PAGE gel (Figure 14), suggested that the folding efficiency was still only about 70%, and the incompletely folded IPNS accounted for the losses occurring during the purification.

3.4.2 Purification of Soluble IPNS:

In order to achieve the maximal recovery of IPNS protein with the fewest number of purification steps, the purification protocol used by Jensen *et al.* (1986a) was modified and simplified for the purification of soluble IPNS. The steps essentially involved fractional precipitation with ammonium sulfate, ion-exchange chromatography using DEAE-Trisacryl, size-exclusion chromatography using Superose 12 and ion exchange chromatography using Mono Q. The purification results obtained are summarized in Table 7. Specific activity measurements of the crude extract obtained from expression at 20^{o}C indicated a value of 23 x10^{-3} units per mg protein which is 15 fold higher than the activity seen in *S. clavuligerus* crude cell extracts. As a result of this higher starting specific activity, only a 10 fold purification was required to attain a near homogeneous preparation. It was observed that the salt precipitation step failed to provide any improvement in purification over the crude extract, and so this step could

Table 6: Purification of solubilised IPNS

Purification step	IPNS activity (units x 10^3 per mg protein)	Protein (mg)	IPNS specific activity (units x 10^3 per mg protein)	Recover (%)
Particulate	---	---	---	---
Crude renatured extract (Dialysis)	2943	16.1	182.7	100
DEAE-Trisacryl eluate	2139	8.9	240.6	55

Figure 14: SDS-PAGE analysis of samples from solubilised IPNS purification. Lane 1 contained 10 μg of protein from the particulate fraction (after solubilisation). Lane 2 contained 15 μg of protein from the concentrated fractions of DEAE-Trisacryl ion-exchange chromatography. The positions of molecular weight marker proteins, bovine serum albumin (66,000), ovalbumin (45,000) and trypsinogen (24,000) are indicated with arrows.

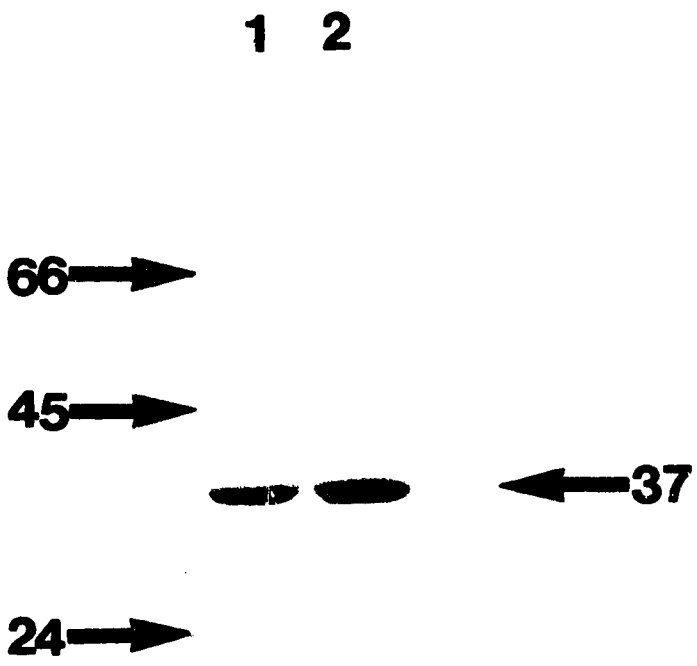

1 2
66
45
37
24

Table 7: Purification of soluble IPNS

Purification step	Total IPNS activity (units x 10^3)	Total protein (mg)	Specific activity (units x 10^3 per mg protein)
Crude extract	4856	211.2	22.9
$(NH_4)_2SO_4$ precipitate	2377	102.0	23.3
DEAE-Trisacryl eluate	2070	12.9	160.3
Superose 12 eluate	1664	7.5	221.5
Mono Q eluate	1201	4.9	243.9

possibly be eliminated when large scale preparation of soluble IPNS is attempted. DEAE-Trisacryl anion exchange chromatography removed the majority of contaminating proteins, and the IPNS protein eluted at around 0.24 M Tris-HCl (Figure 15). The concentrated material from the ion exchange chromatography was then passed through a Superose 12 column to remove several contaminating proteins larger than the IPNS protein, but a minor contaminating protein of molecular weight 46,000 daltons remained. Addition of 100 mM KCl to the elution buffer used for Superose 12 chromatography made it possible to resolve the IPNS from the 46,000 dalton contaminating protein. However, the IPNS material obtained from the Superose 12 step still contained a few low molecular weight contaminants which required chromatography on a Mono Q column to obtain homogeneous IPNS material. The SDS-PAGE analysis of the samples from the various purification stages has been shown in Figure 16.

3.5 Functional analysis of Isopenicillin N synthase:

Isopenicillin N synthase from *S. clavuligerus* contains four cysteine residues at positions 37, 104, 142 and 251 from the N-terminal methionine residue. The cysteine residues at positions 104 and 251 are conserved in all of the fungal and bacterial *pcbC* proteins which have been studied to date. These cysteine groups have been implicated in the function of IPNS both by biochemical studies, and by site-directed mutagenesis studies involving the replacement of cysteine residues with serine residues in the IPNS from *C. acremonium* (Samson *et al.*, 1987b; Kriauciunas *et al.*, 1991).

In addition to their importance in the function of proteins, free thiol groups from cysteine residues have also been observed to promote oligomer formation via intra and inter disulphide linkages in proteins with multiple cysteine residues. During the process of crystallisation, proteins reach very high concentrations and oligomer formation can block the crystallisation process making structural analysis by x-ray crystallography impossible (M. James, personal communication). In order to remove this possible impediment to crystallisation of IPNS and to clarify the role of the cysteine residues in

Figure 15: DEAE-Trisacryl ion exchange chromatography-purification of soluble IPNS. IPNS extract from the $(NH_4)_2 SO_4$ fractional precipitation after dialysis was loaded into the DEAE-Trisacryl ion exchange column and eluted using a linear Tris-HCl gradient with 250 mL each of starting buffer [200 mM Tris-HCl (pH 8.0), 0.01 mM EDTA, 1 mM DTT and 10% glycerol] and limiting buffer [300 mM Tris-HCl (8.0), 0.01 mM EDTA, 1 mM DTT and 10% glycerol]. IPNS elutes around 0.24 M Tris-HCl and has been labelled by an arrow.

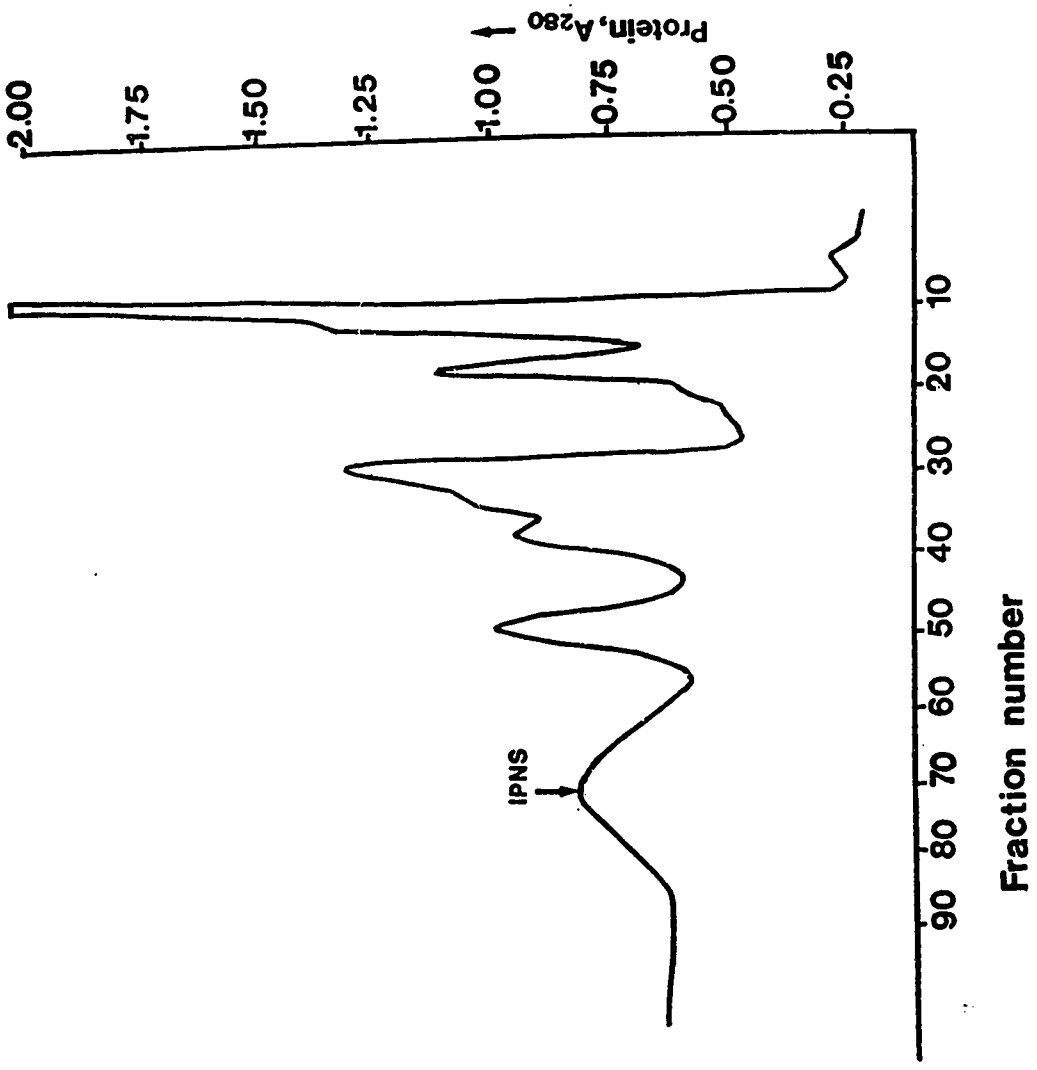
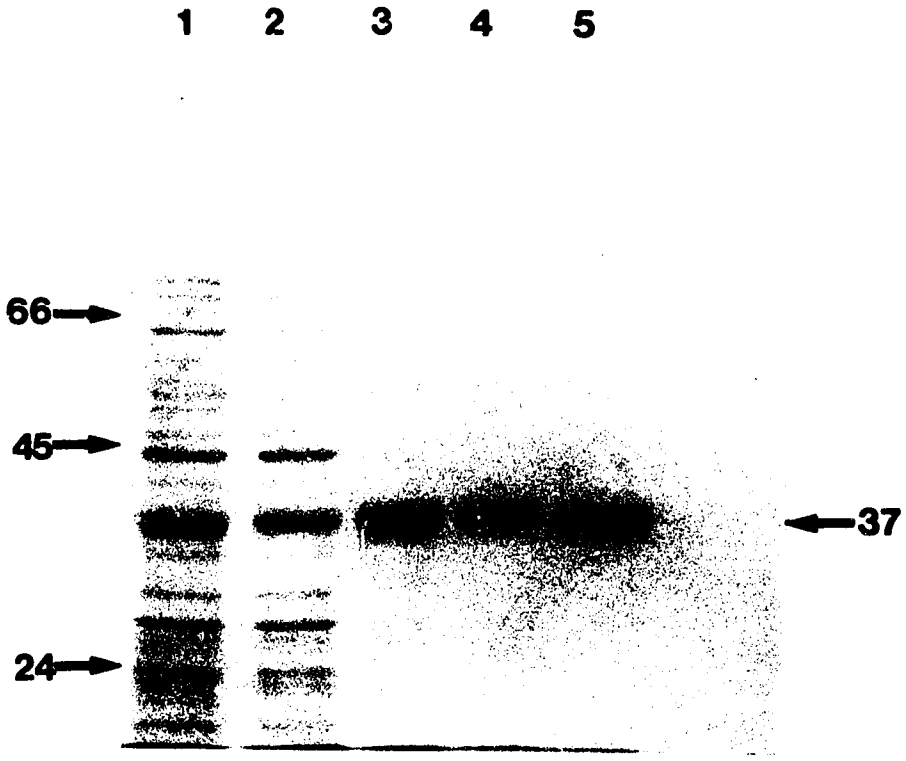

Protein, A_{280}
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
IPNS
10
20
30
40
50
60
70
80
90
Fraction number

Figure 16: SDS-PAGE analysis of soluble IPNS purification. Lane 1 contained 50 μg of protein from the crude extract. Lane 2 contained 30 μg of protein from the $(NH_4)_2 SO_4$ fractional precipitation step. Lane 3 contained 20 μg of protein from the concentrated fraction of DEAE-Trisacryl ion exchange chromatography. Lane 4 contained 10 μg of protein from the concentrated fraction of Superose 12 gel filtration chromatography. Lane 5 contained 15 mg of protein from the Mono Q ion-exchange chromatography. The positions of molecular weight marker protiens bovine serum albumin (66,000), ovalbumin (45,000) and trypsinogen (24,000) are indicated.


1
2
3
4
5
66
45
24
37

the function of IPNS, cysteine residues were exchanged for serine residues by site-directed mutagenesis. The cysteine residues at positions 37, 142 and 251 are encoded by **UGC** codons, and the cysteine residue at position 104 is encoded by a **UGU** codon in the *pcbC* mRNA from *S. clavuligerus*. Serine residues are encoded by six possible codons, **UCU, UCC**, UCA, UCG, AGC and AGU, but the last four codons are not normally found in genes encoding proteins which are highly expressed in *E. coli* and were not considered for further mutagenesis. **UCU** and **UCC** codons involves a single base change, and they are also commonly found in highly expressed proteins. Therefore, the cysteine residues at 37, 142 and 251 represented by **UGC** codons were changed into **UCC** codons and the cysteine residue at 104 designated by a **UGU** codon was changed into a **UCU** codon. The base involved in the change has been underlined and the codons being interchanged have been highlighted in bold letters. These mutations were introduced into cloned fragments of the *pcbC* gene, and the intact mutant *pcbC* gene expression vector derivatives were then reconstructed by splicing the various mutant fragments. From here on, the mutant forms of the IPNS protein will be referred to using the standard 1-letter code for amino acids. Hence, the C37S IPNS mutant would refer to that mutant in which the cysteine residue at position 37 was changed to a serine residue. Figure 17 illustrates the location of four cysteine residues in the wild type *pcbC* gene sequence and the corresponding serine mutant expression vector derivatives.

3.5.1 Introduction of C104S and C251S mutations by PCR:

The region of nucleotide sequence within the *pcbC* gene which encodes the cysteine 104 residue is flanked near its 5' end by a *Sma*I site, and the cysteine 251 residue is flanked near its 3' end by a *Kpn*I restriction enzyme site. The fortuitous proximity of these two restriction sites to the regions to be mutated facilitated the introduction of the C104S and C251S mutations by PCR using OCS-2 and OCS-4

Codon Position

	33	**37**	41
pMD0P7VW (wild type):	ATA CAC GGG GCC T**G**C CGC GGC TCG GGC		
pMD0P7V1(C37S):	ATA CAC GGG GCC T**C**C CGC GGC TCG GGC		

	100	**104**	108
pMD0P7VW (wild type):	GTC GAG TCC TTC T**G**T TAC CTC AAC CCG		
pMD0P7V2 (C104S):	GTC GAG TCC TTC T**C**T TAC CTC AAC CCG		

	138	**142**	146
pMD0P7VW (wild type):	TTC CGG CCG TTC T**G**C GAG GGC TAC TAC		
pMD0P7V3 (C142S):	TTC CGG CCG TTC T**C**C GAG GGC TAC TAC		

	247	**251**	255
pMD0P7VW (wild type):	TTC CTG GTC AAC T**G**C GGT ACC TAC ATG		
pMD0P7V4 (C251S):	TTC CTG GTC AAC T**C**C GGT ACC TAC ATG		

Figure 17: Location of the four cysteine residues encoded by the *pcbC* gene of pMD0P7VW (wild type), and the corresponding mutant serine residues encoded by the *pcbC* gene sequences of pMD0P7V1 (C37S), pMD0P7V2 (C104S), pMD0P7V3 (C142S) and pMD0P7V4 (C251S). Relevant cysteine / serine codons have been highlighted in bold letters. Nucleotides involved in the changes have been underlined.

OCS-2: 5' TCCCGGGCAGGAAGGCCGTCGAGTCCTTCTCTTACCTC 3'

OCS-4: 5' ATGTAGGTACCGGAGTTGACCAGGAAGTCGTTCT 3'

as mutagenic primers. These primers contain the altered bases needed to cause the cysteine to serine mutations as well as these two restriction enzyme sites. The altered bases have been denoted by underlining, and the *Sma*I and *Kpn*I recognition sequences have been highlighted in bold letters. The restriction enzyme sites were included for further genetic manipulations of the resulting PCR products and regeneration of the intact *pcbC* gene containing the altered bases. Introduction of the C104S and C251S mutations was achieved in a single step by PCR using pIPS-1 as the plasmid template and OCS-2 (upstream primer) and OCS-4 (downstream primer). The resulting 484 bp DNA fragment was purified from 5% polyacrylamide gels and cloned into the pCR^{TM}II vector yielding pCRII24 (Figure 18).

3.5.1.1 Construction of pMD0P7V4, an expression vector encoding the C251S IPNS mutant:

Construction of expression vector containing the *pcbC* gene carrying the C251S mutation necessitated the construction of pCRIIK24, which was accomplished by replacing the *Kpn*I fragment of pCRII24 (vector sequence) with the 510 bp *Kpn*I fragment isolated from the pIPS-1 plasmid. The pCRIIK24 vector was digested with *Sal*I and *Hin*dIII and the resulting 0.55 Kb fragment carrying the C251S mutation was ligated to pMD0P7VW**, to generate pMD0P7V4 (Figure 18). pMD0P7VW** is a derivative of pMD0P7VW (see Figure 11) which is devoid of the corresponding wild type *Sal* I/*Hin*dIII fragment.

3.5.1.2 Construction of pMD0P7V2, an expression vector encoding the C104S IPNS mutant:

The pCRIIK24 vector was digested with *Sma*I and *Sal*I and the resulting 0.41 Kb *pcbC* gene DNA fragment containing the C104S mutation was ligated to pUC07VW* to generate pUC07V2. pUC07VW* is a derivative of pUC07VW which is devoid of the

Figure 18: Construction of pMD0P7V4, an expression vector encoding the C251S IPNS mutant.

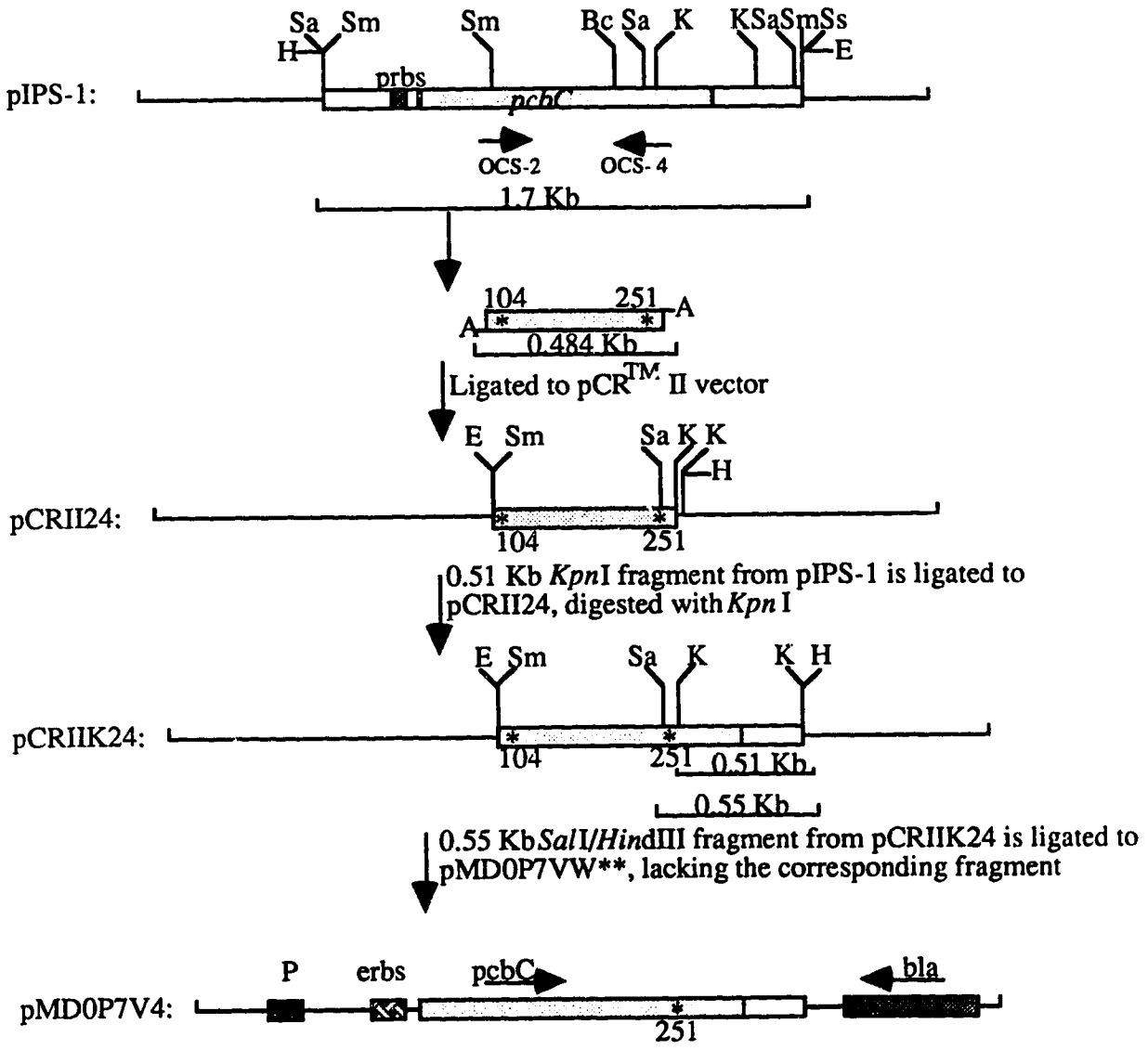
A pIPS-1 phagemid template was subjected to PCR amplification to introduce C104S and C251S mutations into the *pcbC* gene, using *Taq* DNA polymerase and the oligonucleotide primers

5' TCCCGGGCAGGAAGGCCGTCGAGTCCTTCT**C**TTACCTC 3' (OCS-2) and

5' ATGTAGGTACC**GG**AGTTGACCAGGAAGTCGTTCT 3' (OCS-4). The nucleotide triplet encoding the C104S and C251S mutations are shown in bold print and the bases introducing the alterations are shown underlined. The resulting 0.484 Kb PCR product carrying both C104S and C251S mutations denoted by * was cloned into the pCR^{TM}II vector to yield pCRII24.

In order to facilitate the construction of expression vector containing the intact *pcbC* gene carrying the C251S mutation, pCRII24 was digested with *Kpn*I and a 0.51 Kb *Kpn* 1 fragment isolated from pIPS-1 was inserted to generate pCRIIK24.

In the third step, a 0.55 Kb *Sal*I / *Hin*dIII fragment containing the C251S mutation only was isolated from pCRIIK24 and ligated to pMD0P7VW**, a derivative of pMD0P7VW (see Figure 11) lacking the corresponding *Sal*I / *Hin*dIII fragment, to yield pMD0P7V4. The *pcbC* gene has been shown as a light shaded box, and the β-lactamase gene (*bla*) has been shown as a crosshatched box. P denotes the T7 gene 10 promoter; erbs stands for the ribosome binding site optimised for *E. coli*; Bc, E, H, K, Sa and Sm stand for sites for *Bcl*I, *Eco*RI, *Hin*dIII, *Kpn*I, *Sal*I and *Sma*I restriction enzymes.


Sa Sm
H
Sm
Bc Sa K
KSaSmSs
E
pIPS-1:
prbs
pcbC
OCS-2
OCS-4
1.7 Kb
104
251
A
A
0.484 Kb
Ligated to pCR™ II vector
E Sm
Sa K K
H
pCRII24:
104
251
0.51 Kb KpnI fragment from pIPS-1 is ligated to
pCRII24, digested with Kpn I
E Sm
Sa K
K H
pCRIIK24:
104
251
0.51 Kb
0.55 Kb
0.55 KbSalI/HindIII fragment from pCRIIK24 is ligated to
pMD0P7VW**, lacking the corresponding fragment
P
erbs
pcbC
bla
pMD0P7V4:
251

corresponding wild type *SmaI/SalI* fragment. pUC07V2 was then digested with *NdeI/BclI* and the resulting 0.65 Kb *pcbC* gene DNA fragment was ligated to pMD0P7VW* to generate pMD0P7V2 vector (Figure 19). pMD0P7VW* is a derivative of pMD0P7VW which is devoid of the corresponding wild type *NdeI/Bcl* I fragment.

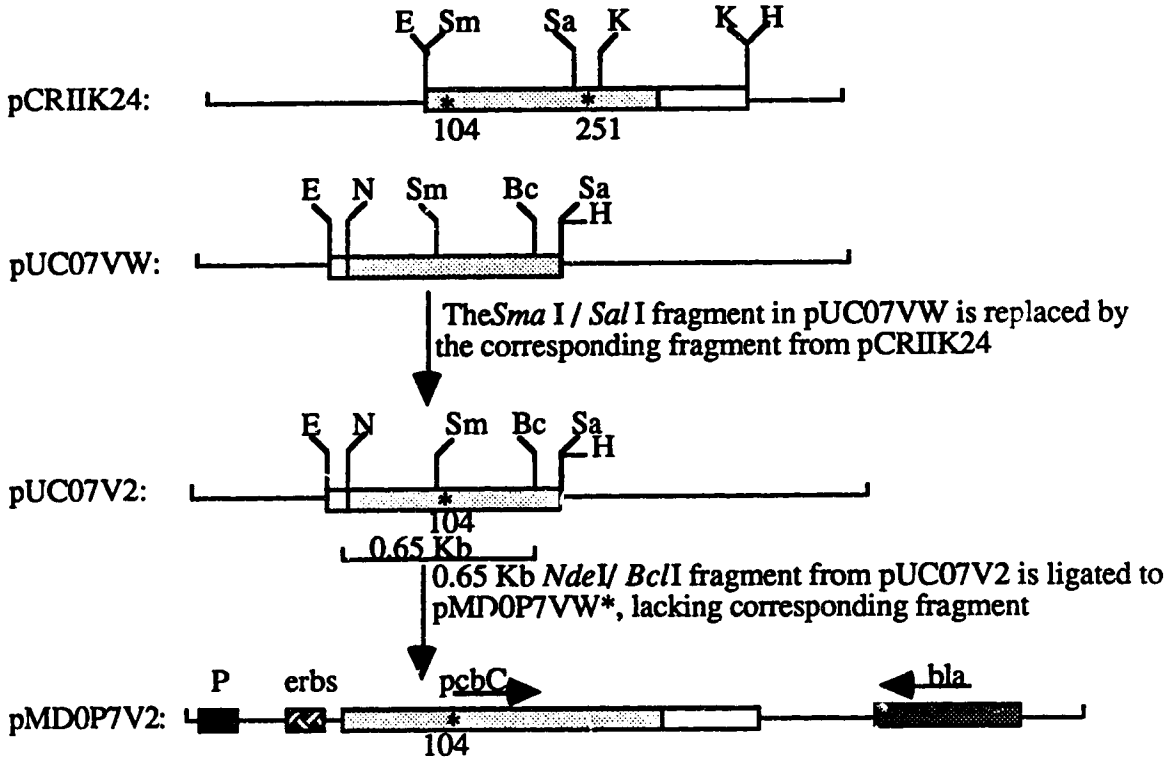
3.5.2 Introduction of C37S and C142S mutations:

The *pcbC* gene contains a strong transcription terminator at the end of its open reading frame and the stem loop structure formed by the dyad symmetry terminator is likely to reduce the elongation efficiency in site directed mutagenesis reactions which must extend across this region. In order to avoid this possible problem, pUC07VW (see Figure 11) was used as the template for single-strand mutagenesis to introduce C37S and C142S mutations. This plasmid contains a 0.7 Kb *pcbC* gene fragment, starting from the 5' ATG start codon and ending at an internal *SalI* site, which encompasses both of the regions to be mutagenised. The high G+C content of the *pcbC* gene required the use of a DNA polymerase with high processivity and therefore a T7 DNA polymerase derivative (Sequenase version 2) was chosen for the mutagenesis reactions. OCS-1 (5' ACGGGGCCTCCCGCGGCTCG 3') and OCS-3 (5' GGCCGTTCTCCGAGGGCTAC 3') with the altered nucleotides underlined were used in separate mutagenesis reactions together with the pUC07VW template to generate the C37S and C142S mutants respectively. For these reactions the pUC07VW template was isolated from an *E. coli dut⁻ ung⁻* strain. Screening of transformants resulting from the C142S mutagenesis reactions by single-stranded DNA sequencing indicated a mutagenesis efficiency of 60-80%. It was also noticed that 20-30% of the positive isolates were heterogenous mixtures containing both the wild type as well as mutant strands. However, screening of transformants from the C37S mutagenesis reactions by single-stranded DNA sequencing indicated that the mutagenesis efficiency was low. In order to screen a large number of colonies, colony hybridisation was adopted. The T_m of the OCS-1 oligonucleotide

Figure 19: Construction of pMD0P7V2, an expression vector encoding the C104S IPNS mutant.

A 0.41 kb *Sma*I / *Sal*I DNA fragment of the *pcbC* gene from pUC07VW was replaced with the corresponding 0.41 Kb *Sma*I / *Sal*I fragment from pCRIIK24 carrying the C104S mutation to generate pUC07V2.

In the second step, pUC07V2 was digested with *Nde*I and *Bcl*I and the resulting 0.65 Kb DNA fragment was cloned into pMD0P7VW*, a derivative of pMD0P7VW which is devoid of the corresponding *Nde*I / *Bcl*I fragment, to generate pMD0P7V2. P stands for T7 gene 10 promoter; erbs stands for ribosome binding site optimised for *E. coli* ; The *pcbC* gene has been shown by light shaded box. The β-lactamase gene (*bla*) is denoted by crosshatched box; Bc, E, H, K, N, Sa and Sm indicate sites for *Bcl*I, *Eco*RI, *Hin*dIII, *Kpn*I, *Nde*I, *Sal*I and *Sma*I restriction enzymes.


pCRIIK24:
E Sm Sa K K H
104
251
pUC07VW:
E N Sm Bc Sa H
The Sma I / Sal I fragment in pUC07VW is replaced by the corresponding fragment from pCRIIK24
pUC07V2:
E N Sm Bc Sa H
104
0.65 Kb
0.65 Kb NdeI/ BclI fragment from pUC07V2 is ligated to pMD0P7VW*, lacking corresponding fragment
pMD0P7V2:
P
erbs
pcbC
bla
104

calculated based on the formula [(A+T) x 2+ (G+C) x 4] was 74^{o}C, and due to the presence of salt in the hybridisation solution, it was estimated to be around 75^{o}C. In general, it is believed that a mismatch located in the middle of an oligonucleotide is likely to reduce the T_m by 12^{o}C. Hence, to ensure complete hybridisation of the oligonucleotide, 55^{o}C was chosen for hybridisation. However, the OCS-1 oligonucleotide also contains 17 (G+C) bases in a 20 mer and the mutagenic base involves a change from G to C. In view of the strong hydrogen bonding between G and C, it was thought that it might be difficult to find a hybridisation temperature which would allow the differentiation of the single base pair change between the wild type and mutant sequence by differential thermal hybridisation. Therefore, reducing the salt concentration gradually during washing while maintaining constant temperature was chosen as the differentiating parameter. During the first wash in 2x SSPE, 0.1% SDS, the T_m was expected to be lowered to 68.5^{o}C while in the second wash in 1x SSPE, 0.1% SDS, the T_m was estimated to be around 48^{o}C. It was expected that during the second wash, a heteroduplex sequence would be destabilised significantly while the homoduplex of the mutant sequence would be stable. When colonies were screened using this hybridisation and washing procedure, it was found that a small fraction of the colonies gave distinct hybridisation signals while the rest of the colonies were negative. The positive isolates identified by hybridisation were further verified by single-strand sequencing and were found to be authentic mutants. The efficiency of mutagenesis was around 10%. The low efficiency could not be explained since the C37S mutagenesis reactions contained the same template as was used for the C142S mutagenesis and the reactions were carried out under identical conditions. It is possible that the OCS-1 primer or the local nucleotide sequence in the priming region could have contributed to the difference. The colony hybridisation and the sequencing reactions for screening and verification of mutations were carried out as described in Sections 2.11.2.1 and 2.11.2.2 respectively. The positive isolates containing the C37S and C142S mutations derived

from the single-strand site directed mutagenesis were labelled pUC07V1 and pUC07V3 respectively.

3.5.2.1 Construction of pMD0P7V1 and pMD0P7V3, expression vectors encoding the C37S and C142S IPNS mutants:

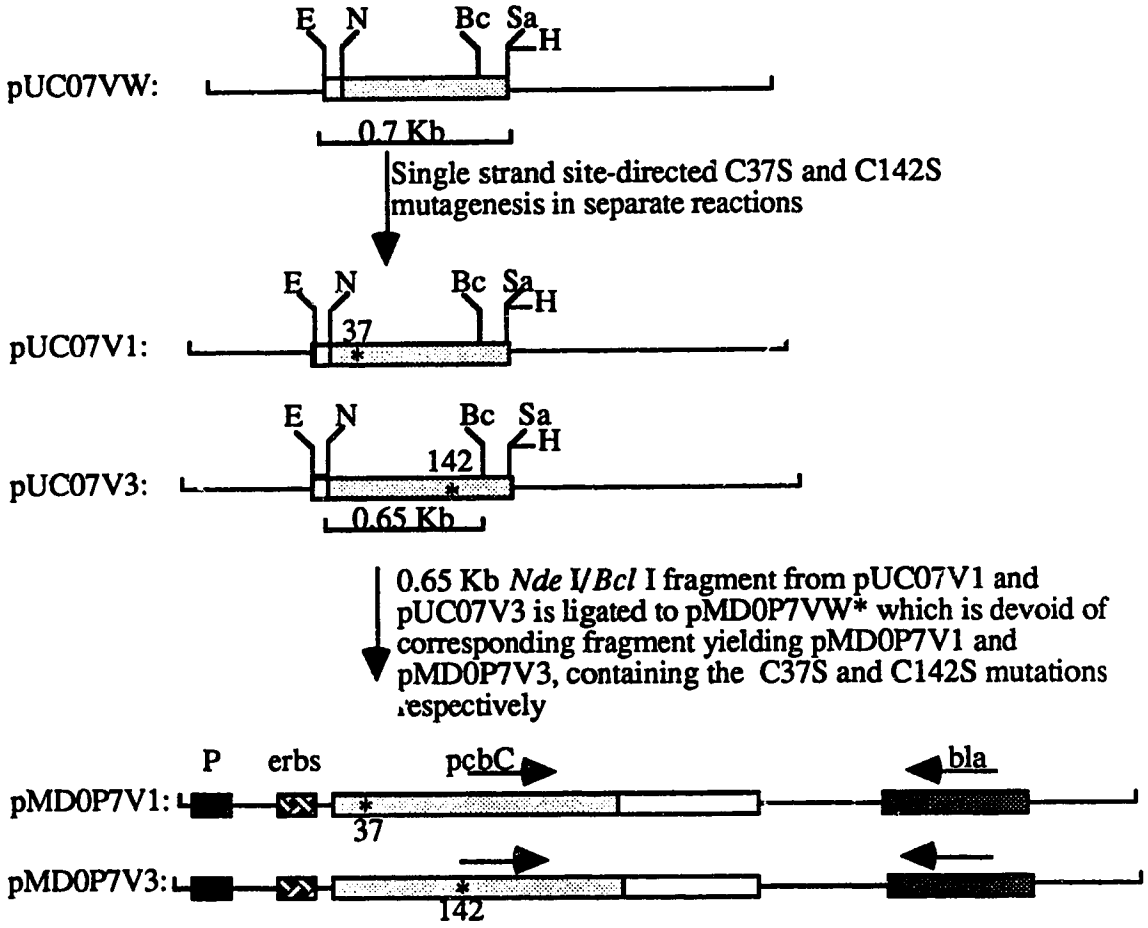
pUC07V1 and pUC07V3 are phagemids derived from pUC07VW by site-directed mutagenesis to introduce C37S and C142S mutations respectively in the *pcbC* gene. The two phagemids were digested with *NdeI/ BclI* and the resulting 0.65 Kb *pcbC* gene DNA fragments were purified and ligated to pMD0P7VW* as discussed earlier to generate pMD0P7V1 and pMD0P7V3 (Figure 20).

3.6 Expression of wild type and mutant *pcbC* genes in *E. coli*:

E. coli strain K38 carrying the pGP1-2 plasmid which contains the T7 RNA polymerase gene necessary for high level gene expression was transformed with pMD0P7VW, pMD0P7V1, pMD0P7V2, pMD0P7V3 and pMD0P7V4 vectors separately in the presence of ampicillin and kanamycin antibiotics as discussed in Materials and Methods. Transformants containing each of the expression plasmids together with pGP1-2 were verified by plasmid isolation followed by restriction digestion and electrophoresis in 0.7% agarose gels. Overnight cultures of the positive transformants were used as seed cultures for further inoculation of 100 mL amounts of expression medium to an initial optical density at 600 nm of 0.05. The cultures were cultivated at 30°C for approximately 5 hours until the optical density at 600 nm reached 1.5. The cultures were then shifted to 42°C for the induction of T7 RNA polymerase for 30 minutes. Following induction, they were cultivated for an additional 2 hours at 37°C. Under these conditions, IPNS proteins are expressed predominantly in the insoluble, inactive form. IB suspensions were prepared from the wild type and the four mutant IPNS cultures and treated with denaturant solution (6.25 M urea, 62.5 mM DTT, 50 mM Tris-HCl (pH 7.5), 1.25 mM EDTA) giving a final concentration of 5 M urea, 50 mM DTT, 50 mM Tris-HCl (pH 7.5) and 1 mM EDTA and approximately 1 mg / mL of protein, for 2 hours at

Figure 20: Construction of expression vectors pMD0P7V1 and pMD0P7V3, encoding the C37S and C142S IPNS mutants.

pUC07V1 and pUC07V3 are phagemids derived from pUC07VW after single-strand site-directed mutagenesis to introduce the C37S and C142S mutations respectively into the *pcbC* gene. The two phagemids were digested with *Nde*I and *Bcl*I and the resulting 0.65 Kb *pcbC* gene DNA fragments were ligated to pMD0P7VW*, a derivative of pMD0P7VW (see Figure 11) devoid of the corresponding *Nde*I / *Bcl*I fragment. The resulting pMD0P7V1 and pMD0P7V3, carry the C37S and C142S mutant forms of the *pcbC* genes, respectively. * indicates the location of the C37S and C142S mutations in the *pcbC* gene. P stands for the T7 gene 10 promoter. erbs stands for the ribosome binding site optimised for *E. coli*. The *pcbC* gene has been indicated by the light shaded box. *bla* denotes the β-lactamase gene and shown by the dark crosshatched box. Bc, E, H, N and Sa stand for *Bcl*I, *Eco*RI, *Hin*dIII, *Nde*I and *Sal*I restriction enzyme sites.


E N Bc Sa H
pUC07VW:
0.7 Kb
Single strand site-directed C37S and C142S mutagenesis in separate reactions
E N Bc Sa H
37
pUC07V1:
E N Bc Sa H
142
pUC07V3:
0.65 Kb
0.65 Kb Nde I/Bcl I fragment from pUC07V1 and pUC07V3 is ligated to pMD0P7VW* which is devoid of corresponding fragment yielding pMD0P7V1 and pMD0P7V3, containing the C37S and C142S mutations respectively
P
erbs
pcbC
bla
pMD0P7V1:
37
pMD0P7V3:
142

room temperature. IPNS preparations were then refolded by dialysis against a 250 fold excess of IPNS assay buffer (50 mM Tris-HCl (pH 7.5), 1 mM DTT, 0.01 mM EDTA and 10% glycerol). The solubilised IPNS preparations from the wild type as well as the four mutants were further purified by DEAE-Trisacryl ion exchange chromatography as described for the purification of solubilised IPNS, and concentrated 20 fold by ultrafiltration using an Amicon PM10 membrane. IPNS activity were then determined and listed in Table 8.

Activity comparisons of the cysteine to serine mutants of IPNS showed that the C104S mutant lost more than 96% of its activity, when compared to the wild type IPNS, and the C251S mutant lost about 50% of its activity. The same pattern of results was also seen when the solubilised IPNS extracts were assayed before purification (data not shown). The 20 - 30 % activity loss seen in the C37S and C142S mutants implies that these cysteine residues are less critical for activity than the cysteine 104 residue, and more similar to the cysteine 251 residue.

3.6.1 Effect of N-ethylmaleimide on IPNS activity:

In addition to probing the role of the cysteine residues of IPNS by site directed mutagenesis, the various preparations of IPNS were also treated with the thiol specific inhibitor, N-ethylmaleimide, to determine the importance of thiol groups on IPNS activity. Solubilised IPNS preparations from the wild type and the four single mutants were treated with 1 mM N-ethylmaleimide at room temperature for 30 minutes. It was found previously that the wild type IPNS loses more than 90% of its activity when incubated with 0.1 mM N-ethylmaleimide. In order to ensure that alkylation of cysteine residues was complete, a 10-fold higher concentration of inhibitor was used. After removing the excess unreacted inhibitor by gel filtration chromatography, as discussed in Materials and Methods, residual activity was determined (Table 9). Cysteine residues were found to be important for IPNS activity as evidenced by the reduction of activity seen when IPNS preparations were treated with N-ethylmaleimide. The wild type IPNS

Table 8: IPNS activity of cysteine to serine IPNS mutants

Plasmid constructs	Nature of IPNS mutation	IPNS activity remaining (%)
pMD0P7VW	wild type	100.0
pMD0P7V1	C37S	81.7
pMD0P7V2	C104S	3.7
pMD0P7V3	C142S	76.0
pMD0P7V4	C251S	52.3

Table 9: Effect of N-ethylmaleimide on mutant IPNS activity

Plasmid construct	Nature of IPNS mutation[a]	Percentage of residual activity remaining
pMD0P7VW	wild type (100.0)	0.0
pMD0P7V1	C37S (81.7)	0.0
pMD0P7V2	C104S (3.7)	25.0
pMD0P7V3	C142S (76.0)	0.0
pMD0P7V4	C251S (52.3)	0.0

a numbers in parentheses indicate the percentage of IPN activity remaining prior to NEM treatment

and the C37S, C142S and C251S mutant IPNS lost all detectable activity when treated with N-ethylmaleimide at 1 mM. In contrast, the C104S mutant, which had already lost 95% of its activity as a result of the mutation, still retained some of that activity upon treatment with N-ethylmaleimide.

3.6.2 Construction of pMDOP7V134, an expression vector encoding the C37S, C142S, C251S IPNS triple mutant:

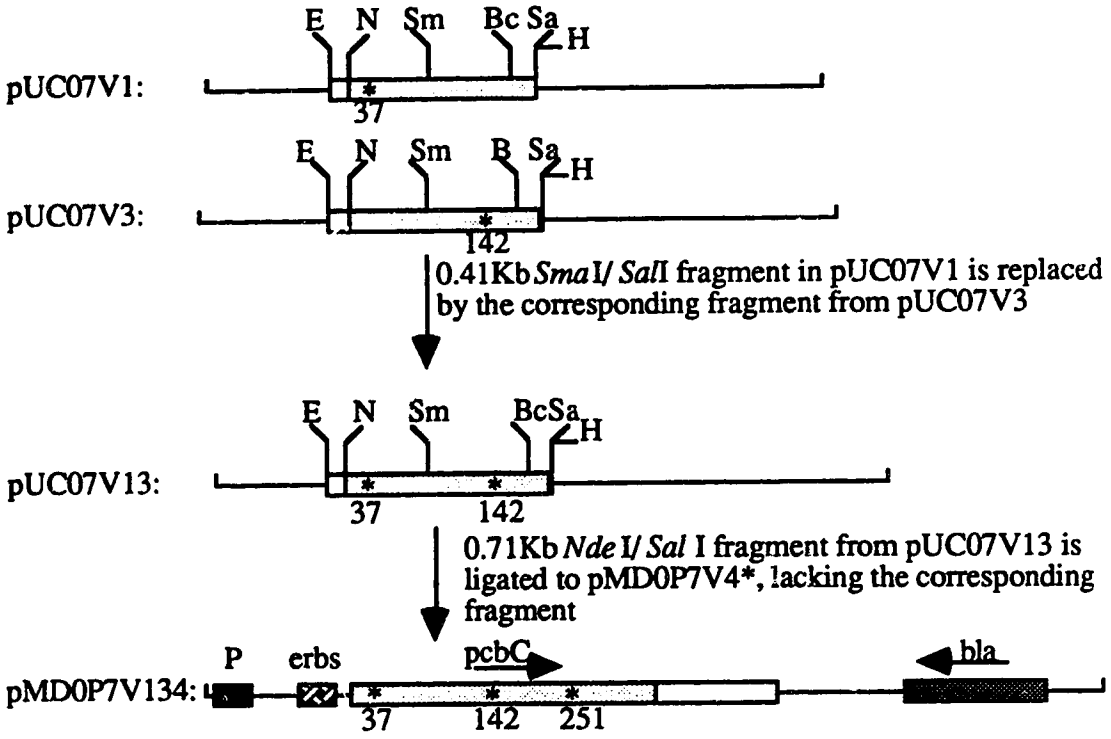
Based on the activity data of the single mutants, the cysteine residues at positions 37, 142 and 251 were concluded to be less important to IPNS activity while the residue at 104 was finally very important. To circumvent the anticipated problem of oligomer formation through disulfide bond formation promoted by thiol groups, a *pcbC* gene triple mutant in which codons for three cysteine residues least important for activity were converted to serine codons, was constructed as follows. pUC07V3 is a recombinant plasmid which contains a 0.71 Kb fragment of the *pcbC* gene extending from the translation initiation site to a *Sal*I site located 0.71 Kb within the gene, and is identical to pUC07VW except that it carries the C142S mutation. Digestion of pUC07V3 with *Sma* I and *Sal* I liberated a 0.41 Kb *pcbC* gene fragment containing the C142S mutation. This fragment was ligated to pUC07V1*, yielding pUC07V13 which carries both the C37S and C142S mutations. pUC07V1* is a derivative of pUC07V1 (see Figure 20) which lacks the corresponding *SmaI/SalI* fragment. pUC07V13 was then digested with *Nde*I and *Sal*I to release a 0.71 Kb *pcbC* DNA fragment containing both the C37S and C142S mutations. This fragment was ligated to pMD0P7V4*, a derivative of pMD0P7V4 which contains the C251S mutation and which is devoid of the corresponding wild type *NdeI/SalI* fragment, leading to the generation of pMD0P7V134. pMD0P7V134 is the final expression vector construct which contains all three of the desired cysteine to serine mutations (Figure 21).

After transforming *E. coli* K38 cells with the pMD0P7V134 construct, expression of the triple IPNS mutant was carried out using conditions favoring IB formation, as

Figure 21: Construction of pMD0P7V134, an expression vector encoding the C37S, C142S, C251S IPNS triple mutant.

A 0.41 Kb *Sma*I / *Sal*I *pcbC* gene fragment from phagemid pUC07V1 was removed and replaced by the corresponding *pcbC* gene fragment from pUC07V3 containing the C142S mutation to yield pUC07V13 carrying both C37S and C142S mutations.

In the second step, a 0.71 Kb *pcbC* gene fragment was released from pUC07V13 by *Nde*I / *Sal*I digestion and cloned into pMD0P7V4*, a derivative of pMD0P7V4 (see Figure 18) which is devoid of the corresponding fragment, to generate pMD0P7V134, a triple *pcbC* gene mutant carrying C37S, C142S and C251S mutations. P stands for T7 gene 10 promoter; erbs indicates the ribosome binding site optimised for *E. coli*; *bla* denotes the β–lactamase gene and is indicated by a dark crosshatched box; the *pcbC* gene has been shown as a light shaded box; Bc, E, H, N, Sa and Sm designate sites for *Bcl*I, *Eco*RI, *Hin*dIII, *Nde*I, *Sal*Iand *Sma*I restriction enzymes.


E N Sm Bc Sa H
pUC07V1:
37
E N Sm B Sa H
pUC07V3:
142
0.41Kb SmaI/ SalI fragment in pUC07V1 is replaced by the corresponding fragment from pUC07V3
E N Sm BcSa H
pUC07V13:
37
142
0.71Kb NdeI/ SalI fragment from pUC07V13 is ligated to pMD0P7V4*, lacking the corresponding fragment
P
erbs
pcbC
bla
pMD0P7V134:
37
142
251

discussed previously for the single mutants. The IB suspension was denatured and then refolded by dialysis, purified, and IPNS activity and protein concentrations were determined. Despite the fact that the three cysteine to serine mutations were chosen to minimize the effect of the mutations on IPNS activity, the purified IPNS from the triple mutant had lost 95% of its activity. In order to find out whether the loss of activity was due to observable structural changes in the IPNS protein, samples of purified solubilised IPNS from the triple mutant and from the wild type were analysed by circular dichroism.

3.6.3 Circular Dichroism Studies:

Circular dichroism studies have been extensively used to study the conformation of proteins in solution. Soluble and solubilised preparations of wild type IPNS, and solubilised preparation of triple mutant IPNS were purified and structural comparisons were made by circular dichroism studies as described in Materials and Methods (Figure 22). It was found that the wild type IPNS prepared in both the solubilised and soluble forms displayed very similar CD profiles, with a characteristic double minimum indicating the presence of α-helix structure. Secondary structure predictions based on the deduced amino acid sequence of the IPNS protein using the SEQSEE software (Wishart *et al.*, 1994), indicated a structure dominated by α-helix (Figure 23), in agreement with the experimental CD spectra. In contrast to the wild type IPNS, the minimum near 222 nm is nearly absent in the triple mutant indicating that the solubilised triple mutant lacks the ordered structure present in the wild type, which probably can explain the loss in activity. Differences between the wild type and triple mutant samples are also evident from examination of the near ultraviolet region of the spectra (Figure 24), although the nature of the structural changes associated with difference in this region of the spectrum are not well defined. The observed discrepancy between the experimentally observed data, which is more reliable (Figure 22), and the theoretically predicted data (Figure 23), in the contents of various secondary structural elements indicates that the algorithm for prediction of secondary structures needs further improvement.

Figure 22: Structural comparison of solubilised wild type IPNS, soluble wild type IPNS and the solubilised triple mutant IPNS by CD Spectroscopy: far ultraviolet spectra.

Solubilised, wild type IPNS: 0.295 mg/mL, mrw 111.535, 0.0502 cm cell, 0.05 M Tris-HCl, 1 mM DTT, 0.01 mM EDTA.

Soluble, wild type IPNS: 0.383 mg/mL, mrw 111.535, 0.0489 cm cell, 0.05 M Tris-HCl, 1 mM DTT, 0.01 mM EDTA.

Solubilised, triple mutant IPNS: 0.338 mg/mL, mrw 111.535, 0.0502 cm cell, 0.05 M Tris-HCl, 1 mM DTT, 0.01 mM EDTA.

Molar ellipicities were claculated from the equation:

$[O] = O_{obs} / (10xlxc)$

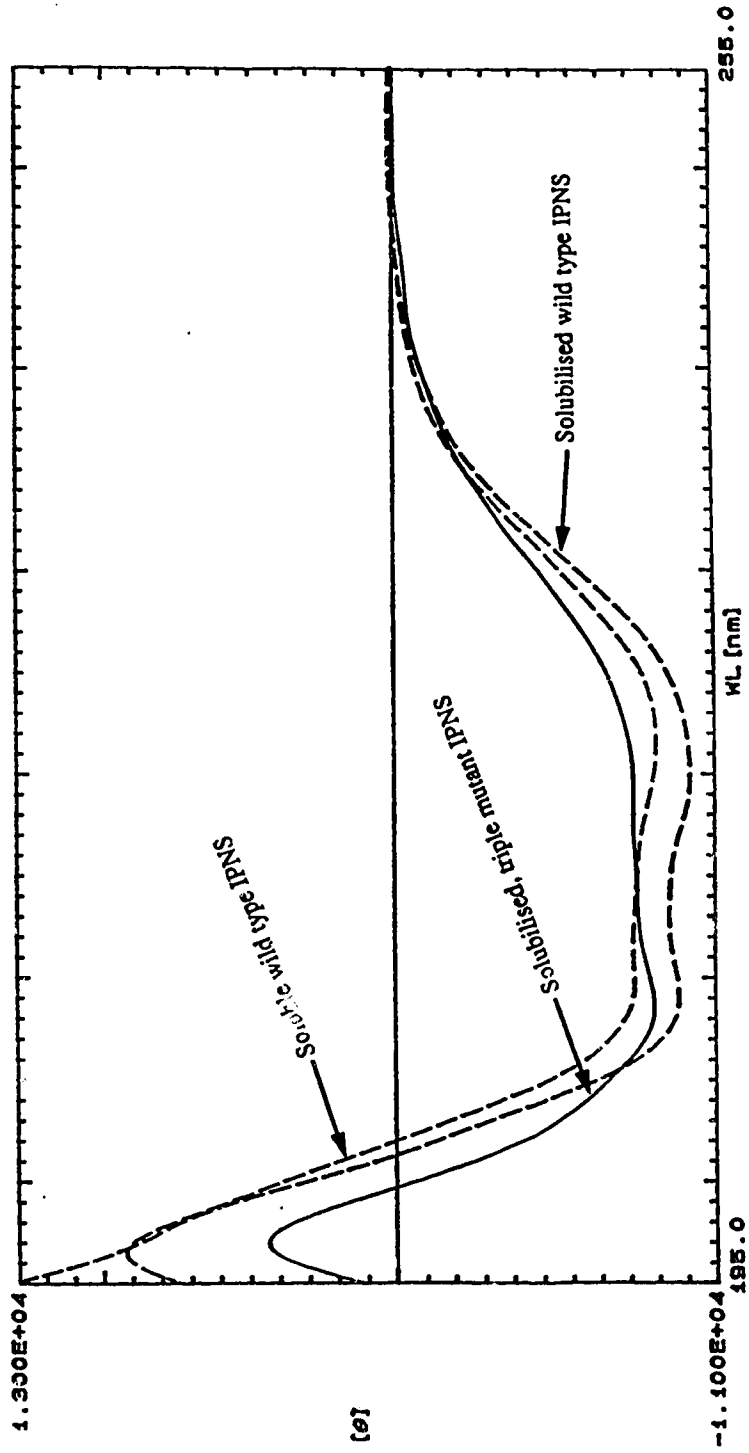
O_{obs} = in degree (in millidegrees)

l = pathlength in cm

c = concentration in mg/mL/mrw (mean residue weight)

The unit for molar ellipicities was degree-centimeter squared per decimole

Sample:	α-helix	β-sheet	β-turn	remainder	Scale factor
Soluble wild type IPNS	0.27	0.25	0.16	0.32	1.00
Soluble wild type IPNS(tfe)	0.47	0.13	0.02	0.38	1.00
Solubilised Wild type IPNS	0.28	0.37	0.25	0.10	1.00
Solubilised Wild type IPNS(tfe)	0.57	0.08	0.11	0.24	1.00
Solubilised Triple mutant IPNS	0.24	0.40	0.10	0.26	1.00
Solubilised Triple mutant IPNS(tfe)	0.62	0.31	0.07	-0.00	1.00


1.330E+04
-1.100E+04
[θ]
195.0
255.0
WL [nm]
Solubilised wild type IPNS
Solubilised, triple mutant IPNS
Soluble wild type IPNS

Program: SEQSEE

Description: Structure prediction/analysis

Sequence name: Isopenicillin N synthase from *Streptomyces clavuligerus*

Amino acids: 329

Membrane spanning regions: No membrane spanning region found

Folding class: IPNS belongs to MIXED folding class

Expected % beta sheet content: 16.4

Expected % coil content: 41.0

Expected % alpha helix content: 42.6

```
        10         20         30         40         50
MPVLMPSAHV PTIDISPLFG TDAAAKKRVA EEIHGACRGS GFFYATNHGV
CCBBBCCCCC CBBBCCCCCH HHHHHHHHHH HHHHHCCCCC CBBBBCCCCC

        60         70         80         90        100
DVQQLQDVVN EFHGAMTDQE KHDLAIHAYN PDNPHVRNGY YKAVPGRKAV
HHHHHHHHHH HHHHHHHHHH HHHHHHHHCC CCCCCCCCCH HHHCCCCCCB

       110        120        130        140        150
ESFCYLNPDF GEDHPMIAAG TPMHEVNLWP DEERHPRFRP FCEGYYRQML
BBBBBCCCCC CCCCCHHHCC CCCCBBBBCC CCCCCCCCCC CHHHHHHHHH

       160        170        180        190        200
KLSTVLMRGL ALALGRPEHF FDAALAEQDS LSSVSLIRYP YLEEYPPVKT
HHHHHHHHHH HHHHHCCHHH HHHHHHHHHC CCBBBBBBBC CCCCCCCCCC

       210        220        230        240        250
GPDGQLLSFE DHLDVSMITV LFQTQVQNLQ VETVDGWRDI PTSENDFLVN
CCCCCBBBHH HHHHHBBBBB BBHHHHHHHH BBBCCCCCCC CCCCCCBBBB

       260        270        280        290        300
CGTYMAHVTN DYFPAPNHRV KFVNAERLSL PFFLNGGHEA VIEPFVPEGA
CCCHHHHHCC CCCCCCCCHH HHHHHHHHHB BBBBCCCCCB BBBBCCCCCC

       310        320        329
SEEVRNEALS YGDYLQHGLR ALIVKNGQT
HHHHHHHHHH HCCHHHHHHH HHHHCCCCC
```

Figure 23: Secondary structure prediction of IPNS from *S. clavuligerus*. IPNS amino acid sequence and its corresponding predicted secondary structure are shown. C:coil; B:beta sheet; H:α-helix

Figure 24: Structural comparison of solubilised wild type IPNS, soluble wild type IPNS and solubilised triple mutant IPNS by CD Spectroscopy: near ultraviolet spectra.

Solubilised wild type IPNS: 0.295 mg/mL, mrw 111.535, 1 cm cell, 0.05 M Tris-HCl, 1 mM DTT, 0.01 mM EDTA.

Soluble wild type IPNS: 0.383 mg/mL, mrw 111.535, 1 cm cell, 0.05 M Tris-HCl, 1 mM DTT, 0.01 mM EDTA.

Solubilised triple mutant IPNS: 0.338 mg/mL, mrw 111.535, 1 cm cell, 0.05 M Tris-HCl, 1 mM DTT, 0.01 mM EDTA.

Molar ellipicities were claculated from the equation:

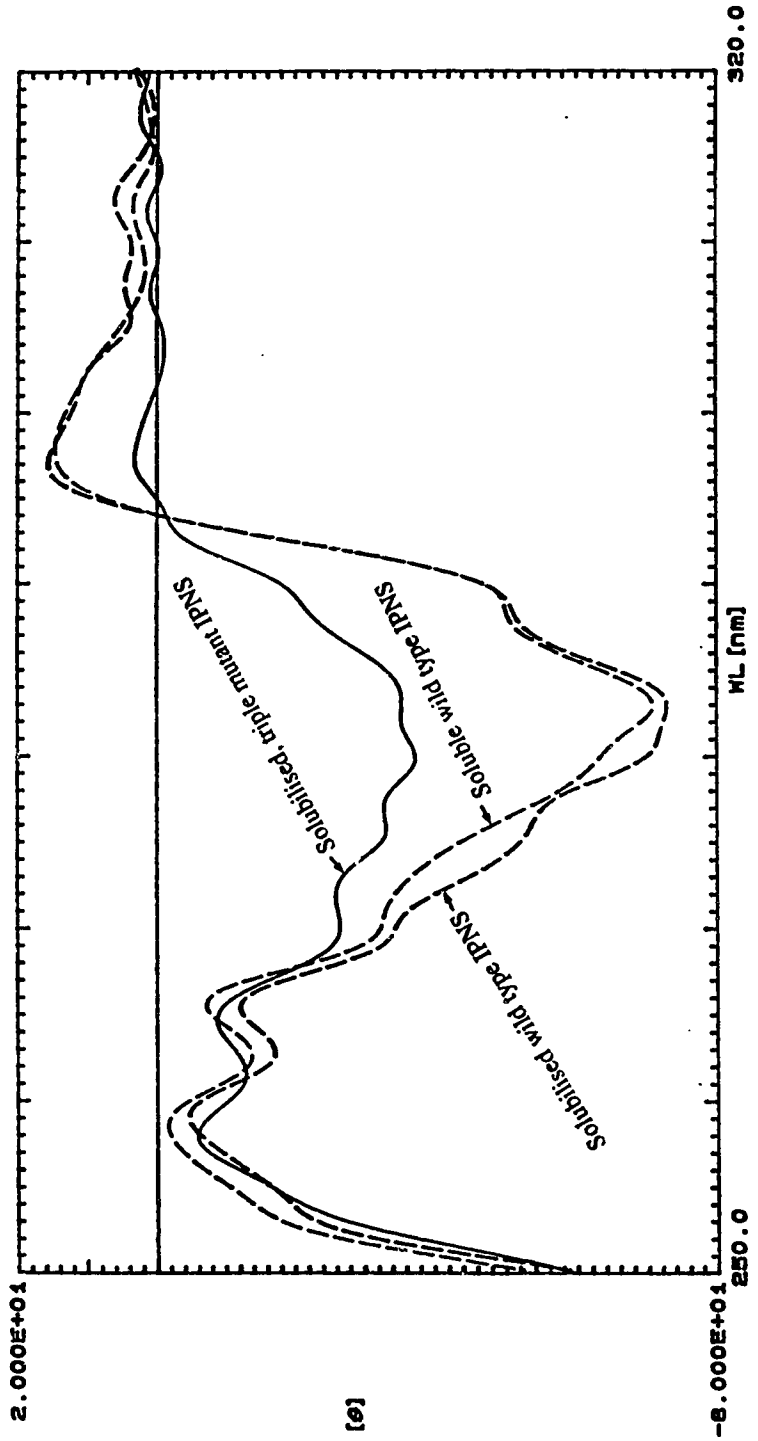
$[O] = O_{obs} / (10xlxc)$

O_{obs} = in degree (in millidegrees)

l = pathlength in cm

c = concentration in mg/mL/mrw (mean residue weight)

The unit for molar ellipicities was degree-centimeter squared per decimole


2.000E+01
-8.000E+01
250.0
320.0
WL [nm]
[θ]
Solubilised wild type IPNS
Soluble wild type IPNS
Solubilised, triple mutant IPNS

Although CD studies on the triple mutant were only conducted on solublised IPNS preparations, soluble preparations were compared to solubilised preparations with respect to IPNS activity. When the triple mutant IPNS was produced in the soluble form by growth at 20^{o}C after induction, it retained about 50% of wild type activity. This contrasts with the 5% retention of activity observed when the same protein was produced in a solubilised form. This indicates that the triple mutant does not refold as well after denaturation as the wild type IPNS does, and suggests that the reduced activity seen in the triple mutant may be due to structural changes affecting the folding of the protein rather than effects on the active site.

4. DISCUSSION

IPNS has been purified to homogeneity, and the corresponding gene has been cloned and sequenced from a number of species, but a detailed understanding of the mechanism of the reaction catalysed remains elusive. This understanding will likely not be achieved until the structure of the protein has been determined by x-ray crystallography, and several groups are actively pursuing this goal using IPNS from different species (Fujishima *et al.*, 1994). Structural characterisation of any protein by x-ray crystallography requires relatively large amounts of pure protein. Obtaining these amounts of IPNS from *S. clavuligerus* would be impractical and tedious because of the complex purification scheme required to yield pure material. Therefore, the *pcbC* gene, encoding IPNS, has been expressed in *E. coli* to simplify the purification procedure.

SDS-PAGE analysis of cell extracts from *E. coli* cultures carrying the expression plasmid pJD004, which contains the complete *pcbC* gene as well as 300 - 400 bp stretches of upstream and downstream sequence, failed to detect any production of IPNS. Secondary structure analysis of the nucleotide sequence upstream of the *pcbC* gene revealed that the transcript is likely to form an imperfect stem-loop structure sequestering the RBS. Occlusion of the RBS through the formation of secondary structure could interfere with the binding of the translation initiation machinery of the organism and therefore limit translation initiation. Stanssens *et al.* (1986) showed that there is a strong relationship between the rate of ribosomal loading and continued transcriptional activity through the coding region of *lacZ* gene in *E. coli*. These authors indicated that in the absence of efficient translation initiation, premature transcription termination is favored, which prevents expression of genes in *E. coli*. Similar observations have been reported by Munson *et al.* (1984) and Schottel *et al.* (1984).

In order to remove this possible interference, the construct pMD004 was created which contains the *pcbC* gene and its associated RBS, but lacks the upstream stem-loop

structure described above. Analysis of expression extracts from pMD004 by SDS-PAGE still did not indicate any production of IPNS protein, implying that some problem other than occlusion of the RBS was preventing expression of the *pcbC* gene. The presence of a prominent protein band of molecular weight 29,000 in the expression extract, likely to be due to β-lactamase, indicated that there was good transcription of the β-lactamase gene. Since the *pcbC* gene was positioned between the T7 promoter and the β-lactamase gene, this implied that there must also have been good transcription of *pcbC*. However, β-lactamase was selectively translated, and therefore the factor limiting expression of *pcbC* was likely to be acting at the post-transcriptional level.

Inefficient translation initiation, biased codon usage in the *pcbC* gene resulting in poor translation elongation, and instability of the transcript or the protein could all have been reasons for the lack of expression (Kurland, 1991). The *pcbC* gene from *S. clavuligerus* does show a different pattern of codon usage from that of *E. coli*, because of the very high G+C content of the *S. clavuligerus* genome. However, heterologous genes including *pcbC* genes from other *Streptomyces* and fungal β-lactam producers with high proportions of rare codons, have been expressed at high-level in *E. coli* (Landman *et al.*, 1991; Makoff *et al.*, 1989). Therefore, rare codon usage was unlikely to have been the main factor limiting the expression of *pcbC*. Analysis of codon usage of the *pcbC* gene did not indicate any concentration of rare codons in the region of the translation initiation site, which has been reported to have particularly deleterious effects on translation by reducing ribosomal loading and interfering with efficient translation initiation (Leskiw *et al.*, 1988; Liljenstrom and Von Heijine, 1987).

Prokaryotic mRNAs are inherently less stable than eukaryotic mRNAs and have short half lives. However, if instability of the transcript was a problem, this would be expected to reduce the level of expression rather than prevent expression of *pcbC* completely. When recombinant proteins are expressed at high-level, *E. coli* has also been shown to induce a variety of heat shock proteins including *lon* and other proteases which

can degrade the expressed proteins (Goff and Goldberg, 1985; Allen *et al.*, 1992). Careful examination of SDS-PAGE gels did not give any indication of degraded products. High-level expression of *pcbC* genes from fungal and other *Streptomyces* species in *E.coli* also argued against the instability of the IPNS protein.

Using these lines of reasoning, the factor limiting expression of *pcbC* was narrowed down to the translation initiation region of the gene. Comparison of the RBS sequence (AGGAGG) of the *pcbC* gene in the pMD004 construct with the consensus RBS of *E. coli* showed a perfect match. However, the spacing between the RBS and the initiator codon was only 5 bp, much less than the optimal 9 bp preferred in *E. coli*. Reduction of 1 bp in the case of human fibroblast and leukocyte interferon genes from the optimal 9 bp spacing led to a reduction of 50% in the expression level in *E. coli* (Shepard *et al.*, 1982). It appeared that a spacing of less than 7 bp reduced the level of expression drastically. The translation initiaton codon of the *pcbC* gene was also found to be preceeded by a C residue which has been shown to be detrimental for efficient translation initiation (Makoff and Smallwood, 1988; Hwang *et al.*, 1990).

It is also possible, although less likely, that the translation initiation signals of the β-lactamase gene support translation initiation so efficiently as to deprive the *pcbC* mRNA of the protein synthesizing machinery necessary for translation. With a view to verifying whether the preferential translation of *bla* transcripts over *pcbC* transcripts could be limiting the level of expression, pMD006 was constructed. It was analogous to pMD004 except that the orientation of the β-lactamase gene was reversed in relation to the *pcbC* gene. Hence, the transcription of *bla* was controlled by its own weak promoter and was not likely to compete with the large amounts of *pcbC* transcript produced from the T7 promoter for the expression of *pcbC* gene. Examination of the SDS-PAGE profile of the expression extracts of pMD006 revealed that it was no different from that of pMD005, a negative control for pMD006. As expected, the prominant band due to β–lactamase observed in pMD004 was absent in pMD006, but no IPNS band was seen. The

failure to detect the expression of *pcbC* was therefore not likely due to the preferential translation of the *bla* transcript over the *pcbC* transcript, and was likely due to poor translation initiation.

In order to verify whether the translation initiation signals of the *pcbC* gene were capable of supporting efficient translation initiation, pMD0N7 was constructed. This construct was designed to provide two different RBS and yield an IPNS fusion protein, authentic IPNS, or a mixture of the two, depending on which of the two translation initiation signals was used for expression. Examination of the SDS-PAGE profile of pMD0N7 indicated the presence of a prominant fusion protein band with a molecular weight of 39,700 dalton while authentic IPNS protein with molecular weight of 37,000 was not observed. The two RBS are separated by nearly 70 bp, and ribosomal complexes typically occupy only 30-40 bp of sequence (De Boer and Hui, 1990). It also has been demonstrated that the spacing between translating ribosomes is insensitive to the efficiency of translation initiation (Guillerez *et al.*, 1991). Therefore, it was unlikely that the spacing between the two RBS would have negatively affected translation initiation from the downstream *pcbC* gene. These results confirmed that the translation initiation signals of the *pcbC* gene were not capable of supporting efficient translation initiation. However, once the translation was initiated using the *E. coli* RBS, protein synthesis proceeded through the upstream sequence region of *pcbC* gene.

For crystallographic studies, authentic IPNS protein was preferred rather than an IPNS fusion protein and hence pMD0P7 was constructed. When expression extracts were analysed by SDS-PAGE, a prominant protein band due to IPNS could be detected, but the IPNS protein was associated predominantly with the insoluble fraction of the cell extracts. When cell extracts from the various expression vector derivatives were assayed for activity, only low levels of activity could be detected in the soluble fraction of extracts from the pMD0N7 and pMD0P7 constructs. Although the IPNS protein bands observed in the insoluble fractions of pMD0N7 and pMD0P7 extracts were prominent, no activity

could be detected, suggesting that the IPNS was produced in the form of inactive insoluble inclusion bodies (IB). This phenomenon has also been observed for numerous other heterologous proteins when expressed at high-level in *E. coli* (Gribskov and Burgess, 1983; Kane and Hartley, 1988).

Creation of the pMD0P7 expression vector, which yielded high-level expression of IPNS, required the introduction of an *Nde*I site at the initiator codon of the *pcbC* gene, and this was accomplished using PCR. *Taq* DNA polymerase used in the PCR procedure, is known to be capable of introducing unintended non-template encoded mutations. Automated DNA sequencing of the whole 990 bp open reading frame of the *pcbC* gene present in the pMD0P7 construct indicated the presence of four unintended mutations in addition to the desired introduction of the *Nde*I site. Analysis of the mutations indicated that they occured randomly and that there was no specificity or bias towards any particular base. A high incidence of mutations was also detected in PCR products prepared for other purposes using *Taq* DNA polymerase. As a result it is essential that PCR products be verifed for sequence errors prior to further genetic manipulation of the products, particularly when the PCR products are used for the production of proteins.

pMD0P7VW was constructed to correct these mutations, and except for the unintended mutations, it was identical to pMD0P7. IPNS produced in pMD0P7VW expression cultures was also present in the form of inactive IB. Expression of *pcbC* genes from other fungal and *Streptomyces* spp. has also resulted in IB formation depending on the nature of the promoters used. The *pcbC* gene from *C. acremonium,* when expressed using the *trp* promoter, resulted in the formation of soluble IPNS, but when expressed using the bacteriophage lambda promoter λP_L, resulted exclusively in IB formation. The bacteriophage lambda P_L promoter is a stronger promoter than the *trp* promoter. In general, the stronger the promoter used to drive expression, the greater the likelihood that the protein will be produced in an insoluble, inactive form. The rate of

transcription initiation compared to the rate of translation in a transcription/translation coupled system may be critical for proper folding of newly synthesised polypeptides. When the rate of transcription initiation dictated by promoter strength and RNA polymerase binding exceeds the rate of translation, the newly formed gene transcripts and the nascent protein get precipitated along with the protein synthesising machinery. Expression of the *pcbC* gene from *A. nidulans* using the λP_L promoter resulted in IB formation (10-15% of total protein) while expression using the *lac* promoter resulted in soluble IPNS being formed (40% total soluble protein) (Baldwin *et al.*, 1991). The T7 promoter is a much stronger promoter than either of the *lac* or *trp* promoters, and the T7 RNA polymerase has been shown to be very processive for transcription (Tabor and Richardson, 1985). *pcbC* genes from *S. jumonjinensis* and *Flavobacterium* spp. resulted in IB formation when expressed using the T7 promoter (Landman *et al.*, 1991). This evidence provides credence to the link betwen the rate of transcription and the formation of IB.

IB are protein aggregates containing incorrectly folded proteins along with other cellular components. Plasmid encoded β-lactamase was the major component of the insoluble fraction of the pMD004 extract confirming that the plasmid encoded proteins are the predominant components of IB. Obtaining active IPNS protein from IB required disaggregation and denaturation by treatment with strong denaturants and refolding by gradual removal of the denaturant. Urea was chosen to denature the IPNS protein aggregates, and when IB suspensions were denatured with 6M urea and then renatured by passage through a Sephadex G-25 column, active IPNS protein with a specific activity of approximately 80 $x10^{-3}$ units/mg protein was recovered. Purification of this refolded protein material by DEAE-Trisacryl chromatography gave preparations with high specific activity, but the recovery of total IPNS was very poor. When the concentration of the limiting buffer was increased, more IPNS protein was recovered, but the specific activity was only about 140 x 10^{-3} units/mg protein. This is considerably lower than the reported

value of 204.1 x 10^{-3} units/mg protein for IPNS purified from the natural host, *S. clavuligerus*. These discrepancies indicated that the refolding of IPNS was not efficient, and that improperly folded material eluted from the column under different conditions from those which elute pure, active IPNS.

When solubilization and renaturation of proteins from IB is carried out at high protein concentrations, the recovery of active protein can be reduced drastically. At high concentrations of proteins, aggregate formation may be favored over proper folding. Therefore, it was necessary to optimise the conditions used in the solubilisation and renaturation of IB containing IPNS.

When IB suspensions were treated with denaturant at different protein concentrations, and then renatured either by passing through a G-25 Sephadex column or by dialysis, an initial concentration of 1 mg / mL of protein yielded maximum IPNS activity for both methods of renaturation. However, the dialysis procedure consistently yielded IPNS with higher specific activities (130 -134 x 10^{-3} units/mg of protein) than the specific activities (approximately 117 x 10^{-3} units/mg protein) obtained from the gel filtration procedure. The recovery of active IPNS was also higher at any given protein concentration when renaturation was carried out by the dialysis procedure than when the gel filtration procedure was used. These results suggested that the protein concentration was not as important a factor as the method of refolding, in affecting the yield of active IPNS protein. At high protein concentrations (5.8 mg / mL), recovery of active IPNS was much better using the dialysis procedure, than that obtained with the gel filtration procedure at the same protein concentration. Comparable recovery of IPNS activity could only be obtained by using a much lower protein concentration of 1.8 mg/mLwhen the gel filtration method was used. Studies aimed at identifying the mechanism of protein folding have indicated that proteins tend to fold in characterised intermediate stages. These intermediates resemble the native proteins in their secondary structures but lack the tertiary structure, and are called " molten globule intermediates " (Haase-Pettingel and

King, 1988; Kuwajima *et al.*, 1989). Under favorable conditions, the molten globule intermediates follow a constructive folding pathway leading to the fully functional structures. However, under unfavorable, non-physiological conditions (e.g., high concentration of proteins, absence of appropriate level of chaperonins, etc.,) these intermediates are trapped, leading to IB formation. It is likely that in the gel filtration procedure, the high localised concentrations of denatured proteins together with concomitant rapid removal of urea, facilitated aggregate formation rather than proper folding. In the dialysis procedure, removal of urea was more gradual which possibly enhanced the recovery of active proteins. Similar procedures involving denaturation with urea followed by dialysis have been used to recover active IPNS from IB formed when *C. acremonium* IPNS was expressed in *E. coli* using the bacteriophage lambda promoter P_L (Kriauciunas *et al.*, 1991).

Recent advances in the study of protein folding pathways have led to the identification of hydrophobic interactions as the major driving force governing the folding pathway of proteins. However, electrostatic interactions also play a significant role in the folding and stabilisation of proteins, especially at extremes of pH for proteins with isoelectric points in the acidic or alkaline regions. The calculated isoelectric point of IPNS from *S. clavuligerus* is 5.16, similar to the experimentally determined isoelectric points of IPNS proteins from *P. chrysogenum* and *C. acremonium*. When proteins fold in an *in vivo* environment, chaperonins like GroEL may shield charged groups and guide the proteins through the constructive folding pathway. In an *in vitro* environment, the absence of chaperonins may increase the importance of electrostatic effects on folding. Therefore, the effect of pH on the solubilisation and refolding of IPNS was investigated. Dramatic increases were observed in the efficiency of folding as the pH of the Tris-HCl buffer was increased from 7.22 (specific activity 117 x 10^{-3} units/mg of protein) to 8.0 (specific activity 184 x 10^{-3} units/mg of protein). This supported the proposal that charged groups could play a significant role in the folding process. As a result of the

various refolding studies, a 120% improvement was observed in the yield of active IPNS protein, from an initial specific activity of 80 x 10^{-3} units/mg protein to a final value of 184 x 10^{-3} units/mg protein. Based on these studies, the dialysis method was chosen for the solubilisation and renaturation of IB suspensions and Tris-HCl buffer at pH 8.0 was used instead of pH 7.5. Our initial attempts at purification of solubilised IPNS which had been refolded by passing through a Sephadex G-25 column after denaturation, also resulted in very low recovery (less than 10% of the refolded material). Hence, by optimising the refolding conditions, marked improvements in the efficiency of folding (70%) were achieved from an initial refolding efficiency of less than 30%, which led to improved recovery of purified IPNS (48-60%) following DEAE-Trisacryl chromatography.

The specific activity of the solubilised, recombinant IPNS was comparable to the native IPNS protein isolated from *S. clavuligerus*. Since the recovery process involved treatment with high concentration of urea and refolding of the denatured protein *in vitro*, it was desirable to compare the formation of the refolded protein with that of the native IPNS. However, since native IPNS from *S. clavuligerus* was not available in the quantities needed for such a comparison, soluble IPNS obtained from *E. coli* was used instead. When the *pcbC* gene was expressed at 37°C, IPNS was produced predominantly in the IB form, and the level of soluble IPNS produced was comparable to the level observed in its native host. Hence, purification of soluble IPNS from this system would be equally as impractical as from *S. clavuligerus*. Therefore, alternate high-level expression conditions were explored which could provide IPNS in the soluble form.

Expression of heterologous genes in *E. coli* at low temperature has been reported to result in a tendency for the proteins to be produced in a soluble form (Schein, 1989). However, this general trend is not universal, and expression of mutant tryptophan synthase α-subunit at 22°C failed to result in any formation of soluble protein. (Lim *et al.*, 1989). It is unclear why the solubility of some recombinant proteins is sensitive to

temperature, but not others. Expression of IPNS in a soluble form by growth at low temperature was a particularly suitable system for this study, since it only involved a shift in the temperature of expression. Therefore, the same expression system and identical genetic background could be maintained as was used for production of the solubilised IPNS, making it easier to compare the two forms of IPNS. Assay of IPNS activity in crude extracts prepared from cultures grown at different temperatures indicated that as the temperature was lowered, the soluble IPNS activity gradually increased. This temperature dependent solubility was most dramatic when expression at 20°C was compared with that at 37°C. SDS-PAGE analysis of the soluble crude extract and the insoluble pellet extract agreed with the activity data. The effect of temperature on the formation of IB was also monitored by examining cultures grown at different temperatures by transmission electron microscopy. The size of IB gradually decreased as the temperature was lowered, which was in accordance with the decreased prominence of the IPNS protein bands observed in SDS-PAGE gels of insoluble material as temperature decreased. It was estimated by SDS-PAGE that nearly 80% of the expressed IPNS protein was present in the soluble form at 20°C while more than 90% of the expressed IPNS was present in IB at 37°C. Although total soluble activity observed at 20°C was relatively low compared to the total solubilised IPNS obtained at 37°C, this could be explained by the fact that growth slowed dramatically after cultures were transferred to 20°C. At the time of harvest, expression cultures incubated at 37 °C had an OD_{600} of approximately 4, while the OD_{600} of expression cultures incubated at 20°C was around 2. In this study, it was observed that increasing the duration of induction beyond 30 minutes did not increase the level of expression of *pcbC* nor did the addition of rifampicin after induction. Future studies aimed at optimising the levels of insoluble and soluble IPNS produced in expression cultures could involve variations in the cell density to which cultures are grown before induction, as well as variations in the time of incubation for expression after induction.

IB formation provides a major advantage during purification of recombinant proteins. Isolation of IB pellets after repeated sonication and centrifugation to remove associated cell debris typically gives rise to material that is more than 50% recombinant protein. In our studies, based on the intensity of the IPNS protein bands compared to the contaminants on a SDS-PAGE gels, it appeared that IB were more than 80% pure prior to solubilisation. However, IB formation may not be desirable if the expressed protein fails to refold upon denaturation or if the efficiency of refolding is low. In the case of the *S. jumonjinensis pcbC* gene, although the expression level was very good, at more than 50% of the total cellular protein, the overall recovery of active IPNS protein was only 3-4%, and the major loss was presumably due to inefficient refolding (Landman *et al.*, 1991).

Expression of IPNS in a soluble form eliminates the need for refolding. Fifteen fold higher specific activity of IPNS in *E. coli* expression extracts was obtained than that seen in *S. clavuligerus* crude cell extracts by altering the temperature of expression. This study is the first report of high-level expression of soluble IPNS from a *Streptomyces* spp. and also the first report of influencing solubility of the expressed IPNS protein by manipulating culture conditions. Although IPNS proteins from different β-lactam producers share considerable similarities, the expression conditions and the host strains for expression are widely different, and therefore each requires a different purification protocol. The large amount of IPNS expressed in the soluble form made it possible to achieve complete purification in just a few steps and required only 10 fold purification overall to obtain nearly homogenous IPNS protein material. As a result, the entire purification could be completed in just 2 days yielding milligrams of protein. Though the IPNS protein was expressed predominantly in the soluble form at 20°C, some was also expressed in the insoluble form, making it difficult to estimate the amount of IPNS produced, expressed as a fraction of total cell protein. *pcbC* genes from *P. chrysogenum* and *C. acremonium* have been expressed in *E. coli* using the *trp* promoter, but the IPNS that resulted was exclusively in the soluble form. In those cases it was possible to

estimate the level of IPNS produced to be about 10 and 20% of the total soluble protein, respectively (Carr *et al.*, 1986; Samson *et al.*, 1985). It is estimated that the level of expression of soluble IPNS in this study was around 15-20% of the total soluble protein.

The specific activities of purified wild type solubilised IPNS (241 x 10^{-3} units per mg protein) and soluble IPNS (244 x 10^{-3} units per mg protein) indicated that the solubilised and soluble IPNS were very similar and there are not likely to be any major differences between the two forms. Therefore, either form should be equally suitable for crystallisation studies, making it possible to use the more readily purified solubilised IPNS. Crystallisation studies of recombinant, solubilised human hemoglobin found this material to be identical to the native hemoglobin in its structure (Nagai *et al.*, 1988) showing that the solubilisation process did not introduce any observable or detrimental defects in the structure of the recombinant protein. However, N-terminal sequence analysis of the solubilised and soluble IPNS suggested that the initiator methionine was not removed in the case of the soluble IPNS while the solubilised IPNS was shown to contain correctly processed IPNS. This interpretation was based on the fact that no sequence information was obtained for the soluble IPNS preparation, suggesting that the N-terminus was blocked. However, blocked N-termini can result from non-physiological modifications occurring during the purification process, and so it is not possible to say unequivocally that the N-terminal methionine is still present. It could also not be ascertained whether the solubilised IPNS preparation might contain some non-processed IPNS, since the sequence analysis procedure was not quantitative. In general, it is believed that the N-terminal residues are not properly processed and not removed in over-expressed proteins in *E. coli* (Marston, 1986). However, these residues do not seem to alter the biochemical properties of IPNS (Baldwin *et al.*, 1987). Since the activity values were as high or higher than the activity of native IPNS from the *S. clavuligerus* (204 x 10^{-3} units per mg protein) this also indicated that the N-terminal methionine residues if present, did not affect activity and that the recombinant IPNS was not likely to be

different in terms of its biochemical properties. Concentrated preparations of purified IPNS were found to retain their IPNS activity much better than dilute preparations, indicating that IPNS is prone to inactivation when kept at low protein concentration, particularly when it is pure. Glycerol was also found to stabilise IPNS activity, but no detailed studies were conducted to allow this effect to be quantified.

Circular dichroism studies have been extensively used to study protein conformation in solution. These studies require relatively low concentrations of protein (0.1 mg / mL) as compared with x-ray crystallographic studies. The secondary structures predicted by CD have been found to be 60% accurate when compared with corresponding x-ray crystallographic data. Analyses using model peptides and proteins whose structures are known from x-ray crystallography have revealed that CD spectra can indicate the presence of specific secondary structural entities. An α-helix has a characteristic double minimum in the far ultraviolet region at 222 and 208-210 nm and a maximum at 191-193 nm. Protein in a β-form conformation has a negative minimum near 216-218 nm and a positive maximum near the 195-200 nm region (Yang *et al.*, 1986). Conformational comparisons of the solubilised and soluble wild type IPNS by CD analysis showed nearly identical profiles in the far ultraviolet region with a characteristic double minimum indicating the presence of α-helix structure in the IPNS. This similarity in CD spectra was consistent with the similarity in specific activity values of the purified proteins and again suggested that the two forms of IPNS may be interchangeable. Secondary structure predictions using the SEQSEE (Wishart *et al.*, 1994) software program also indicated that the structure of the IPNS protein is likely to be dominated by an α-helix conformation, in agreement with the experimental CD observations.

Cysteine residues present in proteins are capable of promoting oligomer formation through intra- and inter-molecular disulphide linkages. Recombinant human fibroblast interferon, when expressed in *E. coli*, formed oligomers through inter-molecular disulphide linkages, even in reducing environment, and those oligomers were inactive.

When the cysteine residue involved was changed to a serine residue by site-specific mutagenesis, oligomer formation was prevented and the mutant protein regained its activity. This also suggested that the cysteine residue was not essential for activity (Mark *et al.*, 1984). Baldwin *et al.* (1985) have observed that the IPNS from *C. acremonium* can exist in a reduced form (active) as well as in an oxidised (inactive) form, involving an intra-molecular disulphide linkage similar to that described above for interferon, although interferon forms inter-molecular disulphide linkages. In contrast to the *C. acremonium* IPNS which has only two cysteine groups, IPNS from *S. clavuligerus* contains four cysteine residues at positions 37, 104, 142 and 251 from the N-terminal methionine residue. During the process of crystallisation, proteins reach very high concentrations and the possibility of oligomer formation via intra- and inter-molecular disulphide linkages exists for proteins with multiple cysteine residues. Oligomer formation in turn, can block the crystallisation process.

In preliminary studies, when preparations of solubilised recombinant *S. clavuligerus* IPNS were provided to Dr. M. James, Department of Biochemistry for crystallisation, no crystal formation was achieved. Although this failure to see crystal formation could not necessarily be attributed to oligomer formation, studies were initiated to prepare mutant forms of IPNS in which one or more of the cysteine residues was converted to a serine residue to eliminate any possibility of oligomerisation. These studies also provided information about the importance of the cysteine residues for IPNS activity. Treatment of IPNS with thiol group specific reagents like 2-pyridyl sulphide or N-ethylmaleimide was known to completely inactivate the enzyme. Inhibition by 2-pyridyl sulphide can be reversed by dithiothreitol whereas the alkylation by N-ethylmaleimide treatment is an irreversible reaction (Jensen *et al.*, 1986b; Samson *et al.*, 1987b). These studies which employed chemical blocking procedures demonstrated that cysteine residues were important for activity but they could not distinguish between the different cysteine residues to determine which were more important for activity. Instead,

atomic substitution of sulphur with oxygen by site-directed mutagenesis was used to shed light on the importance of the individual cysteine residues and their roles in the function of the enzyme. Oligonucleotide directed single-strand site-directed mutagenesis and site-directed mutagenesis by PCR was used to alter the cysteine groups into serine groups in the *pcbC* gene of *S. clavuligerus*.

Functional analysis of C37S, C104S, C142S and C251S mutants of *S. clavuligerus* showed that the C104S mutant had the greatest impairment of activity (96.3% activity loss) and therefore, that the cysteine 104 residue was the most important in terms of IPNS activity. The next most severely affected was the C251S mutant, which lost 47.7% activity. Similar results were obtained from site-directed mutagenesis of the corresponding cysteines residues (at positions 106 and 255) in the *C. acremonium* IPNS enzyme, which resulted in losses of 97% and approximately 50% of IPNS activity, respectively (Samson *et al.*, 1987b). This remarkable similarity in activity values despite major differences in the sources of the enzymes, one coming from a prokaryote and the other from an eukaryote, emphasized the importance of these cysteine residues. In contrast, Kriauciunas *et al.* (1991) used the very same *C. acremonium* IPNS mutants to study the effect of these C106S and C255S mutations, and concluded that the effect of these mutations was to decrease IPNS activity only marginally compared to the wild type. These authors suggested that the partially purified preparations of IPNS, and the bioassay procedures used to estimate the activity of these mutants in the earlier study by Samson *et al.* (1987b), led to an underestimate of their activity. However, in the present study, assays of IPNS activity by HPLC, which is very accurate for estimation of IPNS activity, using crude solubilised, purified solubilised, crude soluble, and purified soluble preparations of the wild type and analogous mutants of *S. clavuligerus* all supported the observations of Samson *et al.* (1987b). The most notable differences between the methods used by Samson *et al.* (1987b) and by Kriauciunas *et al.* (1991) are that the former group measured IPNS activity using a bioassay procedure and analysed crude

extracts containing soluble recombinant IPNS, whereas the latter group measured IPNS activity using a polaragraphic assay which quantitates oxygen consumption, and their studies were conducted on purified solubilised recombinant IPNS preparations. It is not immediately apparent why these technical differences should cause such different results to be obtained. Although, it is not possible to understand the source of discrepncy in the Kriauciunas *et al.* (1991) study, it is my view that the cysteine 104 residue is very important for the function of IPNS.

With a view to further confirming the above observations, the wild type and all four of the single mutants of the *S. clavuligerus* IPNS were treated with N-ethyl maleimide, and IPNS activity was assayed with these chemically blocked mutants. Residual activity could be detected only in the C104S mutant while the wild type and the rest of the mutants all lost their activity completely. The concentration of inhibitor used was nearly 10 fold higher than that required for the complete inactivation of the enzyme and hence, it was considered unlikely that incomplete alkylation was the cause of the residual activity seen in the C104S mutant. Since the C104S mutant had already lost 96.3% of its activity compared to the wild type, by virtue of the mutation, the fact that treatment of the C104S mutant with N-ethylmaleimide was unable to eliminate the small amount of residual activity suggested that the loss of activity observed when wild type IPNS was treated with N-ethylmaleimide was predominantly through the alkyation of the cysteine 104 residue. These observations support the mutation studies described earlier which suggested that the cysteine 104 residue was functionally important. The inhibitor treated C104S mutant did however lose much of its residual IPNS activity upon treatment with N-ethylmaleimide. This showed that other cysteine groups in addition to cysteine 104 are required for a fully functional enzyme. IPNS is known to require ferrous ion, oxygen and ascorbate for maximal activity (Jensen *et al.*, 1982a). It was originally speculated that one or more of the cysteinyl sulphur groups might coordinate to the ferrous ion [illegible] and actively participate in the catalytic reaction. The ability of the

C104S mutant and the other single mutants to catalyse the conversion of ACV into isopenicillin N, although at reduced efficiencies, implies that none of the cysteine groups are essential for the catalytic reaction. This view was developed based on the premise that the energetics of an electron transfer between oxygen and a metal ion centre (as would be required in a serine mutant) would be far less favorable compared to the transfer reaction between sulphur and the metal ion centre (as would occur in the wild type). However, enzyme catalysed rections mediated through metal ion centres may not show the same properties as would the uncatalysed chemical reactions in terms of their energy requirements. Enzyme catalysed reactions, in general, have much lower activation energies than uncatalysed reactions. Hence, even if the energy for electron transfer appeared to be prohibitively high between oxygen and the metal ion centre, it might not be impossible, and a low level of transfer could account for the very low level of activity seen in the C104S mutant. In order to completely eliminate such a remote possibility, it would be desirable if the cysteine 104 residue was replaced by a non-bulky alanine or glycine residue and the functional analysis repeated. A cysteine residue essential for the function of anthranilate synthase of *Serratia marcescens* was identified by substitution with glycine by site-directed mutagenesis, where the enzyme was shown to have lost more than 99.97% activity (Paluh *et al.*, 1988).

Kinetic analysis of the C106S and C255S mutant IPNS from *C. acremonium* showed that K_M values are 5 fold higher for the C106S mutant, while the C255S mutant does not differ from the wild type (Samson *et al.*, 1987b). In view of these observations, Kriauciunas *et al.* (1991) speculated that cysteine 106 is involved in substrate binding while cysteine 255 could be involved in the maintenance of structure. In order to assess the roles of the cysteine residues in iron or substrate binding, the wild type and the serine-substituted enzymes of *C. acremonium* were compared by spectroscopic analysis in the presence of iron, ACV and NO. These results showed that the iron bound in the enzyme complex was coordinated to three histidine groups and one aspartic acid present in the

enzyme, as well as to oxygen and ACV suggesting an octahedral symmetry (Jiang *et al.*, 1991; Ming *et al.*, 1991; Orville *et al.*, 1992; Scott *et al.*, 1992). The identities of the histidine and aspartic acid ligands, in terms of their locations within the IPNS sequence, remain to be elucidated. Although binding of an endogenous cysteine residue from IPNS along with ACV to the metal ion centre has not been absolutely disproved, it now appears more likely that the cysteinyl groups of IPNS are not involved in either coordination to the metal ion or in covalent substrate binding.

Since the cysteine residues at positions 37, 142 and 251 appeared to be less important for activity of the enzyme than the cysteine residue at 104, an IPNS mutant carrying mutations in all three of these cysteine residues was constructed (pMD0P7V134). By replacing these cysteine residues with serine residues which are similar in size and charge to cysteine residues, it was hoped that an IPNS mutant that was structurally similar to wild type, but unable to form oligomers, would be generated. X-ray crystallographic analysis of cysteine to serine mutants of solubilised human hemoglobin have been found to show structures identical to the native hemoglobin except for minor differences in the vicinity of the mutation (Luisi and Nagai, 1986). However, functional analysis of the solubilised triple mutant IPNS indicated that it had lost approximately 95% of activity compared to the wild type. In order to investigate whether this severe loss of activity might reflect major structural changes in the triple mutant in relation to the wild type, solubilised and soluble wild type IPNS and the solubilised triple mutant were compared with respect to their secondary structural characteristics. While the CD spectra were similar for the various forms of the wild type IPNS proteins, the triple mutant looked considerably less structured than the wild type proteins. The lack of ordered structure in the solubilised triple mutant IPNS could be due to instability in the native structure, or to an inability to refold properly following denaturation. Serine residues are very similar to cysteine residues in their ability to form hydrogen bonds and their electronic configurations. When wild type and C106S/C255S mutant IPNS of *C.*

acremonium, solubilised and purified from *E. coli*, were analysed with respect to their global structural properties, they were found to display identical CD spectra (Kriauciunas *et al.*, 1991). Since the cysteine residue at position 251 in *S. clavuligerus* IPNS is analogous to cysteine 255 in *C. acremonium* IPNS, the C251S mutation in the *S. clavuligerus* triple mutant IPNS was unlikely to have contributed to the disorganised structure. Hence, the disorganisation may have been due primarily to the replacement of the cysteine 37 and cysteine 142 residues in the triple mutant, possibly because these residues are involved in the folding of IPNS. These two residues are not as strongly conserved as the cysteine 104 and cysteine 251 residues in IPNS from various species. Nevertheless, when present, their positions are always restricted to only these two sites which might reflect their importance for the enzyme.

When the triple mutant IPNS was produced in the soluble form by growth at 20°C, it was much more active than when produced in an insoluble form and refolded. If the three mutated cysteine residues were involved in stability of the enzyme, the triple mutant IPNS obtained in the soluble form would be expected to be as unstable as the solubilised IPNS, and activities of both forms would be affected to the same extent. Since this was not observed, it seemed more likely that these groups were involved in folding or in other as yet undefined roles. Structural perturbations created by a single mutation, particularly by a conservative change like a cysteine to serine mutation, are not generally detectable by CD analysis. However, if the mutated residue plays a dominant role in the folding of the protein, then even a change in a single residue could significantly alter the structure, and even greater alterations would be expected when more than one residue are altered. It would have been useful if soluble IPNS from the triple mutant had been available in sufficient quantity to provide a direct comparison between the solubilised and soluble triple mutant IPNS by CD analysis, but time did not permit the complete purification of the various enzyme forms to be repeated.

IPNS from *C. acremonium* has been shown to convert several hundred ACV analogues into a variety of products, indicating that the enzyme has broad substrate specificity. Similar studies, but with a fewer number of substrate analogues, have indicated that IPNS from *S. clavuligerus* can also convert unnatural substrate analogues. IPNS also has been shown to catalyse the conversion of single substrates into a variety of products. Analysis of the structure of the products showed that the enzyme behaves as a desaturase, its typical activity, but also as an oxygenase which is atypical for IPNS (Baldwin *et al.*, 1991). Such a broad spectrum of product formation may indicate that the enzyme has a fairly flexible active site conformation. It is also conceivable that these unusual traits, such as the ability to catalyse more than one type of reaction chemistry, can be accounted for only if the protein part of the enzyme plays a passive intermediary role by providing an appropriate environment, rather than an active role in catalysis. Evidence from the site-directed mutagenesis studies and spectroscopic analysis by other investigators of the wild type and mutant IPNS enzymes supports such a passive role. IPNS from *C. acremonium* is the best characterised of the IPNS enzymes with regard to biochemical properties. However, all IPNS enzymes, both fungal and bacterial share considerable similarity with each other, and so the same traits are expected in all IPNS enzymes. The differences in the specific activities and kinetic parameters which have been reported for IPNS from different sources could be a reflection of evolutionary divergence between the enzymes but more likely they are due to practical differences in the assay procedures used. Another possible source of different properties between the IPNS derived from different species lies in the greater number of cysteine residues found in some bacterial IPNS enzymes. *S. clavuligerus* IPNS, with four cysteine residues, can be considered as representative of bacterial IPNS enzymes. The large amounts of recombinant *S. clavuligerus* IPNS available as a result of these studies can be used for further characterisation of substrate specificity with substrate analogues. Although IPNS can convert a variety of ACV analogues into bioactive products, the efficiency of

conversion of the analogues is generally low in comparison to the natural substrate ACV. Any possibility of developing a practical process for the production of unnatural antibiotics by the *in vivo* or *in vitro* conversion of ACV analogues, will probably require that IPNS with broader substrate specificity and greater activity against analogues be developed. While site-directed mutagenesis provides the means to create such mutant *pcbC* genes, the types of mutations that could achieve such specificity changes cannot be predicted rationally until information is available on the structure of the enzyme. For example, a spontaneous *C. acremonium* IPNS mutant has been characterised which carries a mutation at amino acid residue 285, resulting in a leucine residue replacing a proline residue. The mutant enzyme is totally inactive, although there was no reason to predict that the mutation would have been deleterious, except that the residue lies in a highly conserved region (Ramsden *et al.*, 1989). On the other hand, expression of the pMD0P7 construct in *E. coli* resulted in production of a mutant form of *S. clavuligerus* IPNS carrying a number of unintended mutations. Two of the mutations resulted in the introduction of charged residues and two of the mutations were also located in highly conserved regions. Nontheless, the enzyme was fully functional. These types of results indicate that until a three dimensional structure is determined for IPNS, very little progress can be made in attempting to alter the substrate specificity or activity of the enzyme, by site directed mutagenesis of conserved residues. Crystallisation of IPNS, as a first step in x-ray crystallographic analysis, will make it possible to determine the three dimensional structure of the enzyme which, in turn, will shed light on the identity of histidine and aspartate ligands coordinating with the metal ion, the location and stereochemistry of the active site, and the mechanism of its unusual reaction chemistry. Elucidation of the structure of the enzyme and the nature of the active site will facilitate the ultimate objective of modifying the enzyme to broaden the substrate specificity and increase the specific activity of the enzyme against unnatural substrates, which may make possible the enzymatic synthesis of potent, unnatural β-lactam antibiotics.

4.1 Future objectives:

Functional analysis of the solubilised and soluble triple mutant IPNS indicated that the cysteine residues at positions 37 and 142 are likely to be involved in the folding of IPNS. If true, a solubilised double mutant IPNS carrying serine residues instead of cysteine residues at positions 37 and 142 would likely be defective in folding and would be expected to be analogous to the solubilised triple mutant IPNS in its activity levels as well as CD spectra. If expressed in the soluble form, the corresponding double mutant IPNS would regain activity similar to the soluble triple mutant IPNS. It would be interesting to see the experimental results of these mutants. Structural comparisons of solubilised wild type IPNS, soluble wild type IPNS, solubilised triple mutant IPNS and soluble triple mutant IPNS (C37S, C142S, C251S) would also be very informative in ascertaining the role of the cysteine 37 and cysteine 142 residues in the folding of IPNS from *S. clavuligerus*. Creation of an IPNS mutant carrying an alanine residue in the place of the cysteine 104 residue might also help to identify the role of the cysteinyl thiol group in the function of IPNS more precisely. Crystallographic analysis of the wild type and the mutant IPNS would shed light on the role of the cysteine 251 residue which is at present unknown.

5. BIBLIOGRAPHY

Abraham, E. P., J. A. Huddleston, G. S. Jayatilake, J. O'Sullivan, and R. L. White. 1981. Conversion of δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine to isopenicillin N in cell free extracts of *Cephalosporium acremonium*, in Recent advances in the chemistry of β-lactam antibiotics. G. I. Gregory (ed), Royal Society of Chemistry, London, p. 125-134.

Aharonowitz, Y., G. Cohen, and J.F. Martin. 1992. Penicillin and cephalosporin biosynthetic genes: structure, organisation, regulation, and evolution. Annu. Rev. Microbiol. 46:461-495.

Aharonowitz, Y., Y. Av-Gay, R. Schreiber, and G. Cohen. 1993. Characterisation of a broad-range disulfide reductase from *Streptomyces clavuligerus* and its possible role in beta-lactam antibiotic biosynthesis. J. Bacteriol. 175:623-629.

Ahmed, A. 1987. A simple and rapid procedure for sequencing long (40 Kb) DNA fragments. Gene. 61:363-372.

Albers-Schonberg, G., B.H. Arison, E. Kazka, F.M. Kahan, J.S. Kahan, B. Lago, W.M. Maise, R.E. Rhodes and J.L. Smith. 1976. Thienamycin: A new β-lactam antibiotic. III. Structures determination and biosynthetic data, Abstract 229, 16th Intrascience Conf., Antimicrobial Agents Chemotherapy, Chicago.

Allen, S.P., J.O. Polazzi, J.K. Gierse, and A.M. Easton. 1992. Two novel heat shock genes encoding proteins produced in response to heterologous proteins expression in *Escherichia coli*. J. Bacteriol. 174:6938-6947.

Amann, E., J. Brosius, and M. Ptashne. 1983. Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. Gene. 25:167-178.

Andersson, S.G.E., and C.G. Kurland. 1990. Codon preferences in free-living microorganisms. Microbiol. Reviews. 54:198-210.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. Science. 181:223-230.

Baker, T.A., A.D. Grossman, and C.A. Gross. 1984. A gene regulating the heat shock response in *Escherichia coli* also affects proteolysis. Proc. Natl. Acad. Sci. USA. 81:6779-6783.

Baker, B.J., J.E. Dotzlaf, and W.K. Yeh. 1991. Deacetoxycephalosporin C hydroxylase of *Streptomyces clavuligerus*: Purification, characterisation, bifunctionality, and evolutionary implication. J. Biol. Chem. 266:5087-5093.

Baldwin, J.E., R.M. Adlington, H.H. Ting, C. Arigone, P. Graf, and B. Martinoni. 1985. Penicillin biosynthesis: the immediate origin of the sulphur atom. Tetrahedran. 41:3339-3343.

Baldwin, J.E., S.J. Killin, A.J. Pratt, I.D. Sutherland, N.J. Turner, M.J.C. Crabbe, E.P. Abraham, and A.C. Willis. 1987. Purification and characterisation of cloned isopenicillin N synthetase. J. Antibiot. 40:652-659.

Baldwin, J.E., and E.P. Abraham. 1988. The biosynthesis of penicillins and cephalosporins. Nat. Prod. Rep. 5:129-145.

Baldwin, J.E., J.W. Bird, R.A. Field, N.M. O'Callaghan, and C.J. Schofield. 1990. Isolation and partial characterisation of ACV synthetase from *Cephalosporium acremonium* and *Streptomyces clavuligerus*. J. Antibiot. 43:1055-1057.

Baldwin, J.E., J.M. Blackburn, J.D. Sutherland, and M.C. Wright. 1991. High-level soluble expression of isopenicillin N synthase isozymes in *E. coli*. Tetrahedron 47:5991-6002.

Blackshear, P.J. 1984. Systems for polyacrylamide gel electrophoresis. Methods Enzymol. 104:237-255.

Blum, P., M. Velligan, N. Lin, and A. Matin. 1992. DnaK-mediated alterations in human growth hormone protein inclusion bodies. Bio/Technology. 10:301-304.

Barredo, J.L., B. Diez, E. Alvarez, and J.F. Martin. 1989a. Large amplification of a 35-kb DNA fragment carrying two penicillin biosynthetic genes in high penicillin producing strains of *Penicillium chrysogenum*. Curr. Genet. 16:453-459.

Barredo, J.L., P. van Soligen, B. Diez, E. Alvarez, J.M. Cantoral, A. Kattevilder, E.B. Smaal, M.A.M. Groenen, A.E. Veenstra, and J.F. Martin. 1989b. Cloning and characterisation of acyl-CoA:6-APA acyltransferase gene of *Penicillium chrysogenum*. Gene. 83:291-300.

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

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

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

Buchner, J., M. Schmidt, M. Fuchs, R. Jaenickle, R. Rudolph, F.X. Schmid, and T. Kiefhaber. 1991. GroEL facilitates refolding of citrate synthase by suppressing aggregation. Biochemistry. 30:1586-1591.

Buell, G., M.F. Schulz, G. Selzer, A. Chollet, N. Rao Movva, D. semon, S. Escanez, and E. Kawashima. 1985. Optimising the expression in *E. coli* of a synthetic gene encoding somatomedin-C (IGF-I). Nucl. Acids Res. 13:1923-1938.

Carlson, J.D., and M.L. Yarmush. 1992. Antibody assisted protein refolding. Bio/Technology. 10:86-91.

Carr, L.C., P.L. Skatrud, M.E. Scheetz, S.W. Queener, and T.D. Ingolia. 1986. Cloning and expression of the isopenicillin N synthetase gene from *Penicilliumchrysogenum*. Gene. 48:257-266.

Carter, P., H. Bedouelle, and G. Winter. 1985. Improved oligonucleotide site-directed mutagenesis using M13 vectors. Nucl. Acids Res. 13:4431-4443.

Castro, J.M., P. Liras, L. Laiz, J. Cortes, and J.F. Martin. 1988. Purification and characterisation of the isopenicillin N synthase of *Streptomyces lactamdurans*. J. Gen. Microbiol. 134:133-141.

Chung, C.T., S.L. Niemela, and R.H. Miller. 1989. One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. USA. 86:2172-2175.

Chung, C.H. 1993. Proteases in *Escherichia coli*. Science. 262:372-374.

Chalmers, J.J., E. Kim, J.N. Telford, E.Y. Wong, W.C. Tacon, M.L. Shuler, and D.B. Wilson. 1990. Effects of temperature on *Escherichia coli* overproducing β-lactamase or human growth factor. Appl. Environ. Microbiol. 56:104-111.

Cleland, J.L, and D.I.C. Wang. 1990. Cosolvent assisted protein refolding. Bio/Technology. 8:1274-1278.

Cleland, J.L., S.E. Builder, J.R. Swartz, M. Winkler, J.Y. Chang, and D.I.C. Wang. 1992. Polyethylene glycol enhanced protein refolding. Bio/Technology. 10:1013-1019.

Cohen, G., A. Argaman, R. Schreiber, M. Mislovati and Y. Aharonowitz. 1994. The thioredoxin system of *Penicillium chrysogenum* and its possible role in penicillin biosynthesis. J. Bacteriol. 176:973-984.

Coque, J.J.R., J.F. Martin, J.G. Calzada, and P. Liras. 1991a. A gene encoding lysine 6-aminotransferase which forms α-aminoadipic acid, a precursor of beta-lactam antibiotics, is located in the cluster of cephamycin biosynthetic genes in *Nocardia lactamdurans*. J. Bacteriol. 173:6258-6264.

Coque, J.J.R., J.F. Martin, J.G. Calzada, and P. Liras. 1991b. The cephamycin biosynthetic genes *pcbAB*, encoding a large multidomain peptide synthetase, and *pcbC* of *Nocardia lactamdurans* are clustered together in an organisation different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*. Mol. Microbiol. 5:1125-1133.

Corpet, F. 1988. Multiple sequence alignment with hierarchial clustring. Nucl. Acids Res. 16:10681-10890.

Creighton, T.E. 1986. Folding of proteins adsorbed reversibly to ion-exchange resins. In : Protein structure, folding, and design (Oxender, D.L., ed.), Alan R. Liss, Inc., New York. p. 249-257.

Creighton, T.E. 1990. Protein folding. Biochem. J. 270:1-16.

De Boer, H.A., L.J. Comstock, and M. Vasser. 1983. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. Proc. Natl. Acad. Sci. USA. 80:21-25.

Demain, A.L., and A.N. Solomon (eds). 1983. Biosynthesis of β-lactam antibiotics: In antibiotics containing the betalactam strucutre. Berlin/ Heidelberg, Springer-verlag. 1:189-228.

Diez, B., S. Gutierrz, J.L. Barredo, P. van Soligen, L.H.M. Van der Voort, and J.F. Martin. 1990. The cluster of penicillin biosynthetic genes: identification and characterisation of the *pcbAB* gene encoding the α-aminoadipyl-cysteinyl-valine synthetase and linkage to the *pcbC* and *penDE* genes. J. Biol. Chem. 265:16358-16365.

Dill, K.A. 1990. Dominant forces in protein folding. Biochemistry. 29:7133-7155.

Doran, J.L., B.K. Leskiw, A.K. Petrich, D.W.S. Westlake, and S.E. Jensen. 1990. Production of *Streptomyces clavuligerus* isopenicillin N synthase in *Escherichia coli* using two-cistron expression systems. J. Ind. Microbiol. 5:197-206.

Durairaj, M., J.L. Doran, and S.E. Jensen. 1992. High-level expression of the *Streptomyces clavuligerus* isopenicillin N synthase gene in *Escherichia coli*. Appl. Environ. Microbiol. 58:4038-4041.

Elson, S.W. and R.S. Oliver. 1978. Studies on the biosynthesis of clavulanic acid. I. Incorporation of ^{13}C-labelled precursors. J. Antibiot. 31:586-592.

Elson, S.W., R.S. Oliver, B.W. Bycroft and E.A. Faruk. 1982. Studies on the biosynthesis of clavulanic acid. III. Incorporation of DL-[3,4-$^{13}C_2$] glutamic acid. J. Antibiot. 35:81-86.

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

Fujishima, Y., P. Nordlund, G. Pelosi, C.J. Schofield, S.C.J. Cole, J.E. Baldwin, and J. Hajdu. 1994. Crystallisation and preliminary x-ray diffraction studies on a recombinant isopenicillin N synthase from *Cephalosporium acremonium*. J. Mol. Biol. 242:712-714.

Garcia-Dominguez, M., P. Liras, and J.F. Martin. 1991. Cloning and characterisation of the isopenicillin N synthase gene of *Streptomyces griseus* NRRL 3851 and studies of expression and complementation of the cephamycin pathway in *Streptomyces clavuligerus*. Antimicrob. Agents Chemother. 35:44-52.

Ghatge, M.S., S.U. Phadatare, A.M. Bodhe, and V.V. Desphande. 1994. Unfolding and refolding of glucose/xylose isomerase from *Streptomyces* sp. NCIM 2730. Enzyme Microb. Technol. 16:323-327.

Glick, B.R., and G.K. Whitney. 1987. Factors affecting the expression of foreign proteins in *Escherichia coli*. J. Ind. Microbiol. 2:277-282.

Goeddel, D.V (ed). 1990. Gene Expression Technology. Methods Enzymol. 185:3-195.

Goldberg, M.E. 1985. The second translation of the genetic message: protein folding and assembly. Trends Biochemi Sci. 10:388-391.

Goff, S.A., and A.L. Goldberg. 1985. Production of abnormal proteins in *E. coli* stimulates transcription of *lon* and other heat shock genes. Cell. 41:587-595.

Gottesman, S. 1990. Minimizing proteolysis in *Escherichia coli*: genetic solutions. Methods Enzymol. 185:119-129.

Gragerov, A., E. Nudler, N. Komissarova, G.A. Gaitanaris, M.E. Gottesman, and V. Nikiforov. 1992. Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 89:10341-10344.

Gramajo, H.C., J. White, C.R. Hutchinson, and M.J. Bibb. 1991. Overproduction and localisation of components of the polyketide synthase of *Streptomyces glaucescens* involved in the production of the antibiotic tetracenomycin C. J. Bacteriol. 173:6475-6483.

Gribskov, M., and R.R. Burgess. 1983. Overexpression and purification of the sigma subunit of *Escherichia coli* RNA polymerase. Gene. 26:109-118.

Guillerez, J., M. Gazeau, and M Dreyfus. 1991. In the *Escherichia coli lacZ* gene the spacing between the translating ribosomes is insensitive to the efficiency of translation initiation. Nucl. Acids Res. 19:6743-6750.

Gutierrez, S., B. Diez, E. Montenegro, and J.F. Martin. 1991. Characterisation of the *Cephalosporium acremonium pcbAB* gene encoding α-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide ssynthetase: linkage to the *pcbC* gene as a cluster of early cephalosporin-biosynthetic genes and evidence of multiple functional domains. J. Bacteriol. 173:2354-2365.

Haase-Pettingell, C.A., and J. King. 1988. Formation of aggregates from a thermolabile in vivo folding intermediate in P22 tailspike maturation: a model for inclusion body formation. J. Biol. Chem. 263:4977-4983.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557-580.

Hoess, A., A.K. Arthur, G. Wanner, and E. Fanning. 1988. Recovery of soluble, biologically active recombinant proteins from total bacterial lysates using ion exchange resin. Bio/Technology. 6:1214-1217.

Hollander, I.J., Y.Q. Shen, J. Heim, A.L. Demain, and S. Wolfe. 1984. A pure enzyme catalysing penicillin biosynthesis. Science. 224:610-612.

Hargrove, J.L., and F.H. Schmidt. 1989. The role of mRNA and protein stability in gene expression. FASEB J. 3:2360-2370.

Hartley, D.L. and J.F. Kane. 1988. Properties of inclusion bodies from recombinant *Escherichia coli*. Biochem. Soc. Trans. 16:101-102.

Hui, A., and H.A. de Boer. 1987. Specialised ribosome system: Preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 84:4762-4766.

Hwang, L.S., B. Chen, H. Lin, L. Soong, Y. Jou, W. Chang, and S. Liu. 1990. The striking effects of spacer and terminator sequences on the synthesis of porcine growth hormone in *Escherichia coli*. Biotech. Lett. 12:335-340.

Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. 146:1-21.

Innis, M.A., D.H. Gelfand, J.J. Sninsky, and T.J. White. (ed). 1990. PCR protocols. A guide to methods and applications, Academic Press Inc., New York.

Jacob, W.F., M. Santer, and A.E. Dahlberg. 1987. A single base change in the Shine-Dalgarno region of 16S RNA of *Escherichia coli* effects translation of many proteins. Proc. Natl. Acad. Sci. USA. 84:4757-4761.

Jensen, S.E., D.W.S. Westlake, and S. Wolfe. 1982a. Cyclization of δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine to penicillins by cell-free extracts of *Streptomyces clavuligerus*. J. Antibiot. 35:483-490.

Jensen, S.E., D.W.S. Westlake, and S. Wolfe. 1982b. High performance liquid chromatographic assay of cyclization activity in cell-free systems from *Streptomyces clavuligerus*. J. Antibiot. 35:1026-1032.

Jensen, S.E., D.W.S.Westlake and S. Wolfe. 1983. Partial purification and characterisation of isopenicillin N epimerase activity from *Streptomyces clavuligerus*. Can. J. Microbiol. 29:1526-1531.

Jensen, S.E., D.W.S. Westlake and S.Wolfe. 1985. Deacetoxycephalosporin C synthetase and deacetoxycephalosporin C hydroxylase are two separate enzymes in *Streptomyces clavuligerus*. J. Antibiot. 38:263-265.

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

Jensen, S.E., B.K.Leskiw, L.C. Vining, Y. Aharonowitz, D.W.S. Westlake, and S. Wolfe. 1986a. Purification of isopenicillin N synthetase from *Streptomyces clavuligerus*. Can. J. Microbiol. 32:953-958.

Jensen, S.E., B.K. Leskiw, J.L. Doran, A.K. Petrich, and D.W.S. Westlake. 1989. Expression of the *Streptomyces clavuligerus* isopenicillin N synthase gene in *Escherichia coli* and *Streptomyces lividans*. *In* C. L. Hershberger, S.W. Queener, and G. Hegeman (ed.), The genetics and molecular biology of industrial microorganisms. American Society for Microbiology, Washington, D. C, p. 239-245.

Jensen, S.E., A. Wong, M. Rollins, and D.W.S. Westlake. 1990. Purification and partial characterisation of δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Streptomyces clavuligerus*. J. Bacteriol. 172:7269-7271.

Jensen, S.E. and A.L. Demain. 1995. Beta-Lactams. In: genetics and biochemistry of antibiotic production. L.C. Vining and C. Stuttard (eds), Butterworth-Heinemann, p. 239-268.

Jiang, F., J. Peisach, L.J. Ming, L. Que, and V.J. Chen. 1991. Electron spin echo envelope modulation studies of the Cu(II)-substituted derivative of Isopenicillin N synthase: a structural and spectroscopic model. Biochemistry. 30:11437-11445.

Kadima, T.A. 1993. ACV synthetase from S. *clavuligerus*. Ph.D. thesis, University of Alberta, Edmonton, Canada.

Kane, J. F., and D. L. Hartley. 1988. Formation of recombinant protein inclusion bodies in *Escherichia coli*. Trends Biotechnol. 6:95-101.

Kern, B.A., D. Hendlin and E. Inamine. 1980. L-lysine ε-aminotransferase involved in cephamycin C synthesis in *Streptomyces lactamdurans*. Antimicrob. Agents Chemother. 17:676-685.

Kleinkauf, H. and H. von Dohren. 1990. Nonribosomal biosynthesis of peptide antibiotics. Eur. J. Biochem. 192:1-15.

Kohsaka, M., and A.L. Demain. 1976. Conversion of penicillin N to cephalosporin(s) by cell-free extracts of *Cephalosporium acremonium*. Biochim. Biophys. Res. Commun. 70:465-473.

Konomi, T., S. Herchen, J.E. Baldwin, M. Yoshida, N. A. Hunt, and A.L. Demain. 1979. Cell-free conversion of δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine into antibiotic with the properties of isopenicillin N in *Cephalosporium acremonium*. Biochem. J. 184:427-30.

Kovacevic, S., M.B. Tobin and J.R. Miller. 1990. The β-lactam biosynthesis genes for isopenicillin N epimerase and deacetoxycephalosporin C synthetase are expressed from a single transcript in *Streptomyces clavuligerus*. J. Bacteriol. 172:3952-3958.

Kramer, W., K. Schughat, and H.J. Fritz. 1982. Directed mutagenesis of DNA cloned in filamentous phage: influence of hemimethylated GATC sites on marker recovery from restriction fragments. Nucl. Acids Res. 10:6475-6485.

Kramer, W.,V. Drutsa, H.W. Jansen, B. Kramer, M. Pflugfelder, and H.J. Fritz. 1984. The gapped duplex DNA approach to oligonucleotide-directed mutation construction. Nucl. Acids Res. 12:9441-9456.

Kriauciunas, A., C.A. Frolik, T.C. Hassell, P.L. Skatrud, M.G. Johnson, N.L. Holbrook, and V.J. Chen. 1991. The functional role of cysteines in Isopenicillin N synthase-correlation of cysteine reactivities toward sulfhydryl reagents with kinetic properties of cysteine mutants. J. Biol. Chem. 266:11779-11788.

Kunkel, T.A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA. 82:488-492.

Kupka, J., Y.Q. Shen, S. Wolfe and A.L. Demain. 1983. Partial purification and properties of the α ketoglutarate-linked ring-expansion enzyme of β-lactam biosynthesis of *Cephalosporium acremonium*. FEMS Microbiol. Lett. 16:1-6.

Kurland, C.G. 1991. Codon bias and gene expression. FEBS Letters. 285:165-169.

Kuwajima, K., M. Mitani, and S. Sugai. 1989. Characterisation of the critical state in protein folding: effects of guanidine hydrochloride and specific Ca^{2+} binding on the folding kinetics of α-lactalbumin. J. Mol. Biol. 206:547-561.

Laiz, L., P. Liras, J.M. Castro, and J.F. Martin. 1990. Purification and characterisation of the isopenicillin N epimerase from *Nocardia lactamdurans*. J. Gen. Microbiol.136:663-671.

Laminet, A.A., T. Ziegelhoffer, C. Georgopoulos, and A. Pluckthun. 1990. The *Escherichia coli* heat shock proteins GroEL and GroES modulate the folding of the β-lactamase precursor. EMBO. J. 9:2315-2319.

Landman, O., D. Shiffman., Y. Av-Gay, Y. Aharonowitz, and G. Cohen. 1991. High level expression in *Escherichia coli* of isopenicillin N synthase genes from *Flavobacterium* and *Streptomyces*, and recovery of active enzyme from inclusion bodies. FEMS Microbiol. Lett. 84:239-244.

LaVallie, E.R., E.A. Diblasio, S. Kovacic, K.L. Grant, P.F. Schendel, and J.M. McCoy. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. Bio/Technology. 11:187-193.

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

Leskiw, B.K., E.J. Lawlor, J.M. Fernandez-Abalos, and K.F. Chater. 1991. TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative *Streptomyces* mutants. Proc. Natl. Acad. Sci. USA. 88:2461-2418.

Liljenstrom, H., and G. Von Heijine. 1987. Translation rate modification by preferential codon usage: intragenic position effects. J. Theor. Biol. 124:43-55.

Lim, W.K., H.E. Smith-Somerville, and J.K. Hardman. 1989. Solubilisation and renaturation of overexpressed aggregates of mutant tryptophan synthase α-subunits. Appl. Environ. Microbiol. 55:1106-1111.

London, J., C. Skrzynia, and M.E. Goldberg. 1974. Renaturation of *Escherichia coli* tryptophanase after exposure to 8 M urea: evidence for the existence of nucleation centres. Eur. J. Biochem. 47:409-415.

Luisi, B.F., and K. Nagai. 1986. Crystallographic analysis of mutant human haemoglobins made in *Escherichia coli*. Nature. 320:555-556.

MacCabe, A.P., M.B.R. Riach, S.E. Unkles, and J.R. Kingdom. 1990. The *Aspergillus nidulans* npeA locus consists of three contiguous genes required for penicillin biosynthesis. BMBO J. 9:279-287.

MacCabe, A.P., H. van Liempt, H. Palissa, S.E. Unkles, M.B.R. Riach, E. Pfeifer, H. von Dohren, and J.R. Kinghorn. 1991. δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Aspergillus nidulans*: molecular characterisation of the acvA gene encoding the first enzyme of the penicillin biosynthetic pathway. J. Biol. Chem. 266:12646-12654.

Madduri, K., C. Stuttard and L.C. Vining. 1989. Lysine catabolism in *Streptomyces* spp. is primarily through cadaverine: β-lactam producers also make α-aminoadipate. J. Bacteriol. 171:299-302.

Makoff, A.J., and A. E. Smallwood. 1988. Heterologous expression in *Escherichia coli*. Effects of alterations in the sequence 5' to the initiation codon. Biochem. Soc. Trans. 16:48-49.

Makoff, A.J., M.D. Oxer, M.A. Romanos, N.F. Fairweather, and S. Ballantine. 1989. Expression of tetanus fragment C in *E coli*: high level expression by removing rare codons. Nucl. Acids Res. 17:10191-10202.

Malmberg, L.H., W.S. Hu, and D.H. Sherman. 1993. Precursor flux control through targeted chromosomal insertion of the lysine ε-aminotransferase (*lat*) gene in cephamycin C biosynthesis. Bacteriol. 175:6916-6924.

Mark, D.F., S.D. Lu, A.A. Creasey, R. Yamamoto, and L.S. Lin. 1984. Site-specific mutagenesis of the human fibroblast interferon gene. Proc. Natl. Acad. Sci. USA. 81:5662-5666.

Martin, J.F., and P. Liras. 1989. Enzymes Involved in penicillin, cephalosporin and cephamycin biosynthesis. Adv. Biochem. Eng. 39:153-187

Martin, J., T. Langer, R. Boteva, A. Schramel, A.L. Horwich, and F. U. Hartl. 1991. Chaperonin-mediated protein folding at the surface of GroEL through a " molten globule"-like intermediate. Nature. 352:36-42

Marston, F.A.O. 1986. The purification of eukaryotic polypeptides synthesised in *Escherichia coli*. Biochem. J. 240:1-12.

Mead, D.A., N.K. Pey, C. Herrnstadt, R.A. Marcil, and L.M. Smith. 1991. A universal method for the direct cloning of PCR amplified nucleic acid. Bio/Technology. 9:657-663.

Ming, L.J., L. Que, A. Kriauciunas, C.A. Frolik, and V.J. Chen. 1991. NMR studies of the active site of isopenicillin N synthase, a non-heme iron(II) enzyme. Biochemistry. 30:11653-11659.

Misoka, F., T. Miyake, K. Miyoshi, M. Sugiyama, S. Sakamoto, and T. Fuwa. 1989. Overproduction of human insulin-like growth factor-II in *Escherichia coli*. Bioteh. Lett. 11:839-844.

Munson, M.L., G.D. Stormo, R.L. Niece, and W.S. Reznikoff. 1984. *LacZ* translation initiation mutations. J. Mol. Biol. 177:663-683.

Newton, G.G.F., and E.P. Abraham. 1955. Cephalosporin C, a new antibiotic containing sulphur and D-α-aminoadipic acid. Nature. 175:548.

Nagai, K., H.C. Thogersen, and B.F. Luis. 1988. Refolding and crystallographic studies of eukaryotic proteins produced in *Escherichia coli*. Biochem. Soc. Trans. 16:108-110.

Olins, P.O., and S.H. Rangwala. 1990. Vector for enhanced translation of foreign genes in *Escherichia coli*. Methods Enzymol. 185:115-119.

Orville, A.M., V.J. Chen, A. Kriauciunas, M.R. Harpel, B.G. Fox, E. Munck, and J.D. Lipscomb. 1992. Thiolate ligation of the active site Fe^{2+} of isopenicillin N synthase derives from substrate rather than endogenous cysteine: spectroscopic studies of site-specific cys-ser mutated enzymes. Biochemistry. 31:4602-4612.

Pang, C.P., B. Chakravarti, R.M. Adlington, H.H. Ting, R.L. White, G.S. Jayatilake, J.E. Baldwin, and E.P. Abraham. 1984. Purification of isopenicillin N synthetase. Biochem. J. 222:789-795.

Paluh, J.L., H. Zalkin, D. Betsch, and H.L. Weith. 1985. Study of anthranilate synthase function by replacement of cysteine 84 using site-directed mutagenesis. J. Biol. Chem. 260:1889-1894.

Perry, D., E.P. Abraham, and J.E. Baldwin. 1988. Factors affecting the isopenicillin N synthetase reaction. Biochem. J. 255:345-351.

Petrich, A.K., B.K. Leskiw, A.S. Paradkar and S.E. Jensen. 1994. Transcriptional mapping of the genes encoding the early enzymes of the cephamycin biosynthetic pathway of *Streptomyces clavuligerus*. Gene. 142:41-48.

Ptashne, M., K. Backman, M.Z. Humayun, A. Jeffrey, R. Maurer, and R.T. Sauer. 1976. Autoregulation and function of a repressor in bacteriophage lambda. Science. 194:156-161.

Ramon, D., L. Carramolino, C. Patino, F. Sanchez, and M.A. Penalva. 1987. Cloning and characterisation of the isopenicillin N synthetase gene mediating the formation of the β-lactam ring in *Aspergillus nidulans*. Gene. 57:171-181.

Ramos, F.R., M.J. Lopez-Nieto, and J.F. Martin. 1985. Isopenicillin N synthetase of *Penicillium chrysogenum* , an enzyme that converts δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine to isopenicillin N. Antimicrob. Agents Chemother. 27:380-387.

Ramsden, M., B.A. McQuade, K. Saunders, M.K. Turner, and S. Harford. 1989. Characterisation of a loss-of-function in the isopenicillin N synthetase gene of *Acremonium chrysogenum*. Gene. 85:267-273.

Randall, C.R., Y. Zang, A.E. True, and L. Que. 1993. X-ray absorption studies of the ferrous active site of isopenicillin N synthase and related model complexes. Biochemistry. 32:6664-6673.

Rinas, U., L.B. Tsai, D. Lyons, G.M. Fox, C. Stearns, J. Fieschko, D. Fenton, and J.E. Bailey. 1992. Cysteine to serine substitutions in basic fibroblast growth factor: effect on inclusion body formation and proteinolytic susceptibility during in vitro refolding. Bio/technology. 10:435-440.

Rinas, U., and J.E. Bailey. 1993. Overexpression of bacterial hemoglobin causes incorporation of pre-β-lactamase into cytoplasmic inclusion bodies. Appl. Environ. Microbiol. 59:561-566.

Rosenberg, A.H., B.N. Lade, D. Chui, S. Lin, J.J. Dunn, and F.W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene. 56:125-135.

Russel, M., and P. Model. 1984. Replacement of the *fip* gene of *Escherichia coli* by an inactive gene cloned on a plasmid. J. Bacteriol. 159:1034-1039.

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

Samson, S.M., R. Belagaje, D.T. Blankenship, J.L. Chapman, D. Perry, P.L. Skatrud, R.M. Van Frank, E.P. Abraham, J.E. Baldwin, S.W. Queener, and T.D. Ingolia. 1985. Isolation, sequence determination and expression in *E. coli* of the isopenicillin N synthetase gene from *Cephalosporium acremonium*. Nature 318:191-194.

Samson, S.M., J.E. Dotzlaf, M.L.Slisz, G.W. Becker, R.M. van Frank, L.E. Veal, W.K. Yeh, J.R. Miller, S.W. Queener, and T.D. Ingolia. 1987a. Cloning and expression of the fungal expandase/hydroxylase gene involved in cephalosporin biosynthesis. Bio/Technology. 5:1207-1216.

Samson, S.M., J.L. Chapman, R. Belagajie, S.W. Queener, and T.D. Ingolia. 1987b. Analysis of the role of cysteine residues in isopenicillin N synthase activity by site-directed mutagenesis. Proc. Natl. Acad. Sci. USA. 84:5705-5709.

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

Sawada, Y., J.E. Baldwin, P.D. Singh, N.A. Solomon, and A.L. Demain. 1980. Cell-free cyclisation of δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine to isopenicillin N. Antimicrob. Agents Chemother. 18: 465-470.

Schein, C.H. 1989. Production of soluble recombinant proteins in bacteria. Bio/Technology. 7:1141-1149.

Scott, R.A., S. Wang, M.K. Eidsness, A. Kriauciunas, C.A. Frolik, and V.J. Chen. 1992. X-ray absorption spectroscopic studies of the high-spin iron (II) active site of isopenicillin N synthase: Evidence for Fe-S interaction in the enzyme-substrate complex. Biochemistry. 31:4596-4601.

Schottel, J., J.J. Sninsky, and S.N. Cohen. 1984. Effects of alterations in the translation control region on bacterial gene expression: use of *cat* gene constructs transcribed from the *lac* promoter as a model system. Gene. 28:177-193.

Shepard, H.M., E. Yelverton, and D. Goeddel. 1982. Increased synthesis in *E. coli* of fibroblast and leukocyte interferon through alterations in ribosome binding sites. DNA. 1:125-131.

Shiffman, D., M. Mevarech, S.E. Jensen, G. Cohen, and Y. Aharonowitz. 1988. Cloning and comparative sequence analysis of the gene coding for isopenicillin N synthase in *Streptomyces*. Mol. Gen. Genet. 214:365-377.

Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to non-sense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA. 71:1342-1346.

Simon, L.D., B. Randolph, N. Irwin, and G. Binkowski. 1983. Stabilisation of proteins by a bacteriophage T4 gene cloned in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 80:2059-2062.

Simons, G., E. Remaut, B. Allet, R. Devos, and W. Fiers. 1984. High-level expression of human interferon gamma in *Escherichia coli* under control of the P_L promoter of bacteriophage lambda. Gene. 28:55-64.

Singh, P.D., M.G. Young, J.H. Johnson, C.M. Cimarusti and R. Sykes. 1984. Bacterial production of 7-formamidocephalosporins, isolation and structure determination. J. Antibiot. 37:773-780.

Skatrud, P.L., A.J. Teitz, T.D. Ingolia, C.A. Cantwell, D.L. Fisher, J.L. Chapman, and S.W. Queener. 1989. Use of recombinant DNA to improve production of cephalosporin C by *Cephalosporium acremonium*. Bio/Technology. 7:477-486.

Skatrud, P.L., and S.W. Queener. 1989a. An electrophoretic molecular karyotype for an industrial strain of *Cephalosporium acremonium*. Gene. 78:331-338.

Skatrud, P.L. 1992. Molecular biology of the beta-lactam producing fungi. In more gene manipulations in fungi, T.W. Bennett and L.L. Lashre (ed), New York, Academic.

Skowyra, D., C. Georgopoulos, and M. Zylicz. 1990. The *E. coli dnaK* gene product, the hsp70 homolog, can reactivate heat-inactivated RNA polymerase in an ATP hydrolysis-dependent manner. Cell. 82:939-944.

Smith, D.J., J.H. Bull, J. Edwards, and G. Turner. 1989. Amplification of the isopenicillin N synthetase gene in a strain of *Penicillium chrysogenum* producing high levels of penicillin. Mol. Gen. Genet. 216:492-497.

Sorensen, M.A., C.G. Kurland, and S. Pedersen. 1989. Codon usage determines translation rate in *Escherichia coli*. J. Mol. Biol. 207:365-377.

Stanssens, P., E. Remaut, and W. Fiers. 1986. Inefficient translation initiation causes premature transcription termination in the *lacZ* gene. Cell. 44:711-718.

Tabor, S., and C.C. Richardson. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA. 82:1074-1078.

Taylor, J.W., W. Schmidt, R. Cosstick, A. Okruszek, and F. Eckstein. 1985a. The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA. Nucl. Acids Res. 13:8749-8764.

Taylor, J.W., J. Ott, and F. Eckstein. 1985b. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. Nucl. Acids Res. 13:8765-8785.

Tobin, M.B., M.D. Fleming, P.L. Skatrud, and J.R. Miller. 1990. Molecular characterisation of the acyl-coenzyme A:isopenicillin N acyltransferase gene (*penDE*) from *Penicillium chrysogenum* and *Aspergillus nidulans* and activity of recombinant enzyme in *Escherichia coli*. J. Bacteriol. 172:5908-5914.

Tobin, M.B., S. Kovacevic, K. Madduri, J.A. Hoskins, P.L. Skatrud, L.C. Vining, C. Stuttard, and J.R. Miller. 1991. Localisation of lysine ε-aminotransferase (*lat*) and δ-(L-α-aminoadipyl) L-cysteinyl-D-valine (ACV) synthetase (*pcbAB*) genes from *Streptomyces clavuligerus* and production of lysine ε- aminotransferase activity in *Escherichia coli*. J. Bacteriol. 173:6223-6229.

Usui, S., and C.A. Yu. 1989. Purification and properties of isopenicillin N epimerase from *Streptomyces clavuligerus*. Biochim. Biophys. Acta. 999:78-85.

Van Liempt, H., H. von Dohren, and H. Kleinkarf. 1989. δ-(L-α-aminoadipyl -L-cysteinyl-D-valine synthetase from *Aspergillus nidulans*. J. Biol. Chem. 264:3680-3684.

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

Velasco, J., S. Gutierrez, F.J. Fernandez, A.T. Marcos, C. Arenos and J.F. Martin. 1994. Exogenous methionine increases levels of mRNAs transcribed from *pcbAB*, *pcbC*, and *cefEF* genes, encoding enzymes of the cephalosporin biosynthetic pathway, in *Acremonium chrysogenum*. J. Bacteriol. 176:985-991.

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

Wishart, D.S, R.F. Boyko, L. Willard, F.M. Richards, and B.D. Sykes. 1994. SEQSEE: a comprehensive program suite for protein sequence analysis. CABIOS. 10:121-132.

Wolfe, S., A.L. Demain, S.E. Jensen, and D.W.S. Westlake. 1984. Enzymatic approach to synthesis of unnatural beta-lactams. Science. 226:1386-1392.

Xiao, X., G. Hintermann, J. Piret, A. Hausler and A.L. Demain. 1991. Purification of cephalosporin 7-α-hydroxylase from *Streptomyces clavuligerus* and cloning of its gene. Int. Symp. Biol. Actinomycetes, 8th, Madison, Wisconsin, p. 60 (Abstr.).

Yang, J.T., C.S.C. Wu, and H.M. Martinez. 1986. Calculation of protein conformation from Circular Dichroism. Methods Enzymol. 130:208-269.

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

Zhang, J.Y., and A.L. Demain. 1990. Purification of ACV synthetase from *Streptomyces clavuligerus*. Biotech. Lett. 12:649-654.

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

Zuker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucl. Acids Res. 9:133-145.