Studies of the glyoxylate bypass of *Streptomyces coelicolor* A3(2): purification of isocitrate lyase and cloning of the *icl* gene.

by

PAUL G. CHAPMAN

Thesis submitted to the Faculty of Science, University of Glasgow, for the degree of Doctor of Philosophy.

Department of Genetics.

May 1994

The research reported in this thesis is my own original work except where otherwise stated and has not been submitted for any other degree.

Paul G. Chapman, May 1994.

Dedicated to my mum

CONTENTS

Page No.

List of contents i
List of figures vii
List of tables x
Abbreviations xi
Acknowledgements xii

SUMMARY xiii

CHAPTER 1: General Introduction

1.0 Introduction 2
1.1 Introduction to *Streptomyces* 2

1.2 Growth studies in *Streptomyces* 5
1.2.1 Glycerol utilisation in *Streptomyces coelicolor* 5
1.2.2 Galactose utilisation in *Streptomyces lividans* 7

1.3 The glyoxylate bypass 10

1.4 Fatty acid catabolism in *E.coli* 14
1.4.1 The FAO system 14
1.4.2 The ato system and short-chain fatty acid metabolism 15

1.5 Control of IDH in *E.coli* 18

1.6 IDH kinase/phosphatase 20

1.7 Isocitrate lyase and the *ace* operon 21
1.7.1 Icl's from various organisms 22
1.7.1.1 Activity of Icl 22
1.7.1.2 Important amino acids for activity 23

1.7.2 The *ace* operon 24
1.7.2.1 IclR and FadR 27

1.8 Aims of the project 31

CHAPTER 2: Materials and methods

2.0 Introduction 34

2.1 Bacterial strains, vectors and chemicals 34
2.1.1 Bacterial strains. 34
2.1.3 Plasmid and phage vectors. 34
2.1.4 Chemicals and biochemicals. 35
2.1.5 Enzymes, proteins and kits. 35

2.2 Standard media and microbiological techniques 36
2.2.1 Media used for growth of *S.coelicolor*. 36
2.2.1.1 Minimal Media. 36
2.2.1.2 Complex media. 36
2.2.2 Media for propagation of *E.coli*. 37
2.2.3 Growth of *Streptomyces* mycelia in liquid media. 37
2.2.4 Harvesting of mycelia. 38
2.2.5 Growth of *E.coli*. 38
2.2.8.1 Preparation of competent cells. 40
2.2.8.2 Transformation procedure. 40
2.2.8.3 Selection of pUC derived recombinant clones. 40
2.2.9 Introduction of plasmid DNA into *Streptomyces* 41
2.2.9.1 Preparation of protoplasts 41
2.2.9.2 Transformation of protoplasts 42
2.2.9.3 Regeneration of transformed protoplasts 42
2.2.10 Conjugation between *E.coli* and *Streptomyces* 42

2.3 General DNA methods. 43
2.3.1 Commonly used buffers. 43
2.3.2 Preparation of plasmid DNA. 44
2.3.2.1 Reagents for isolation of plasmid DNA from cells. 44
2.3.2.2 Large scale plasmid preparations 44

2.3.2.3 Small scale plasmid preparations 45
2.3.4 Organic solvent extraction. 46
2.3.5 Precipitation of DNA using ethanol or isopropanol. 46
2.3.6 Spectrophotometric measurement of nucleic acid. 46
2.3.7 Digestion of DNA with restriction enzymes. 46
2.3.9 Removal of the 5' phosphate from linearised DNA 47
2.3.10 Agarose gel electrophoresis. 47
2.3.11 Photography of agarose gels. 48
2.3.12 Recovery of DNA from agarose gels by electroelution. 48
2.3.13 Recovery of DNA from agarose using SPIN-X tubes 48
2.3.14 Techniques with E.coli bacteriophage λ 49
2.3.14.1 Preparation of plating bacteria for infection with bacteriophage λ. 49
2.3.14.2 Infection of bacteriophage l, plating and titre. 49
2.3.14.3 Isolation of phage particles from a plaque. 49
2.3.15.1 Transfection and plating of M13. 50
2.3.15.2 Isolation of plaques. 50
2.3.15.3 Preparation of M13 single stranded DNA. 50
2.3.16 Labelling of DNA with ^{32}P or Digoxigenin. 51
2.3.17 Southern blotting. 52
2.3.17.1 Reagents. 52
2.3.17.2 Procedure. 52
2.3.18 Hybridisation of oligonucleotides to filter bound nucleic acid. 53
2.3.18.1 Prehybridisation. 53
2.3.18.2 Hybridisation and washing. 53
2.3.19 Screening of plasmid clones by colony hybridisation 53
2.3.20 Screening of a phage λ library 54
2.3.21 Hybridisation and detection of DIG-labelled DNA hybrids 54
2.3.22 DNA Sequencing techniques. 55
2.3.22.1 Sequencing reactions. 55
2.3.23 Denaturing PAGE for DNA sequencing. 55
2.3.23.2 Preparation of glass plates and pouring the gel. 56
2.3.23.3 Electrophoresis of sequencing gels 56
2.3.24 Polymerase chain reaction 56

2.4 General protein methods 57
2.4.1 Preparation of crude protein extracts. 57
2.4.2 Buffers used in purification of ICL. 57
2.4.3 Chromatography media. 58

2.4.4 General biochemical methods. 58
2.4.5 Subunit Mr. 59
2.4.6 N-terminal sequence analysis by gas-phase sequencing 60
2.4.6.1 Tricine-SDS PAGE for proteins to be sequenced from blots 60
2.4.6.2 Electrophoresis conditions 61
2.4.6.3 Electroblotting 61
2.4.6.4 Amido black staining 62
2.4.7 Proteolytic digestion of ICL 62
2.4.8 Amino acid sequencing 62

CHAPTER 3: Attempts to clone *S.coelicolor icl* using primers designed against consensus sequences

3.1 Introduction 64

3.2 Results and discussion 64
3.2.1 Design of an oligonucleotide against conserved regions of the Icl proteins (oligo *icl*) 64
3.2.2 Probing of genomic DNA from *S.coelicolor* with oligo *icl* 66
3.2.3 Cloning of size selected *Pst* I digested genomic DNA from *S.coelicolor* 66
3.2.4 Restriction mapping of pAUL1 67
3.2.5 Exonuclease III deletions of pAUL1 70
3.2.6 Sequencing of the 1kb hybridising region 70
3.2.7 Amplification of *S.coelicolor* DNA using oligonucleotides designed against other *icl* gene sequences 73
3.2.8 N-Terminal sequence analysis of a purified protein, thought to be *S.coelicolor* icl 78

CHAPTER 4: Purification and characterisation of ICL from *S.coelicolor*

4.1 Introduction 82

4.2 Results and Discussion 83
4.2.1 Growth of S.coelicolor on Tween-containing compounds 83
4.2.2 Purification of ICL from *S. coelicolor* 83
4.2.2.1 Introduction 83

4.2.2.2 Purification procedure 84

4.3 Amino acid sequence determination 96
4.3.1 N-terminal sequence determination 96
4.3.2 Proteolytic digestion of the *S. coelicolor* ICL and sequencing of the polypeptide fragments 96
4.3.3 Comparison of sequences obtained to that of the other sequenced ICLs 98

4.4 Characterisation of the ICL from *S. coelicolor* 98
4.4.1 Km for Isocitrate 98
4.4.2 pH dependence on Vmax and Km 103
4.4.3 Effect of DTNB and Iodoacetate on the ICL activity 103

CHAPTER 5: Cloning of the *icl* gene from *S.coelicolor*

5.1 Introduction 106

5.2 Results and discussion 106
5.2.1 Design of oligonucleotides 106
5.2.2 PCR with genomic *S.coelicolor* DNA as template 110
5.2.3 Sequencing of the PCR product 110
5.2.4 Screening of an amplified λ library 112
5.2.5 Screening of genomic DNA 115
5.2.6 Purification of size selected hybridising DNA 117
5.2.7 PCR of size selected genomic *S.coelicolor* DNA as template 117
5.2.8 Construction of a plasmid sub library 117
5.2.9 Restriction mapping of pICL92 120

CHAPTER 6: DNA sequencing and analysis of *icl*

6.1 Introduction 125

6.2 Results and Discussion 126
6.2.1 Construction of sub-clones used for DNA sequencing 126
6.2.2 Sequencing strategy 126
6.2.3 Identification of Open Reading Frames ORF's 129

6.2.4 Translation of the *icl* gene from *S.coelicolor* 135
6.2.5 Database searching 135
6.2.6 Identification of putative ribosome binding sites RBS's upstream from *icl* and *ms* 137
6.2.7 Identification of possible promoter sequences 138
6.2.8 Identification of a stem-loop structure downstream of *icl* 141
6.2.9 Codon usage 143
6.2.10 Protein analysis 144
6.2.11 Search for consensus sequences 144
6.2.12 Alignment of the available ICL and MS enzyme sequences 147
6.2.13 Proposal of new signature motifs for ICL and MS 155
6.2.14 Comparison of the *S.coelicolor* ICL to other ICL's 155

CHAPTER 7: Overexpression and attempted disruption of the *S.coelicolor* IDH

7.1 Introduction 163
7.2 Results 164
7.2.1 Overexpression of *S.coelicolor* IDH in *S.lividans* 164
7.2.1.1 Construction of the sub-clone pCBSidh1 165
7.2.2 SDS-PAGE of crude extracts from *S.lividans* containing plasmids pidh486 and pCBSidh 167
7.2.3 Construction and assay of *S.coelicolor* 213 overexpressing IDH 167
7.2.4 Attempts at overexpression of *S.coelicolor* IDH in E.coli 169
7.2.4.1 The pT7-7 vector 169
7.2.4.2 Mutagenesis of the DNA encoding the N-terminus of IDH from *S.coelicolor* 170
7.2.5 Attempts to disrupt the genomic copy of the *S.coelicolor idh* 173
7.2.5.1 Construction of pIDH-erm E1 173
7.2.6 Conjugation of *E.coli* S17-1 with *S.coelicolor* 1147 174
7.3 Discussion 174

CHAPTER 8: General discussion

8.1 Introduction 178
8.2 Attempts to clone *S.coelicolor icl* using primers designed against consensus sequences. 178

8.3 Purification and characterisation of ICL from *S.coelicolor* 181

8.4 Cloning and sequencing of the icl gene from *S.coelicolor* 181
8.5 Overexpression and attempted disruption of the *S.coelicolor idh*. 182
8.6 Future work 182

REFERENCES 184

List of figures

CHAPTER 1

1.1 Life cycle of *Streptomyces* 4
1.2 Organisation of the *Streptomyces* glycerol and galactose operons 9
1.3 The TCA cycle and the glyoxylate bypass 12
1.4 Reactions catalysed by isocitrate lyase and malate synthase 13
1.5 The fatty acid oxidation system of *E.coli* 16
1.6 Reaction mechanisms of isocitrate lyase 26
1.7 The *ace* operon of *E.coli* 30

CHAPTER 3

3.1 Design of the oligonucleotide, oligo*icl* 65
3.2 Colony hybridisation 68
3.3 Restriction map of pAUL1 and determination of the oligo*icl* binding region 69
3.4 Exonuclease III deletion clones from pAUL1 71
3.5 Diagramatic representation of exo deletions 72
3.6 Sequence of the oligo binding region of pAUL1 73
3.7 PCR primers designed against regions of various ICL's which show a high degree of similarity 75
3.8 Amplification of genomic DNA from *S.coelicolor* using primers designed against an alignment of various ICL protein sequences 76
3.9 Sequences from the PCR products obtained using primers designed against other sequences 77

3.10 Partial purification of the putative ICL from *S.coelicolor*.....79
3.11 N-terminal sequence of the purified putative *S.coelicolor* ICL and alignment with the N-terminal regions of various lipoamide dehydrogenases.....80

CHAPTER 4

4.1 Phenyl sepharose hydrophobic interaction chromatography.....85
4.2 Mono Q anion exchange chromatography on the FPLC.....87
4.3 SDS PAGE of samples around the peak of ICL activity which eluted from the Mono Q column.....88
4.4 SDS PAGE of samples from the peak of ICL activity eluting from the phenyl sepharose column.....89
4.5 SDS PAGE of samples from stages involved in the purification of *S.coelicolor* ICL.....92
4.6 SDS PAGE calibration curve.....94
4.7 N-terminal amino acid sequence of *S.coelicolor* ICL.....97
4.8 N-terminal sequence of the 35kDa peptide.....99
4.9 N-terminal sequence of the 14kDa peptide.....100
4.10 Comparison of the N-terminal sequences of ICL's.....101
4.11 Alignment of internal regions of the ICL's.....102

CHAPTER 5

5.1 Design of oligonucleotide ICLNEW1 used in cloning of *icl*.....108
5.2 Design of oligonucleotide ICLNEW2 used in probing for *icl*.....109
5.3 Amplification of genomic *S.coelicolor* 1147 DNA using primers ICLNEW1 and ICLNEW2.....111
5.4 Sequencing of the PCR product using primers ICLNEW1 and ICLNEW2.....113
5.5 Comparisons with the peptide sequence deduced from the amplified PCR product with purified ICL from *S.coelicolor* and deduced sequence of ICL of *E.coli*.....114
5.6 Restriction digests and Southern hybridisation of genomic DNA of *S.coelicolor* 1147.....116
5.7 Size selection of DNA for analysis by PCR.....118
5.8 Analysis of size-selected genomic DNA of *S.coelicolor* 1147 by PCR.....119

5.9 Colony hybridisation....121
5.10 Restriction map of plasmid clones and subclones used in this study....122
5.11 Southern analysis of pICL92B....123

CHAPTER 6

6.1 Summary of the construction of M13 sub-clones used for sequencing of the *S.coelciolor icl*....127
6.2 Regions of sequences generated from sequencing of the various M13 subclones....130
6.3 DNA and amino acid sequences for the *icl* gene and surrounding regions....134
6.4 Analysis of the sequenced DNA for open reading frames - ORF's....135
6.5 Putative ribosome binding sites for *icl* and *ms*....139
6.6 Possible promoter sequences found upstream of *icl*....140
6.7 Putative terminator of the *icl* gene....142
6.8 Comparison of codon usage in the *icl* gene to that used by streptomycete genes in general....145
6.9 Alignment of the known ICL sequences....150
6.10 Alignment of the known MS sequences....154
6.11 Comparison of some ICL sequences using "DOTPLOTS"....159
6.12 Comparison of some MS sequences using "DOTPLOTS"....161

CHAPTER 7

7.1 Agarose gel to check which transformants contained inserts....166
7.2 SDS PAGE of crude extracts of *S.lividans* containing plasmids pIDH486 and pCBSidh1....168
7.3 Representation of the stages in the construction of an overexpressing copy of *S.coelicolor idh* in *E.coli*....171
7.4 PCR amplification of part of the *S.coelicolor idh* gene....172
7.5 Conceptual representation of a method to disrupt the *S.coelicolor idh* gene....175

CHAPTER 8

8.1 Binding of designed oligonucleotides to the *S.coelicolor icl* sequence....180

List of Tables

CHAPTER 2

2.1 Bacterial strains.....34

CHAPTER 4

4.1 ICL activity from *S.coelicolor* grown on glucose and Tween-containing compounds.....85
4.2 Purification table showing main steps of ICL purification.....93
4.3 Comparison of some characteristics of various ICL's.....95

Abbreviations.

AMP	ampicillin
ATP	adenosine triphosphate
BSA	bovine serum albumin
(k)Da	(kilo) daltons
DMSO	dimethylsulphoxide
DNAse	deoxyribonuclease
DTNB	5,5'- dithio- bis (2- nitrobenzoic acid)
DTT	dithiothreitol
EDTA	ethylene diamine tetra acetate
EtBr	ethidium bromide
FPLC	fast protein liquid chromatography
HPLC	high pressure liquid chromatography
ICL	Isocitrate lyase (EC 4.1.3.1)
IDH	Isocitrate dehydrogenase (EC 1.1.1.42)
IPTG	isopropyl-β-D-thiogalactoside
kb	kilo base pairs
MS	Malate synthase (EC 4.1.3.2)
MES	2-(N-Morpholino) ethanesulphonic acid
MOPS	Morpholino propane sulphonic acid.
M_r	molecular weight
NAD^+	nicotinamide adenine dinucleotide
$NADP^+$	nicotinamide adenine dinucleotide phosphate
NADPH	reduced nicotinamide adenine dinucleotide phosphate
OAA	Oxaloacetate.
PAGE	polyacrylamide gel electrophoresis
PEP	phospho*enol* pyruvate
3-PG	3-phosphoglycerate
pfu	plaque forming units
PMSF	phenylmethanesulphonylfluoride
R_f	Mobility of a protein compared to the dye front
RNAse	ribonuclease
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
TAE	Tris-acetate/EDTA buffer

TBE	Tris-borate/EDTA buffer
TEMED	N,N,N'N'-tetramethylethylene diamine
TES	N-tris (Hydroxymethyl) methyl-2-aminoethane sulphonic acid
Tris	Tris (hydroxymethyl) aminomethane
U	units of enzyme activity
UV	ultra violet
X-gal	5-bromo-4-chloro-3-indoyl-β-galactoside

Acknowledgements

Many thanks to my supervisor, Iain Hunter, for his helpful advice during the lab work of this project and an extra special thank you for the proof reading of this thesis.

A special thank you to "Auntie" Marion Stone for her duties in keeping the lab running like a well oiled vacuum pump and finally getting Charlie and Phil to wear their lab coats. My gratitude also goes out to the prep-room ladies for their excellent work.

Thanks to everyone in the lab (past and present), especially to Kenny McDowell, Richard, Charlie, Kenny Linton, and Jo, my drinking partners overs the years. A further hello to my fellow footballers and cricketers.

My appreciation goes to Javier Rua for showing me how run protein columns. Also a special thank you to Bryan Dumbar for his amino acid sequencing.

Cheers, to Linda, who will be glad to get her computer back to herself once more. Well, at least until Phil starts writing up.

Thanks to Colin Smith for his inspiration !

This work was supported by a grant from the SERC and the writing up was supported by my wages from Cruikshank and Fairweather.

Finally, my extra special thanks to Gill and Phil. Two friends who have kept me amused, fed, drunk, through the highs and the lows. My best wishes go out to them both.

Finally, finally, if at first you don't succeed.....

Summary

This thesis is concerned with the glyoxylate bypass of *Streptomyces coelicolor* A3(2) and in particular the enzyme isocitrate lyase (ICL) and its corresponding gene (*icl*).

At the start of this project oligonucleotides were designed against regions of other ICL proteins which were found to be highly conserved across a variety of species. These oligonucleotides were used for Southern analysis of genomic DNA from *S.coelicolor* and also as primers for use in PCR amplification of genomic DNA from *S.coelicolor.* Sequencing of DNA which was cloned using these techniques revealed that the *S.coelicolor icl* gene had not been cloned.

In order to facilitate in cloning of the *S.coelicolor icl* gene, it was decided to purify the *S.coelicolor* ICL protein. Acetate was found to be an extremely poor carbon source, resulting in poor growth and low ICL activity. Growth on Tween (polyoxyethylensorbitan) containing compounds was found to result in high activities of the ICL enzyme and purification of the enzyme was facilitated using a five step protocol employing hydrophobic chromatography and anion exchange chromatography. Amino acid sequence from the purified protein and also from chymotrypsin digested peptides was obtained. This amino acid sequence was then used to design oligonucleotides, which were used as probes to facilitate in cloning the *S.coelicolor icl* gene.

An amplified fragment of DNA from *S.coelicolor* was obtained using the oligonucleotides and sequencing of this DNA revealed that it encoded part of the *icl* gene. The amplified DNA was then used in Southern analysis in order to clone the entire *S.coelicolor icl* gene. Sequencing of the gene was carried out and the sequence of the regions surrounding the gene were also determined. The *icl* gene sequence shows that it is similar to the previously sequenced *icl's* and in particular to the *E.coli icl.*, notably lacking the similar internal region as the *E.coli* ICL protein. Sequencing downstream from the *icl* has revealed the *ms* gene from *S.coelicolor* This is the opposite to *E.coli,* where the gene order is *ms/icl.* It remains to be determined if these genes form part of an operon in S.coelicolor.

Some work on the *S.coelicolor idh* gene was carried out and 30-fold overexpression of the *idh* gene in *S.coelicolor* was achieved. Preliminary experiments to overexpress *idh* in *Escherichia coli* and an attempt to disrupt the *S.coelicolor idh* were also carried out, but had still to be completed by the end of this project.

CHAPTER 1

General Introduction

1.0 Introduction

This chapter provides a background to some of the topics discussed in this thesis. A very general introduction to *Streptomyces* and their importance is presented first (section 1.1). This is followed by a discussion of carbon catabolism in *Streptomyces*, in particular carbon catabolic operons.

Little work has been done on the genetics of growth of *Streptomyces* on fatty acids or acetate, and an objective of this project was to study an enzyme central to this, namely isocitrate lyase.

A great deal of work on the genetics of fatty acid and acetate catabolism has been carried out using *Escherichia coli* and some of this is presented in this introduction. A discussion of the enzymes concerned with the control of carbon flux between the TCA cycle and the glyoxylate bypass is also presented.

The final section of this chapter describes the initial aims of this project and the strategies used to achieve them.

1.1 Introduction to *Streptomyces*

Streptomyces are filamentous, chemoheterotrophic, aerobic, spore-forming, gram-positive eubacteria. They possess a complex life cycle of morphological and physiological development and are adapted to survive in relatively hostile environments. Their characteristic DNA base composition has a high G+C bias of between 69-73% and they have genomes of between 6-9 megabases of DNA, based on pulsed-field gel electrophoresis data (Woese, 1987; Hopwood and Kieser, 1990).

The morphology, growth and formation of metabolites from *Streptomyces* under certain conditions bear resemblance to soil fungi and it may be more pertinent to compare *Streptomyces* to soil fungi rather than to bacteria such as *E.coli*. However *Streptomyces* are most certainly bacteria and not fungi.

It does appear that certain features of *Streptomyces* are very similar to those found in soil fungi, in particular their habitat and growth cycle. At the protein level there are also similarities: for example, the *Streptomyces coelicolor* dehydroquinase gene is highly similar at the amino acid level and in functionality to the *Neurospora crassa* dehydroquinase and may reflect convergent evolution or gene transfer (White *et al*, 1990).

It also appears from amino acid sequence data, that two antibiotic pathway enzymes have been transferred horizontally from *Streptomyces* to fungi (Miller and Ingolia, 1989a&b).

Typically, when authors describe the life cycle of *Streptomyces*, they begin with spore germination. However it should be noted that this is just a convenient place with which to start and that in a growing streptomycete colony there may be many stages of growth taking place at once.

Usually following an enviromental stimulus, spore germination and outgrowth occurs. This is followed by apical extension and branching of vegetative mycelia, to enable colonisation of the solid and particulate substratum. Hybrophobic aerial mycelia form, which subsequently differentiate to produce long chains of spores (Fig. 1.1). The formation of the aerial mycelia is typically associated with antibiotic production. The timing of antibiotic production and its association with sporulating organisms, suggest that an important role of antibiotics may be concerned either with the regulation of differentiation or with protection of the producer at stages of its development when it is vulnerable to invasion by competing microbes. However, it has been found that antibiotic production is not usually an essential part of the differentiation process.

Streptomyces produce over 70% of all known antibiotics (Berdy, 1980); this includes 80% of all commercially produced antibiotics. Indeed much of the work concerned with the species has centred around the production of these and other secondary metabolities, which have such diverse functionalities as anti-cancer agents, immunosuppressants and anti-helminthics. Each species may produce more than one secondary metabolite. For example *S.coelicolor* is known to produce four different antibiotics (Hopwood and Wright, 1983), whereas some species (eg. *S.clavuligerus*) produce over thirty structures.

Industrially *Streptomyces* are grown on either glucose, from sugar beet or cane, or from a wide variety of oils (fatty acids) such as soya or rape seed, depending on the most economical nutrient source at the time. Carbon from these various feed stocks will enter central metabolism at different levels and therefore will probably require different pathways.

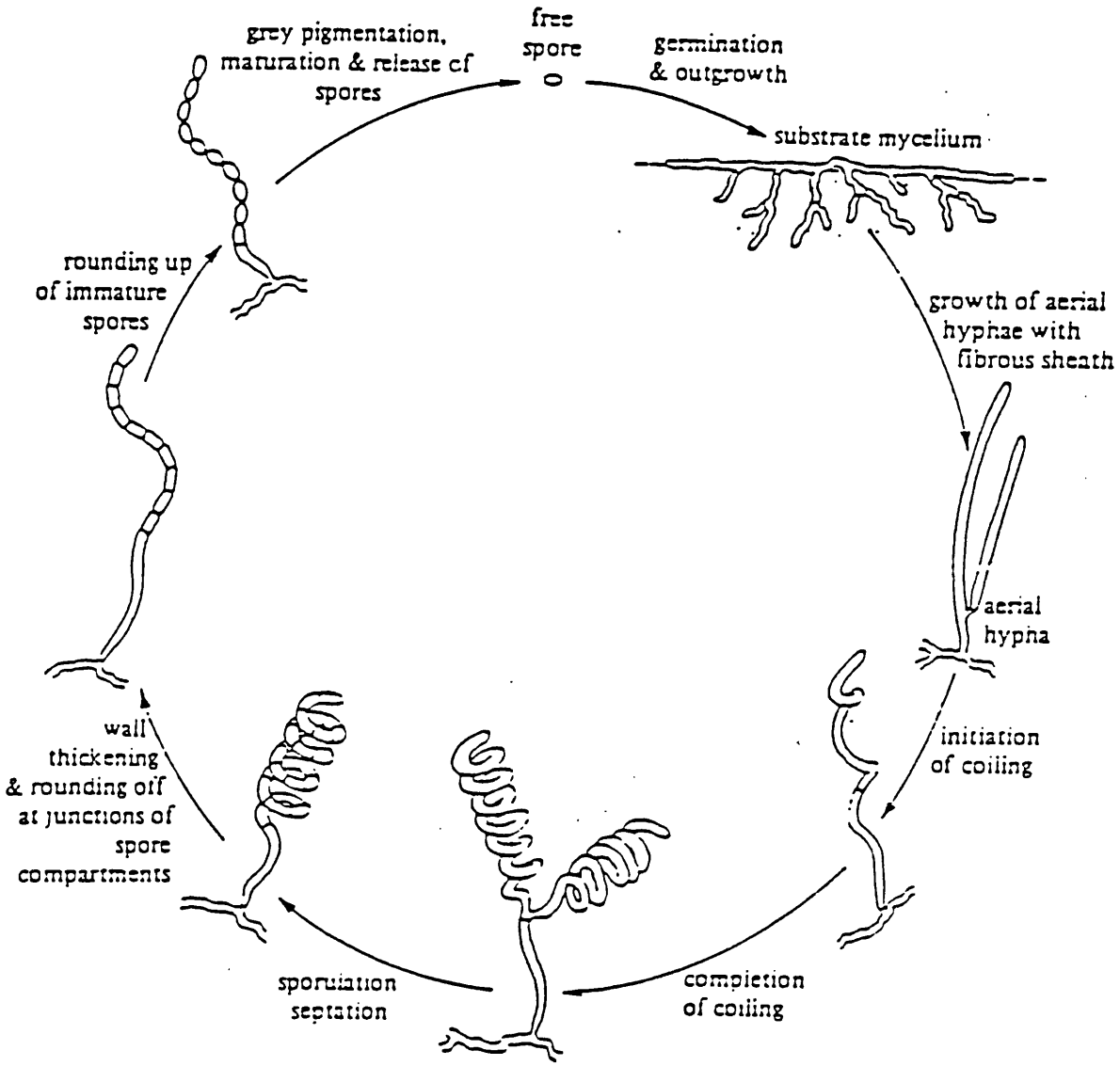


Figure 1.1 Life cycle of streptomycetes.

This idealised diagram summarises the life cycle of *S.coelicolor* from germination to sporulation. Modified from Chater and Merrick, 1979.

1.2 Growth studies in *Streptomyces*

Most early work on metabolic regulation had been concerned with the control of antibiotic biosynthesis and in many cases there was evidence of glucose repression (Martin and Demain, 1980). However, up until then very little work had been carried out to elucidate the regulation of primary carbon source uptake and utilisation.

Hodgson (1982), observed that glucose repressed a number of uptake mechanisms for other carbon sources. He found that glucose repressed the uptake of glycerol and galactose. Further studies found that there were two glycerol uptake mechanisms, one which was an active transport mechanism and was glycerol inducible and glucose repressible. The other appeared to be glucose-inhibited and probably operated using a facilitated diffusion process. The galactose uptake system appeared to be an active transport mechanism, which did not require induction by galactose.

In *E.coli*, expression of many carbon catabolic pathways is controlled in part by cAMP and its associated protein cAMP receptor protein (CRP). The cell responds to the availability of glucose by regulating the intracellular concentration of cAMP, which in concert with CRP then affects the expression of many catabolic pathways. In *Streptomyces*, cAMP has been observed, but the intracellular concentration of cAMP does not change in response to a repressing carbon source (Hodgson, 1982). Thus it appears that cAMP does not have the same role in *Streptomyces* as it does in *E.coli*.

The systems for the utilisation of galactose and glycerol in *Streptomyces* have generated much interest and a great deal of work has been carried out to elucidate the enzymes necessary for their catabolism and how their respective genes are controlled.

1.2.1 Glycerol utilisation in *Streptomyces coelicolor*

Glycerol utilisation has been studied in *S.coelicolor* and two enzymes have been found to be essential for growth on glycerol as a sole carbon source, a soluble glycerol kinase and an apparently membrane-associated L-glycerol-3-phosphate dehydrogenase (Seno and Chater, 1983). These enzymes were found to be induced during growth on glycerol and repressed during growth on glucose. Seno and Chater (1983) identified mutants (*gyl*$^-$) which lacked glycerol kinase and G3P dehydrogenase activities. It was suggested that this was either as a result of a mutation in a regulatory region, or a mutation in the glycerol kinase gene.

Seno *et al* (1984) mapped all the *gyl* mutants to a small region of the *S.coelicolor* genome, near to the *arg*A locus and showed that a small 3kb fragment of DNA could restore many of the *gyl* mutations. Further studies were carried out on the small fragment of DNA and complementation analysis and insertional mutagenesis concluded that the genes glycerol kinase (*gyl*A) and G3P dehydrogenase (*gyl*B) were part of an operon which was transcribed from left to right, with *gyl*A transcribed before *gyl*B.

This work did show that the entire operon had not been cloned and it was the work of Smith and Chater (1988a,b) which led to the cloning of the entire glycerol operon. It was shown that the *gyl* operon was contained on a 5.4kb fragment of DNA. The organisation of the operon is *gyl*ABX. *gyl*X had not been shown by Seno *et al* (1984) to be necessary for growth on glycerol, so it was thought that *gyl*X may be a glycerol facilitator protein for glycerol uptake, which would not be essential when cultures were grown in high levels of glycerol, as glycerol would enter the cells by passive diffusion.

Northern analysis showed that the *gyl* operon was transcribed by two discrete mRNA species of 5.4knt and 4.3knt, which was probably due to transcription termination between *gyl*B and *gyl*X. A smear of lower molecular weight mRNA was also observed. It was also shown that *gyl* mRNA was not detected in cultures grown on arabinose or glucose. The 4.3knt fragment identified would contain just enough information to encode the two inducible polypeptides of 101kDa and 51kDa S1 nuclease mapping studies helped to further identify the extent of the *gyl* transcripts. These studies identified a 0.9knt transcript, upstream from *gyl*A. This transcript was detected in glycerol-grown cultures, but not in glucose-grown ones. The S1 studies also showed that the low molecular weight smear observed from the Northerns, was due to degradation of 2.9knt transcripts originating from the 5' end of the *gyl* operon and a terminator site after the *gyl*A sequence.

Biro and Chater (1987) cloned the *gyl* operons from *S.griseus* and *S.lividans* and also found that they contained *gyl*A and *gyl*B genes, as well as the 0.9kb upstream sequence.

Smith and Chater (1988) sequenced ~1500bp of the upstream region of the *gyl*ABX to identify possible promoters and also the *gyl*R gene which was transcribed as the 0.9knt transcript and which encoded a polypeptide of 254 amino acids (Mr = 27640).

Immediately downstream of *gylR*, a number of clustered inverted repeats were identified which coincided with the 3' end of the *gylR* transcript. The calculated free energy of formation of the larger stem-loop was -113kJ/mol, which is comparable to that of rho-independent transcriptional terminators.

Analysis of the *gylR* revealed its similarity to the transcriptional activator *AsnC* (de Wind *et al*, 1985) and in particular the DNA binding domain. It was observed that the *gylR* sequence contained the general criteria for forming a α-helix-turn-helix structure that is found in many DNA binding proteins (Pabo and Sauer, 1984). These results were in accord with the proposal that *gylR* is the transcriptional regulator of the *gylABX* operon (Smith and Chater, 1988b).

High resolution S1 mapping upstream of *gylABX* identified two distinct promoters which were 50bp apart. Further studies clearly showed that transcription from the two promoters was induced by growth on glycerol, while repressed during growth on glucose. Studies on the *gylR* promoter also revealed that this promoter is induced by glycerol and repressed during growth on glucose. All of the promoter regions identified show similarity to the general consensus promoter of *E.coli* (Harley and McClure, 1983). A comparison of the *gylR* and *gylABX* promoter regions identified a similar 17bp palindromic sequence which resembles typical bacterial operator elements. Another region of dyad symmetry was observed in the *gylABX* promoter region, which resembled several prokaryotic *cis*-acting elements that bind regulatory proteins.

A representation of the *gyl* operon is shown in Fig. 1.2a

1.2.2 Galactose utilisation in *Streptomyces lividans*

Studies on *E.coli* had revealed that the enzymes required for galactose utilisation were part of a metabolic operon (Shapiro and Adhya, 1969; Lemarie and Miller, 1986). Growth studies of *Streptomyces* also revealed that glucose represses galactose uptake (Hodgson, 1982).

Work was therefore carried out to try to isolate the genes responsible for galactose utilisation. The approach taken was to attempt to complement an *E.coli* galactose kinase mutant, using a *S.lividans* expression library. Initially a *S.lividans galK* clone was identified, which complemented an *E.coli galK* mutant (Adams *et al*, 1988). Further studies revealed that the DNA cloned from *S.lividans* could complement an *E.coli* mutant for which the entire *gal* operon had been deleted.

The gene organisation and structure of the *S.lividans gal* genes showed that they were transcribed in the order *galT*, *galE* and *galK* and also that they were transcribed in the same direction. This was unexepected as the gene order of the *E.coli* operon is *galE*, *galT* and *galK*.

Translational coupling is observed in the *E.coli* operon, with only 3 nucleotides separating *galT* from *galK*, which helps to ensure coordinate expression (Schumperli *et al*, 1982). This does not appear to be the case in *S.lividans*, as there are quite long intercistronic regions between the genes.

Fornwald *et al* (1987) studied the transcription from the *gal* operon of *S.lividans* and confirmed that the galactose genes are part of a polycistronic operon. Interestingly two promoters of the operon were identified. One promoter, *galP1*, appeared to be induced by galactose and results in transcription of all three *gal* genes, while the other, *galP2*, was totally separate and resulted in constitutive expression of the *galE* and *galK* genes. This helped to answer why significant levels of galactokinase were observed even in the absence of induction by galactose.

It was unknown why *S.lividans* should express galactokinase in the absence of galactose. However Adams *et al* (unpublished), have evidence to show that galactokinase may be involved in the induction of the *gal* genes.

The promoter organisation and the gene structure of the *S.lividans* operon did appear quite different to that of *E.coli*. However in both organisms, a constitutive promoter was found upstream of the *galE* gene. It is known that in *E.coli* the product of the *galE* gene, UDP galactose 4-epimerase, is required for cell wall biosynthesis, regardless of galactose being the sole carbon source or not. This may also be the case for *S.lividans* and would explain the function of constitutive promoter *galP2*.

Studies of the sequences of the two *gal* promoters showed that there was little similarity between them, which suggests that different factors, such as different forms of holo RNA polymerases, may be responsible for transcription from these promoters.

The entire nucleotide sequence of the *S.lividans gal* operon has been determined (Adams *et al*, 1988), and comparison of the deduced amino acid sequences to those of the other gene products of organisms has been carried out. All of the gene products of the *S.lividans gal* operon show similarity to regions of products of the *E.coli* and *Saccharomyces carlbergensis gal* genes

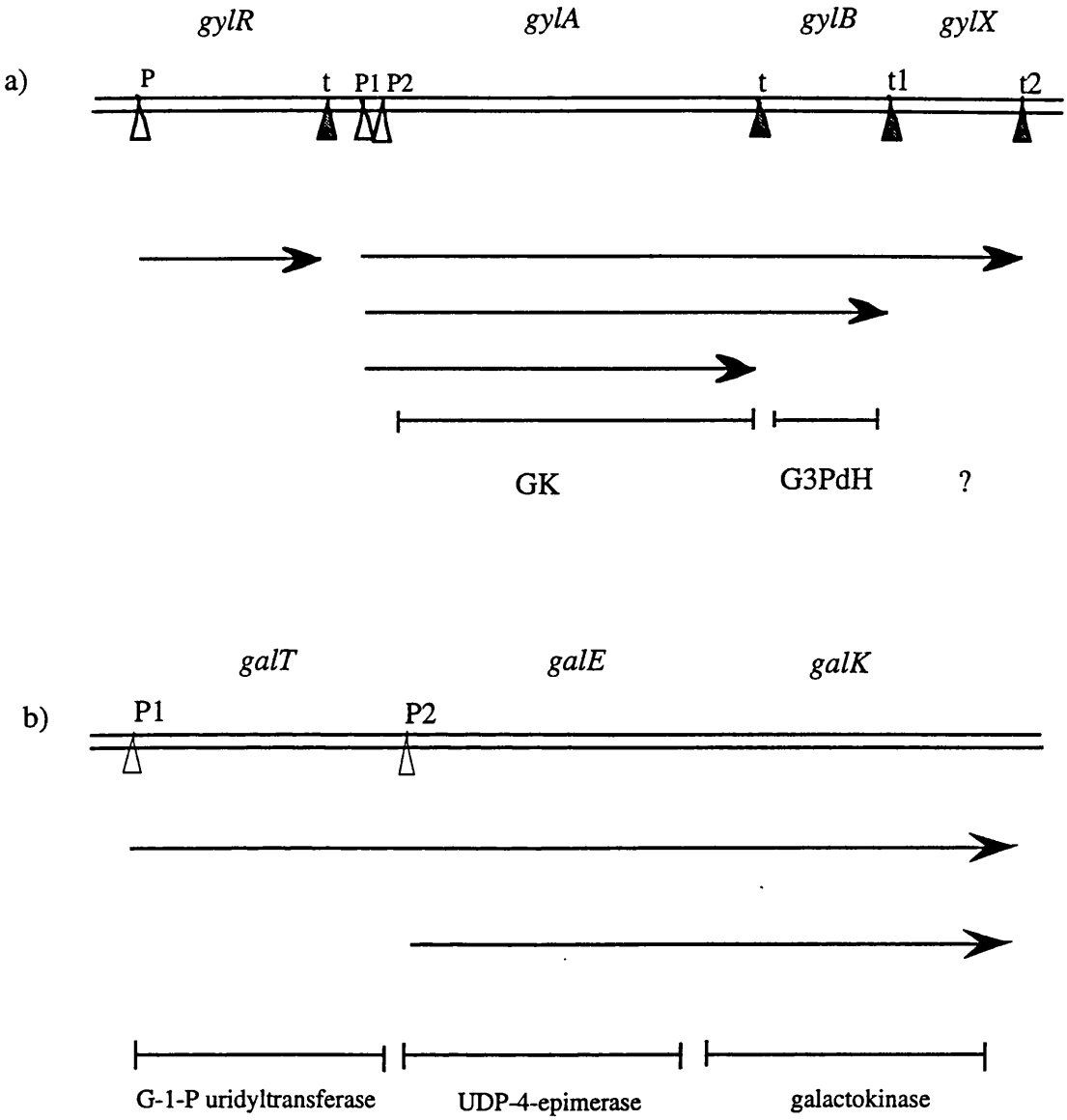


Figure 1.2 Organisation of the Streptomyces glycerol and galactose operons

a) Diagrammatic representation of the operon for glycerol utilisation from *S.coelicolor*. The two structural genes, *gylA* and *gylB* are shown, as well as the repressor, *gylR*. The transcripts obtained from the promoters (P, P1, P2) are also shown. The terminators are represented by t,t1 and t2

b) Diagrammatic representation of the operon for galactose utilisation from *S.lividans*. The three structural genes, *galT, galE* and *galK* are shown. The transcripts from the constitutive promoter P2 and the galactose inducible promoter, P1, are shown.

Since the gene products of the *S.lividans gal* operon can complement an *E.coli gal* mutant, it has been suggested that some of the regions of similarity between the amino acid sequences could be substrate binding sites. Indeed there appears to be a conserved ATP-binding site found at the carboxy terminus of the galactokinase gene.

Recent work by Mattern *et al* (1993), has involved mutagenising upstream sequences which are thought to be involved in regulation of transcription. Inspection of the sequence upstream from *gal*P1 revealed the presence of two overlapping potential regulatory motifs and a series of six hexamers, which have the sequence TNTNAN and have been suggested by Ebright (1986), to interact with the helix-turn-helix motif of many DNA binding proteins. Mattern *et al* (1993) disrupted 4 of the hexamers, by using oligonucleotide directed mutagenesis. The mutations in some of these sites resulted in a massive increase in the expression from the *gal*P1 promoter and suggested that these DNA sequences may be a binding site for a negative regulator of transcription. This hypothesis has been supported by the fact that multiple copies of the wild-type *gal*P1 region induce expression, as the multiple copies presumably titrate out the regulator and allow overexpression of the operon.

This work helped show that the control of the *S.coelicolor* and *S.lividans gal* operons are quite different from that of the *E.coli* operon. The *E.coli* gal repressor binds to two sites of dyad symmetry and modulates transcription of both of the *gal* promoters, which are next to each other. In *S.coelicolor* the two promoters *gal*P1 and *gal*P2 are separated by the *gal*T gene and regulation is essentially separate. The *gal*P1 promoter also appears to be controlled in part by *cis*-acting elements. It was also observed that a 20-fold increase in the concentration of *gal*P1 promoter sequences resulted in only a 2-fold increase in expression of *gal*P1.

A representation of the galactose operon can be seen in Fig. 1.2b

1.3 The glyoxylate bypass

As mentioned previously, *Streptomyces* are often grown on fatty acids as a sole carbon source. Studies on a great many organisms have shown that growth on fatty acids or acetate, yields acetyl-CoA which can then be metabolised by the TCA cycle.

However if the TCA cycle was to operate by catabolising acetyl-CoA, both of the carbons of the acetyl unit would be lost as CO_2 and alone would not allow net accumulation of carbon. Therefore growth on fatty acids or acetate, which are catabolised to acetyl-CoA, requires the operation of a separate anaplerotic pathway, namely the glyoxylate bypass. The glyoxylate bypass diverts carbon away from the two CO_2 evolving steps of the TCA cycle (Fig. 1.3). There are two unique enzymes of the glyoxylate bypass, isocitrate lyase (ICL) and malate synthase A (MS) and these enzymes are induced when *E.coli* is grown on fatty acids or acetate as the sole carbon source (Kornberg, 1966).

ICL was first discovered in sonic extracts of *Pseudomonas aeruginosa* and was found to catalyse the reversible formation of succinate and glyoxylate on the addition of either citrate or cis-aconitate (Campbell *et al*, 1953). This result was confirmed using fungi and other bacteria and it was established that Ds-(+) isocitrate was the true substrate (Olson, 1954; Saz, 1954), Fig. 1.4.

The second enzyme of the pathway, MS was discovered by Wong and Ajl (1956) and was shown to catalyse the condensation of glyoxylate with acetyl-CoA, to form malate (Fig. 1.4).

The overall reaction of ICL and MS is the reaction:

acetyl-CoA + isocitrate + H_2O = succinate + malate + CoA.

This gives a net gain of one C_4-dicarboxylic acid which can be used by the TCA cycle to produce oxaloacetate (OAA). OAA can then be used to supply precursors required for biosynthesis. OAA can be converted to phosphoenolpyruvate (PEP) using PEP carboxykinase. The PEP can then be used to generate the phosphorylated biosynthetic precursors.

Other biosynthetic precursors are required which are generated from 2-oxoglutarate and succinyl-CoA (reducing power is also required), all of which are generated from the TCA cycle. This means that some of the carbon must be channelled through IDH and not ICL. This creates a branchpoint where IDH and ICL compete for the same substrate, isocitrate. It is because of this in *E.coli*, that tight regulation of ICL and IDH activities takes place, to ensure efficient use of the available acetyl-CoA.

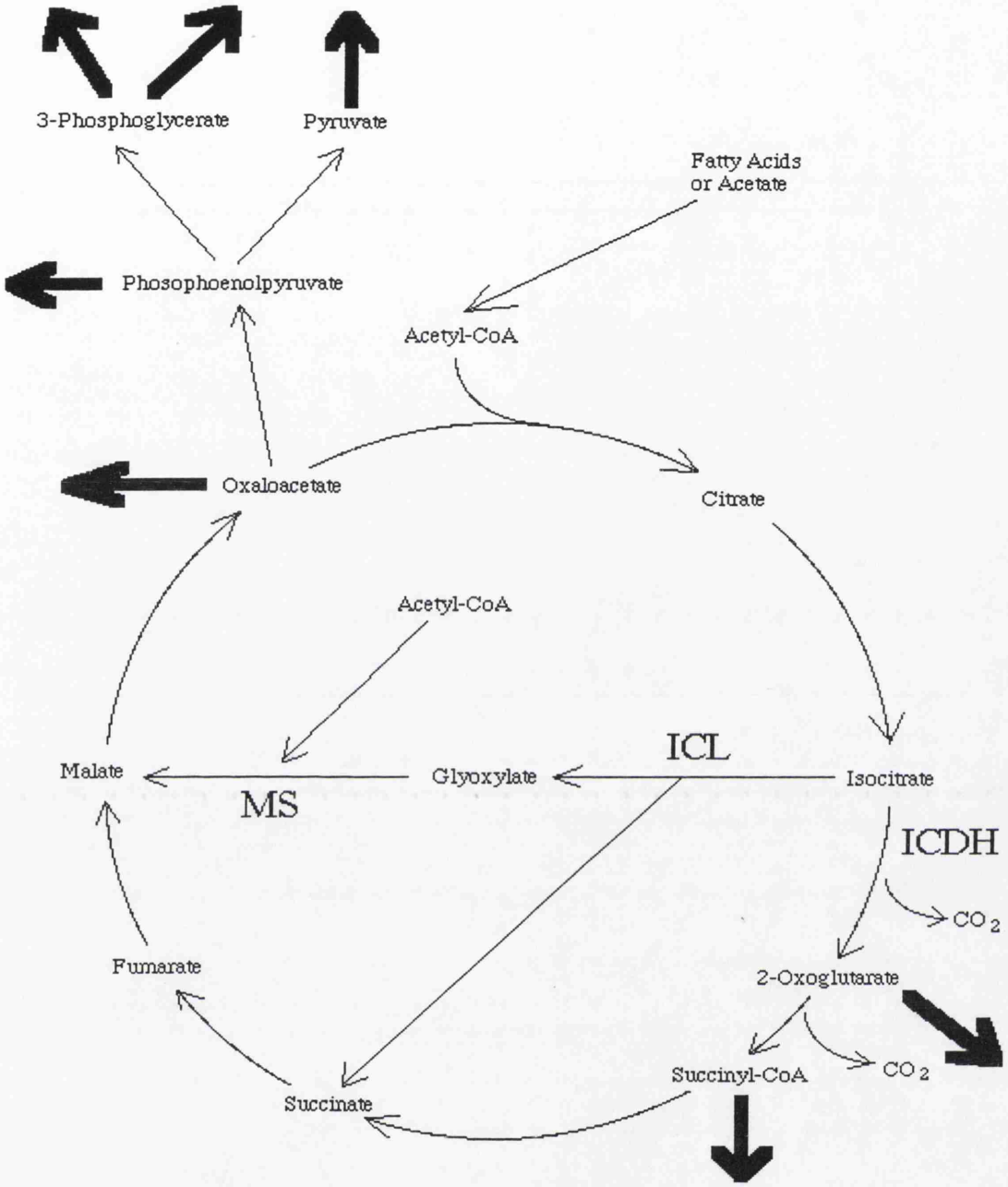


Figure 1.3 The TCA cycle and the glyoxylate bypass

Diagrammatic representation of the reactions of the TCA cycle and those of the glyoxylate bypass.
The heavy arrows indicate fluxes to biosynthesis.

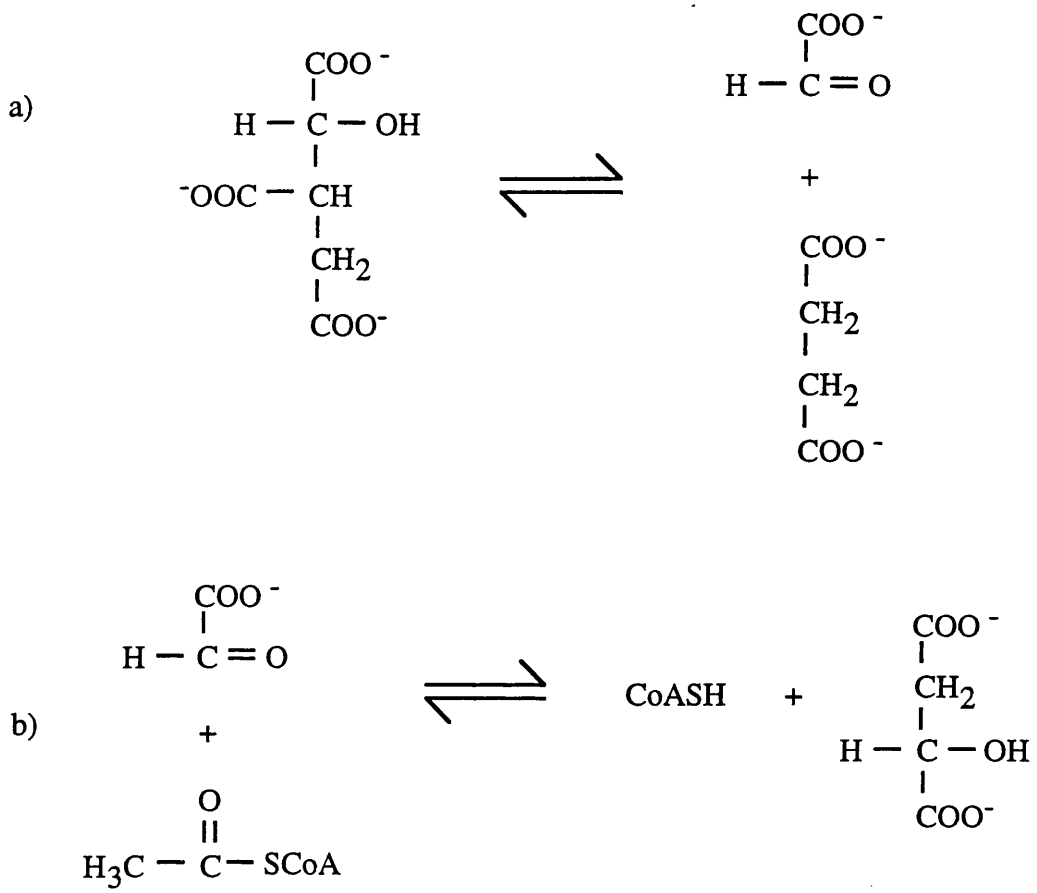


Figure 1.4 Reactions catalysed by Isocitrate lyase and Malate synthase

a) Isocitrate lyase catalyses the reversible cleavage of isocitrate to give succinate and glyoxylate.

b) Malate synthase catalyses the condensation of glyoxylate with acetyl-CoA to give malate and CoASH.

1.4 Fatty acid catabolism in *E.coli*

Many bacteria can grow on fatty acids or acetate as their sole carbon source, but the greatest amount of work has been carried out on *E.coli* and the systems it employs to transport the fatty acids or acetate and metabolise them to acetyl-CoA.

There are principally two different systems which *E.coli* uses to transport and catabolise fatty acids. The first system is concerned with the transportation and catabolism of medium chain fatty acids (C_7-C_{11}) and long chain fatty acids (C>12) and is called the fatty acid oxidative system (FAO). The second uses, in addition to the FAO system, further enzymes and is concerned with the transportation and catabolism of short-chain fatty acids (C_4-C_6) and is termed the ATO system.

In either system the object is to produce acetyl-CoenzymeA (CoA). The acetyl-CoA produced is then further metabolised to yield energy and precursors for biosynthesis.

1.4.1 The FAO system

Firstly the fatty acids must enter the cell through the cell membrane. Fatty acids are hydrophobic compounds and it is generally thought the outer membrane of *E.coli* is impermeable to hydrophobic compounds (Nikaido and Vaara, 1985). Therefore it was postulated that a mechanism for transporting long-chain fatty acids must be present.

Black *et al* (1987) purified the *fad*L gene product, an outer membrane-bound protein (FLP), which was shown to be involved in the transport of long-chain fatty acids. Previous reports had shown that FLP was an inner-membrane component, but this was thought to be due to cross-contamination of the membrane fractions.

Another protein had been shown to be involved in the transport of fatty acids - the *fad*D gene product acyl-CoA synthetase (Kameda and Nunn 1981). A proposed model for transport suggests that long-chain fatty acids first bind to FLP and are then transported across the outer membrane, possibly by FLP acting as a permease, to the inner membrane. Here FLP interacts with the inner-membrane bound acyl-CoA synthetase, where the fatty acids are activated and released into the cytoplasm.

Medium-chain fatty acids are thought to be transported either by FLP, or by diffusion.

Once inside the cell, fatty acids are degraded in a similar manner to the β-oxidative pathways of eukaryotic organisms. The first step is the activation to fatty-acyl-CoA, catalysed by the membrane bound acyl-CoA synthetase. Two oxidation steps follow which yield one molecule each of $FADH_2$ and NADH. This is then followed by thiolytic cleavage to give a fatty-acyl-CoA which is two carbons shorter than the original fatty-acyl-CoA and an acetyl-CoA. The shortened fatty acid then reenters the cycle and repeats the sequence (Overath *et al,* 1969, see Fig. 1.5).

The genes responsible for fatty acid degradation have been mapped to different regions of the *E.coli* linkage map. However the genes for *fad*A and *fad*B have been cloned and are found to be transcribed together in the direction of *fad*A to *fad*B (Spratt *et al*, 1984). The *fad*A gene was found to encode a 42kDa protein and the *fad*B gene, a 78kDa protein. These two gene products in fact form a multienzyme complex which catalyses five enzyme steps. The multienzyme complex has been purified and its structure identified (Binstock et al, 1977; Pawar and Schultz, 1981; Pramanik *et al* 1979). The complex was found to constitute a $\alpha_2\beta_2$ structure, where α=78kDa *fad*B gene product and β= 42kDa *fad*A gene product. These studies also indicated that the *fad*B product contained both isomerase and epimerase activities, which would allow for growth on unsaturated fatty acids. Another gene involved in fatty acid degradation is the *fad*R gene, which controls both the negative regulation of fatty acid degradation and positively regulates fatty acid synthesis.

1.4.2 The ato system and short-chain fatty acid metabolism

The transport of short-chain fatty acids is different from the FAO system. Work carried out by Frerman (1973) showed a link between acetoacetate-CoA transferase and short-chain fatty acid transport. Frerman (1973) showed that acetoacetate-CoA transferase was associated with membranes and that short-chain fatty acid uptake was inhibited by butyryl-CoA and acetate, which are the products of acetoacetate-CoA transferase.

Acetoacetate-CoA transferase is the product of the *ato*D/*ato*A genes, but mutations in *ato*B and *fad*AB genes also reduce short-chain fatty acid uptake. Short-chain fatty acid transport may be more complex than the simple interpretation of acetoacetate-CoA transferase, controlling transport.

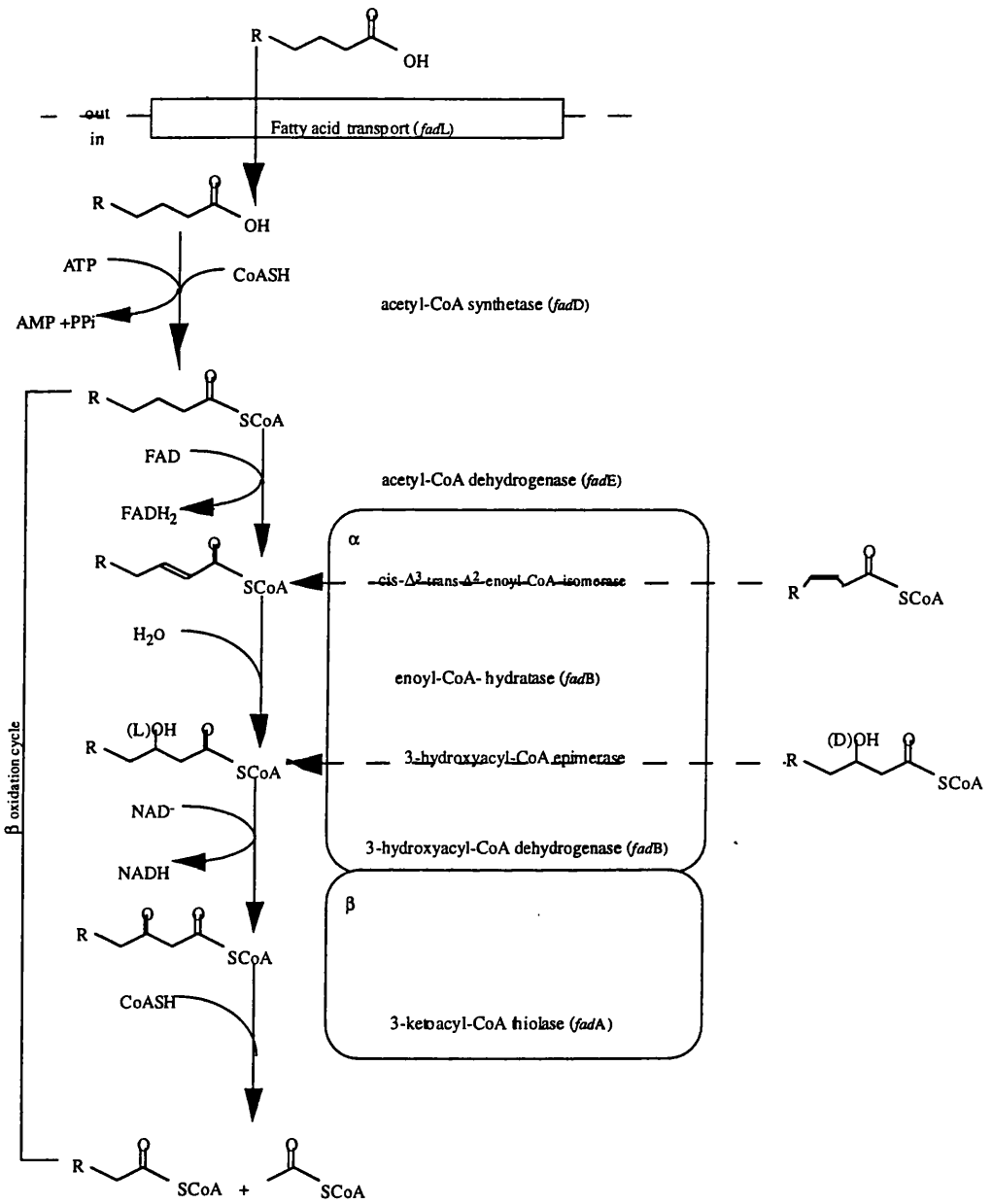


Figure 1.5 The Fatty oxidation system of *E.coli*

Cyclic pathway of fatty acid degradation. Principle enzymes of the pathway are shown on the right. Acetyl-CoA is further metabolised in the TCAcycle (Redrawn from Nunn, 1986)

The other genes involved in short-chain fatty acid catabolism are found to map together and a 6.2kb region of DNA was able to complement mutations in the *ato* genes (Jenkins and Nunn, 1987a). The structural genes are *ato*D/*ato*A encoding acetyl-CoA:acetoacetate-CoA transferase and *ato*B encoding thiolase II. Another *ato* gene, *ato*C, is present in this locus, which encodes a regulatory protein.

Jenkins and Nunn (1987a), identified the gene products to be *ato*C= 48kDa; *ato*D= 26.5kDa; *ato*A= 26kDa and *ato*B= 42kDa. It appears that the *ato*D and *ato*A gene products combine to form a tetrameric protein, with an $\alpha_2\beta_2$ structure. The studies also suggested that the *ato*DAB locus is transcribed as an operon. Further studies (Jenkins and Nunn, 1987b), showed that the *ato*C gene product acts as a transcriptional activator of the *ato*DAB genes, in the presence of acetoacetate.

For acetoacetate to be metabolised by *E.coli*, it is degraded to acetyl-CoA in a two step reaction, in which acetoacetate is first activated to acetoacetate-CoA by acetyl-CoA:acetoacetate-CoA transferase and then cleavage by thiolase II to yield acetyl-CoA.

Saturated short-chain fatty acids cannot be utilised by wild-type *E.coli*, as the sole carbon source, as they also require the products of *fad*E and *fad*B, which are not induced by short-chain fatty acids.

E.coli is also able to grow on acetate as a sole carbon source. It is commonly thought that acetate enters the cell by passive diffusion before it is metabolised (Kell et al, 1981; Baronofsky *et al*, 1984). Additionally however work by Sheu *et al* (1972) showed that addition of high concentrations of acetate (200mM) to growing cultures causes a breakdown of the proton motive force.

More recent work has suggested that an active transport mechanims may be involoved in the transport of acetate. Work by Ebbighausen *et al* (1991) suggests that acetate transport in *Corynebacterium glutamicum* is by a specific secondary carrier, acetate/proton symport mechanism. This is supported by the fact that acetate uptake shows saturation kinetics. Mallinder and Moir (1991) obtained *Bacillus stearothermophilus* mutants which were unable to grow on acetate as the sole carbon source. 13 mutants were defective in either ICL or MS, 3 were auxotrophic for isoleucine or valine, but one mutant was defective specifically in acetate uptake. At the same time work on *E.coli* suggested that *E.coli* possesses a specific transport system and that transport by simple diffusion alone is unlikely.

Once acetate has entered the cell it can then be metabolised to acetyl-CoA by two identified routes. The first route involves the acetylation of CoA and cleavage of ATP to AMP by acetyl-CoA synthetase (Berg, 1956). The second route involves acetate kinase which converts acetate to acetyl-phosphate, with the cleavage of ATP to ADP and then phophotransacetylase transfers the acetyl group from acetyl-phosphate to CoA, liberating phosphate. (Stadtman and Barker, 1980; Levine *et al,* 1980). It is thought that this second route is favoured in *E.coli.*

1.5 Control of IDH in *E.coli*

As mentioned previously, growth on fatty acids or acetate results in induction of the glyoxylate bypass, but it was not known how carbon flux was controlled between IDH and ICL.

Studies on the activities of different TCA cycle enzymes, malate dehydrogenase, 2-oxoglutarate and IDH in *E.coli* helped to elucidate the control of IDH. Holms and Bennett (1971) grew *E.coli* on limiting glycerol and observed that the specific activities of the three enzymes in the stationary phase of growth remained constant over a four hour period. However when *E.coli* was grown on limiting glucose, IDH activity was observed to drop to 20% of its original level, for two hours, before rising to 75% by the end of the four hour period.

The reason for this was that when *E.coli* is growing on glucose, it excretes acetate into the medium (Britten, 1954). Once all the glucose has been utilised, the glyoxylate bypass is induced and acetate can be metabolised. Thus while the cells were growing on acetate, the IDH level dropped and when the acetate was exhausted, the IDH level rose again.

Bennett and Holms (1975) also found that the specific activity of IDH from acetate-grown *E.coli*, was 30% of that found in glucose-grown or pyruvate-grown cultures. Addition of 1mM pyruvate to acetate-growing cells resulted in an increase in IDH activity, which dropped when the pyruvate was exhausted. Addition of chloramphenicol to acetate-grown cultures, before the addition of pyruvate, did not stop the increase in IDH activity.

This showed that the synthesis of IDH was not responsible for the increase in IDH activity (as chloramphenicol inhibits protein synthesis). Bennett and Holms (1975), suggested that a reversible covalent modification of the enzyme may be responsible for the differring levels of activity.

The first evidence to support this theory came from Garnak and Reeves (1979a,b). It was shown that the addition of ^{32}P to cultures, during the period when IDH decreased, resulted in the co-purification of a phosphorylated protein along with IDH. Biochemical and immunological studies showed that the protein was identical to IDH, suggesting that it was in fact ^{32}P-IDH. Partial hydrolysis of the purified protein revealed that it was phosphorylated at a serine residue.

Borthwick *et al* (1984a) purified partially active IDH from acetate-grown *E.coli* cells. Non-denaturing gel electrophoresis of the IDH revealed two bands , only one of which appeared on gels stained for IDH activity. Further studies revealed that this was due to charge differences and not due to a difference in molecular weight. Also, a phosphate group was observed after alkali treatment, showing that active and inactive IDH were dephosphorylated and phosphorylated respectively.

Although different levels have been reported for the Km of ICL for isocitrate, it is clear that the Km is higher than that of IDH for isocitrate (Nimmo *et al*, 1987). This results in IDH metabolising the available isocitrate when the intracellular concentration is low. ICL cannot compete for the low levels of the isocitrate (Nimmo *et al,* 1984). This is in fact the situation found when *E.coli* is grown on glucose, when the intracellular concentration of isocitrate is too low even to detect (El-Mansi *et al*, 1985). However, during growth on acetate, IDH is partially inactivated (phosphorylated), which reduces the flux of carbon through IDH and results in the level of isocitrate rising to a level where ICL can compete with IDH (Holms, 1987)

The phosphorylation/dephosphorylation of IDH is affected by a single bifunctional enzyme, IDH kinase/phosphatse (LaPorte and Koshland, 1982); 1.6.

Studies showed that IDH was phosphorylated at a single position, serine 113 of the *E.coli* IDH (Borthwick *et al,* 1984b;Thorsness and Koshland, 1987). Borthwick *et al* (1984c) also showed that the phosphorylation of IDH resulted in complete inactivation of the enzyme, which is unusual, as other phosphorylated enzymes often show some activity.

Thorsness and Koshland (1987) carried out mutagenesis of the serine 113 residue to see what effect other amino acids would have on the enzyme activity. Substitution of serine 113 with a negatively-charged aspartate, resulted in complete inactivation of IDH.

It was postulated that the negative charge of the bound phosphate group would be responsible for the complete inactivation of IDH. Serine 113 was replaced with threonine, and tyrosine and no detectable phosphorylation was observed, suggesting that the serine residue had a critical interaction with the IDH kinase.

X-ray crystallographic studies by Hurley *et al* (1990) showed this to be the case. Replacing serine113 with either aspartate or glutamate resulted in very little conformational change in the protein, and yet resulted in rendering the protein inactive. Calculations suggested that the change in electrostatic potential resulting from the negative charge caused by either the site-directed mutagenesis or by phosphorylation, is sufficient to cause inactivation of the enzyme. Dean and Koshland (1990) showed from kinetic studies that electrostatic repulsion and stearic hindrance between the phosphate group and isocitrate are the major cause for inactivation of IDH.

1.6 IDH kinase/phosphatase

Attempts to purify the IDH kinase, were undertaken to investigate how this enzyme phosphorylates IDH. LaPorte and Koshland (1982) purified the enzyme, but found to their suprise that the IDH phosphatase was purified concurrently. Fractions from various chromatographic purification steps contained both kinase and phosphatase activities. A single band of 66kDa was eluted from an IDH bound affinity column which contained both kinase and phosphatase activities. This led LaPorte and Koshland to conclude that the two activities were probably associated with the same polypeptide. Nimmo *et al* (1984), reported the native molecular weight to be 135kDa, suggesting that the IDH kinase/phosphatase was a dimer.

It was previously mentioned that IDH is controlled by IDH kinase/phosphatase. In consequence, IDH kinase/phosphatase is controlled *in vitro* by a number of effectors. Increased levels of various metabolites, PEP, pyruvate, 3-phosphoglycerate, and the TCA cycle intermediates, isocitrate, OAA and 2-oxoglutarate, result in higher phosphatase levels. ADP and AMP also result in an increase of IDH phosphatase (Nimmo *et al,* 1987).

LaPorte and Chung (1985) confirmed that both activities were associated with one protein, when they isolated a clone which restored growth on acetate to an *ace*K mutant. This clone expressed a protein of 66kDa and had both IDH kinase and IDH phosphatase activities.

Later sequencing of the *ace*K gene (Cortay *et al* 1988; Klumpp *et al* 1988) confirmed that the *ace*K gene encoded a protein of 577 amino acids, with a molecular weight of 66.5 kDa. Analysis of the deduced amino acid sequence did not reveal any of the consensus motifs normally associated with protein kinases (Celenza and Carlon, 1986). However one site similar to the consensus for ATP binding was observed. It is still unclear if both activities are catalysed by the same or distinct active sites of the protein (Ikeda *et al*, 1991).

Cortay *et al* (1988), observed two consecutive long dyad symmetrical sequences almost identical in sequence, just upstream from *ace*K. These sequences were calculated to form very stable stem-loop structures, similar to other intercistronic regions of other operons found in *E.coli* and *S.typhimurium*. Stern *et al* (1984) proposed that these elements would effect a decrease in the expression of distal genes in an operon. Cortay *et al* (1988), suggested that these sequences would strongly decrease expression of *ace*K. This was supported by LaPorte andChung (1985), who observed *ace*A to be expressed 100-1000 fold more than *ace*K. However Klumpp et al (1988) also observed these repeated sequences and altered them using oligonucleotide-directed deletion mutagenesis, removing the region comprising the stem-loop structure. This alteration had little effect on the expression of *ace*K, indicating that the stem-loop does not inhibit translation of the operon.

1.7 Isocitrate lyase and the *ace* operon

So far it has been seen that much of the work surrounding the control of the glyoxylate bypass has centred around IDH and its regulation *in vivo* by reversible phosphorylation. However properties of ICL also play an important role in determining the sensitivity of carbon flux control at the branchpoint (LaPorte *et al*, 1984).

It is thought that carbon flux through *E.coli* ICL is controlled at least in part by the intracellular concentration of isocitrate, which itself is affected by the phosphorylation state of IDH. Robertson *et al* (1987a,b) suggested that ICL may itself be controlled by phosphorylation of the enzyme at a histidine residue. Another report shows *Acinetobacter* ICL to be phosphorylated (Hoyt and Reeves, 1992). Clearly ICL also plays an important role in the control of carbon flux through the glyoxylate bypass.

The rest of this chapter will be concerned with the genes of the glyoxylate bypass operon, their control, and some of the work that has been carried out on the ICL protein.

1.7.1 ICLs from various organisms

ICL has been found in a great variety of organisms, incuding bacteria; algae; ferns; gymnosperms; angiosperms; fungi; protozoa; and worms (Cioni *et al*, 1981). There are a number of reports that ICL is also present in vertebrates, but this seems a matter of debate, as some of the levels reported are extremely low. In plants, ICL is found in a peroxisome, known as a glyoxysome in which the glyoxylate cycle takes place.

ICL has been purified from a number of sources: bacteria; plants; fungi; and yeast. The bacterial subunit molecular weight of ICL has been found to be around 48kDa (McFadden *et al*, 1968; Chell *et al*, 1978; Robertson and Reeves, 1987c; Matsuoka and McFadden, 1988). However the plant, yeast and fungal ICLs generally have a higher subunit molecular weight of between 62-70kDa (Johanson, 1974; Roberts and Lord, 1981; Uchida *et al*, 1986; Ruchti and Widmer, 1986).

The native molecular weights have been determined for most purified ICLs and suggests that most ICLs are tetramers. However ICLs from *Candida tropicalis* (Uchida *et al*, 1986) and *Glycine max* (Ruchti and Windmer, 1984) exist as dimers, while the *Pinus densiflora* ICL is reported to be a trimer (Tsukamoto *et al*, 1986).

The amino acid sequence of the *E.coli* ICL has been determined and this suggests that the subunits are identical *i.e.* ICL is a homo-tetramer (Mackintosh, 1987).

1.7.1.1 Activity of ICL

The specific activity of ICL has been found to vary considerably from organism to organism, with a reported low of 0.6U/mg protein, for *Anemia phyllitidis* (Gemmrich, 1979), to a high of 35U/mg protein for *Chlorella pyrenoidosa* (John and Syrett, 1967). ICL has also found to constitute as much as 8.3% of the total extracted protein (Robertson and Reeves, 1987c).

The K_m for isocitrate has been calculated for a great many ICLs. However the values were often obtained using different buffering systems and pH values, so little information can be taken from this. The effect of pH on the K_m of ICL is certainly noticeable, with the K_m of ICL from *E.coli* ML308 decreasing with pH: 63μM at pH7.3; 32μM at pH6.8 and 7μM at pH6.3 (Mackintosh and Nimmo, 1988).

Other studies on the effect of pH have been carried out on *N.crassa* (Rogers and McFadden, 1977) and *P.indigofera* (Rogers and McFadden, 1976). For both organisms studied, a plot of pKm against pH gave a straight line, with a gradient of -1. This indicates that the Km is dependent on a single dissociating group. Other work showed that the dissociating group is on the free enzyme, and not on the enzyme-substrate complex.

One noticeable feature of ICL which has been studied, is the requirement for magnesium. It appears that all ICLs require Mg^{2+} for catalysis (Cioni *et al*, 1981) and several reports have shown that when magnesium is left out of the assay mixture, no activity is detected. Hoyt *et al* (1988), tested *E.coli* ICL activity with various metal ions. No activity was observed without any metal ion, while the highest activity was found with Mg^{2+}, followed by Mn^{2+} (54%); Co^{2+} (17%); Ni^{2+} (7%) and Sr^{2+} (3%). Mackintosh (1987) determined the optimum concentration of magnesium of the *E.coli* ICL to be 5mM and observed that concentrations above this were inhibitory.

Studies carried out by Giachetti *et al* (1988), have shown that the Mg^{2+}-isocitrate complex is the true substrate for ICL. Other reports have also shown that magnesium can have a stabilising effect on the enzyme and protect the enzyme from thermal inactivation (Pinzauti *et al*, 1983), or inactivation by proteolysis (Galassi *et al*, 1988). Work has suggested that there are two magnesium binding sites, one with high affinity and a second catalytic site with lower affinity.

1.7.1.2 Important amino acids for activity

Studies with 3-bromopyruvate have indicated that a cysteine residue is involved in the activity of ICL. This glyoxylate analogue has been shown to inactivate ICL from a number of sources (McFadden *et al*, 1968; Khan and McFadden, 1982).

Chemical modification by diethylpyrocarbonate has indicated that a histidine residue is involved at the active site of ICL. (Khan and McFadden,1982; Jameel *et al*, 1985). Khan and McFadden (1982), removed the C-terminal histidine from *L.usitatissimum* ICL, which resulted in complete inactivation of the enzyme. It is not known however if this is the histidine involved at the active site.

Work carried out by Malhorta and Singh (1992) also demonstrated that a histidine residue is involved in the catalytic activity of ICL from castor seedling endosperm. Previously Malhorta *et al* (1991) had shown that glyoxylate binds closer to the active site sulphydryl group than succinate.

On the basis of the proximity and the properties of the functional groups -CHO and -SH, it was proposed that a hemi-thioacetyl linkage was formed between the aldehydic group of glyoxylate and the active site -SH group. It was proposed that there was ion pairing by intra-molecular proton transfer from a methylene group of succinate to the active-site imidazole group in the enzyme-succinate complex.

Malhorta *et al* (1992) proposed that, in a ternary complex, glyoxylate and succinate are bound close together to allow the condensation reaction to take place. A possible chemical reaction was outlined (Fig. 1.6).

1.7.2 The *ace* operon

The structural genes for ICL (*aceA*) and MS (*aceB*) were mapped by Brice and Kornberg (1968) and found together at 90 min on the *E.coli* K-12 chromosome.

Vanderwinkel and DeVlieghere (1968) showed that the gene *metA* mapped alongside *ace*A and *ace*B and found the gene order to be *metA-aceB-aceA*. The expression of the *ace* genes appeared to be negatively controlled by an adjacent regulatory gene *iclR* (Brice and Kornberg, 1968).

Evidence that the *ace* genes were part of an operon came first from Maloy and Nunn (1982) when they created transposon mutants in the *aceA* and *aceB* genes. Mutants generated by Tn10 are polar, preventing expression of genes downstream, in the same transcriptional unit. When *aceA*::Tn10 insertions were created, the expression of ICL was completely eliminated; however MS was expressed in a normal manner.

The *glc* gene encodes a second MS which can substitute for the *aceB* product (*glc* is induced during growth on glyoxylate or compounds which are metabolised to glyoxylate, such as glycolate); therefore *aceB* mutants were examined in a *glc*$^-$ background. When aceB::Tn10 insertions were constructed in a *glc*$^-$ background , both ICL and MS activities were eliminated. This not only suggested that *aceA* and *aceB* were part of an operon, but implied that transcription occured in the direction of *aceB* to *aceA*.

LaPorte and Chung (1985) confirmed that *aceK* was also part of the *ace* operon by creating *ace*::Mu insertions. One of the Mu lysogens created resulted in loss of all three enzyme activities, while one resulted in loss of only IDH kinase/phosphatase activity. This result confirmed that *aceK* was part of the operon and that it was located downstream from *aceA*.

The *ace* operon from *E.coli* has now been cloned by several groups (elMansi *et al*, 1987; Chung *et al*, 1988) and the seqences of *aceB* (Byrne *et al*, 1988; Rieul *et al*, 1988; Matsuoka and McFadden, 1988); *aceA* (Byrne *et al*, 1988); *aceK* (Cortay *et al*, 1988; Klumpp *et al*, 1988) and *iclR* (Sunnarborg *et al*, 1990; Negre *et al*, 1991) have all been determined (Fig. 1.7).

aceA was found to consist of 1,302 nucleotides, encoding a 47,700Da protein (Rieul *et al*, 1988) and *aceB* is a gene of 1,599 nucleotides, encoding a protein of 60,205Da (Byrne *et al*, 1988).

Each of the three structural genes of the *ace* operon is preceded by a ribosome binding site (Shine and Dalgarno, 1974): AGAGG for *aceB*, GGAG for *aceA* and GAGG for *aceK*. There is an intergenic region of 32 bases between *aceB* and *aceA*, with no particular structural motifs. *aceA* and *aceK* are separated by 184 nucleotides which contains 2 consecutive long dyad symmetries (1.7).

Chung *et al* (1988) identified a single promoter for the operon using high resolution S1 mapping. Sequences immediately upstream from the promoter site have -35 and -10 regions which match the *E.coli* consensus (Harley and McClure, 1983) at four out of six positions in the -35 region and 3 out of 6 positions in the -10 region. The two regions are separated by 16 nucleotides which is consistent with the consensus of 17± 1bp and the -10 is located 6bp away from the proposed transcriptional start site.

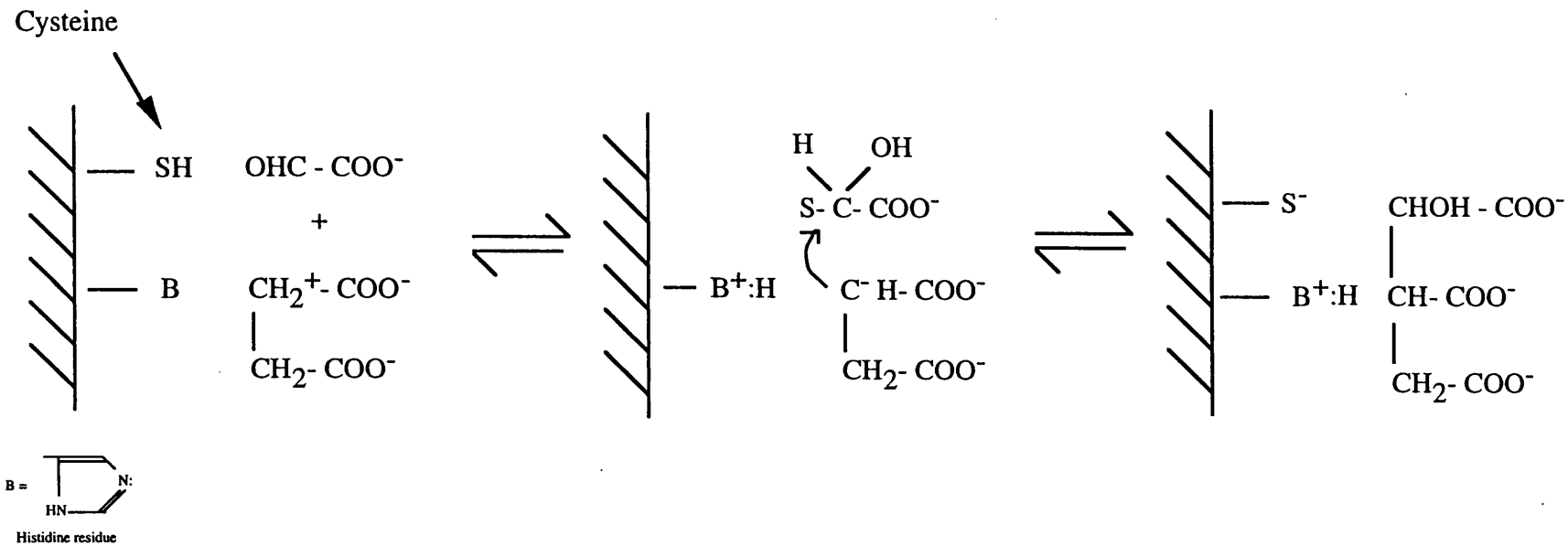


Figure 1.6 Reaction mechanism of Isocitrate lyase

The model of isocitrate lyase action presented by Malhorta and Singh (1992) proposes that in a ternary complex glyoxylate and succinate bind close together to allow the condensation reaction to take place

As mentioned in 1.7, LaPorte and Chung (1985) had observed *aceK* to be expressed 100-1000 fold less than *aceA*. Chung *et al* (1993) set out to determine whether the differential expression of the *aceBAK* operon was due to transcriptional or translational effects. They created gene fusions (which are dependent on both transcriptional and translational activities from upstream sequences), with each of the enzymes, linked to the *lacZ* gene and assayed for β-galactosidase activity. The levels were found to be 0.2: 1.0: 0.03, when linked to *aceB*, *aceA* and the 5' end of the *aceK* gene respectively, which is quantitatively similar to that found in the native operon. However for corresponding operon fusions (which reflect only transcriptional effects), all the levels of β-galactosidase were found to be the same. This result indicated that the relative expression of the operon was due mainly to differences in translational efficiency.

Work carried out to assess the reason for the striking downshift in expression between *aceA* and *aceK* has focussed on the repetitive sequences between *aceA* and *aceK* (1.7).

Deletion of these sequences had little effect on the expression of *aceK*. They did on the other hand result in two-fold decrease in the relative expression of *aceA*. It thus seems that the repeat sequences may differentially stabilise the *aceA* mRNA.

The deletion work localised the region responsible for inefficient translation to be between -26 and +3 of the *aceK* gene. Replacement of the *aceK* ribosome binding site with that of the *aceA* RBS, resulted in a dramatic increase in expression of β-galactosidase in a *aceK* gene fusions. It thus seems, that the dramatic decrease in translational efficiency of *aceK* is due to a poor RBS. This was unexpected as the RBS of *aceK* (GAGG), is similar to that of *aceA* (GGAG). Analysis of the region around this site also failed to reveal any other sequences which may affect translation.

1.7.2.1 *iclR* and *fadR*

Brice and Kornberg (1968) suggested that expression of *aceA* and *aceB* was regulated coordinately and controlled by an adjacent regulatory gene, *iclR*. Further work by Maloy *et al* (1980) suggested that the glyoxylate bypass enzymes are also controlled by the *fadR* gene product. The *fadR* gene maps at 25 mins on the *E.coli* K-12 linkage map and is also involved in the regulation of the regulon for fatty acid degradation (Simons *et al*, 1980). Mutations in either the *iclR* or the *fadR* genes resulted in elevation in the expression of the glyoxylate bypass enzymes under non-inducing conditions, suggesting that the *ace* operon is under transcriptional control of both *iclR* and *fadR*.

Maloy and Nunn (1982) created *aceA*::Mu*d*(Ap *lac*) fusions. These strains lacked isocitrate lyase, but malate synthase was expressed coordinately with β-galactosidase. These fusions placed the *lac* operon under control of the *ace* promoter. The fusions were placed in *iclR*$^-$ and *fadR*$^-$ backgrounds and resulted in increased levels of β-galactosidase. However, when the operon fusions were placed in an *iclR*$^-$ /*fadR*$^-$ background, even higher levels of β-galactosidase were observed. These results suggested that *iclR* and *fadR* both act separately at the level of transcription. A model proposed that separately, *iclR* and *fadR* each cause partial repression of the *ace* operon, but together they cause complete inactivation.

To determine whether the *iclR* and *fadR* genes acted in a *cis* or *trans* manner, Maloy and Nunn (1982) constructed merodiploid strains for *iclR*, or *fadR*. The results on expression of the *ace* enzymes showed that *iclR* and *fadR* acted in a *trans* manner, suggesting that the *iclR* and *fadR* gene products are acting as repressor proteins.

This dual repressor system may confer a metabolic advantage to cells growing on either fatty acids or acetate as the sole carbon source. Cells growing on fatty acids generate a great deal of energy during β-oxidation of the fatty acids, thus less energy is required from the TCA cycle and so more carbon can be fluxed through the glyoxylate bypass, resulting in induction of the *ace* operon. Growth on acetate on the other hand requires energy to be produced via the TCA cycle as well as the anaplerotic role of the glyoxylate bypass, perhaps explaining why the *ace* operon is expressed to a lesser extent.

The nucleotide sequences of the *iclR* genes of both *E.coli* and *S.typhimurium* have now been determined. (Galanier *et al*, 1991; Sunnarborg *et al*, 1990; Negre *et al*, 1991). Sunnarborg *et al* (1990) cloned the gene by defining the region of DNA which complemented an *iclR* mutant. This was found to be a 0.9kb fragment, which was subsequently sequenced. The gene was identified to be a 872 nucleotides in length, which corresponded to a protein of 29,741Da. The sequence also revealed putative -10 and -35 regions, which matched the consensus sequences in 5 out of 6 and 3 out of 6 positions respectively.

Analysis of the amino acid sequence revealed a putative helix-turn-helix domain at the amino terminus of the protein, supporting the model of a repressor protein.

Cortay *et al* (1991) purified overexpressed IclR protein, in a one step purification, using cation exchange chromatography. Further experiments were then carried out to identify its specific interaction with the promoter region of the *ace* operon. This was achieved by carrying out gel retardation assays and DNaseI footprinting experiments. The gel retardation studies showed that IclR bound to the promoter region of the *ace* operon. The specific region was localised by footprinting. It was found that IclR bound to a 35bp imperfect repeat. This repeat includes the -35 region of the *ace* promoter. Thus the repressor protein would block and not allow binding of RNA polymerase to the promoter (Fig.1.7).

Gel retardation experiments were carried out in the presence of various metabolites, to examine their effects on the IclR-DNA interaction. Out of PEP; acetate; acetyl-CoA; pyruvate and oxaloacetate, only PEP disrupted the IclR-DNA interaction.

Further work by Negre *et al* (1992) has revealed that IclR is a dimer, which is common for binding proteins involved in transcription. It was suggested that each monomer of the protein binds to one half of the palindrome-like structure.

Recent work carried out by Ramseier *et al* (1993) using the pleiotropic transcriptional regulatory protein FruR, has presented some interesting results. FruR is the fructose repressor, but it has been shown to control expression of a number of operons concerned with carbon metabolism. It had been shown previously that FruR mutants of *S.typhimurium* were impaired for the glyoxylate bypass (Chin *et al*, 1989). Ramseier *et al* (1993) set out to determine whether the FruR protein may act as a positve regulator of the *ace* operon.

Gel retardation assays involving the FadR protein and the *ace* operator/promoter region showed that the FruR protein did bind to this region of DNA. DNaseI footprinting was carried out to determine the exact location of the FruR binding. This revealed a 16bp protected region at about 170bp upstream from the transcriptional start site of the *ace* operon (Byrne *et al*, 1988). Further work revealed that the FruR protein also binds to the promoter region of the IDH gene and that Fructose-1-phosphate prevents binding of the FruR to both the *ace* and *idh* promoter regions.

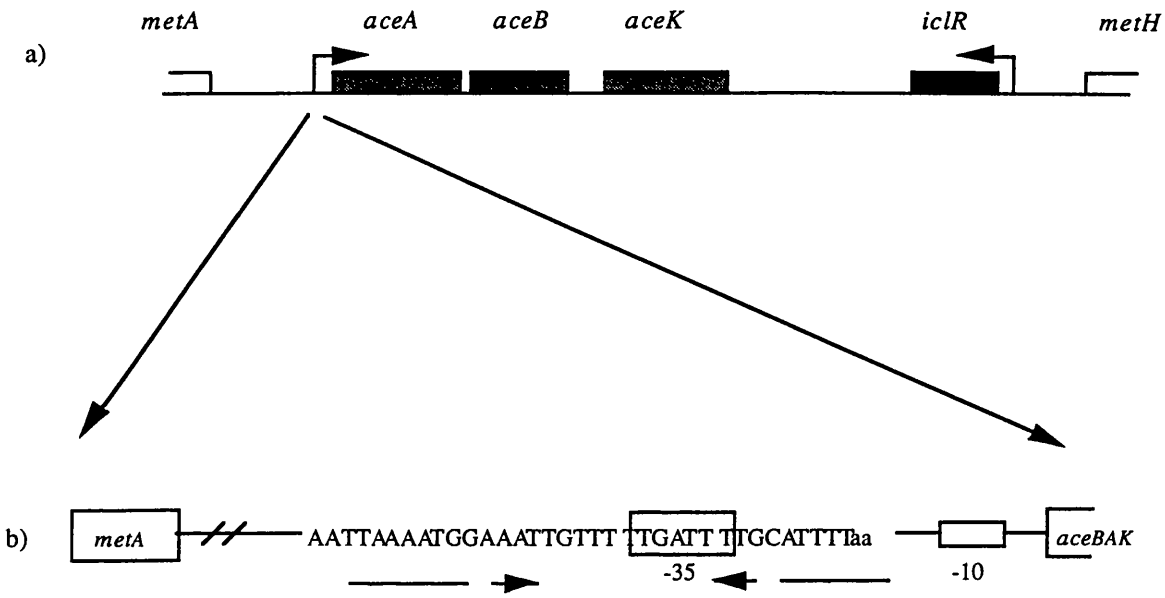


Figure 1.7 The *ace* operon of *E.coli*

a) A diagrammatic representation of the *ace* operon of *E.coli*. The *aceA*, *aceB* and *aceK* genes are transcribed from the same promoter and the *iclR* gene is transcribed in the other direction, from its own promoter.

b) A detailed representation of the operator/promoter region of the *ace* operon. The two arrows represent the imperfect palindromes to which the ICL repressor protein binds.

1.8 Aims of the project

As mentioned in 1.1, *Streptomyces* produce a great number a commercially important antibiotics. Of significant importance are the polyketide antibiotics, which are derived from acetyl-CoA. The interest is how *Streptomyces* control control the flux of carbon between the biosynthetic routes and the antibiotic biosynthetic pathways.

This chapter has dealt with how the glyoxylate bypass (in particular, that of *E.coli*), allows for growth on fatty acids and acetate. It has shown how IDH is controlled by reversible phosphorylation by IDH kinase/phosphatase and that this enzyme is part of the *ace* operon. It thus seemed important to study these enzymes in *Streptomyces,* to elucidate how the utilisation of acetyl-CoA is controlled between primary metabolism and secondary metabolism.

The *S.coelicolor* IDH has been studied by another member of this group and characterisation and cloning of this enzyme was already underway at the start of this project (Taylor, 1992)

It has since been observed that the *S.coelicolor* IDH is the first of the type II IDHs to be cloned and sequenced. Purified IDHs have been found to constitute two different classes. The first class (Type I) are dimeric enzymes, with subunit molecular weights of ~50kDa. The second class (Type II) are monomers of ~80kDa.

The *E.coli* IDH is a type I enzyme, which is different to the *S.coelicolor* IDH. It seems possible that the *S.coelicolor* IDH is controlled by a different manner to that of the *E.coli* IDH. Indeed studies by Taylor (1992), revealed that the *S.coelciolor* IDH does not appear to be phosphorylated. Further work is being carried out to study this enzyme.

This project set out to study the *S.coelicolor icl* gene in order to see if it was part of an operon and indeed to study the other genes of the operon.

It was envisaged that if the *icl* gene could be cloned, then the enzyme could be overexpressed, purified and studied. It would be possible to study how different levels of the enzyme, at different stages of growth, influenced antibiotic production. Metabolic control analysis and modulation of the gene, requires that the gene be cloned..

At the start of this project, 3 gene sequences of ICL were available. Alignments of these sequences revealed a good deal of similarity. In particular there was a highly conserved region of similarity centred around the putative active site cysteine. The initial aim was to use this information to design and synthesise oligonucleotides (which had *Streptomyces* codon bias built in) and then to use these oligonucleotides in hybridisation studies, with genomic *S.coelicolor* DNA, to attempt to clone the *S.coelicolor icl* gene.

CHAPTER 2

Materials and methods.

2.0 Introduction

This chapter contains the general procedures used in the experiments which were the basis of this thesis. The chapter is divided into 4 main sections for convenience; (2.1) bacterial strains, vectors and chemicals, (2.2) microbiological techniques and standard media, (2.3) general DNA methods and (2.4) general protein methods.

2.1 Bacterial strains, vectors and chemicals

2.1.1 Bacterial strains.

The bacterial stains used are listed in table 2.1

TABLE 2.1 Bacterial strains

STRAIN	GENOTYPE	REFERENCE/SOURCE
***Escherichia coli* strains**		
DS941	*recF*143, *proA7*, *str*31, *thr*1, *leu*6, *tsx*33, *mtl*2, *his*4, *argE*3, *lacY*$^+$, *lacZ*ΔM15, *lacI*^{q}, *galK*2, *ara*14, *supE*44, *xyl*5.	Dave Sherratt
TG1	*supE*, *hsdΔβ*, *thi*, Δ(*lac-proAB*), F' [*traD*36, *proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15].	Gibson (1984)
S17-1		Mazodier et al (1989)
***Streptomyces lividans* strains**		
TK24	*str*6.	Hopwood *et al.*, (1985a)
TK64	*pro*2, *str*6.	Hopwood *et al.*, (1985a)
***Streptomyces coelicolor* strains**		
GLW 209	(SCP2$^-$, NF)	G.Hobbs UMIST
GLW 1147	(SCP1$^+$,SCP2$^+$)	G.Hobbs UMIST

2.1.3 Plasmid and phage vectors.

The plasmid pUC18 was obtained from Pharmacia. Plasmid **pBluescript^{TM} II (KS$^+$ and SK$^+$)** and phage M13mp 18/19 was obtained from Stratagene (La Jolla, CA, USA). Plasmid pIBI 24/25 was obtained from Dr. M Anderson, Department of Biochemistry, Universitiy of Glasgow.

The *Streptomyces coelicolor* genomic library (Taylor, 1992) was constructed using the **λGEM-11** replacement vector supplied as "*Bam* HI arms" by Promega Corporation, Madison, USA.

pDK101 is a pGEM5zf(+) based vector, which has been altered, to aid cloning of PCR products (Kovalic *et al* 1991).

pRTB1 is a pBluescipt-derived plasmid containing the wild type *icd* gene from *S.coelicolor* (Taylor, 1992).

pIJ486/7 is a general streptomycete cloning vector (Ward *et al* 1986).

pCBS7 is a pIJ487 derived plasmid which contains the *veg* promoter of *B.subtlis* inserted into the polylinker region (Binnie personal communication; Moran *et al* 1982).

2.1.4 Chemicals and biochemicals.

Ampicillin, benzamidine, blue dextran, bromophenol blue, Coomassie brilliant blue, and ethidium bromide, were obtained from Sigma Chemical Co., Poole, UK.

Bactotryptone, yeast extract and Bactotryptone (agar) were obtained from Difco, Detroit, USA.

ATP, DTT, NAD^+, $NADP^+$, NADPH, pig heart lactate dehydrogenase, phenylmethanesulphonyl fluoride (PMSF), N,N,N',N'- tetramethylethylene diamine (TEMED), and Tris buffer were obtained from Boehringer Mannheim, Lewes, UK.

DMSO, phenylalanine, polyethylene glycol 8000, and tryptophan were obtained from BDH Chemicals, Poole, UK.

Agarose, isopropyl-β-thiogalactoside (IPTG), phenol (ultrapure), and 5-bromo-4-chloro-3-indolyl-ß-galactoside (X-gal) were obtained from BRL, Gibco Ltd., Paisley, UK.

Acrylamide, bisacrylamide and SDS were obtained from FSA Laboratory Supplies, Loughborough, UK. Junlon PW110 was a gift from Glaxo Group Research, Middlesex, UK.

Oligonucleotides were synthesised on an Applied Biosystems Model 280A DNA synthesiser at the Institute of Genetics, University of Glasgow, using reagents from Cruachem, Science Park, Glasgow.

All other chemicals were of analytical reagent grade and were obtained from one of the following suppliers: BDH Ltd., Poole, UK ; Formachem Ltd., Strathaven, UK; FSA Laboratory Supplies, Loughborough, UK; Koch-Light Ltd., Haverhill, UK; Sigma Chemical Co., Poole, UK.

2.1.5 Enzymes, proteins and kits.

A kit for molecular weight determination of proteins was obtained from Sigma Chemical Co., Poole,UK.

All restriction enzymes, T4 DNA ligase, and T4 Polynucleotide kinase were obtained from BRL, Gibco Ltd., Paisley, UK and Promega Corporation, Madison, USA.

Sequencing kits used were, TaqTrack™ (Promega Corporation, Madison, USA), T7 sequencing kit (Promega Corporation, Madison, USA), and Sequenase™ sequencing kit from USB Biochemicals (La Jolla, USA).

2.2 Standard media and microbiological techniques

2.2.1 Media used for growth of *S.coelicolor.*

All growth media were sterilised by heating to 120°C for 15 mins in an autoclave. Supplements and buffer solutions were heated to 108°C and $CaCl_2$ to 114°C for 10 mins. Heat-labile solutions, such as antibiotics, were sterilized by filtration through Nalgene 0.22μm pore membranes.

2.2.1.1 Minimal Media.

The minimal media below was used to allow growth on defined media with different carbon sources (developed by Hobbs *et al.*, 1989).

a) Glucose NMM.

4g glucose, 4.5g $NaNO_3$, 5g NaCl, 5g Na_2SO_4, 1g $MgSO_4.6H_2O$, 0.5g $CaCl_2$, 0.01g $ZnSO_4$, 1.2g Tris buffer, 20g KH_2PO_4 pH 7.2. Made to 1 litre with distilled water. The phosphate was autoclaved separately and added prior to inoculation. Also added was 1 ml of a filter-sterilized trace salts solution containing per litre ; 2.04g $ZnCl_2$, 1.015g $MnCl_2.4H_2O$, 0.310g H_3BO_3, 0.425g $CuCl_2.2H_2O$, 0.242g $Na_2MoO_4.2H_2O$, 0.238g $CaCl_2.6H_2O$, 8.775g $FeCl_3$ and 0.415g NaI. Approximately 10 small glass beads were added in order to stop clumping of the mycelia.

b) Tween NMM.

Same as above except;- Glucose was replaced by 1% (w/v) of Tween 20, 40, or 60 . The pH of the mixture was adjusted to pH 7.3 with NaOH prior to autoclaving.

2.2.1.2 Complex media.

a) Soya Mannitol Agar (SM)

This was used as a general plating media for *Streptomyces*, particularly for production of spores. It consists of 20g mannitol, 20g soya bean flour, 16g agar, made up to 1 litre using tap water.

b) Yeast extract-Malt extract (YEME).

This media was used for most purposes. It consists of 3g Difco yeast extract, 5g Difco bacto peptone, 3g Oxoid malt extract, 10g glucose, 340g sucrose, made up to 1 litre in distilled water.

2.2.2 Media for propagation of *E.coli.*

Chemicals of good quality were used in the preparation of the growth media and solutions; AnalaR grade when available. The sources of many of the chemicals varied during the course of this work. The most common suppliers were BDH Chemicals Ltd, Poole, Dorset; Difco Laboratories, Detriot, Michigan, USA and Sigma Chemical Co. Ltd, Poole, Dorset.

a) L-broth

10g tryptone, 5g yeast extract, 5g NaCl, 1g glucose, 20mg thymine, made up to 1 litre in distilled water and adjusted to pH 7.0 with NaOH.

b) L-agar

As L-broth without glucose and the addition of 15 g/l bacto-agar.

c). 2xYT medium

16g bacto-tryptone, 10g bacto-yeast extract, 5g NaCl, made up to 1 litre in distilled water and adjusted to pH 7.0 with NaOH.

d) Top agar.

As L-agar but with only 10 g/l bactoagar.

e) Top agarose.

As L-Agar but with 6.5 g/l agarose.

f) Minimal agar

Agar was made to 17.5 g/l with water. To every 75mls of this 25mls of DM salts were added and glucose and thiamine were added to give concentrations of 2 mg/ml and 20 µg/ml, respectively. Other supplements were added if necessary.

g) Davis and Mingoli (DM) Salts (x4)

28g K_2HPO_4, 8g KH_2PO_4, 4g $(NH_4)_2SO_4$ 1g trisodium citrate, $MgSO_4.7H_2O$, made up to 1 litre with distilled water.

h) Minimal top agar

6.5 g/l bacto-agar in dH_2O.

2.2.3 Growth of *Streptomyces* mycelia in liquid media.

Cultures were grown in 2 litre conical flasks containing 400 ml of medium at 30°C on an orbital shaker at 200 rpm. Alternatively, small cultures (~5ml) could be grown in large boiling tubes.

a) Minimal media (NMM).

For growth on NMM, flasks were inoculated from a frozen spore suspension with $2.5x10^5$ spores/ml media. Cells grown on NMM were harvested after 60 hours or just as the mycelia started to produce undecylprodigiosin (red) antibiotic.

b) Complex media (YEME).

Most growth experiments employed this complex medium which produced reproducible and rapid growth. For growth on YEME, spores from frozen suspensions could be used although it was found to be more convenient to use suspensions that were freshly-prepared from a frozen slope. The spores obtained from the preparation of suspension from one slope were used to inoculate the medium. Cells grown on YEME could be harvested after 48 hours and stored as a cell pellet at -20°C indefinitely until required.

2.2.4 Harvesting of mycelia.

After growth, mycelia were recovered from the media by centrifugation (20 minutes 7,500 g) and resuspended in extraction buffer (100 mM MOPS-NaOH pH 7.3, 2 mM-DTT, 1.2 mM-PMSF, 1 mM-benzamidine, 1mM EDTA), if protein was to be extracted, or in dH20 if DNA was to be extracted. Cells grown on YEME (section 2.5b) were diluted in an equal volume of distilled dH_2O prior to centrifugation.

2.2.5 Growth of *E.coli.*

Liquid cultures of *E. coli* strains from which plasmids were to be isolated were grown in L broth with the appropriate antibiotic selection (usually ampicillin at 50µg/ml). The volume of broth inoculated depended on the quantity of plasmid required. Routinely, 5 ml and 200 ml cultures were used for small and large scale plasmid preparations, respectively (2.19.3 and 2.19.4). For the preparation of competent cells, liquid cultures of *E. coli* DS941 were grown in L broth while *E. coli* TG1 was grown in 2YT (section 2.2.2). To maximise aeration of the culture, the volume of the Ehrlenmeyer flask was at least five times that of the broth. All cultures were incubated at 37°C in an orbital shaker at ca. 250 rpm.

2.2.6 Production of *S.coelicolor* spores.

a) Preparation of suspensions and storage.

Concentrated spore suspensions were required for inoculating liquid cultures of *S. coelicolor,* or *S.lividans.* The protocol described by Hopwood *et al.* (1985a) was followed with minor modifications.

A boiling tube containing a slant of SM agar (produced by pouring ca. 15ml of molten agar into the tube and allowing it to solidify with the tube held a +5° from the horizontal) was inoculated with 150µl of a spore or mycelial fragment suspension and incubated at 30°C. After 10 days the surface of the culture was covered in a dark grey mass of spores.

The slant was then sealed using parafilm and frozen at -20°C. Spores could then be harvested immediately or stored at -20°C indefinitely. The spores were harvested by adding 5ml of dH_2O to the frozen slant and rubbing the surface of the slant with a 10ml glass pipette. When the surface had been scraped clean of spores, the spore suspension was decanted into a sterile universal. The slant was rinsed with an additional 5ml of dH_2O to remove any spores remaining from the first wash. The spore suspension (c.10ml) thus produced was dark grey in colour. Contaminating agar or mycelial fragments were removed by a single passage through a cotton wool filter, as described in Hopwood *et al.* (1985a). The filtered spore suspension was then either used fresh to inoculate YEME media or frozen at -20°C after the addition of glycerol to 20% (v/v).

b) Spore counts.

Colony forming units were determined by plating suitably diluted spore samples on SM plates. Counts of the number of colonies were made after incubation at 30°C for 5 days.

2.2.7 Preservation of *E.coli* strains.

E.coli strains were stored in glycerol. An 800 µl aliquot of an overnight culture was mixed with an equal volume of 40% (v/v) glycerol, 2% peptone (w/v) and frozen at -70°C. The strains were revived by scraping the surface of the frozen suspension with a toothpick and either inoculating liquid broth or streaking onto agar to isolate a single colony.

2.2.8 Introduction of plasmid DNA into *E.coli.*

2.2.8.1 Preparation of competent cells.

a) $CaCl_2$ method.

An overnight culture of the recipient strain was diluted 1 in 100 into 30ml L-broth and incubated for 90-120 min to a density of approximately $10^8.ml^{-1}$ cells (OD_{600} 0.45-0.55). The cells were harvested using a centrifuge (12000g, 5 min, 4°C) and resuspended in 10ml of ice-cold 50mM $CaCl_2$. The cells were pelleted again, resuspended in 1ml of ice-cold 50mM $CaCl_2$ and kept on ice for at least 15 min before use.

b) Hanahan method.

For cells with high transformation efficiencies, the following steps were carried out (Hanahan, 1983). Cells were prepared and harvested as above but were then resuspended in 2.5 ml of ice cold TFB (10 mM-MES/KOH pH 6.3, 100 mM-RbCl, 45 mM-$MnCl_2$, 10 mM-$CoCl_2$, 3 mM-hexaminecobaltic chloride) and incubated on ice for 15 minutes.

100 µl of DMSO was added and the cells were incubated on ice for 5 minutes. Next 100 µl of 2.25 M-DTT, 40 mM-potassium acetate pH 6.0 was added and the cells further incubated on ice for 10 minutes. Finally, 100 µl of DMSO was added, then the cells were kept on ice and used on the day of preparation.

2.2.8.2 Transformation procedure.

Transformations were carried out in sterile 1.5 ml microfuge tubes. An aliquot of ligation mix (or plasmid) containing up to 25 ng plasmid DNA was added to 100 µl aliquots of competent cells and the mixture was incubated on ice for at least 30 minutes. The DNA/cell mix was then heat shocked at 42°C for 2 minutes, before being placed back on ice. 1 ml of L-broth was added to the tubes and they were incubated without shaking at 37°C for 30 mins. The cells were then plated onto LB plates containing the appropriate antibiotic/chromogenic substances and incubated overnight at 37°C.

2.2.8.3 Selection of pUC derived recombinant clones.

a) Ampicillin

Stock solutions (20mg/ml made up in water) were added to molten agar (cooled to 55°C) to a final concentration of 50µg/ml.

b) X-gal (5-bromo-4-chloro-3-indolyl-ß-galactosidase).

This was used in conjunction with IPTG to identify *E. coli* strains DS941 and TG1 containing pUC or M13mp vectors with inserts in their multiple cloning sites. Recombinants containing inserts are generally white while those lacking inserts are blue. X-gal was stored at a concentration of 20mg/ml in dimethylformamide (DMF) at -20°C while IPTG was stored at a concentration of 24mg/ml in dH_2O at -20°C. X-gal and IPTG were added to L-agar plates to a final concentration of 20µg/ml and 50µg/ml, respectively.

2.2.9 Introduction of plasmid DNA into *Streptomyces*

Plasmids were introduced into *Streptomyces spp.* by genetic transformation using the polyethylene-glycol-mediated protocol described by Hunter (1985).

Reagents: Trace element solution; Per litre, 40mg $ZnCl_2$, 200mg $FeCl_3.6H_2O$, 10mg $CuCl_2.2H_2O$, 10mg $MnCl_2.4H_2O$, 10mg $Na_2B_4O_7.10H_2O$ and 10mg $(NH_4)_6Mo_7O_{24}.4H_2O$. **Medium P**; 5.73g N-tris[Hydroxymethyl]methyl-2-aminoethanesulfonic acid (TES), 103g sucrose, 2.93g $MgCl_2.7H_2O$, 0.5g K_2SO_4, 3.68g $CaCl_2.2H_2O$, 2ml trace element solution; adjusted to pH 7.4 with NaOH and made up to 1 litre in distilled water.

Lysozyme solution; 10% (w/v) sucrose, 25mM TES buffer (pH 7.2), 2.5mM K_2SO_4, 2ml trace elements (Hopwood et al., 1985a), 2.5mM $MgCl_2$, 2.5mM $CaCl_2$. KH_2PO_4 (0.005% [w/v]) and lysozyme (0.3mg.ml^{-1}) were added immediately prior to use. **PEG solution;** 1g of polyethylene glycol 1540 (supplied by BDH) was melted in a microwave (600W) for 10s on the reheat setting and then mixed with 3ml of medium P.

2.2.9.1 Preparation of protoplasts

30ml cultures were grown in the appropriate medium at 30°C. The optimal time to harvest the mycelium in order to recover the most competent protoplasts was species-dependent i.e. *S. lividans* mycelium were harvested after 65hr and *S. rimosus* after 48hr, respectively. The mycelium was pelleted at 12000g for 10 minutes and washed twice in 10.3% (w/v) sucrose. The pellet was then resuspended in 4ml of lysozyme solution and incubated at 37°C for 15-30 min. The formation of protoplasts was monitored using a microscope, the reaction terminated by adding 5ml of P medium and the protoplasts triturated twice.

The protoplasts were then filtered through cotton wool (Hopwood *et al.*, 1985a), pelleted using a centrifuge (12000g for 10 minutes) and washed twice in P medium. Finally, they were resuspended in 4ml of medium P, dispensed into 200µl aliquots and frozen at -70°C.

2.2.9.2 Transformation of protoplasts

The protoplasts were thawed on ice. DNA was added in a volume of less than 10µl and the mixture incubated on ice for 30s. PEG solution (400µl) was added, the solution incubated for a further 1 min on ice and finally, medium P (800µl) was added. Dilutions of the transformation mix were then made in medium P.

2.2.9.3 Regeneration of transformed protoplasts

The method used to prepare the regeneration medium for the protoplasts was standardised. The medium was stored in two parts, RA and RB ; the former solid and the latter liquid. Both parts were placed in a steam bath until the RA portion melted. The two components were allowed to cool to 50°C before they were combined and 1ml of 1%(w/v)KH_2PO_4 added. The complete regeneration medium was mixed by swirling and then poured into petri dishes (diameter 9cm). On average 8 plates were obtained from 200ml of medium. The plates were dried in a laminar flow hood to minimise air borne contamination by leaving the lids half open for 45min. They were then rotated 180°C and their relative positions reversed so that those at the front of the hood were placed at the back. After a further 45min, the plates were removed and left overnight at 30°C. The next day, any plates which showed any signs of contaminating growth were discarded and the rest used for the regeneration of protoplasts.

The only drug resistance used in this work for plasmid selection in *Streptomyces* was thiostrepton (obtained from E.R. Squibb, New Jersey, USA). It was dissolved in DMSO to make a 1% (w/v) stock solution. Transformed protoplasts were selected after 16-22 hrs of non-selective growth at 30°C by overlaying the regeneration plates with 1ml of a 220µg.ml^{-1} thiostrepton solution in 10.3% (w/v) sucrose solution.

S. lividans and *S.coelicolor* protoplasts were regenerated on R2 agar plates.

2.2.10 Conjugation between *E.coli* and *Streptomyces*

A vector capable of being maintained in *E.coli* was first cloned into *E.coli* S17-1. A transformant was obtained and a 5ml culture grown in 2xYT.

A slope of *Streptomyces* spores was resuspended in 5ml of dH_2O. This was then heat shocked at 50°C for 10min, before being cooled under a cold tap and 5ml of Pre-germination medium added (Mazodier *et al,* 1989). This was then placed at 37°C with shaking for 2-3 hrs. The germinating spores were then spun down at 12 000g for 10min and resuspended in 1ml of dH_2O.

1ml of the overnight S17-1 culture was spun down at 12 000g for 30secs and resuspended in 1ml of fresh 2xYT.

100µl of germinating spores and 100µl of S17-1 were plated and spread onto an L-agar plate and left at 30°C for 12hrs. The plates were then washed with L-broth and scraped with a pasteur pipette, before being left to dry. The plates were then overlayed with the selective drug and naladixic acid (to kill the S17-1 cells).

Exconjugants grew up in 3-4 days.

2.3 General DNA methods.

2.3.1 Commonly used buffers.

a) TE buffer (10x), contained

100 mM Tris.HCl pH 8.0, 10 mM EDTA. Sterilised using an autoclave and stored at room temperature. It was used as a 1x solution for most applications.

b) Phage buffer (1x), contained

20 mM Tris.HCl pH7.4, 100 mM NaCl, 10 mM $MgSO_4$.

c) TBE buffer (10x) pH8.3, consisted of

109g Tris, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$ made up to 1 litre in distilled water.

d) TAE Buffer (10x), consisted of

48.4g Tris, 16.4g Na acetate, 3.6g $Na_2EDTA.2H_2O$, made up to 1 litre in distilled water, pH adjusted to 8.2 with glacial acetic acid.

e) Agarose gel loading buffer (10x) pH 7.4, contained

0.5% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol, 50% (w/v) ficoll, 1% (w/v) SDS, 100 mM EDTA.

f) λ–*Hind* III DNA markers.

λ c857s7 DNA was obtained from BRL, Gibco Ltd., Paisley, UK. This DNA was cleaved with the restriction enzyme *Hind* III and resulting DNA fragments made to a final concentration of 27 ng/µl in TE (final concentration 1x) with loading buffer added to 1x concentration. 10 µl was used on agarose gels as markers for comparing the size and concentration of bands in samples.

Sizes of fragments are 23,130 bp, 9,416 bp, 6,557 bp, 4,361 bp, 2,322 bp, 2,027 bp and 590 bp.

2.3.2 Preparation of plasmid DNA.

2.3.2.1 Reagents for isolation of plasmid DNA from cells.

Protocols based on the alkaline lysis method (Birnboim and Doly, 1979) were used for the isolation of plasmid from small (5ml) or large (50-200ml) cultures of *E. coli.*

a) Birnboim Doly I (BDI) contained

50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA. Lysozyme was added immediately before use to a final concentration of 1-4mg/ml, if streptomycete plasmids were being isolated.

b) Birnboim Doly II (BDII) contained

0.2M NaOH, 1% (w/v) SDS which was stored in a plastic container.

c) Birnboim Doly III (BDIII) contained

5M KOAc (pH 4.8); prepared by mixing equal volumes of 3M CH_3COOK and 2M CH_3COOH.

d) DNAase-free RNAase.

Pancreatic RNAase (RNAase A) was dissolved at a concentration of 10mg/ml in dH_2O, heated to 100°C for 15min and allowed to cool slowly to room temperature. The RNAase was then aliquoted and stored at -20°C.

2.3.2.2 Large scale plasmid preparations

a) Caesium chloride gradient

200ml cultures of stationary phase cells were harvested using a centrifuge (12000g, 5min at 4°C). The pellet was resuspended in 4ml of Birnboim-Doly I solution and incubated on ice for 5min. Then 8ml of Birnboim-Doly II solution were added and the solution left on ice for 5-10 min before 6ml of cold Birnboim-Doly III solution were added. The suspension was mixed gently and left on ice for 15-30 min. The cell debris and most of the chromosomal DNA were removed by centrifugation (32000g, 5 min at 4°C). The remaining nucleic acid was precipitated by the addition of an equal volume of isopropanol and then harvested by centrifugation (39200g, 15 min).

The nucleic acid pellet was washed with 70% (v/v) ethanol. The plasmid DNA was further purified by equilibrium density centrifugation on a caesium chloride/ethidium bromide (CsCl/EtBr) gradient. The nucleic acid pellet was redissolved in 1 ml of dH_2O and 4.5g of CsCl dissolved in 3.5ml of dH_2O. The DNA and CsCl solutions were combined with 250µl of EtBr (10 mg/ml), creating a solution with a density of 1.58 g/ml. The nucleic acid-CsCl solution was spun in a Beckman Ti70 angled rotor at 289,000g for 16 hours at 20°C. Two bands were visible in the gradients after centrifugation, a lower supercoiled plasmid band and an upper chromosomal and open circle plasmid DNA band.

The lower band was removed using a 1ml syringe and the EtBr removed by repeated extractions with water-saturated butanol. After dilution with 3 volumes of dH_2O, 9 volumes of absolute ethanol were added. The precipitate was pelleted by centrifugation (27000g, 4°C for 30min). The resulting plasmid pellet was washed twice with 70% (v/v) ethanol and dried *in vacuo* before being redissolved in 1ml dH_2O. This procedure yielded very large amounts of pure plasmid DNA (up to 1mg from *E. coli* cultures) suitable for all *in vitro* manipulations.

b) PEG precipitation

The protocol was the same as that for (a), up to the resuspension of the DNA, in dH_2O and CsCl. Instead the DNA was resuspended in 1.6ml of dH_2O. 0.4ml of 4M NaCl was added and mixed. 2ml of 13% PEG was added,the tube mixed and left on ice, for 60 minutes. The preciptitated DNA was removed by centrifugation at 12,000g for 20 minutes.The pellet was washed with 70% EtOH, before being dried under vacuum and redissolved in dH_2O.

2.3.2.3 Small scale plasmid preparations

Routinely, plasmids were isolated from 1.5ml of *E. coli* cultures. The cells were pelleted by centrifugation in a 1.5ml microfuge tube (12000g for 30secs) and resuspended in 100 µl of BDI, containing lysozyme at a concentration of 1 mg/ml, using a vortex mixer. This was followed by the addition of 200µl of BDII and repeated inversion of the microfuge tube to mix thoroughly the suspension. Immediately afterwards, 150 µl of prechilled BDIII was added to the viscous bacterial lysate, mixed gently on the vortex mixer and placed on ice for 5-10 min. The cell debris and most of the chromosomal material were harvested by centrifugation (12000g, 4°C for 10min) in a microfuge. The supernatant was transferred to a fresh tube and extracted with half volumes of phenol/chloroform and chloroform. The nucleic acid was then precipitated by the addition of 2 volumes of ethanol and allowed to stand at room temperature for 3-5min. The precipitate was harvested by centrifugation in a microfuge (12000g, 4°C for at least 15min). The resulting pellet was rinsed twice with 70% (v/v) ethanol before it was allowed to dry by leaving the tube open on the bench. The nucleic acid was then resuspended in 50µl dH_2O containing DNAase-free RNAase (20µg/ml).

The typical yield of high-copy-number plasmids such as pUC from *E. coli* was 2-5µg. Plasmid prepared in this way could be used for most *in vitro* manipulations.

2.3.3 Preparation of *S.coelicolor* genomic DNA.

Total DNA from *S.coelicolor* was prepared by Mrs. M. Stone and Ms. A. Wylie (Department of Genetics, University of Glasgow) essentially as described by Hopwood *et al.*, (1985).

2.3.4 Organic solvent extraction.

Protein was removed from DNA solutions by phenol/chloroform extraction. An equal volume of TE-saturated phenol was added to samples which were then mixed by vortexing and centrifuged in a microfuge for 1-5 minutes. The upper aqueous phase was removed to a fresh microfuge tube and the process repeated this time using phenol/chloroform (1:1 v/v). Finally, traces of phenol were removed by extraction with an equal volume of chloroform in an identical manner. Precipitation with ethanol or isopropanol removed any remaining solvent.

2.3.5 Precipitation of DNA using ethanol or isopropanol.

DNA solutions were precipitated by the addition of 1/50 volume of 5M NaCl and 2 volumes of cold ethanol or an equal volume of isopropanol. After mixing, the DNA was pelleted by centrifugation (27000g, 30 min, 4°C for volumes of 7.5-20ml or 12000g, 15 min, 4°C for small volumes in microfuge tubes). The pellet was washed in 70% (v/v) ethanol and dried briefly in a vacuum desiccator or in an open tube on the bench.

2.3.6 Spectrophotometric measurement of nucleic acid.

Nucleic acid concentrations were determined spectrophotometrically at 260 nm. In a 1 cm path length an absorbance value of 1.0 corresponds to 50 µg/ml for double stranded DNA , 33µg/ml for single stranded DNA and 20 µg/ml for oligonucleotides.

2.3.7 Digestion of DNA with restriction enzymes.

Restriction digests were carried out using the BRL React buffers which were provided with each batch of enzyme. There are ten different React buffers with a range of salt concentrations, each one suitable for a range of enzymes. Analytical digests were carried out in a volume of 10 or 20 µl at 37°C. Preparative digests were carried out in larger volumes. When DNA was digested with two restriction enzymes, the endonuclease requiring the lower salt buffer was used first. After the recommended duration of digestion, the salt concentration was adjusted and the second enzyme added.

2.3.8 Ligation of DNA fragments.

The ligation of DNA fragments was carried out usually at a DNA concentration of 6mg/ml. The molar ratio of insert fragment to vector was 2:1, when the vector could not ligate to itself *e.g.* when using vector that has been dephosphorylated or has been cut with two enzymes). A molar ratio of 10:1 was used when the ends of the vector could ligate to each other. Ligations were performed usually in 10µl of 1x ligation buffer provided by BRL, containing 1U of T4 ligase per µg of DNA. The reactions were incubated for 4 hours at room temperature or overnight at 16°C.

2.3.9 Removal of the 5' phosphate from linearised DNA

10X CIP Buffer contained: 200mM Tris.HCl (pH 8.0), 10mM $MgCl_2$, 10mM $ZnCl_2$ and 0.5mg.ml^{-1} Bovine Serum Albumin.

Procedure: Calf Intestinal Alkaline phosphatase (CIP) was used to remove the 5' phosphate from DNA. Around 5 pmoles of 5'-terminal phosphorylated DNA with 5' protruding ends (approximately 7µg of a 5kb molecule) were incubated in 1X CIP buffer, containing 0.1U of CIP at 37°C for 30min. The reaction was terminated by heating to 65°C in 1X gel loading buffer for 10min. The 5'-terminal dephosphorylated DNA was recovered from an agarose gel after electrophoresis. .

2.3.10 Agarose gel electrophoresis.

DNA was visualized on horizontal neutral agarose gels. Although 0.8% (w/v) gels were most commonly used, 1-2% (w/v) gels were occasionally used to separate fragments of <1.5kb. Gels were routinely prepared and run in TBE buffer. However, TAE buffer was used when DNA fragments were to be isolated from the gels (see section 2.3.12 and 2.3.13). λ-markers were used on all gels as size markers and for quantification of the amount of DNA by comparing the intensity of bands to those of the samples (2.3.1f).

a) Mini gels.

BRL model H6 gel kits were used for the rapid analysis of DNA after digestion with restriction enzymes or precipitation steps. 0.16g agarose was added to 20ml of 1X TBE (or TAE), boiled then cooled to 60°C. EtBr was added to 200ng/ml and the molten agarose poured into a 7.6cm x 5.1cm gel caster with an 8 well slot former (4.1 x 0.8mm wells). After the gel had set, the slot former was removed and the gel placed in the tank with 500ml of 1 x TBE (or TAE).

Depending on the time available and the level of resolution required the DNA was separated by electrophoresis for 30-45min with an applied voltage of 2-10V.cm^{-1}. The separated DNA molecules were visualised on a 302nm UV transilluminator.

b) Large gels.

200ml gels were also used to ensure good separation of DNA fragments for accurate sizing and/or Southern analysis. They were made by pouring 200ml of molten agar containing 200µg EtBr, into a 16.5 x 23cm gel former with a 20 space slot former. The gels were run overnight at 20V in 1 x TAE or TBE buffer in gel tanks with a capacity of 3 litres. DNA samples were mixed with 1/5 volume of 5 x AGL buffer (2.3.1e), heated to 70°C for 2min and cooled on ice before loading onto the gel.

2.3.11 Photography of agarose gels.

Gels stained with ethidium bromide were viewed on a 302nm UV transilluminator and photographed using Polaroid type 67 land film or using a Pentax 35mm SLR loaded with Ilford HP5 film. Both cameras were fitted with Kodak Wratten filters (No. 23A). Alternatively, pictures were obtained using a Mitsubishi video copy processor attached to a UVP video camera

2.3.12 Recovery of DNA from agarose gels by electroelution.

The DNA band of interest was visualised by examination of under a hand-held long-wavelength ultraviolet lamp. A slice of DNA containing the desired DNA band was excised, placed in a piece of dialysis tubing clipped at one end and 1xTAE buffer added. All air bubbles were removed before sealing the bag with another clip. The bag was placed in a shallow layer of 1xTAE in an electrophoresis tank. A voltage of 50 V was applied across the bag for 2 to 3 hours causing DNA to migrate from the gel onto the inner wall of the bag. The current was reversed for 1 minute to release DNA from the wall of the bag into the buffer. Buffer was then transferred from the dialysis bag to sterile Eppendorf tubes and DNA purified by phenol extraction.

2.3.13 Recovery of DNA from agarose using SPIN-X tubes

SPIN-X tubes were obtained from Costar UK Ltd.. These tubes contain a cellulose acetate membrane, which allows the passage of buffer and DNA, but stops agarose from passing through.

The DNA band is excised from the gel and placed in the upper chamber of the SPIN-X tube. This is then placed at -20°C for 15 minutes. The tube is then thawed,by placing at 37°C for 5 minutes, before being centrifuged at 5,600g for 5 minutes. The filtrate contains the DNA, in a state which can be used immediately.

2.3.14 Techniques with *E.coli* bacteriophage λ

2.3.14.1 Preparation of plating bacteria for infection with bacteriophage λ.

50 ml of L-broth supplemented with 0.2% (w/v) maltose, 20 mM-$MgSO_4$ was inoculated with a single colony of the appropriate *E.coli* strain (*e.g.* NM621), and grown overnight at 37°C on an orbital shaker.

The cells were pelleted by centrifugation (12,000.g, 5 minutes, 4°C) and resuspended in 0.5 volumes of sterile, ice cold 10 mM $MgSO_4$. The cell suspension was diluted if the OD_{600} was greater than 2 (*i.e.* 1.6 x 10^9 cells/ml). The cells were stored at 4°C and remained viable for at least 3 weeks.

2.3.14.2 Infection of bacteriophage λ, plating and titre.

Serial 10-fold dilutions of λ phage stock (or packaged λ DNA) were prepared in phage buffer (2.3.1b). Bacteriophage λ infection was achieved by adding 100 µl aliquots of each dilution to 100 µl (1.5 x 10^8 cells) of a suspension of plating bacteria. The samples were incubated at 37°C for 20 minutes. 3 ml of top agarose at a temperature of 45°C was added and the mixture was poured onto plates containing bottom agar. The plates were left to stand for 5 minutes at room temperature to allow the top agarose to harden and then incubated at 37°C overnight. The plaques were counted and the titre determined for each dilution assayed.

2.3.14.3 Isolation of phage particles from a plaque.

The plaque of interest was stabbed out of the plate using the narrow end of a sterile glass Pasteur pipette to form a plug of agar. The plug was left in 1 ml of phage buffer with 70 µl of DMSO (to kill any cells) for two hours at room remperature (or overnight at 4°C) to allow bacteriophage particles to diffuse out of the agar. An average plaque yielded 10^6-10^7 infectious bacteriophage particles, which could be stored indefinitely at -70°C in phage buffer/DMSO without loss of viability.

2.3.15 Techniques for handling *E.coli* bacteriophage M13.

2.3.15.1 Transfection and plating of M13.

Cloning in M13 vectors is essentially the same as cloning in plasmids. M13 produces a double stranded, replicative form (RF) which is isolated and treated just like an ordinary plasmid such as pUC18/19. After introduction of foreign DNA into such vectors by ligation, they can be introduced into a suitable M13 host such as TG1 (*ie.* one that produces pili).

E.coli TG1 was made competent by one of the methods above (2.2.8.1) and the transformation protocol followed up to and including the heat-shock stage. After this step, 200 µl of a fresh exponential TG1 culture were added to the transfected cells, followed by 10µl of IPTG (24 mg/ml) and 50 µl X-gal (20 mg/ml).

The cells were then mixed and added to 2.5ml of molten water-agar (0.6% w/v, pre-cooled to 45°C), which were poured onto thoroughly-dried L-agar plates. Plaques containing recombinant phage appeared white on the agar, whereas non-recombinant phage appeared blue.

2.3.15.2 Isolation of plaques.

An agar plug was pulled from the agar that contained the plaque of interest and phage soaked out into phage buffer similar to described in section 2.3.14.3. However M13 are sensitive to chloroform and DMSO and so these additives were ommited.

2.3.15.3 Preparation of M13 single stranded DNA.

The single-stranded M13 templates were prepared as described in the "M13 Cloning/Dideoxy sequencing Instruction Manual" published by Bethesda Research Laboratories.

a) Minipreparations

A single M13 plaque was used to infect 1.5ml of 2X YT broth containing 15µl of an overnight culture of *E. coli* TG1. This was grown at 37°C for 5-6 hours with vigorous shaking, then transferred to a microfuge tube and harvested by centrifugation at room temperature for 5min. The supernatant, containing the phage particles, was recovered and respun. The supernatant (1ml) that remained was mixed with 200µl of a solution of 20% (w/v) PEG (6000)/2.5M NaCl and left to stand at room temperature for 15min to precipitate the phage particles. These were recovered by centrifugation at room temperature in a microfuge for 15min.

The supernatant was discarded, the pellet respun and all traces of supernatant removed. The pellet was then resuspended in 100µl dH_2O and extracted twice with phenol/chloroform and twice with chloroform. The DNA was then precipitated with ethanol from the aqueous phase and recovered by centrifugation in a microfuge. The ssDNA was then washed with 70% (v/v) ethanol, dried *in vacuo* and redissolved in 20µl dH_2O.

b) Maxipreparations.

The same overall procedure was followed as for the "minipreparations", except that all the volumes were scaled up 20-fold. The cells from a 30ml culture were spun out (14000g for 2min), the supernatant recovered and respun as before. The supernatant (20ml) that remained was then precipitated with 5ml of 20% (w/v) PEG/2.5M NaCl for 10 min at room temperature and the phage harvested by centrifugation (14000g at 20°C for 15min). The phage pellet was resuspended in 1ml of dH_2O, then reprecipitated and processed as for the minipreparations with the volumes scaled up accordingly.

2.3.16 Labelling of DNA with ^{32}P or Digoxigenin.

a) Labelling the 5' terminus of oligonucleotides with bacteriophage T4 polynucleotide kinase.

Bacteriophage T4 polynucleotide kinase catalyzes the transfer to a free hydroxyl group of the γ-phosphate group from ATP on the 5' terminus of DNA. In a total reaction volume of 10 µl the mixture contained :
8 pmoles of purified oligonucleotide, kinase buffer (50 mM-Tris/HCl pH 8.0, 10 mM-$MgCl_2$, 5 mM-DTT, 1 mM-spermidine), 8 pmoles (γ-^{32}P) ATP and 10 units of T4 polynucleotide kinase. The reaction was carried out at 37°C for 30 minutes by which time it had gone to completion. Unincorporated label was removed by gel filtration (section 2.3.16b).

b) Removal of unincorporated radionucleotide.

Unincorporated label from end-labelling reactions was removed by gel filtration chromatography on a 20x1 cm Sephadex G-50 column. Sephadex G-50 was hydrated in 1xTE and poured columns equilibrated in 1xTE. The reaction mixture was mixed with an equal volume of Blue Dextran dye in 1xTE and loaded directly on top of the column. The radioactivity was monitored as it passed down the column and as it approached the bottom, fractions were collected manually. Labelled oligonucleotide was co-eluted first with the Blue Dextran dye, followed by a trough then a second peak of radioactivity corresponding to unincorporated label. The fractions containing the largest number of counts were pooled.

c) Labeling of DNA using the "random priming" technique

The method of "random primed" DNA labeling is based on the hybridisation of a mixture of all possible hexanucleotides to the DNA to be labeled. The complementary strand is synthesised from the 3'-OH termini of the random hexanucleotide primer, with Klenow enzyme.

A random priming kit from Boehringer Mannheim, UK., was used for this purpose. Protocols were followed as per the manufacturer's instructions.

d) Labeling of DNA with digoxigenin-dUTP (DIG-dUTP), using the "random priming" technique.

The method of "random primed" DIG DNA is essentially the same as for labeling with ^{32}P, except that the labeling reaction was carried out overnight and not for 30min.

A random priming kit from Boehringer Mannheim, UK., was used for this purpose. Protocols were followed as per the manufacturer's instructions.

2.3.17 Southern blotting.

After electrophoresis and photography, the resolved DNA fragments were transferred under alkali conditions to Hybond-N^{TM} (adapted from Southern, 1975), as described in "Blotting and hybridization protocols for Hybond-N^{TM} membranes" published by Amersham International plc).

2.3.17.1 Reagents.

a) Denaturing solution.
1.5M NaCl, 0.5M NaOH.

b) Alkali transfer buffer.
0.25M NaOH, 1.5M NaCl.

c) 20X SSC.
3M NaCl, 0.3M tri-sodium citrate.

2.3.17.2 Procedure.

The gel was rinsed in distilled water, placed in enough denaturing solution to immerse it completely and left for 30min. The gel was removed, excess liquid removed by blotting and equilibriated for 10-15min in alkaline transfer buffer. The DNA was then transferred to the nylon membrane in alkali buffer, by capillary action (disposable nappies proved a particularly useful absorbent material for driving the transfer process).

After blotting for a least 4 hours (but usually overnight), the membrane was washed briefly in 2xSSC to remove any adhering agarose. The DNA was fixed to the membrane either by UV crosslinking using a Statagene "Stratalinker", or by baking at 80°C for 2 hours.

2.3.18 Hybridisation of oligonucleotides to filter bound nucleic acid.

Southern analysis was carried out by a method adapted from Southern (1975), and described in "Blotting and hybridization protocols for Hybond-N^{TM} membranes" published by Amersham International plc.

2.3.18.1 Prehybridisation.

Hybond-N filters were not pre-wetted, but were placed directly into a polythene bag containing prehybridisation solution (6xSSC, 0.05% (w/v) sodium pyrophosphate, 200 µg/ml heparin, 0.05% (w/v) SDS); 50 µls of the prehybridisation solution was used per cm^2 of filter surface area. The bag was placed in an agitating waterbath and the filter prehybridised for at least 4 hours at 50-65°C, depending on individual prehybridisations.

2.3.18.2 Hybridisation and washing.

After prehybridisation the bag was opened and hybridisation solution added. The hybridization solution was the same as the prehybridization solution, except that it contained 0.5% (w/v) SDS and the salt concentration was varied according to the conditions required (see results of individual hybridisations).

50 µls of the solution was used per cm^2 of filter surface area. 8.33 pmols (approximately 50 ng) of labelled oligonucleotide (2.3.16) was added to the bag before it was sealed. Hybridisation conditions were as described for each individual hybridisation.

Hybridised filters were washed three times in large volumes of buffer at ionic strengths appropriate to the experimental conditions. The temperature of this buffer was also varied experimentally. After washing, the filters were left damp, wrapped in Saran wrap and exposed to Fuji RX film using intensifying screens at -70°C. Films were developed by a Kodak X-OMAT processor.

2.3.19 Screening of plasmid clones by colony hybridisation

Recombinant pUC clones were screened using a modification of the method described in the protocol described in the Hybond-N protocol manual. Nylon filters (Hybond-N, Amersham) were placed on duplicate agar plates containing the selective antibiotic.

Bacterial colonies were crossed onto a master plate (containing antibiotic) then onto the filter containing plates. The plates were inverted and grown overnight at 37°C. Alignment marks were made on the filters. The filters were removed and placed colony side up on a pad of absorbent filter paper soaked in denaturation solution (1.5 M NaCl, 0.5 M NaOH) and left for 7 minutes.

The filters were then transferred, colony side up, to a pad of filter paper soaked in neutralising solution (1.5 M-NaCl, 0.5 M-Tris/HCl pH 7.2, 1 mM-EDTA). They were left for 3 minutes then this step was repeated with a fresh pad soaked in the same solution. The filters were then washed in 2 x SSC, transferred to dry filter paper and allowed to dry in air, colony side up. Finally, the filters were baked at 80°C for 2 hours. The filter was then hybridised with a nucleic acid probe (2.3.18.2).

2.3.20 Screening of a phage λ library

a) First round screening

Cells from a prepared bacterial suspension were infected (2.3.14.1) with phage from the bacteriophage λ library at a multiplicity of 10^4 pfu/10^8 cells. 2 x 10^3 pfu were plated onto 10 x 10 cm petri dishes using 8 ml of 0.65% (w/v) top agarose in LB. The plates were incubated overnight at 37°C.

Up to six impressions could be taken from one plate onto nylon membranes, provided time was allowed for fresh phage to diffuse to the top agarose surface (Sambrook *et al.*, 1989). Alignment marks were made on the plate and marks on the filter made using a syringe needle. Filters was removed and treated as described for colony filters (2.3.19).

b) Secondary screening

Plaques or single plaques were isolated as described in 2.3.14.3. A lawn of bacteria (10^8 cells) were plated onto a 9 mm circular petri dish and a loopful of phage particles was streaked onto the plate. The plates were then grown up overnight at 37°C. A dozen individual plaques from this plate were picked and crossed onto a 10 x 10 cm square plate overlaid with bacteria then grown up overnight at 37°C. An impression was taken of the plate onto duplicate nylon filters which were prepared as before (2.3.20a).

2.3.21 Hybridisation and detection of DIG-labelled DNA hybrids

Prehybridisation and hybridisation of filters with DIG-labelled random primed DNA was carried out as described in the protocols manual from Boehringer Mannheim, UK..

DIG-bound DNA was detected using AMPPD as a chemiluminescent substrate for alkaline phosphatase. Again, all protocols were followed as per the manufacturer's instructions.

Filters were normally exposed to X-ray film for between 2-16 hours.

2.3.22 DNA Sequencing techniques.

2.3.22.1 Sequencing reactions.

Dideoxy sequencing (Sanger *et al.*, 1977) was carried out on single-stranded M13 templates using either a Sequenase™ kit (supplied by United States Biochemical Corporation), or a TaqTrack kit (supplied by Promega), or a T7 sequencing kit (supplied by Promega).

a) Sequencing with Taq polymerase.

Reactions were mainly performed according to manufacturers instructions. However the extension and labelling reactions as well as termination reactions were carried out at 65°C to help further reduce artefacts caused by secondary structure. ^{35}S-ATP was used for labelling.

b) Sequencing with T7 polymerase.

Extension and labelling reactions were performed as suggested by the manufacturers. Termination was however carried out at 42°C instead of the normal 37°C. ^{35}S-ATP was used for labelling.

c) Autoradiography.

Autoradiography was performed in metal cassettes (medical chest X-ray type) using Kodak X-OMATS film. All autoradiography of sequencing gels was performed at room temperature.

Autoradiography at -70°C produces bands that are diffuse and therefore more difficult to read. The X-ray films were developed using a Kodak X-OMAT automatic processor, Model ME-I.

2.3.23 Denaturing PAGE for DNA sequencing.

A BRL sequencing unit (Model S2) was used for high voltage polyacrylamide gel electrophoresis.

2.3.23.1 Preparation of polyacrylamide gels.

6% (w/v) denaturing polyacrylamide gels were used for sequencing. The gels were prepared from the following stock solutions:

40% (w/v) acrylamide stock	15ml
urea	55g
10X TBE	10ml
dH_2O	35ml

The urea was dissolved by heating the mix to 37°C and then cooled to room temperature. The sequencing gel solution could be stored at 4°C for several weeks without loss of resolution. Before pouring the gel, 300µl of freshly prepared 10% (w/v) ammonium persulphate and 50µl of TEMED were added to 50ml of the stock solution.

2.3.23.2 Preparation of glass plates and pouring the gel.

The plates (40cm X 33cm) were cleaned thoroughly with alcohol and water and assembled using three spacers (0.4mm thick) along the vertical sides and the bottom of the gel. The entire assembly was held in place by clamps. The gel solution was poured from a beaker down one edge of the plates while tilting the plates at an angle of about 30°. The plates were then laid at an angle of 5° and the sharks tooth combs inserted. The gel polymerized usually within 30min at room temperature.

2.3.23.3 Electrophoresis of sequencing gels

The gel was pre-electrophoresed for 30min at a constant power of 60W. Prior to loading, the samples containing sequencing loading buffer were heated to 95°C for 5 mins, placed on ice and loaded on to the gel. 6% (w/v) gels were run for 1.75-2hrs to read the first 100 nucleotides and for 4.5-5hrs to read up to 400 nucleotides.

2.3.24 Polymerase chain reaction

DNA was amplified from a number of sources, but all following the same general procedure, detailed bleow.

Reaction cocktail:
- Unamplified DNA (genomic ~100ng, plasmid ~5ng)
- PCR buffer, containing $MgCl_2$ at concentrations of 0.5-8 mM)
- DMSO (10% v/v)
- dNTP's (1.25 mM)
- Primers (100 pmol each)
- Taq polymerase (1U)

Protocol: 1) Denature DNA (94°C for 1min)

2) Anneal primers (45-65°C depending on the annealing temperature of the primers used).

3) Extension of the DNA (72°C for 30sec-2min, depending on the length of DNA being amplified.

4) Repeat procedure 1-3 for 20-30 times.

A tenth of the PCR reaction was normally run on a 1% agarose gel to determine if amplification of the DNA had occured. DNA could be extracted and purified as described in 2.3.12 and 2.3.13.

2.4 General protein methods

2.4.1 Preparation of crude protein extracts.

a) Cell breakage

Cell pellets were resuspended in ice-cold extraction buffer (1 ml buffer/4g wet weight of cells) and broken by three passages through an automatic French pressure cell at 98 mPa (14 300 psi internal pressure). The cell was pre-cooled on ice before use (cat. no. 4-3398A, American Instruments Company, Maryland, U.S.A.).

b) Removal of cell debris.

The suspension obtained from above was subjected to centrifugation at 10,000g for 1 hour at 4°C.

2.4.2 Buffers used in purification of ICL.

a) Buffer A.

50mM MOPS-NaOH (pH 7.3), 5mM $MgCl_2$, 1 mM EDTA, 1mM benzamidine and 2 mM DTT. DTT was always added immediately prior to use from a freshly-thawed frozen stock.

b) Buffer B.

As buffer A, but with 0.6M ammonium sulphate.

c) Buffer C

As buffer A, but with 1.7M ammonium sulphate.

d) Extraction buffer.

As buffer A but with 100mM MOPS-NaOH (pH 7.3) and 1.2 mM PMSF (made up fresh in absolute alcohol) which was added to inhibit proteases. Benzamidine and PMSF were added to buffer A just before addition to cells.

e) Assay buffer

50mM MOPS-NaOH (pH 7.3), 5 mM $MgCl_2$ and 1mM EDTA was stored at 4°C.

f) Sample buffer (2X).

10 mM sodium pyrophosphate, pH 7.2, 20% (v/v) glycerol, 1% (w/v) SDS, 0.01% (v/v) bromophenol blue, 1% (v/v) 2-mercaptoethanol.

2.4.3 Chromatography media.

Phenyl Sepharose CL-4B was supplied (Pharmacia fine chemicals) pre-swollen in 20% ethanol. A 20ml column was poured and the gel washed with 10 volumes of distilled water, prior to use. Pre-packed Mono Q HR5/5 (1 ml volume) and phenyl-Superose FPLC columns were obtained from Pharmacia Fine Chemicals and used in conjunction with a Pharmacia f.p.l.c. system according to manufacturers recommendations. All buffers were filtered using Millipore-GS filters (pore size, 200 µm; Millipore S.A.Peterborough Road, Harrow U.K.). Samples were all filtered through a Millex™-GV filter unit before being loaded onto the column.

2.4.4 General biochemical methods.

a) pH measurement.

pH measurements were made with a Radiometer Model pH meter, using a combination electrode calibrated at room temperature.

b) Conductivity.

Conductivity measurements were made at 4°C with a Radiometer Model CDM2e conductivity meter.

c) Protein estimation.

Protein was determined by the method of Bradford (1976), with bovine serum albumin as standard from which a calibration curve was determined.

d) Denaturing polyacrylamide gel electrophoresis of proteins.

Electrophoresis in the presence of 0.1% (w/v) SDS was performed by the method of Laemmli (1970), with a 3% stacking gel and usually a 10% running gel. The ratio of acrylamide : bis-acrylamide was 30 : 0.8 and polymerization was induced by addition of 0.03% (v/v) TEMED and 0.05% (w/v) ammonium persulphate (freshly made). An equal volume of sample buffer (2X) was added to the sample and boiled for 2 mins. The amount of sample loaded was usually based on the protein content of the sample- ~20µg for crude samples, down to ~4µg for pure samples.

e) Protein staining.

Protein was localised on gels by staining with Coomassie blue (0.1% Coomassie brilliant blue G250 in dH_2O), for 10 minutes at room temperature. The gel was then destained in 10% acetic acid, 10% methanol at room temperature overnight. Gels were then dried down for storage.

f) Drying down destained polyacrylamide protein gels.

Slab gels were dried down for easy storage. Destained gels were placed on top of 2 pieces of Whatman no.3MM chromatography paper cut to size. Heat proof SaranWrap™ was placed over the top of this and the sandwich dried down for 30-60 mins on a Biorad Slab Gel Dryer model 1125.

Alternatively gels were dried down on to cellophane, using a kit supplied by Promega Corporation, Madison, USA. The manufacturers protocols were followed.

g) Preparation of dialysis tubing.

New dialysis tubing was boiled in 1 mM EDTA for 10 mins and stored at 4°C until needed. Tubing was rinsed in distilled water before use.

h) Assay for ICL activity.

The formation of glyoxylate by ICL was coupled to the oxidation of NADH using lactate dehydrogenase (El-Mansi, *et al.* 1987) and the corresponding decrease in absorbance at 340 nm followed at 30°C. Reactions were carried out in 50 mM MOPS/NaOH pH 7.3, 5 mM $MgCl_2$, 1mM EDTA, 5 mM DL-isocitrate, 0.2 mM NADH, 0.1 mg/ml pig heart lactate dehydrogenase. The reaction was initiated by the addition of substrate.

i) Assay for IDH activity.

The assay used was essentially the same as that of Reeves *et al.* (1972) but was performed at the growth temperature of *S.coelicolor* (30°C) instead of 37°C.

j) Standard spectrophotmetric assays.

Substrate was normally added last at an appropriate dilution to 1ml of assay buffer in a quartz cell with a 1 cm path length. The production of NADPH or oxidation of NADH was monitored at 340 nm using a Phillips PU 8720 UV/VIS spectrophotometer.

Unless otherwise stated, one unit of enzyme activity is defined as the amount of enzyme required to catalyse the disappearance of 1 µmol of substrate (or the production of 1µmol of product) per minute. Specific activity is defined as the number of units of activity per milligram of protein.

2.4.5 Subunit M_r.

SDS PAGE was used to estimate the subunit M_r of the purified protein. A molecular weight marker kit, purchased from Sigma Chemical Co., was used to provide the protein standards for calculation of the subunit M_r. The proteins used to produce standard curves of Rf against log M_r are listed below. Rf is defined as the distance travelled by the protein divided by the distance travelled by the dye front.

Protein	Subunit M_r
Rabbit muscle myosin	205,000
E. coli β-galactokinase	116,000
Rabbit muscle phosphylase B	97,400
Bovine albumin	66,000
Egg albumin	45,000
Bovine carbonic anhydrase	29,000

2.4.6 N-terminal sequence analysis by gas-phase sequencing

Outline Method:

A sample of partially-purified protein was subjected to Tricine-SDS PAGE. The gel had been pre-run in the presence of glutathione. After electrophoresis the entire gel was electroblotted to Problott membrane (Applied Biosystems) and stained with amido black. After destaining, the relevant band was cut out and presented to the sequencer. The aim was to have 100-200 pmol of protein in the excised band but 30pmol was more than enough, with 10 pmol getting close to the limit.

*CAUTION: When gel tanks had been used for glycine SDS page, they were cleaned thoroughly before use.

2.4.6.1 Tricine-SDS PAGE for proteins to be sequenced from blots

Anode Buffer:	0.2M Tris.HCl pH 8.9
Cathode Buffer:	0.1M Tris, 0.1M Tricine, 0.1% (w/v) SDS pH 8.25, tricine was used to pH the buffer.
Gel buffer:	3M Tris.HCl pH8.45, 0.3% (w/v) SDS
Acrylamide solution:	48 gram acrylamide,1.5 gram bisacrylamide in 100ml water
Sample buffer:	SDS (2% w/v), 2-mercaptoethanol (2% v/v), bromophenol blue (0.01% w/v), glycerol (10% V/V)
Separating gel:	2ml acrylamide solution, 3.3ml gel buffer, 1.3 gram glycerol (1ml 98% glycerol), adjusted to 10ml with 3.4ml distilled water. 75µl 10% (w/v) ammonium persulphate and 7.5µl TEMED was then added, to faciltitate gel polymerisaton.
Stacking gel:	0.35ml acrylamide solution,1.03 ml gel buffer,adjusted to 4.17ml with 2.79ml distilled water. 40µl 10% (w/v) APS and 7.5µl TEMED was then added, to facilitate polymerisation of the gel.

Running Conditions: 100 V (constant), until dye reaches the bottom of the gel.

2.4.6.2 Electrophoresis conditions

1. Prerun electrophoresis: 600ml of glutathione solution (10mM, stored at -20^{o}C) was added to 120 ml of electrophoresis buffer in the upper reservoir. The power supply was set at 6 mA constant current and operated for 1 hour.
2. Sample preparation: One volume (1-10µl) of sample solution was mixed with one volume of sample preparation solution and heated in boiling H_2O for 2 min. 1-3µl of 0.1% bromophenol blue was added and the sample loaded into the sample well.
3. Electrophoresis: The prerun buffer was decanted off and the reservoirs filled with fresh buffer. 120ml of sodium thioglycollate (100mM) was added to 120 ml of electrophoresis buffer in upper reservoir. The power was set on constant current 100 V and run for about 1 hour or until dye reached the bottom of the gel.

2.4.6.3 Electroblotting

Stock Buffer:(100 mM, pH 11.0) CAPS (3-[cyclohexylamino]-1- propanesulfonic acid. 22.13g CAPS was dissolved in 900 mL of dH_20. 2N NaOH was then added dropwise to bring the pH to pH11.0. dH_20 was then added to make a final volume of 1L. The solution was stored at 4°C.

Electroblotting buffer: 10 mM CAPS buffer in 10% Methanol.

2 litres of buffer were prepared by mixing 200ml of the 10X CAPS buffer with 200mL of methanol and 1.6L of dH_20.

Procedure:

1. Gel removed from the electrophoresis cell and soaked in 100mL of electroblotting buffer for 5 minutes.
2. ProBlott^{TM} prewetted with 100% methanol for a few minutes and transfered to a dish containing the blotting buffer.
3. The sponges and filter papers were soaked in a separate container of blotting buffer before starting to assemble the transblotting sandwich.
4. The transblot sandwich was assembled in the following order starting from anode side:
   sponge, filter paper, 1 sheet of ProBlott^{TM}, gel, filter paper, sponge.
5. 1L of electroblotting buffer was poured into the transblot cell and the transblot sandwich inserted. The proteins were electroblotted at around 250 mA constant, at room temperature for 75 minutes.
6. The ProBlott^{TM} membrane was removed from transblotting sandwich and rinsed with deionised water before staining.

2.4.6.4 Amido black staining

0.1g of Amido Black was dissolved in 40mL of methanol. This solution was stired one hour. 1mL of acetic acid and 59mL of D.I. water, was then added and the solution stired for another 30 minutes. The solution was then filtered with a Nalgene filter (0.45μm pore size).

Staining Procedure:

1. The ProBlott™ membrane was removed from the transblot sandwich, and rinsed with deionised water.
2. The ProBlott™ membrane was saturated with 100% MeOH for a few seconds.
3. The ProBlott™ membrane was placed in the staining solution using constant orbital shaking. Protein bands appeared within one minute.
4. The ProBlott™ membrane was then destained by soaking in deionised water. The destaining solution was changed several times, before the membrane was hung up to dry.

2.4.7 Proteolytic digestion of ICL

Protein was first run on a conventional glycine SDS gel, to seperate ICL from any other contaminating proteins. The gel was stained with coomassie to visualise the proteins. The band corresponding to ICL was excised in order that ICL could be subjected to proteolysis.

A second gel was then made. Two gel slices containing ICL were loaded into each well. To this was added 10μl of a solution containg 1vol gel buffer (2.4.6.1), 1vol glycerol and 3vol proteolytic enzyme (chymotrypsin, V8 protease, elastase, or trypsin).

The gel was then run until the protein reached the stacking/ running gel boundary. The voltage was then switched off and the protein allowed to be digested for 30min. The gel was then restarted and run until the dye reached the bottom of the gel.

The proteolytic fragments could then be visualised by staining, or blotted as in 2.4.6.3, before being visualised and sent for amino sequencing.

2.4.8 Amino acid sequencing

Amino acid sequencing was carried out by Mr. Bryan Dumbar at the University of Aberdeen by automated Edman degradation on an Applied Biosystems gas phase sequencer.

Chapter 3

Attempts to clone *S.coelicolor icl* using primers designed against consensus sequences

3.1 Introduction

At the start of this project, sequences of 3 *icl* genes were available. After alignment, it was observed that there were regions of considerable similarity at the amino acid level. The initial task was to design an oligonucleotide (incorporating streptomycete codon bias) encoding such a similar region. It was then envisaged that this oligonucleotide could be used to facilitate the cloning of the *S.coelicolor icl*, by hybridisation of the oligonucleotide to genomic DNA from *S.coelciolor.*

This chapter describes the design of such an oligonucleotide and its use as a probe in attempts to clone a region of hybridising genomic DNA. As a result of sequencing the cloned DNA, it was deduced that the DNA did not encode *icl.*

The chapter then goes on to describe the design of two oligonucleotide primers encoding two conserved regions of the ICL proteins and how these primers were used in PCR studies to attempt to amplify a region of the *S.coleicolor icl.* Again, sequencing of the PCR products showed that the amplified DNA did not encode *icl.*

Lastly the chapter details the N-terminal sequencing of a partially-purified protein initially thought to be the ICL of *S.coelicolor.*

3.2 Results and discussion

3.2.1 Design of an oligonucleotide against conserved regions of the ICL proteins (oligo *icl*)

As mentioned above, 3 *icl* sequences were available at the commencement of this project. These were from *E.coli* (Matsuoka and McFadden, 1988), Castor bean (Beeching and Northcote, 1987) and *Aspergillus nidulans* (G.Turner, personal communication).

These sequences were aligned by eye and a region around a proposed active site cysteine (Nimmo *et al*, 1989) was oserved to be particularly conserved (Fig 3.1). An oligonucleotide was designed using the codon usage tables compiled from streptomycete genes previously sequenced (Seno and Baltz, 1989). A more detailed description of the rationale behind the design of oligonucleotides for streptomycete DNA can be found in 5.2.1.

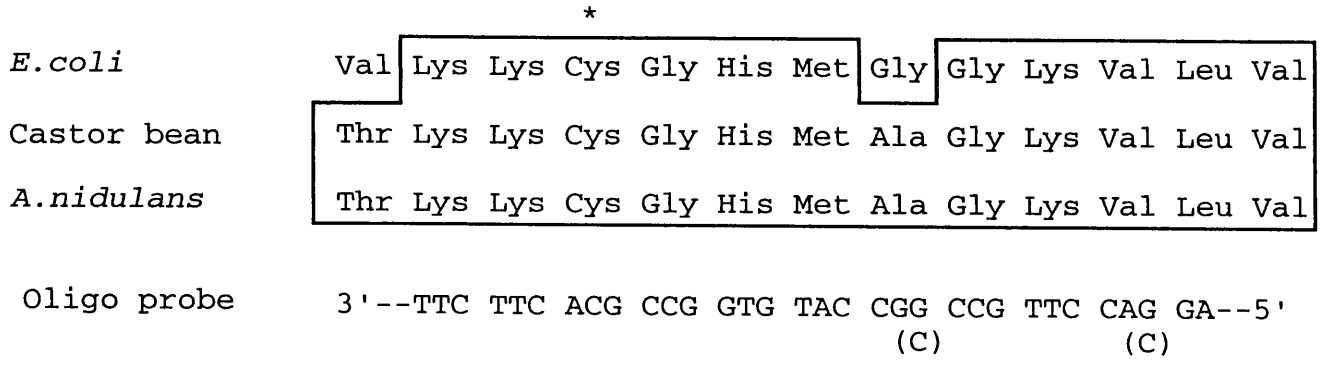


Figure 3.1 Design of the oligonucleotide, oligo *icl*

This diagram shows regions of the three icl gene sequences which were available at the start of this project. The three gene sequences were studied to identify if they contained regions of similarity. The region shown is a region of particular similarity. The putative active site cysteine is highled by an asterix.

The codon usage table of Seno and Baltz (1989) was used to identify the codons for each amino acid, which were used most frequently in streptomycete genes. Using the rationale detailed in section 5.2.1, an oligonucleotide was designed and synthesised, to be used in hybridisation studies.

3.2.2 Probing of genomic DNA from *S.coelicolor* with oligo *icl*

10µg of *S.coelicolor* genomic DNA was digested to completion using a selection of restriction enzymes. The digested DNA was then separated on a 1% agarose gel. DNA was then transfered to Hybond-N nylon membrane using alkaline transfer and the DNA fixed to the membrane by baking at 80°C for 2 hours. The membrane was then prehybridised as described in 2.3.18.1.

After prehybridisation, the prehybridisation solution was removed and replaced with hybridisation solution containing radiolabelled oligonucleotide- oligo *icl*. Hybridisation was carried out for 2 hours before the filters were washed three times (2.3.18.2).

Different hybridisation and washing conditions had to be tried out. Washing conditions were gradually altered by altering the salt concentration and temperature, until conditions were found for which only a single hybridising DNA fragment was observed for each restriction enzyme digest.

The optimised conditions for the oligonucleotide-oligo *icl*, were found to be a hybridisation temperature of 55°C, 5xSSC, and washing at 60°C, in 4xSSC.

This work was originally started by Miss Alison Wylie and then carried on in conjunction with her, until a suitable band for cloning was identified, at which time the work was passed onto myself.

3.2.3 Cloning of size selected *Pst* I digested genomic DNA from *S.coelicolor*

Hybridisation studies revealed a single *Pst*I hybridising band of approximately 4.5kb. The expected size of ICL was between 48-70kDa, which would be encoded by a DNA fragment of between 1.2-1.9kb. It was predicted that there was a resonably high probability that the 4.5kb *Pst*I fragment could encode the entire *icl* gene.

10µg of *S.coelicolor* DNA was digested with *Pst*I and separated on a 1% agarose gel. DNA corresponding to approximately 4.3-5kb was then excised from the gel and purified using electroelution (2..3.12).

The plasmid pIBI24 was digested with *Pst* I, to linearise the plasmid. The DNA

was then subjected to electrophoresis and the linearised plasmid band excised from the gel. The DNA was then purified by electroelution.

A ligation was set up with an insert to vector concentration ratio of 5:1. *E.coli* DS941 was made competent by the method of Hanahan (2.8.1.1) and the ligated DNA used to transform the cells. Recombinant colonies were detected by the white colour of the colony when plated onto media containing X-gal.

Since the plasmid was only digested with a single restriction enzyme, a high proportion of the plasmid religated without any insert DNA. Thus, a second ligation was set up with plasmid which had first been dephosphorylated using alkaline phosphatase (2.3.9).

400 colonies were picked and plated onto square L-agar plates containing ampicillin, as a 10x10 array on 4 plates. At the same time, duplicate arrays were also set up on plates which had Hybond-N placed on the surface of the media.

After the colonies had grown, the plates containing the Hybond-N were subjected to further experimentation. The colonies were lysed and the DNA denatured, neutralised and fixed to the membrane, according to 2.3.19.

The filters were subjected to prehybridisation, followed by hybridisation using oligo *icl*. The optimised conditions described in 3.2.2 were followed.

3 colonies out of the 400 gave positive signals (Fig. 3.2). One recombinant was chosen for further study. The plasmid from this colony was prepared (pAUL1) for further experimentation.

3.2.4 Restriction mapping of pAUL1

pAUL1 was digested with various restriction enzymes which were known to cut within the pIBI24 polylinker. The DNA framents were separated on an agarose gel (Fig. 3.3a) and a restriction map determined (Fig. 3.3b).

To determine which fragments should be sub-cloned for sequencing purposes, southern hybridisation studies were carried out on various digests of pAUL1. These studies revealed that the oligo *icl* hybridised to the 2.5kb *Bam*HI/*Bst*EII fragment (Fig. 3.3).

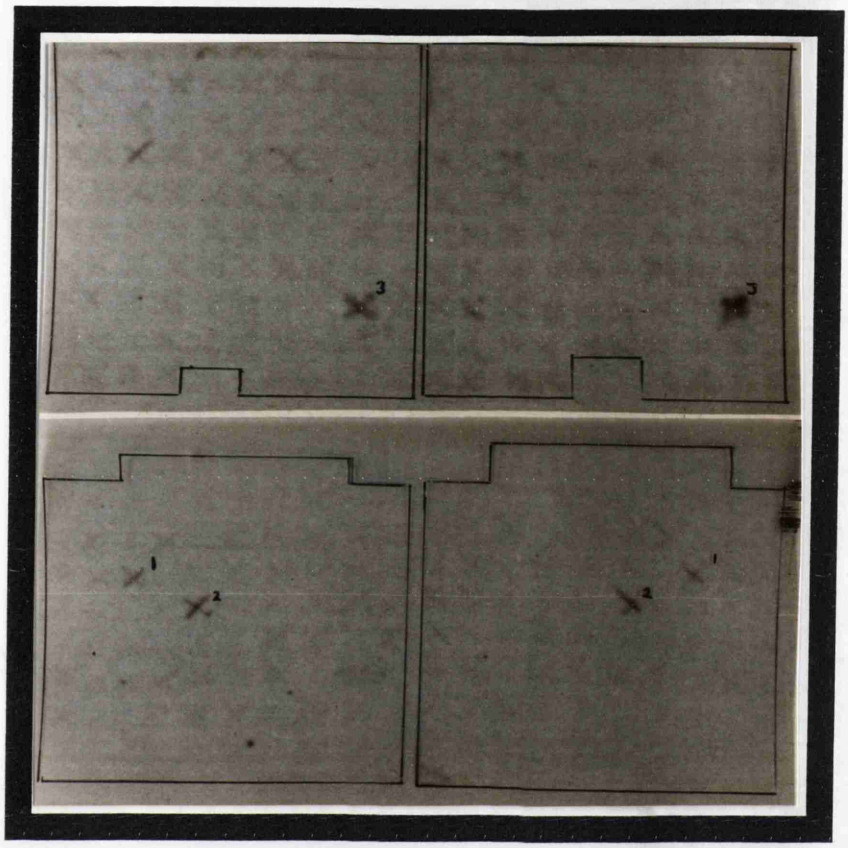


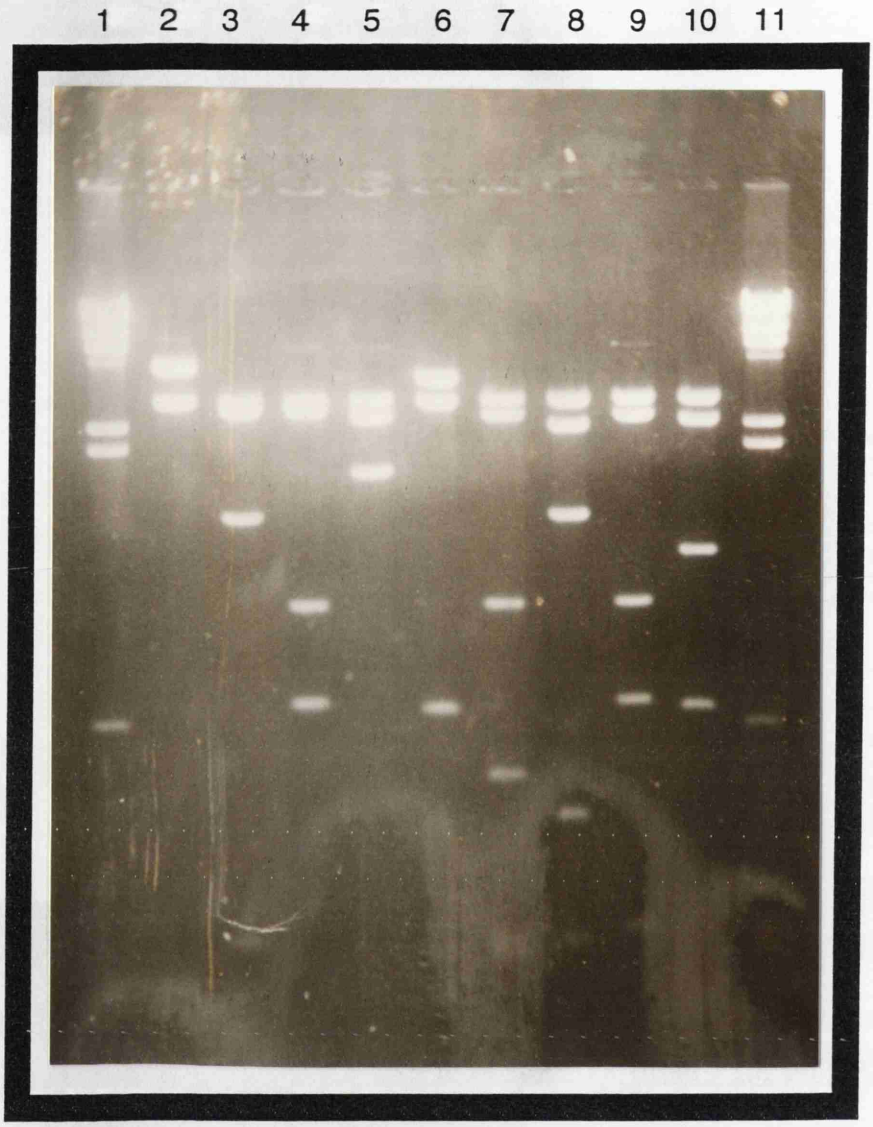
Figure 3.2 Colony hybridisation

400 colonies of *E.coli* were grown in duplicate on Hybond-N. The cells were then lysed and the DNA fixed onto the membrane.The filters were then probed with ^{32}P-labelled oligo*icl*.

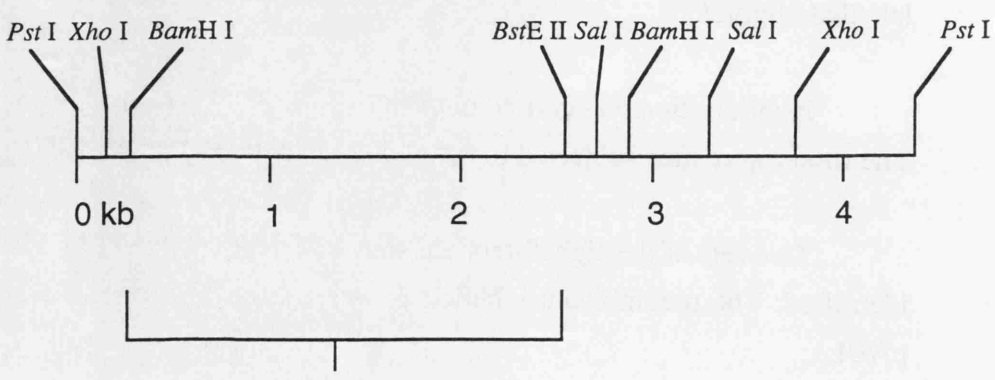
Hybridisation was carried out at 55°Cin 5xSSC. Washing was at 60°C in 4xSSC. The filters were then subjected to autoradiography overnight at -70°C.

Two sets of the eight filters are shown. Three positively hybridising colonies were identified. The recombinant labelled 3, was chosen forfurther studies and was named pAUL1.

a)



b)



Region to which oligonucleotide remains bound

Figure 3.3 Restriction map of pAUL1 and determination of the oligo *icl* binding region

a) Plasmid pAUL1 was digested with various restriction enzymes, which were known to cut within the pIBI24 polylinker (see below). The resulting fragments were separated on a 1% (w/v) agarose gel.

Lane	Digests
1	λ *Hind*III molecular weight markers
2	*Pst*I
3	*Pst*I/*Bam*HI
4	*Pst*I/*Sal*I
5	*PstI*/*Bst*EII
6	*Pst*I/*Xho*I
7	*Pst*I/*Bam*HI/*Sal*I
8	*Pst*I/*Bam*HI/*Bst*EII
9	*Pst*I/*Sal*I/*Bst*EII
10	*Pst*I/*Bst*EII/*Xho*I
11	λ *Hind*III molecular weight markers

b) A restriction map was determined from the above digests. Southern hybridisation was carried out on a blotted gel and a region was identified which still allowed hybridisation of the oligo *icl*, using the conditions detailed in 3.2.2..

3.2.5 Exonuclease III deletions of pAUL1

To identify a smaller region of DNA which hybridised to oligo *icl* and reduce the size of DNA to be sequenced, exonuclease III digests were carried out on pAUL1. It was envisaged that a series of clones reducing in size would be obtained. The protocol followed was that found in the Promega protcols and applications guide, using the reagents of the Erase-a-base system. The temperature used to generate the deletions was 37°C.

A gel of the various deletion time points revealed that the deletion reactions had proceded correctly. However the susequent transformation yielded very few transformants. This was probably due to inefficient blunt-ending of the DNA using S1 nuclease and Klenow enzymes. The transformants were screened and six deletion subclones were obtained.

Plasmid DNA was prepared from each of the deletion sub-clones and the DNA digested to separate the plasmid DNA from the cloned *S.coelicolor* DNA. This DNA was separated on a 1% agarose gel and visualised by staining with ethidium bromide staining (Fig. 3.4 and Fig. 3.5).

The gel was then blotted onto Hybond-N before being subjected to hybridisation studies using oligo *icl*. Again, the optimised conditions described in section 3.2.2 were followed. This revealed that the oligo *icl* hybridised to the three largest deletion sub-clones. From this it was deduced that the oligo *icl* must bind within the ~1kb of DNA which was present in pDel2, but not in pDel7 (Fig. 3.5).

3.2.6 Sequencing of the 1kb hybridising region

Double stranded sequencing of the plasmids pDel2 and pDel7 was carried out according to 6.2.2. Oligonucleotides were designed against the sequence obtained and synthesised, to enable futher sequencing. This strategy was followed until sequence spanning the entire 1kb region had been determined (Fig. 3.6)

This DNA was then analysed at the level of the putative translation product to identify any possible similarity to the other sequenced *icl* genes. However, no similarity was observed. The best oligo*icl* binding region contained 18bp matches out of a possible 32, which shows a poor homology of the oligonucleotide to the sequenced DNA.

The fact that the other sequenced ICL's showed a great deal of similarity, suggested that the DNA cloned from *S.coelicolor* did not encode *icl.*

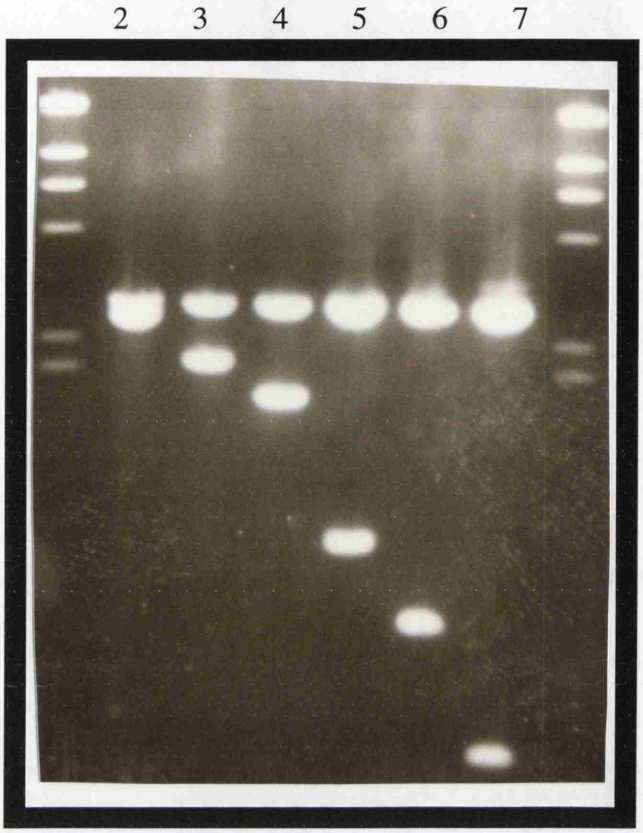


Figure 3.4 Exonuclease III deletion clones from pAUL1

*Sal*I/*Sst*I digested plasmid pAUL1 was subjected to digestion using exonuclease III, for varying lengths of time.

This resulted in the generation of six deletions clones.The clones were digested with *Hind*III and *Eco*RI and separated on an agarose gel

Lane	Exonuclease III deletion clone
2	pDEL 1
3	pDEL 8
4	pDEL 2
5	pDEL 7
6	pDEL 5
7	pDEL 4

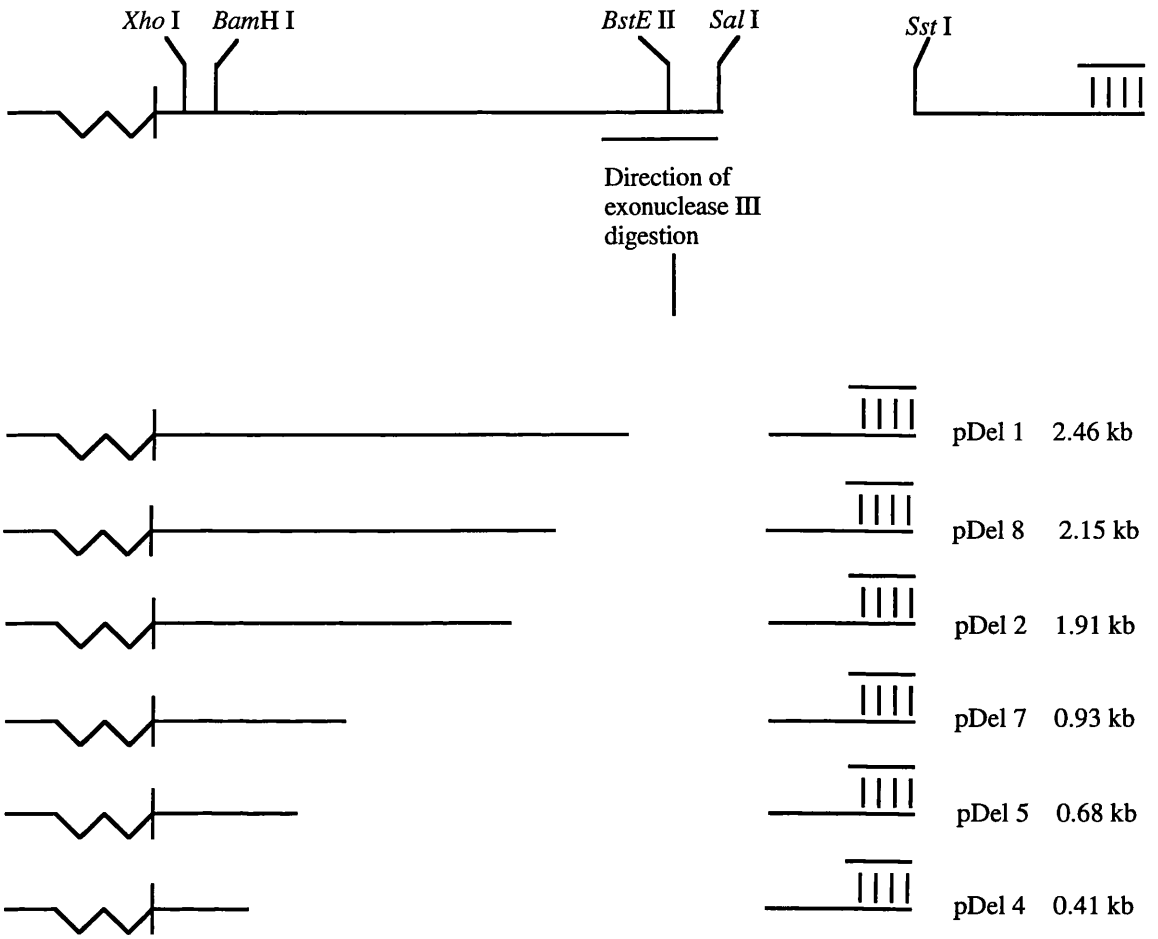


Figure 3.5 Diagramatic representation of exo deletions

Plasmid pAUL1 was digested with *Sst*I and *Sal*I. The *Sal*I digest generates a 5' overhang, which is susceptible to digestion by exonuclease III digestion. The digestion reactions were carried out at 37°C and the reaction stopped at various time points.

This diagram represents the six deletion clones which were obtained and the relative sizes of each insert.

```
1    GGTCATCCTG CTCGACACCT TCGAGGACCT CGGGGACCGT ACCCACCGTG
51   ATCTCCAACG CCTCATCCAG CGCGTGGTGT GGCTCATGCC GAACGCGTTC
101  TTCGTCGTCA CCGGCCGCTC CCGCCTTTAG TGGGCCGTTT CCGCCCTCCA
151  GGGCCAGCTC GACTACACCG GCCCCGCGGC CTGGCCCXGG CCTCCAGCCG
201  GCTCCCTCCC CCACGCCCGC TCCGGCAGCG GCACCGGCGA GCGGACGGGC
251  GGGCGGCAGG TGCTGATCGG GGACTTCTCC CCCXGAGGAC TGCGACGCTA
301  CCTCGCGCGC CGGCTCGCCC GCAACGGCCG GCCCCTCATC CGCGACGAGA
351  TCCGGCAGGT CATCAGCGGC CGGTCCCACG GACTGCCCCT CTACCTGGAC
401  CTGTCGGTAT GCGGTTCCTG GAGATCCGCC GCGCCGGAAC GGGAGCCGGC
451  ATCGGCCGAC TTCGACCACG ACTTCCCCGC CCTGATCGCC CGCACGCTGT
501  CCGACCTCAC CCGCGAGGAG CGGCACGTCC TGCGGTCCGT CAGCCTGCTG
551  GACGCCTTCG ACCTCTCCCT CGCCACCGCG GCCGCGGGCC TGCCCCACCA
601  GGCGCCGGCG CTGCGGCTGA TCGAGCGGCC CTTCGTCCGG GAGAATCCGG
651  TCGGGCTGTG GCCCTTCCAC CTGCACGGGC TGATCCGCTC CACCGTCCGC
701  GGTGCCGACG ACCAGACCGA CGACCGGTGG AGCCCCGGGA CTGGCAGCAG
751  GCCGCCGAGC GAAGCCACGC GGCTCCTCGG CGAGCAGTGG CGTGCCGGCA
801  CCGGCCGGGA CCGGGCCCTG CTGGTCGGCT GCCTGCGTCA GGGCTGGGCC
851  TGGGCCTGGC CCGCGACTTC CGCCTGAATC TGGGCTGGGT TATGGATGCC
901  GCCTGGGACT ACGTCGGCGA CTCCGTGTGG GAACCCGTCG CACCGCCGGG
951  CCGGCAGGAC CGGGCCGCTG TGGAGACCCC CGCGGACGGC CTCGCCGAGC
1001 TTCTCAGTGC CCTGTCCCGA CGCCAGCACG AGCACCGCGC CCGCACCGCC
1051 TACCGGC
```

Figure 3.6 Sequence of the oligo binding region of pAUL1

This represents the nucleotide sequence obtained from sequencing the DNA, which was shown to bind the oligo *icl.* The region underlined shows the best binding region (by homology) of the oligo*icl.*

3.2.7 Amplification of *S.coelicolor* DNA using oligonucleotides designed against other *icl* gene sequences

By the time it had been identified that the cloned *S.coelicolor* DNA did not encode for *icl*, two further *icl* sequences had been published. These further sequences were aligned, along with the previously sequenced *icl*'s. The same region of DNA that had been previously identified (Fig. 3.1), appeared to be highly similar across all 5 sequences. Another region of similarity was also identified near the C-terminus of the enzyme.

Two PCR primers were synthesised; one against the sequence around the active site (iclPCR1), the second around the second region (near the C-terminus) showing a high degree of similarity (iclPCR2; Fig. 3.7). The strategy was to attempt to amplify a region of DNA which encodes an internal fragment of *icl,* and then to use this in hybridisation studies to attempt to clone the entire *icl* gene.

Reactions were set up as described in 2.3.24. The melting temperatures of the two primers were calculated (Fritsch unpublished; cited in Sambrook *et al*, 1989) and were found to be approximately equal; ~54°C. The annealing temperature of the reaction was thus set at 51°C. A series of reactions with magnesium concentrations ranging from 0.5-8mM, were set up.

A set of control reactions were also set up; two single primer controls, one control using *E.coli* genomic DNA and one control using *A.nidulans* genomic DNA.

It was estimated that the size of the PCR product would be between 480-780 nucleotides in length (as determined from the other known *icl*'s). Two amplified DNA fragments within the predicted size range were obtained (Fig. 3.8). It was observed that one of the bands appeared to be amplified in a single primer control. However, both fragments were purified and cloned into pDK101 (Kovalic *et al*, 1991).

The control experiments using the genomic DNA from *E.coli* and *A.nidulans* both yielded bands of the predicted size (Fig. 3.8).

Double stranded sequencing of the cloned PCR fragments was carried out to determine if the amplified DNA did indeed encode an internal fragment of *S.coelicolor icl* (Fig. 3.9). The sequences obtained, revealed that both fragments had been amplified from a single primer (iclPCR1).

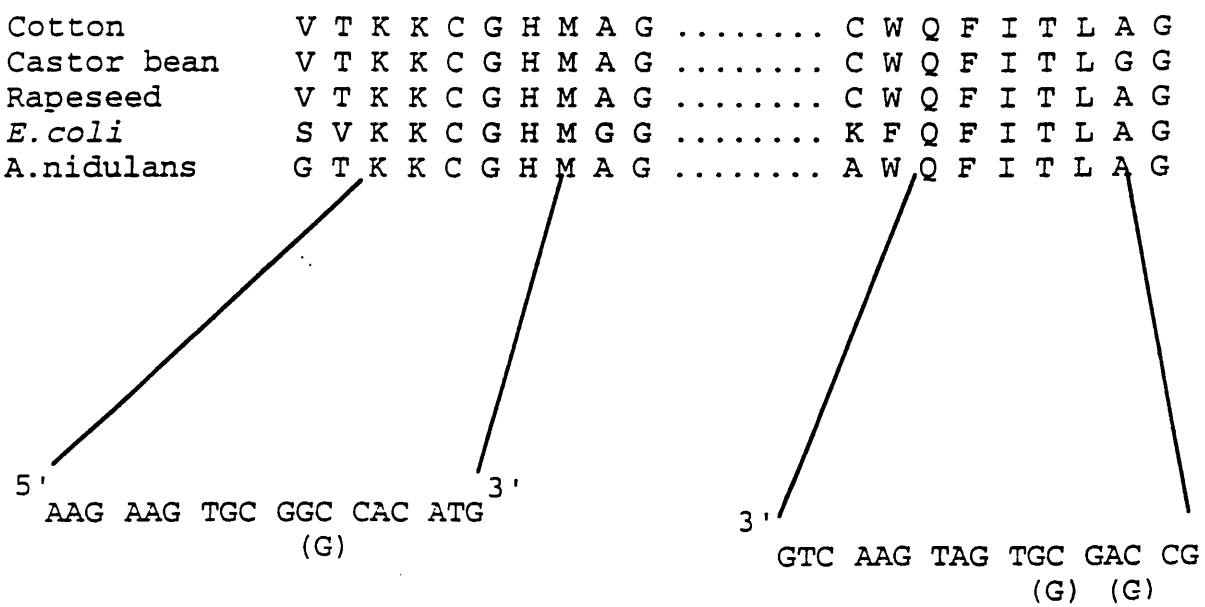


Figure 3.7 PCR primers designed against regions of various ICLs which show a high degree of similarity

Five ICL protein sequences were aligned using the PILEUP program from the GCG sequence analysis package. The sequences were then studied to identify regions of the proteins which were highly similar. Two such regions were observed and two oligonucleotides were designed (which incorporated Streptomycete codon bias), against these regions. The oligonucleotides were synthesised such that they could be used in a PCR, to attempt to amplify a region of the *S.coelicolor icl*

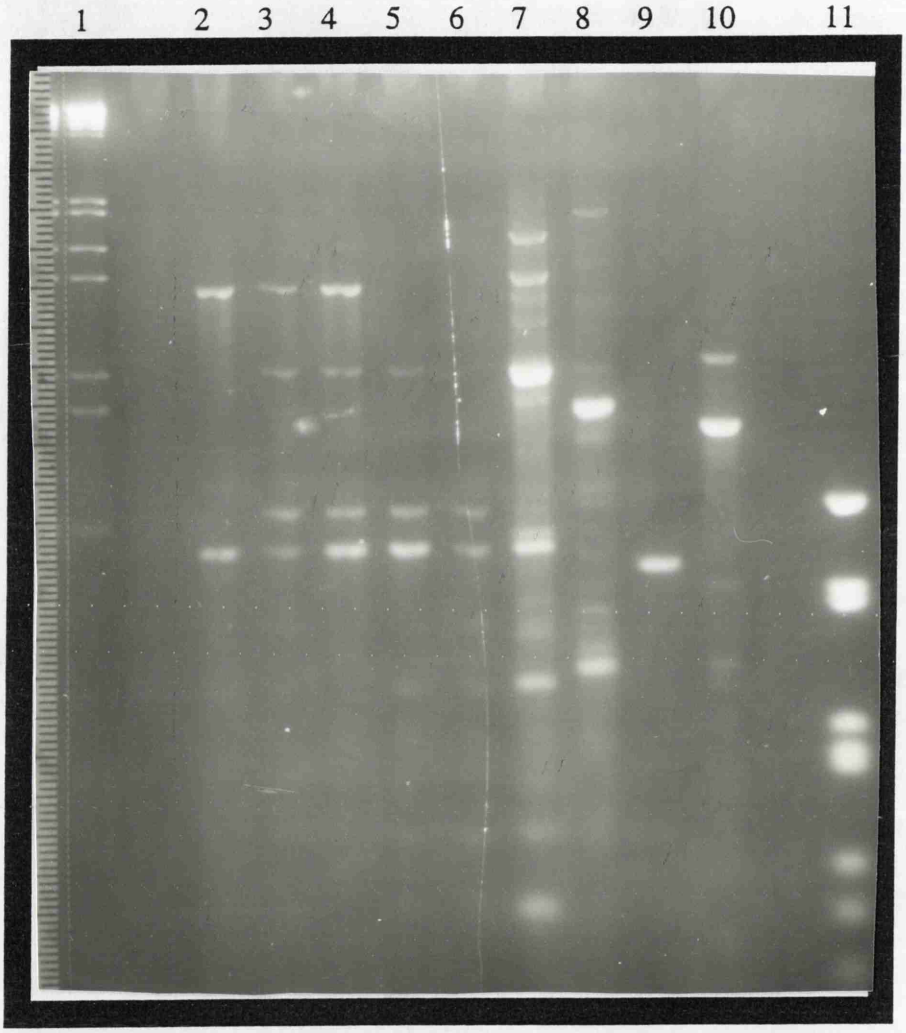

1
2
3
4
5
6
7
8
9
10
11

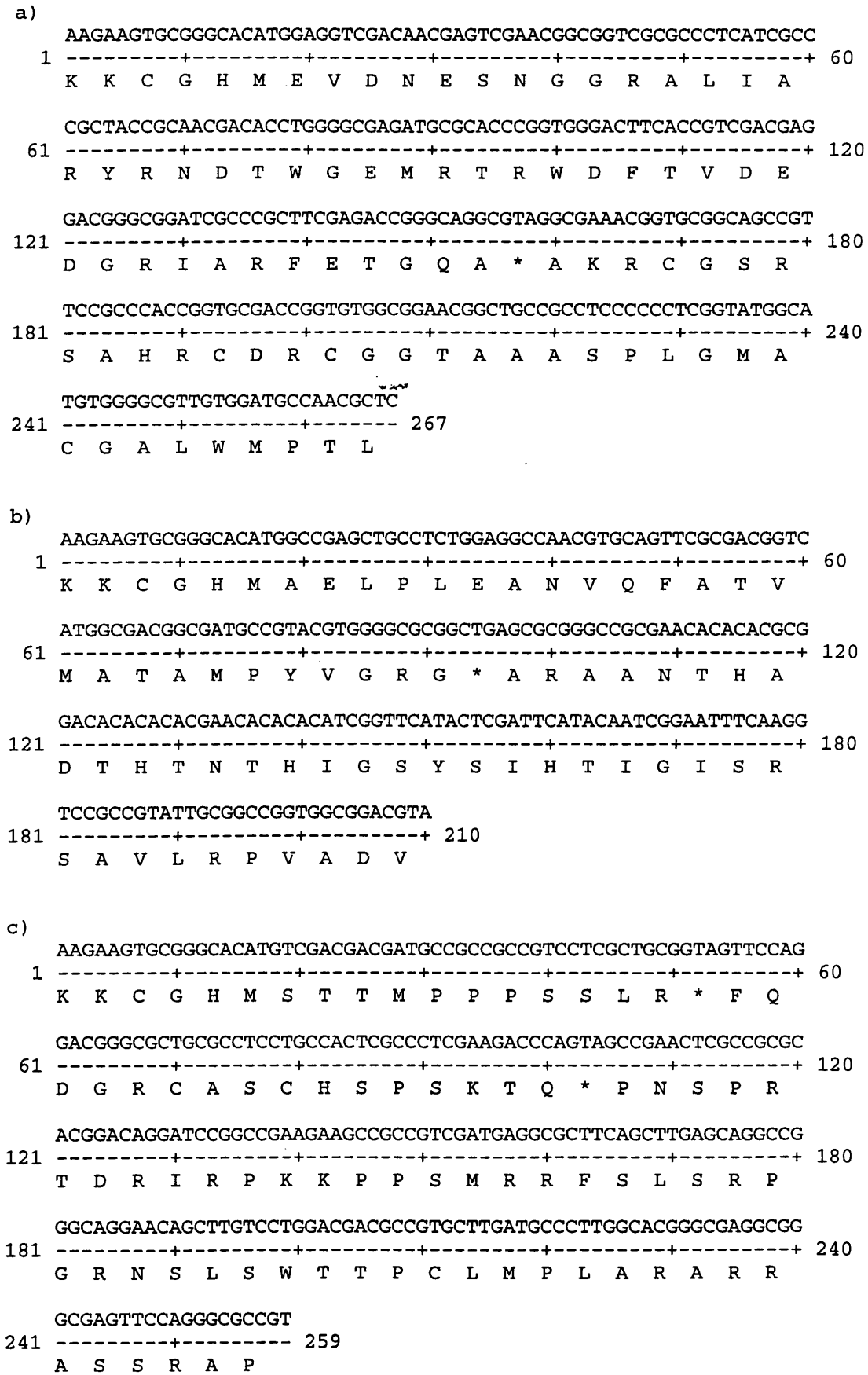
Figure 3.8 Amplification of genomic DNA from *S.coelicolor* using primers designed against an alignment of various ICL protein sequences

100ng of genomic DNA from *S.coelicolor* was subjected to amplification using primers which had been designed against aligned regions of ICLs. The regions had shown a high degree of similarity, when the protein sequences were aligned.

The reactions were set up as described in 2.3.24, but with various magnesium ion concentrations.

Control reactions were also set up (single primer controls and genomic DNA from *E.coli* and *A.nidulans*.).

Lane	Reaction	Mg^{2+} concentration
1	λ *Hind*III/*Eco*RI markers	---
2	*S.coelicolor* DNA	0.5 mM
3	*S.coelicolor* DNA	1mM
4	*S.coelicolor* DNA	1.5 mM
5	*S.coelicolor* DNA	2 mM
6	*S.coelicolor* DNA	4 mM
7	*S.coelicolor* DNA (primer 1)	1.5 mM
8	*S.coelicolor* DNA (primer 2)	1.5 mM
9	*E.coli* DNA	1.5 mM
10	*A.nidulans* DNA	1.5 mM
11	pIBI25 *Hae*III digest	---


a)
AAGAAGTGCGGGCACATGGAGGTCGACAACGAGTCGAACGGCGGTCGCGCCCTCATCGCC
1
60
K K C G H M E V D N E S N G G R A L I A
CGCTACCGCAACGACACCTGGGGCGAGATGCGCACCCGGTGGGACTTCACCGTCGACGAG
61
120
R Y R N D T W G E M R T R W D F T V D E
GACGGGCGGATCGCCCGCTTCGAGACCGGGCAGGCGTAGGCGAAACGGTGCGGCAGCCGT
121
180
D G R I A R F E T G Q A * A K R C G S R
TCCGCCCACCGGTGCGACCGGTGTGGCGGAACGGCTGCCGCCTCCCCCCTCGGTATGGCA
181
240
S A H R C D R C G G T A A A S P L G M A
TGTGGGGCGTTGTGGATGCCAACGCTC
241
267
C G A L W M P T L
b)
AAGAAGTGCGGGCACATGGCCGAGCTGCCTCTGGAGGCCAACGTGCAGTTCGCGACGGTC
1
60
K K C G H M A E L P L E A N V Q F A T V
ATGGCGACGGCGATGCCGTACGTGGGGCGCGGCTGAGCGCGGGCCGCGAACACACACGCG
61
120
M A T A M P Y V G R G * A R A A N T H A
GACACACACACGAACACACACATCGGTTCATACTCGATTCATACAATCGGAATTTCAAGG
121
180
D T H T N T H I G S Y S I H T I G I S R
TCCGCCGTATTGCGGCCGGTGGCGGACGTA
181
210
S A V L R P V A D V
c)
AAGAAGTGCGGGCACATGTCGACGACGATGCCGCCGCCGTCCTCGCTGCGGTAGTTCCAG
1
60
K K C G H M S T T M P P P S S L R * F Q
GACGGGCGCTGCGCCTCCTGCCACTCGCCCTCGAAGACCCAGTAGCCGAACTCGCCGCGC
61
120
D G R C A S C H S P S K T Q * P N S P R
ACGGACAGGATCCGGCCGAAGAAGCCGCCGTCGATGAGGCGCTTCAGCTTGAGCAGGCCG
121
180
T D R I R P K K P P S M R R F S L S R P
GGCAGGAACAGCTTGTCCTGGACGACGCCGTGCTTGATGCCCTTGGCACGGGCGAGGCGG
181
240
G R N S L S W T T P C L M P L A R A R R
GCGAGTTCCAGGGCGCCGT
241
259
A S S R A P

d)

```
    AAGAAGTGCGGGCACATGGGCTACCGCCAGCACCTCGTCCGCTCCATCCTGGCCCTGCGC
  1 ---------+---------+---------+---------+---------+---------+ 60
    K  K  C  G  H  M  G  Y  R  Q  H  L  V  R  S  I  L  A  L  R

    GAACAGGGCGGCCTCGACCTCGGCGACGGCATCGTGCTGTGGCCGGAGCCGGTC
 61 ---------+---------+---------+---------+---------+---- 114
    E  Q  G  G  L  D  L  G  D  G  I  V  L  W  P  E  P  V
```

Figure 3.9 Sequences from the PCR products, obtained using primers designed against other *icl* sequences

Two PCR products (of the predicted size) were obtained, following amplification of genomic DNA from *S.coelicolor* A3(2), using primers designed against previously aligned *icl* sequences.

The PCR products were cloned into pDK101 (Kovalic *et al*, 1991) and regions of the inserts were sequenced using the universal primer and the reverse primer.

a) Sequence of the smaller PCR product (~500bp) obtained using the universal primer.

b) Sequence of the smaller PCR product (~500bp) obtained using the reverse primer.

c) Sequence of the larger PCR product (~600bp) obtained using the universal primer.

d) Sequence of the larger PCR product (~600bp) obtained using the reverse primer.

This result showed that the amplified DNA did not encode part of *S.coelicolor icl.* The DNA was translated from the appropriate frame, as determined using CODONPREFERENCE and then compared on the protein database. Database searching did not reveal any significant similarity to any other known proteins.

3.2.8 N-Terminal sequence analysis of a purified protein, thought to be *S.coelicolor icl*

Javier Rua, a visiting postdoctoral biochemist, had set out to purify the ICL protein from *S.coelicolor.* He found great difficulty in growing *S.coelicolor* on acetate and poor levels of ICL activity were obtained. However purification of ICLwas attempted and a four step protocol resulted in partial purification of a protein, which was thought to be ICL (Fig. 3.10).

The partially purified protein was passed to myself for blotting and subsequent N-terminal sequencing. Approximately 20μg of the protein was separated on a 10% SDS-tricine gel (2.4.6.1), before being blotted onto Problott membrane (2.4.6.2). The blotted protein was then visualised by staining with amido black.

The protein band which was thought to be ICL was then excised from the membrane and sent to Mr. Bryan Dunbar at the SERC protein sequencing facility in Aberdeen. Approximately 100pmol of protein was subjected to automated sequencing using an Applied Biosystems model 470 gas phase sequencer. 30 residues were obtained.

These residues were then compared to the N-terminal sequences of other ICLs, but no similarity was obtained. Database searching revealed that the sequence was very similar to the N-terminal region of lipoamide dehydrogenase from a number of organisms (Fig. 3.11). It was thus considered that the purified protein was probably the lipoamide dehydrogenase from *S.coelicolor.*

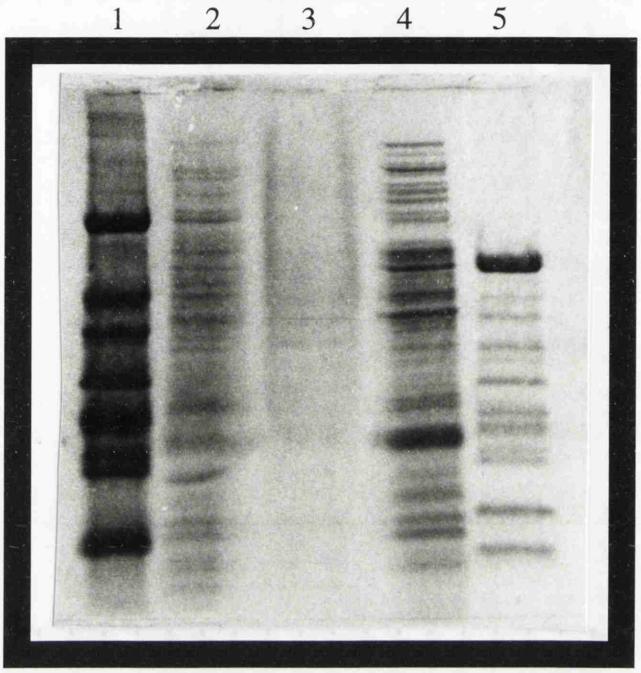


Figure 3.10 Partial purification of the putative ICL from *S.coelicolor*

Javier Rua partially purified a protein which was thought to be the ICL from *S.coelicolor*.

Protein from the various purification steps was run on a 10% SDS-PAGE gel and is shown above. The strongest band, visible in lane 5, was thought to be ICL.

Lane	Purification step
1	Molecular weight markers
2	Crude extract
3	Ammonium sulphate cut
4	Phenyl sepharose
5	Mono Q

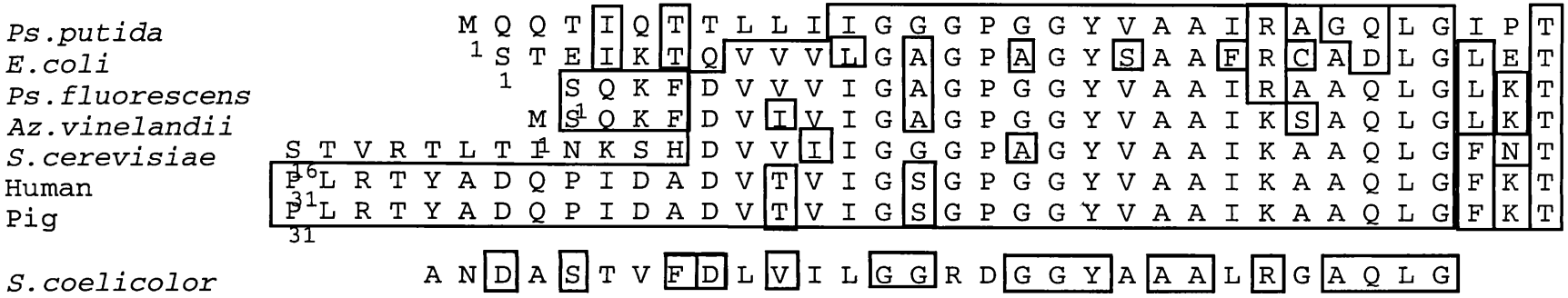


Figure 3.11 N-terminal sequence of the purified putative *S.coelicolor* ICL and alignment with N-terminal regions of various lipoamide dehydrogenases

A protein that was thought to be the *S.coelicolor* ICL was partially purified and subsequently blotted onto Problott membrane. This was then sent to Mr. Bryan Dumbarf or N-terminal sequencing of the protein. Thirty residues were obtained, as can be seen above.

Alignment with the N-terminal regions of the previously sequenced ICLs did not reveal any similarity. On searching the Swissprot protein database, (using the FASTA program from the GCG sequence analysis package) significant similarity was observed to the N-terminal regions of the enzyme lipoamide dehyrogenase, from various organisms.

CHAPTER 4

Purification and characterisation of ICL from *S. coelicolor*

4.1 Introduction

One of the aims of this project was to clone the gene encoding ICL from *S.coelicolor.* Cloning of the gene using oligonucleotide probes, which had been designed against conserved amino acid regions of previously cloned *icl* genes, proved unsuccessful (Chapter 3).

Attempts were made to purify ICL from *S. coeliclolor*, which had been grown on acetate (J. Rua, personal communication). The levels of ICL activity were not as high as had been observered for other prokaryotic ICL's (Vanni *et al*, 1990) and the results from N-terminal peptide sequence analysis of the protein purified by Dr. Rua led us to believe that it was not ICL (3.2.8). It was possible that the assay was measuring another enzyme activity and there was not any ICL. Dekleva and Strohl (1988) stated that *Streptomyces* C5 does not have a glyoxylate bypass and consequently no ICL activity. It was also possible that acetate was not being utilised efficiently by *S. coelicolor*, perhaps because of some transport deficiency.

Another carbon source which would induce ICL (if present) needed to be found, in order that the enzyme might be purified. Once purified, N-terminal amino acid sequence could be obtained, so that an oligonucleotide probe could be constructed to enable identification of the gene by hybridisation to the probe.

Previously *S.coelicolor* had been grown on glycerol trioleate, to try and induce the enzymes of the glyoxylate bypass. However growth was extremely slow and activities were poor (A.Moran, personal communication).

It was suggested that growth on Tween (polyoxyethylensorbitan) containing compounds may be a possible route (Colin Smith, personal communication), and so Tween 20, 40 and 60 containing the fatty acid groups lauric, palmitic and stearic respectively, were tried. This chapter describes the use of Tween compounds for the growth of *S. coelicolor* and the successful purification of the ICL enzyme. A comparison of some of the properties found with those of other ICL's characterised is also discussed. At this time, a good review of previously-purified ICL's was available (Vanni *et al*, 1990). Unless otherwise stated, all comparisons have been made with data derived from the review.

4.2 Results and Discussion

4.2.1 Growth of *S.coelicolor* on Tween-containing compounds

S.coelicolor was grown in YEME (2.2.1.2b), until thick growth was observed. The mycelia were then harvested by centrifugation, after the medium had been diluted 1:1, in dH_2O. The mycelia were then resuspended in NMM (2.2.1.1b) and aliquots used to inoculate fresh NMM, supplemented with 1% v/v of various Tween compounds.

Tween 20, 40 and 60 were tested for their ability to sustain growth of *S.coelicolor*. The cells were grown for 60 hours and then harvested by centrifugation (7,500g for 10 mins.) Crude extracts were then made (4.2.2) ICL activity was assayed using various amounts of crude extract. Protein was estimated using the method of Bradford (1976).

Table 4.1 shows the ICL activity and protein content of crude extracts from cells grown on various carbon sources.

Cells grown in Tween 40 showed the highest specific activity of ICL and this compound was used as the carbon source in subsequent experiments.

4.2.2 Purification of ICL from *S. coelicolor*

4.2.2.1 Introduction

Purification of *S.coelicolor* ICL was facilitated using a five-step protocol incorporating ammonium sulphate fractionation of a crude extract, followed by hydrophobic interaction chromatography using phenyl sepharose, anion exchange chromatography on the FPLC and finally futher hydrophobic interaction using phenyl superose on the FPLC.

Working rapidly, the enzyme could be purified in 48 hours. This was important, due to a high degree of instability (whether this was inherent or due to proteolysis, is unknown) in early rounds of the purification. If the protein was left after the ammonium sulphate fractionation step, at 4°C for 14 hours, then over 60% of the total activity was lost. Many of the ICL's from other organisms were also reported to be unstable at early stages of purification (Vanni *et al*, 1990). It may be that a protease is purified away in subsequent rounds of purification.

It had been reported that ICL was inhibited by HPO_4^{2-} , so MOPS was used as the desired buffer. The extraction buffer contained $MgCl_2$, EDTA and DTT, a composition used by others to purify ICL.

PMSF and benzamidine were also used as protease inhibitors. The use of PMSF was suggested by Pinzauti *et al*, (1983) and Jameel *et al*, (1984).

Previous reports had shown that hydrophobic interaction chromatography was a very effective step to purify ICL (Pinzauti *et al*, 1986 and Robertson, 1992), and thus phenyl sepharose hydrophobic interaction chromatography was used.

S. coelicolor was grown in NMM (2.2.3) supplemented with 1% v/v Tween 40 as the sole carbon source, for 60 hours. The cells were then harvested by centrifugation. (7500g for 10 mins.)

4.2.2.2 Purification procedure

Harvesting of mycelia and all stages of the purification were carried out at 4°C, except for the running of FPLC columns, which was performed at room temperature. Buffers, media and general biochemical methods are described in 2.4.2 - 2.4.4.

step 1 - Preparation of a crude extract

Harvested mycelia (50-60g wet weight) were resuspended in 100ml of extraction buffer, to yield a smooth paste. Cells were disrupted by passage through a French pressure cell, (3 passages) as described (2.4.1a). Cell debris and insoluble proteins were removed by centrifugation at 10,000g for 1 hour. The clear supernatant was then decanted off, and subjected to further purification.

step 2 - Ammonium sulphate fractionation

Powdered ammonium sulphate was gradually added to the crude extract, with stirring, to bring the solution to 40% saturation by adding 242mg of ammonium sulphate to every ml of crude extract. After stirring for a further 30mins, the precipitated protein was then removed by centrifugation at 12,000g for 20mins. The supernatant was then brought to 60% saturation by the addition of 130mg/ml ammonium sulphate. The precipitated protein was again removed by centrifugation.

Carbon source	ICL (rate)	Protein (mg/ml)	Specific activity mU/mg
Glucose	0.00	0.45	0.00
Tween 40	0.011	0.515	34.00
Tween 60	0.0067	0.430	24.65

Table 4.1 ICL activity from *S.coelicolor* grown on glucose and Tween-containing compounds

S.coelicolor was grown for 60 hours in NMM supplemented with either glucose or Tween, as the sole carbon source. Crude extracts were prepared (4.2.2) and the protein concentration (2.4.4c) and ICL activity (2.4.4h) determined.

The rate in this case is defined as the change in absorbance per minute observed from the addition of 100µl of crude extract.

step 3 - Hydrophobic interaction chromatography using a phenyl sepharose column

The column was equilibrated with Buffer B (2.4.2b) overnight, at a constant flow rate of 0.1ml/min. This was equivalent to at least ten column volumes.

The resuspended pellet (from the 40-60% ammonium sulphate cut) was then brought to 0.6M ammonium sulphate and loaded onto the column. The flow rate was increased to 0.5ml/min. After 1 column volume of buffer B, a 100ml linear gradient of 100% buffer B to 100% buffer A (creating a 0.6 - 0M ammonium sulphate gradient at pH 7.3) was applied to the column while maintaining the constant flow rate. Samples (2ml) were collected and assayed for protein content and ICL activity. Typically about 10 samples (Fig. 4.1) which showed the greatest activity were pooled and subjected to dialysis overnight against 1L of buffer A (2.4.2a), with at least two changes. To check that dialysis had gone to completion, the conductivity of the protein-containing solution was compared to that of buffer A alone.

step 4 - FPLC using anion exchange Mono Q column

The total protein eluted from the phenyl sepharose column was typically below 50mg; thus a small scale Mono Q column (HR 5/5) was used. The column was equilibrated by washing with buffer A at a flow rate of 1ml/min before use. Samples were loaded onto the column which was washed until no further protein emerged, as determined by A_{280}. A linear gradient of 0 - 0.5M NaCl in buffer A was applied to the column with a constant flow rate of 1ml/min. Most of the ICL activity eluted at the end of the gradient, at around 0.45M NaCl. (Fig. 4.2)

Samples containing activity were subjected to SDS PAGE, to assess purity. (Fig. 4.3). A number of contaminating bands were present, so a further round of purification was deemed necessary.

step 5 - FPLC using a phenyl superose column

Pooled fractions were adjusted to 1.7M ammonium sulphate and the column was equilibrated in buffer C (2.4.2c).The sample was loaded using a 5ml loop and the column washed with buffer C until no further protein eluted. A linear gradient of 100% buffer C to 100% buffer A was applied to the column, at a constant flow rate of 0.5ml/min. 0.5ml samples were collected.

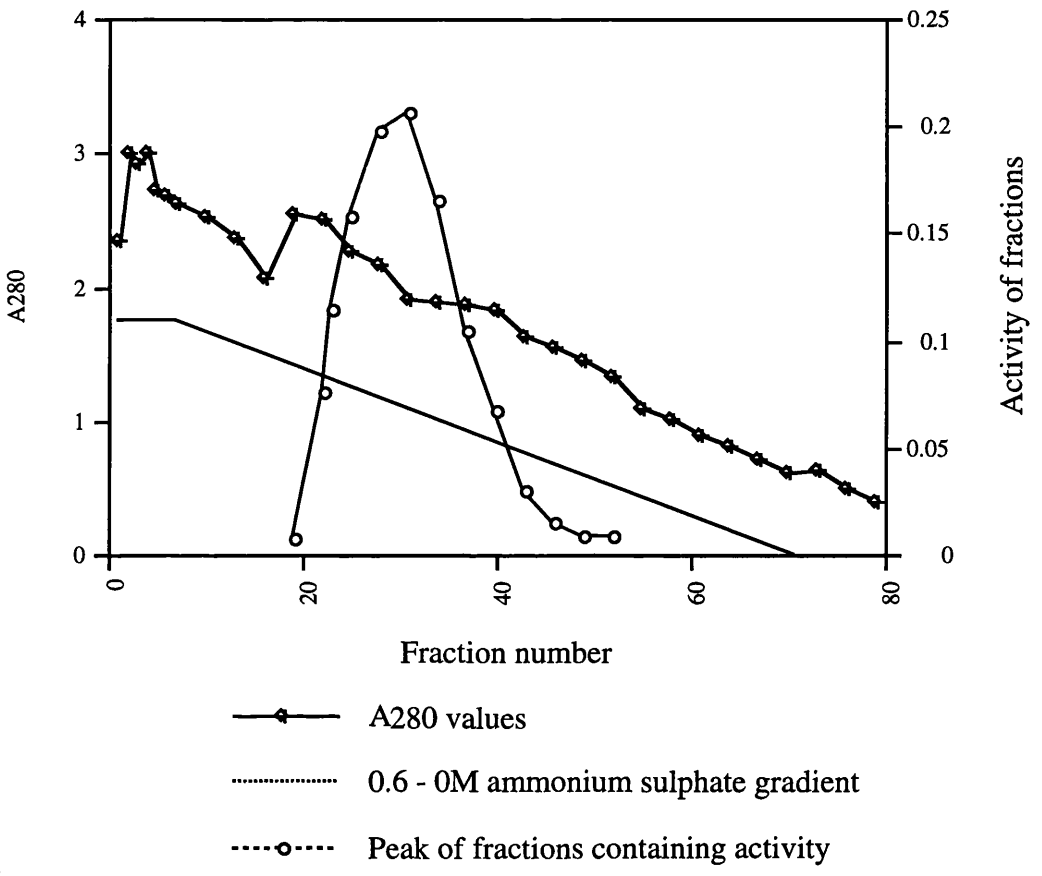


Figure 4.1 Phenyl sepharose hydrophobic interaction chromatography

Protein from the ammonium sulphate fractionation was made to 0.6M ammonium sulphate and applied to a 10ml phenyl sepharose column, which had been equilibrated in buffer B. The solid trace shows the A_{280} of protein eluting from the column, while a gradient of 0.6-0M ammonium sulphate was applied (dotted trace). The peak of ICL activity is shown as the dashed line. Activity was defined as the change in optical density per minute, resulting from assaying 10µl of each fraction.

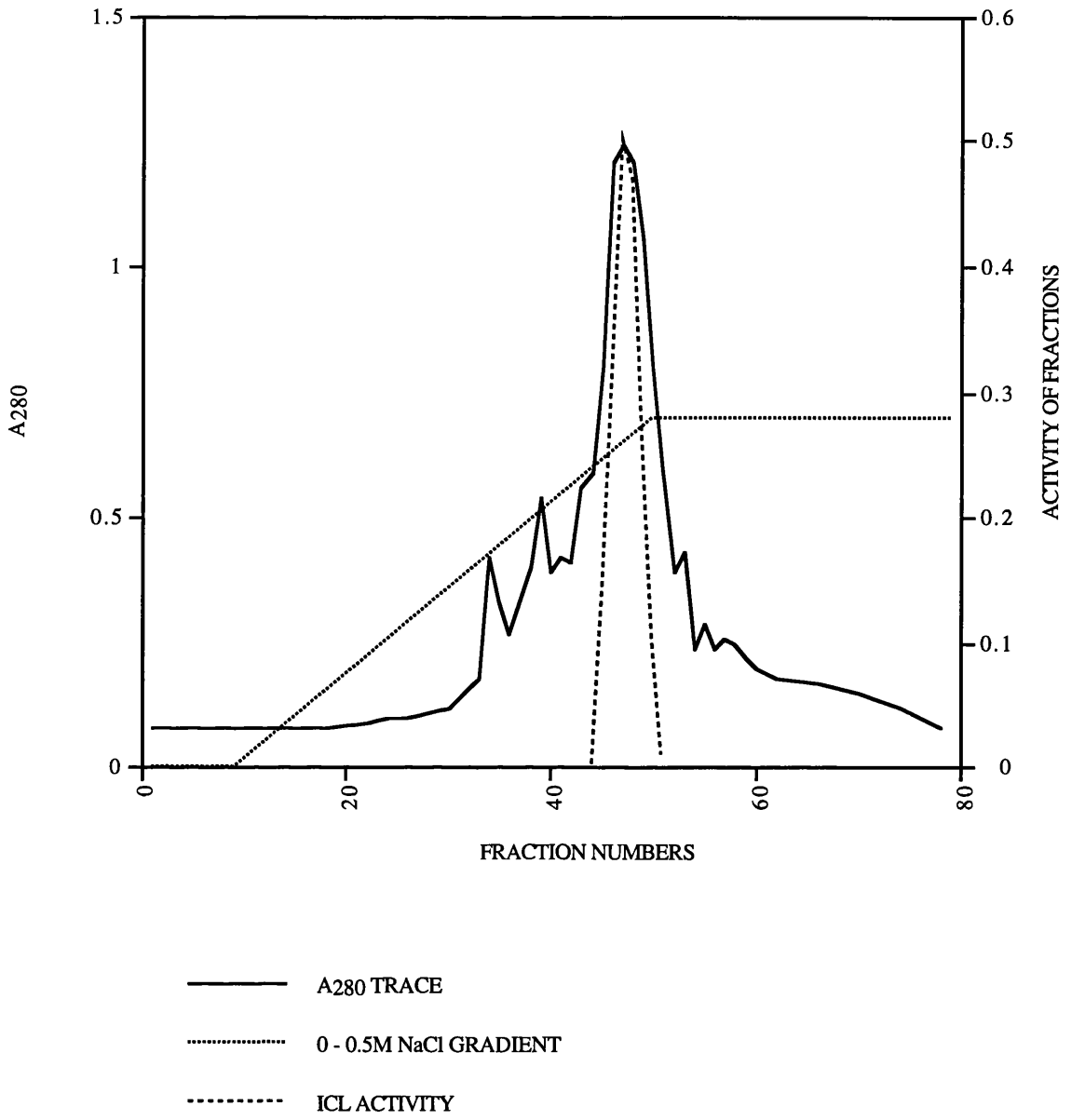


Figure 4.2 Mono Q anion exchange chromatography on the FPLC

Pooled fractions from the phenyl sepharose column which had been dialysed against buffer A were applied to the Mono Q HR 5/5 column and the column was washed with buffer A. The trace shows A280 of protein eluting from the column, while a linear gradient of 0 - 0.5M NaCl in buffer A, was applied. The dotted peak shows the fractions containing ICL activity. Activity was defined as the optical density change per minute, resulting from an assay using 10µl of sample.

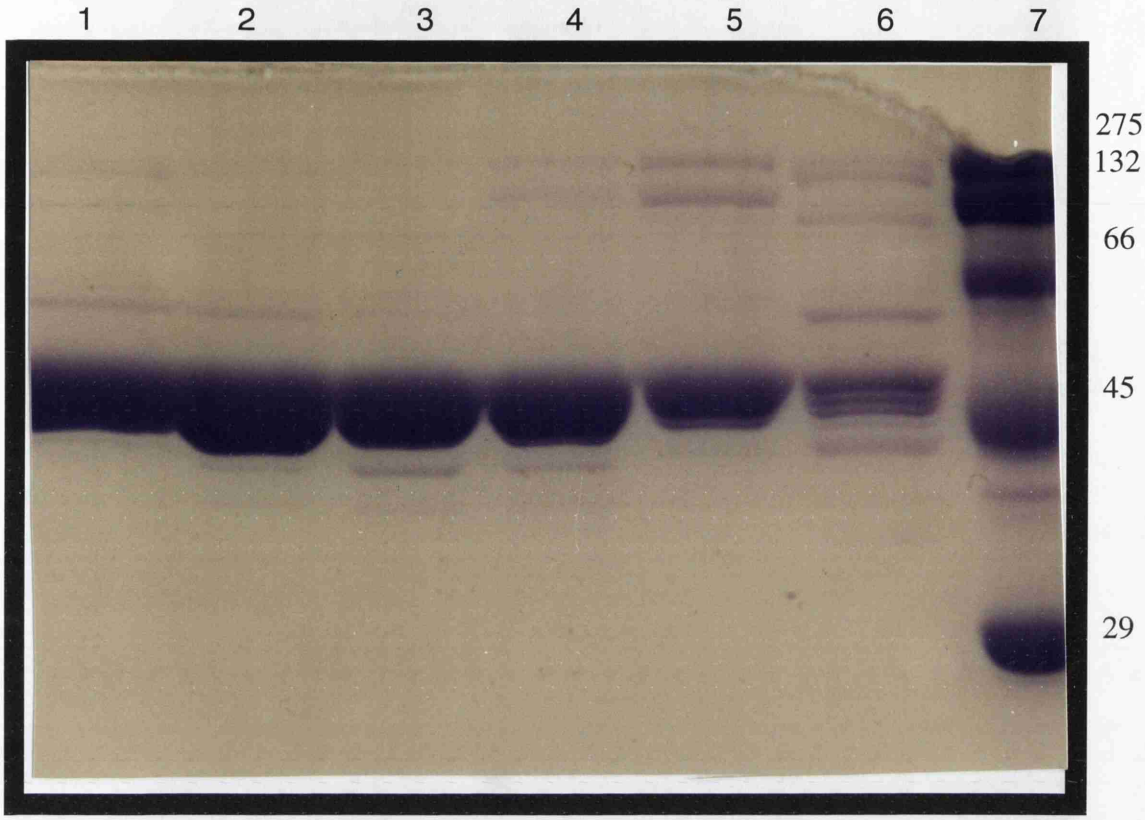


Figure 4.3 SDS PAGE of samples around the peak of ICL activity, which eluted from the MonoQ column

Fractions eluting from a MonoQ column, were assayed for ICL activity. The fractions which showed activity are shown in the order in which they eluted from the column. The four fractions which showed the greatest level of activity, were pooled and subjected to futher purification.

1µl of each fraction was loaded on to the gel. Rate is defined in this case as the change in O.D. 340/ min, resulting from the addition of 10µl of extract. The migration of molecular weight standards (figures in kDa) is shown in track 7.

Track	Fraction No.	Rate
1	20	0.143
2	21	0.500
3	22	0.470
4	23	0.260
5	24	0.105
6	25	0.010
7	Molecular weight standards (2.4.5)	

Only one sharp peak of protein was obtained, which corresponded with the peak of ICL activity. Samples containing activity were again subjected to SDS PAGE and the protein was found to be pure, by coomassie staining (Fig. 4.4).

The overnight dialysis step after the hydrophobic interaction chromatography resulted in a loss of activity, which decreased the factor of purification. However the dialysis step was retained as it allowed the overnight break to be taken at a convenient point. In future purifications, the loss of activity could be reduced by applying the protein to a desalting column, followed immediately by the FPLC steps. However, this would mean that the purification procedure lasts 36 hours without a break.

Using the FPLC resulted in rapid further purification. Anion exchange using a MonoQ column, followed by hydrophobic interaction using phenyl superose resulted in purification of ICL, as determined by coomassie staining of an SDS-PAGE gel (Fig. 4.4). ICL was found to constitute approximately 5% of the total soluble proteins of *S.coelicolor*, when grown using Tween 40 as the sole carbon source. This was calculated by taking the final protein content and converting this to the figure which would have been obtained should the yield of purification have been 100%. This figure was then divided by the total protein content of the crude extract and converted to a percentage.

A purification table summarises the results obtained from a typical purification (Table 4.2) and a gel shows the extent of purification at each stage (Fig. 4.5).

The subunit molecular weight of the ICL was estimated by comparing the relative mobility of the protein, to the mobility of the standard proteins with known molecular weights (Fig. 4.6). The results of several experiments showed that the subunit molecular weight of the *S. coelicolor* ICL is 52,000 daltons +/- 3,000 daltons, which is similar to that of other prokaryotic ICLs (Table 4.3). The value obtained for the native molecular weight of 216,000 (as determined by sizing using a superose 12 gel filtration column), indicates that the enzyme is tetrameric (Javier Rua, personal communication), which is in agreement with many ICLs purified previously.

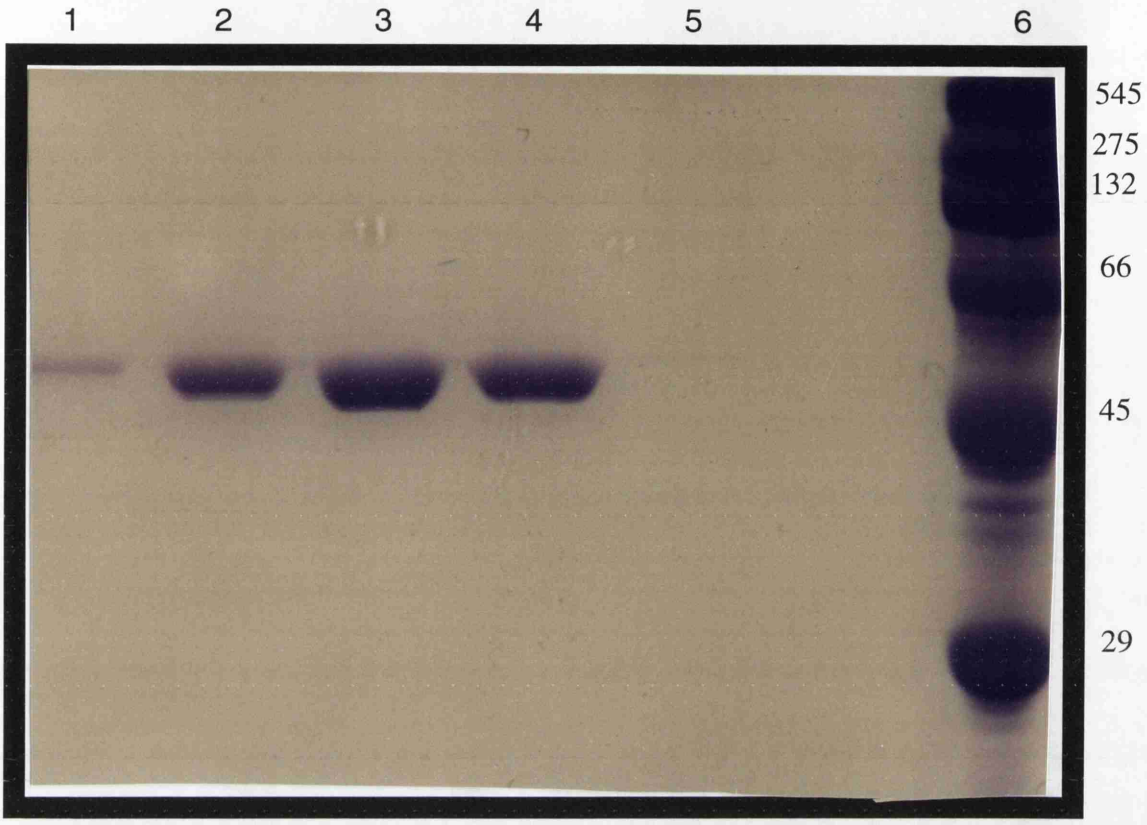


Figure 4.4 SDS PAGE of samples from the peak of ICL activity, eluting from the phenyl superose column

Fractions eluting from a phenyl superose column, were assayed for ICL activity. The fractions which showed activity are shown in the order in which they eluted from the column. The two fractions which showed the greatest level of activity, were pooled and kept for characterisation studies.

1µl of each fraction was loaded on to the gel. Rate is defined in this case as the change in O.D. 340/ min, resulting from the addition of 5µl of extract. Migration of molecular weight markers (in kDa) is shown in track 7.

Track	Fraction No.	Rate
1	22	0.060
2	23	0.152
3	24	0.471
4	25	0.276
5	26	0.046
6	Molecular weight standards (2.4.5)	

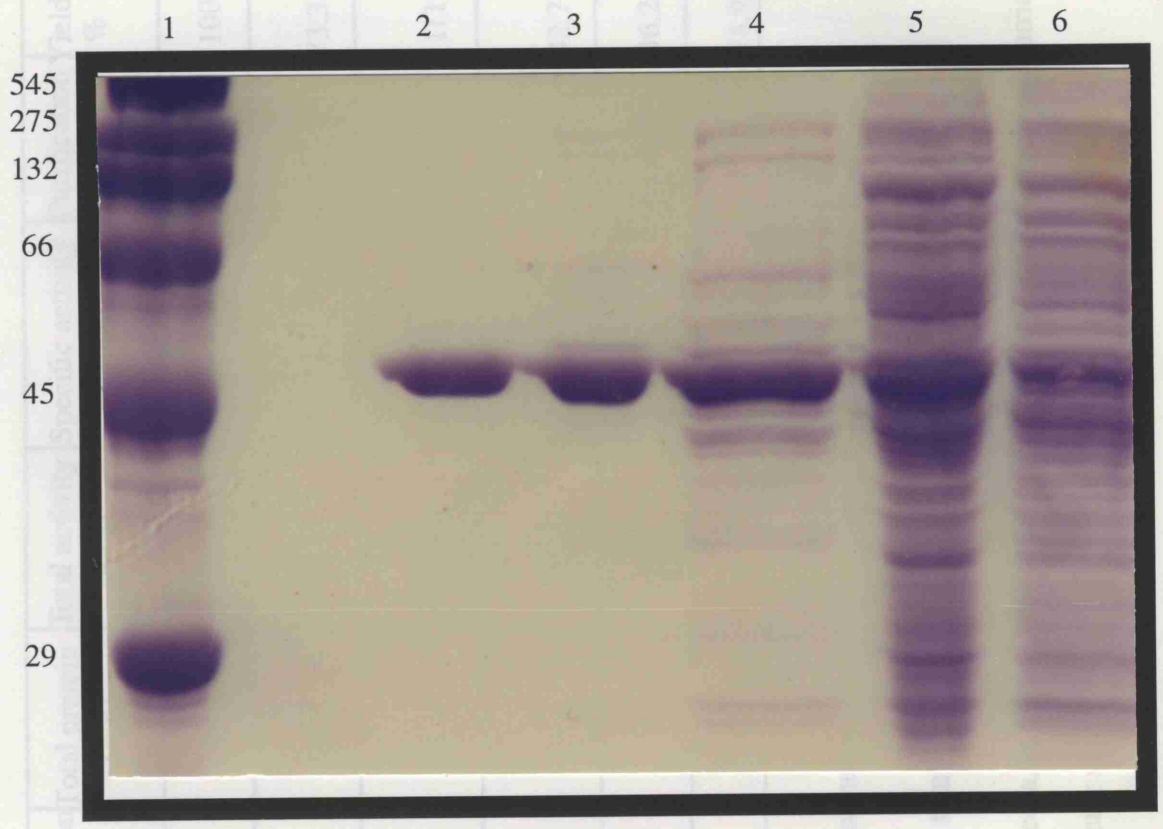


Figure 4.5 SDS PAGE of samples from the stages involved in the purification of *S. coelicolor* ICL

Fractions from each round of the purification were run on an SDS PAGE, to show how the purification proceded.

Protein molecular weight standards (values in kDa) were run in track 1.

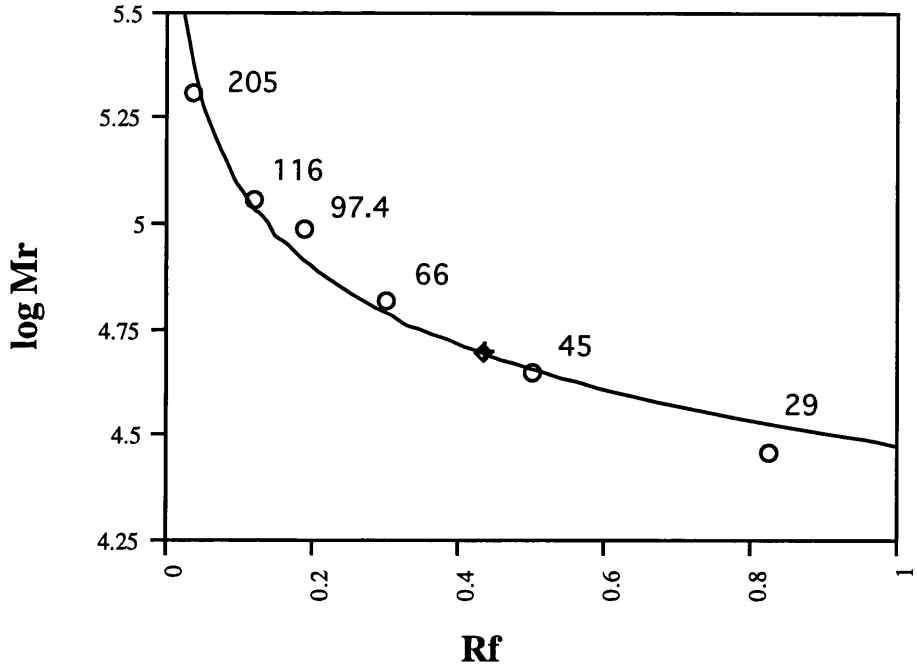
Track	Step	Protein content
6	Crude extract	~20µg
5	Ammonium sulphate	~20µg
4	Phenyl sepharose	~10µg
3	MonoQ	~5µg
2	Phenyl superose	~5µg

Purification step	Volume (ml)	Concentration of protein (mg/ml)	Total protein (mg)	Total activity (units)	Specific activity (units/mg)	Purification (fold)	Yield %
1. Crude extract	128	1.75	224	58.9	0.263	1	100
2. 40-60 % ammonium sulphate cut	5	13.5	67.5	43.15	0.64	2.4	73.3
3. Phenyl sepharose (before dialysis)	22	1.175	25.85	65.56	2.54	9.66	111
4. Phenyl sepharose (after dialysis)	26	1.069	27.8	25.17	0.91	3.44	42.7
5. Mono Q	4.8	1.625	7.8	27.2	3.49	13.27	46.2
6. Phenyl superose	0.7	3.2	2.24	11.11	4.96	18.86	18.9

% ICL of total extracted proteins = approx. 5%

Table 4.2 Purification table showing main steps of ICL purification

Cells were grown for 60 hours on minimal media, supplemented with 1% v/v Tween 40. Purification was carried out as described in 4.2.2.. The purification factor shows the increase in specific activity compared to the specific activity of the crude extract.



Key

○ Standard proteins. Numbers indicate the molecular weight in kilodaltons

◆ ICL

Mr is the molecular weight in daltons

Rf is the mobility of the proteins, defined as the distance travelled by the protein relative to the dye front

Figure 4.6 SDS PAGE calibration curve

The logarithm of the mass of the molecular weight standards (2.4.5) was plotted against the Rf value obtained for each protein. The curve was then fitted by computer.

The Rf value of the *S. coelicolor* ICL was plotted on the curve, enabling the molecular weight to be determined (52 kDa). Repeated experiments gave comparable results, +/- 3 kDa.

Source	Specific Activity U/mg	Km (isocitrate) mM	%ICL of extracted proteins	pH optimum	Buffer	Mol. wt. monomer	No sub	Mol. wt. oligomer	Reference
Thermophilc *Bacillus*	2.9	0.020	6.9	8.0	Imd	48 000	4	180 000	Chell *et al* (1978)
E.coli	2.4	0.008	8.3	7.3	Tris	48 000	4	188 000	Robertson & Reeves (1987)
Ps. indigofera	33.0	0.056	7.3	7.7	MOPS	48 200	4	206 000	McFadden *et al* (1968)
S. coelicolor	4.96	0.034	~5.3	7.5	MOPS	52 000	4	216 000	This work & J. Rua
N. crassa	16.0	0.050	1.3	6.8	MOPS	67 000	4	265 000	Johanson *et al* (1974)
C. vulgaris	16.1	0.250	1.0	---	MOPS	64 000	4	277 000	Jameel *et al* (1984)
G. max	3.2	1.800	0.8	8.0	Pi	70 000	2	130 000	Ruchti & Widmer (1986)
R. communis	2.7	0.290	1.5	7.2 - 7.3	Pi	35 000	4	140 000	Malhortra & Srivastra (1982)
T. aceti	2.4	0.660	3.3	7.6	Tris	123 000	4	480 000	Reiss & Rothstein (1974)
P. densiflora	4.8	0.330	1.0	7.6	Tris	65 000	3	200 000	Tsukamoto *et al* (1986)

Table 4.3 Comparison of some characteristics of various ICL's

Some of the characteristics of both prokaryotic and eukaryotic ICL's are shown above. Some of the values for *S. coelicolor* were kindly supplied by J.Rua.

4.3 Amino acid sequence determination

4.3.1 N-terminal sequence determination

Approximately 3µg of purified *S. coelicolor* ICL was loaded into each of 5 lanes of a Tricine/ SDS gel, which had been prerun with 10 mM glutathione (2.4.6.2). The gel was then run as described (2.4.6.2), before the protein was electroblotted on to Problott^{TM} membrane, stained with amido black and left to dry in air (2.4.6.3 and 2.4.6.4). The protein was then excised and the bands sent to Aberdeen University for sequencing.

Sequencing was carried out by Mr. Bryan Dumbar, by automated Edman degradation on an Applied Biosystems gas phase sequencer. The residues were assigned by comparison with a reference chromatogram of standard amino acids.

A sequence of 25 residues was obtained (Fig. 4.7), before assignment of the sequence became too ambiguous.

Using the GCG sequence analysis package (Devereux *et al* 1984), the program TFASTA was run against the GENEMBL database (for an explanation of TFASTA, see 6.2.5). With the peptide sequence derived from ICL as input, this search confirmed that the N-terminus of the *S. coelicolor* ICL was similar to that of other ICL's which had been sequenced.

4.3.2 Proteolytic digestion of the *S. coelicolor* ICL and sequencing of the polypeptide fragments

To allow a second oligonucleotide to be designed against the protein sequence, digestion of the protein was carried out using chymotrypsin.

Firstly, a conventional SDS PAGE gel was run with 8 lanes each being loaded with ~3µg of purified ICL protein. The gel was then stained in the usual manner and the protein band excised.

A procedure similar to that of Cleveland (1977) was then followed, in order to digest the protein (2.4.7). A Tricine/ SDS gel was prerun, as above. To each of four lanes, two excised fragments were loaded into each of the wells. 10µl of the enzyme cocktail was added to each well and the gel run until the dye reached the stacking/ separating gel boundary.

RESIDUE	1	2	3	4	5	6	7	8	9	10
AMINO ACID	ALA	GLU	ALA	ARG	THR	GLN	ALA	ALA	GLU	ASP
PMOL	137	120	116	30	55	109	92	110	72	97

RESIDUE	11	12	13	14	15	16	17	18	19	20
AMINO ACID	LEU	ALA	ARG	ARG	TRP	ASP	THR	GLU	PRO	ARG
PMOL	58	76	18	24	30	37	20	36	14	4

RESIDUE	21	22	23	24	25
AMINO ACID	PRO	GLN	GLY	ILE	GLY
PMOL	19	10	9	5	8

Figure 4.7 N- terminal amino acid sequence of *S.coelicolor* ICL

Approximately 100pmol of purified ICL was blotted onto Problott and slices of membrane applied to an "Applied Biosystems" gas phase sequencer (2.4.8). The identity of each amino acid is represented in three letter code for amino acids. The yield at each turn of the sequencer is also shown.

The current was then swiched off and the protein digested *in situ* for 30min. The current was then switched back on and the proteolytic fragments separated. The polypeptides were then electroblotted on to Problott^{TM}, as above.

Four different enzymes were tried: elastase, trypsin, V8 protease and chymotrypsin. V8 protease had been successful in the past for other streptomycete proteins (P. White, personal communication). However in this case no digestion was observed. Proteolysis by trypsin and elastase resulted in incomplete digestion and too many partially digested fragments were observed upon staining of a protein gel. Chymotrypsin yielded only two bands of 35kDa and 14kDa respectively and these were blotted and sent for sequencing

36 residues were obtained from the 35kDa peptide (Fig. 4.8). The sequence seems to be from the N-terminus of the intact protein, with two exceptions. Firstly, the sequence starts one residue later than the N-terminal sequence and secondly residues 20 and 24 change from being a proline and a glycine, to a tryptophan and a glutamic acid, respectively. After consulting the chromatograms from each sequencing run, it appeared that the sequence of the 35kDa peptide was more likely to be correct. An oligonucleotide, was thus designed against this sequence (see Chapter 5). 29 residues were obtained from sequencing of the 14kD peptide, with one ambiguity at residue number 2 (Fig. 4.9).

4.3.3 Comparison of sequences obtained to that of the other sequenced ICLs

Both amino acid sequences were compared against the other previously sequenced ICLs, using the "PILEUP" program (for a description of "PILEUP", see 6.2.12) on the GCG sequence analysis package (Figs. 4.10 and 4.11).

Both pileups showed that the *S. coelicolor* sequence had significant similarity to other ICLs, giving enough confidence that the purified enzyme was indeed ICL.

4.4 Characterisation of the ICL from *S. coelicolor*

When the enzyme was found to be in a relatively pure state, it was passed on to J. Rua (Department of Biochemistry, University of Glasgow), for biochemical studies to be carried out. He has kindly allowed me to report some of the work.

4.4.1 Km for Isocitrate

The Km for isocitrate, was found to be 0.034 mM, which is comparable to some of the other prokaryotic ICL's (Table 4.3). The Km of the *S. coelicolor* IDH for isocitrate is much lower, 1.3μM (Taylor, 1992).

RESIDUE	1	2	3	4	5	6	7	8	9	10
AMINO ACID	GLU	ALA	ARG	THR	GLN	ALA	ALA	GLU	ASP	LEU
PMOL	41	27	29	40	15	39	49	36	47	35

RESIDUE	11	12	13	14	15	16	17	18	19	20
AMINO ACID	ALA	ARG	ARG	TRP	ASP	THR	GLU	PRO	ARG	TRP
PMOL	38	18.6	18.9	7	24	25	24	21	5	3

RESIDUE	21	22	23	24	25	26	27	28	29	30
AMINO ACID	GLN	GLY	ILE	GLU	ARG	THR	TYR	SER	ALA	GLN
PMOL	13	11.3	10.8	6.2	2.4	5.0	6.4	1.0	8.0	5.3

RESIDUE	31	32	33	34	35	36
AMINO ACID	ASP	VAL	VAL	PHE	LEU	LEU
PMOL	4.5	4.0	7.2	6.8	6.5	7.0

Figure 4.8 N-terminal sequence of the 35kDa peptide

The 35kD peptide resulting from proteolytic digestion of ICL by chymotrypsin was blotted onto Problott and slices applied to an "Applied Biosystems" gas phase sequencer (2.4.8). Amino acids are represented as the three letter code for amino acids. The yield for each turn of the sequencer is also shown.

RESIDUE	1	2	3	4	5	6	7	8	9	10
AMINO ACID	ASN	?	SER	PRO	SER	PHE	ASN	TRP	LYS	ALA
PMOL	46.9	-	15	33	5	12	14	2.2	2.8	6

RESIDUE	11	12	13	14	15	16	17	18	19	20
AMINO ACID	ALA	LEU	ASP	ASP	ASP	GLN	ILE	ALA	LYS	PHE
PMOL	9	4	6.5	7	7.6	4.5	2.3	2.2	1.6	2.6

RESIDUE	21	22	23	24	25	26	27	28	29
AMINO ACID	GLN	GLU	GLY	LEU	GLY	ALA	MET	GLY	TYR
PMOL	2.7	1.8	2.5	1.5	2.1	1.4	1.7	1.2	1.8

Figure 4.9 N-terminal sequence of the 14kDa peptide

The 14kD peptide resulting from proteolytic digestion of ICL by chymotrypsin was blotted onto Problott and slices applied to an "Applied Biosystems" gas phase sequencer (2.4.8). Amino acids are represented as the three letter code for amino acids. The yield for each turn of the sequencer is also shown.

Arabidopsis thaliana	.....MIDKPNQIMEEEGRFEAEVAEVQTWWSSERFKLTRRPYTARDVVALRGHL
Rapeseed	..MAASFSVPSMIMEEEGRFEAEVAEVQTWWSSERFKLTRRPYTARDVVALRGHL
Cotton	..MAASFSVPSMIMEEEGRFETEVAEVQAWWSNERFKLTRRPYSARDVVALRGSL
Castor bean	..MASSFSGPSMIMEEEGRFEAEVAEVQAWWNSERFKLTRRPYTARDVVALRGSL
Aspergillus nidulans	.........MSYIEEEDQRYWDEVA.VKNWWKDSRWRYTKRPFTAEQIVAKRGNL
Neurosporra crassa	MAANNMVNPAVDPALEDELFAKEVEEVKKWWSDSRWRQTKRPFTAEQIVSKRGNL
Candida tropicalis	.....MAYTKIDINQEEADFQKEVAEIKKWWSEPRWRKTKRIYSAEDIAKKRGTL
Saccharomyces cerevisiae	.MPIPVGNTKNDFAALQAKLDADAAEIEKWWSDSRWSKTKRNYSARDIAVRRGTF
Escherichia coli	..............MKTRTQQIEELQKE.WTQPRWEGITRPYSAEDVVKLRGSV
S. coelicolor	..............A**EARTQA**A**E**D**L**ARR**W**D**TEPRW**Q**G**I**E**R**T**YSA**Q**DVV**F**L**L

Figure 4.10 Comparison of the N-terminal sequences of ICL's

The programme "PILEUP" from the GCG package was used to produce an alignment of the 10 sequnces shown. All sequences except the *S.coelicolor* sequence, were obtained from translation of the DNA sequence. The *S.coelicolor* sequence was obtained by direct sequencing of the protein.

References :- *A.thaliana* ICL - Bernhard and Matile, unpublished. Rapeseed ICL - Conai *et al* , 1989 . Cotton ICL - Turley *et al.*, 1990. Castor bean ICL - Beeching and Northcote., 1987. *A,nidulans* and *N.crassa* ICLs - Gainey *et al.*, 1992. *S.cerevisiae* ICL - Schoeller and Schueller, 1993. *E.coli* ICL - Rieul *et al.*, 1988. *S.coelicolor* ICL - this work.

Arabidopsis thaliana	YNLSPSFNWDASGMTDPLTC.....
Rapeseed	YNLSPSFNWDASGMTDQQMMEFIPRIARLGY
Cotton	YNLSPSFNWDASGMTDEHMRDFIPRIAKLGF
Castor bean	YNLSPSFNWDASGMTDEQMRDFIPRIARLGF
Aspergillus nidulans	YNLSPSFNWK.KAMPRDEQETYIKRLGALGY
Neurosporra crassa	YNLSPSFNWK.TAMGRDDQETYIRRLAKLGY
Candida tropicalis	YNLSPSFNWN.KAMPADEQETYIKRLGQLGY
Saccharomyces cerevisiae	YNLSPSFNWP.KAMSVDEQHTYIQRLGDLGY
Escherichia coli	YNCSPSFNWQ.KNLDDKTIASFQQQLSDMGY
Streptomyces coelicolor	**.N_SPSFNWK.AALDDDQIA**K**FQ**EG**LGA**M**GY**

Figure 4.11 Alignment of internal regions of the ICLs

The programme "PILEUP" from the GCG package was used to produce an alignment of the 10 sequnces shown. All sequences except the *S.coelicolor* sequence, were obtained from translation of the DNA sequence. The *S.coelicolor* sequence was obtained by direct sequencing of the protein.

References :- *A.thaliana* ICL - Bernhard and Matile, unpublished. Rapeseed ICL - Conai *et al* ., 1989 . Cotton ICL - Turley *et al*., .1990. Castor bean ICL - Beeching and Northcote., 1987. *A,nidulans* and *N.crassa* ICLs - Gainey *et al.*, 1992. *S.cerevisiae* ICL - Schoeller and Schueller, 1993. *E.coli* ICL - Rieul *et al.*, 1988. *S.coelicolor* ICL - this work.

ICL and IDH compete for a common pool of isocitrate, when the glyoxylate bypass is operating. Thus, at low isocitrate concentrations, IDH would compete more effectively for isocitrate than ICL. The carbon would be shunted around the TCA cycle and not through the glyoxylate bypass. This is a paradox which remains to be investigated.

In *E. coli*, the competition for isocitrate as a substrate is reduced by rendering IDH inactive (1.5). It remains to be investigated if a similar situation prevails in *S.coelicolor.*

4.4.2 pH dependence on Vmax and Km

Studies have shown that the ICL from *S.coelicolor* has a pH optimum of 7.3 - 7.7 (MOPS-NaOH), with a relatively symetrical peak around the mean. This is in accordance with many of the previously purified ICL's (Table 4.3).

The effect of pH upon Km was studied and the pKm for isocitrate decreased, as the pH of the reaction mixture was raised. The function of pKm vs pH, yielded a straight line with a slope of about -1. This was also noted for the enzymes from *Pseudomonas indigofera* and *Pinus pinea.* The authors postulated that a single dissociable group may be involved in the formation of the enzyme-substrate complex.

4.4.3 Effect of DTNB and Iodoacetate on the ICL activity

ICL activity disappears after 5min incubation with 5μM DTNB or 40min with 2.5μM DTNB.

83% of activity is lost after incubation with 1mM iodoacetate. These experiments imply that -SH groups are involved in the activity of ICL. The DNA sequence (Fig. 6.3) shows there to be only two cysteine residues in the *S. coelicolor* ICL, one of which is conserved across all species so far sequenced, and was reported to be at the active site of the *E. coli* enzyme. It seems most likely this residue is crucial for the activity of ICL. It would be of great interest to try *in vitro* site directed mutagenesis on this cysteine to see if in changing it, activity is altered.

CHAPTER 5

Cloning of the *icl* gene from *S.coelicolor*

5.1 Introduction

One of the aims of this project was to clone the gene encoding isocitrate lyase (ICL) in *S.coelicolor*. Attempts to clone the gene using oligonucleotides designed against other ICL protein sequences, proved unsuccessful (see chapter 3). However the availability of amino acid sequence from the purified ICL from *S.coelicolor* (Chapter 4) provided the opportunity to design new oligonucleotide probes to attempt to clone the gene.

If the gene could be cloned and sequenced, it would be useful to understand why the attempts at cloning the gene using "consensus" oligonucleotides (Chapter 3) had failed.

The gene encoding IDH from *S.coelicolor* had been cloned previously using an oligonucleotide designed against the N-terminal amino acid sequence of the IDH protein. It was hoped that this method of "reverse genetics" could also be used for the cloning of the *icl* gene. However, more sequence information was available for *icl*. This enabled a second oligonucleotide to be designed, which made the task of screening for the *icl* gene much easier.

This chapter describes how the two oligonucleotides were used (as primers in a PCR), to amplify part of the *icl* gene. It then describes how this homologous DNA was used as a probe to clone a segment of DNA encoding the entire *icl* gene.

5.2 Results and discussion

5.2.1 Design of oligonucleotides

To clone the gene encoding ICL, oligonucleotide probes were designed against the amino acid sequences obtained from the purified protein and digested fragments (see section 4.3).

Streptomycete genes show a high G+C content, ranging from 61%-80%, and have a G+C% frequency in the third position of a codon of 76%-98% (Wright and Bibb, 1992). This results in a bias in the codons used in streptomycete genes. The oligonucleotides were therefore designed by taking advantage of this bias, using a codon preference table compiled by Wright and Bibb (1992).

When designing oligonucleotides, regions containing tryptophans and methionines are strongly favoured, as these amino acids are only encoded by a single codon. Lysine, asparagine, phenylalanine, aspartic acid, isoleucine, tyrosine and histidine, are also favoured as these are predominantly encoded in streptomycete genes by one codon. Regions containing leucines, serines and arginines should be avoided, as these amino acids are each encoded by 6 different codons. However because of the bias enforced on streptomycete genes, leucine and arginine are each encoded predominantly by two codons, leucine by CTG and CTC in 94% of cases, arginine by CGG and CGC in 82% of cases. When synthesising an oligonucleotide, it is possible to allow for both possibilities by incorporating a degeneracy in the third position of the codon. Other amino acids, such as alanine, threonine and glycine, are also encoded predominantly by two codons each, with frequencies of 92%, 95% and 83% respectively.

In practice a number of oligonucleotides are looked at. These oligonucleotides are from a series of overlapping windows of the translated amino acid sequence. The window which has the highest probability of being correct is chosen. That is, the oligonucleotide which requires the fewest possible combinations of codons.

With these factors taken into account, two oligonucleotides (ICLNEW1 and ICLNEW2) were designed (Figs. 5.1 and 5.2) and synthesised. Both oligonucleotides contained degeneracies to increase the probability of hybridising to the correct sequence. ICLNEW1 also contained an inosine in the third position of a codon encoding arginine. This decreases the chance of mismatching from 54% to 6%. This occurs because inosine can bind, non-specifically to any of the four bases (it should be noted however that the inosine should not be included in any calulations of the melting temperature, as its binding does not increase the melting temperature).

Although both oligonucleotides were designed to amplify a fragment from genomic DNA by PCR, they were also designed so that they could be used directly as hybridisation probes, should the PCR not work. They were significantly longer than conventional PCR primers at 29 and 45 nucleotides respectively.

```
a)   ALA  GLU  ALA  ARG  THR  GLN  ALA  ALA  GLU  ASP
b)   GCC  GAG  GCC  CGC  ACC  CAG  GCC  GCC  GAG  GAC
     59%  81%  59%  46%  65%  93%  59%  59%  81%  96%

c)   GCC  GAG  GCC  CGI  ACC  CAG  GCC  GCC  GAG  GAC
       G         G         G         G    G

     5'                                          3'
d)   GCC  GAG  GCC  CGI  ACC  CAG  GCC  GCC  GAG  G
       G         G         G         G    G
```

Figure 5.1 Design of oligonucleotide ICLNEW1 used in cloning of *icl*

a) Part of the sequence obtained from amino acid sequencing of the 35kDa polypeptide (Fig. 4.10) and intact protein.

b) The codons most frequently used in streptomycete coding regions. The frequencies with which the codons are used is shown below.

c) By incorporating degeneracies at the third positions of some codons, it was possible to greatly increase the probability of maximising the degree of hybridisation. In one case an inosine was substituted in the third position of the arginine codon, to decrease the degree of mismatching. The frequencies resulting from the changes are shown below.

d) This "best guess" oligonucleotide was designed to anneal to the antisense strand and is therefore an antisense probe.

a)	PHE	ASN	TRP	LYS	ALA	ALA	LEU	ASP	ASP	GLN	ILE	ALA	LYS	PHE	GLN
b)	TTC	AAC	TGG	AAG	GCC	GCC	CTG	GAC	GAC	CAG	ATC	GCC	AAG	TTC	CAG
	99%	96%	100%	95%	59%	59%	55%	96%	96%	93%	92%	59%	95%	99%	93%
c)	TTC	AAC	TGG	AAG	GCC	GCC	CTG	GAC	GAC	CAG	ATC	GCC	AAG	TTC	CAG
					G	G	C					G			
	99%	96%	100%	95%	92%	92%	94%	96%	96%	93%	92%	92%	95%	99%	93%
	3'														
	AAG	TTG	ACC	TTC	CGG	CGG	GAC	CTG	CTG	GTC	TAG	CGG	TTC	AAG	GTC 5'
					C	C	G					C			

Figure 5.2 Design of oligonucleotide ICLNEW2 used in probing for *icl*

a) A region of amino acid sequence obtained from sequencing of the 14kDa polypeptide, generated from partial digestion of *S. coelicolor* ICL (Fig. 4.11).

b) The codons most frequently used by *Streptomyces* in coding regions, are shown. The frequencies with which they are used, are shown below (Wright and Bibb, 1992).

c) Best guess oligonucleotide, designed by incorporating degeneracies into codons which were represented poorly by only one codon .

d) The oligonucleotide ICLNEW2, was designed to hybridise to the antisense strand of *icl*. The sequence was therefore the reverse complement of the antisense strand.

5.2.2 PCR with genomic *S.coelicolor* DNA as template

To obtain an homologous probe for cloning of the ICL gene, the two oligonucleotides were used as primers in a PCR.

100pmol of each primer and 200ng of genomic DNA was used per reaction. The reaction was set up as described in 2.3.24. The Tm's of the two primers, ICLNEW1 and ICLNEW2 were calculated to be 71.7°C and 72.1°C respectively, using the formula of Fritsch (unpublished; cited in Sambrook *et al*, 1989) An annealing temperature of 65°C was used, which allows for a certain degree of mismatching.

The expected size of the PCR fragment was between 1kb and 1.35kb, as estimated from the other ICL sequences. A single amplified band was found at around 1kb, as determined by comparison to DNA molecular weight standards (Fig. 5.3). A $MgCl_2$ concentration of 1.5mM was found to be the optimum and this was used in subsequent reactions.

5.2.3 Sequencing of the PCR product

To determine that the amplified product was an internal region of the *icl* gene, DNA sequencing of the amplified product was carried out.

Asymmetric PCR was carried out to amplify single-stranded DNA, which could then be sequenced. Asymmetric PCR allows a small amount of double-stranded product to be formed, before only single-stranded DNA is amplified. This is achieved by using a limiting amount of one primer and a non-limiting amount of the second primer. When the limiting primer is incorporated into the product, the second primer amplifies the single stranded DNA, in a linear fashion.

The reaction was run using 1pmol of the limiting primer and 50pmol of the second primer. 10ng of the previously amplified product (5.2.2) was used as template . Amplification was carried out as described in 2.3.24.

A tenth of the reaction was run on an agarose gel, to check for single-stranded product. In an agarose gel run without ethidium bromide, single-stranded DNA migrates slower than double-stranded DNA. The converse is true for gels with ethidium bromide.

Figure 5.3 Amplification of genomic *S.coelicolor* 1147 DNA using primers ICLNEW1 and ICLNEW2

200ng of genomic DNA was used as template in a PCR using 100pmol each of primers ICLNEW1 and ICLNEW2 (2.3.24). A single amplified band was obtained at around the expected size (1-1.35 kb) of 1kb. The optimum Mg^{2+} was found to be 1.5mM and no amplification of the band was observed in the single primer control lanes.

LANE	Mg^{2+} conc. mM	Primers	LANE	Mg^{2+} conc. mM	Primers
1	λ *HIND* III markers		5	2.0	1+2
2	0.5	1+2	6	4.0	1+2
3	1.0	1+2	7	1.5	1
4	1.5	1+2	8	1.5	2

The limiting primer was end-labelled with gamma-^{32}P (2.3.16) and 1pmol of this primer was used in a sequencing reaction. Sequencing was carried out using the TaqTrack kit (supplied by Promega) following the manufacturer's instructions, with the exception of the annealing and extension reactions, which were carried out at 65°C, to help reduce aberrant annealing of the primer and formation of secondary structure by the template.

This procedure was carried in turn, for each of the primers being limiting and end-labelled. However only good readable DNA sequence was obtained when ICLNEW1 was used as the limiting and end-labelled primer.

When the DNA sequence obtained was translated from the second frame (Fig. 5.4) and compared to the amino acid sequence obtained from the 35kDa polypeptide, obtained from proteolysis of pure ICL (Fig. 4.10), there was an almost perfect match over part of the amino acid sequence, showing that with a high degree of certainty the cloned DNA sequence encoded part of the ICL of *S. coelicolor* (Fig. 5.5). The DNA when translated from the second frame also showed an extreme G+C% bias in the third position of each codon, indicative of DNA encoding a true protein sequence. Comparison of the translated DNA sequence, with the protein sequence of the ICL of *E. coli*, also showed a high degree of similarity (Fig. 5.5), whch confirmed further that the amplified DNA was part of the *icl* gene.

5.2.4 Screening of an amplified λ library

A λ library of *S. coelicolor* genomic DNA had been constructed previously (Taylor, 1992). This library was screened using the homologous ICL/PCR product to attempt to try to clone the DNA segment containing the entire *icl* gene.

It had been calculated (assuming a average insert size of 11kb), that 3,347 plaques had to be screened to have a 99% probability that a DNA sequence was present in the library (Taylor, 1992) . Approximately 10,000 plaques were grown on a lawn of *E. coli* and the DNA transfered to Hybond-N nylon membrane, as described in 2.3.17. Plaque lifts were processed as described in 2.3.20 and hybridisation was carried out using the amplified ICL/PCR product, labelled by DIG, as a probe. Chemiluminescent DIG detection was then carried out according to the manufacturer's instructions.

Several attempts were made to obtain positive signals. Plaques which gave putative positive plaque signals were obtained, but on secondary screening they were found to be false.

```
    1         10         20         30         40         50
A)  GAGGTGGCAGGGCATCGAGCGCACCTACAXXXCCCAGGACGTCGTCCGGCTCTC

B)  E  V  A  G  H  R  A  H  L  ?  ?  P  G  R  R  P  A  L
C)   R  W  Q  G  I  E  R  T  Y  ?  ?  Q  D  V  V  R  L  S
D)    G  G  R  A  S  S  A  P  T  ?  P  R  T  S  S  G  S  R


         60         70         80         90         100
A)  GGGCAGTGTCCGTAAGGAGCACACCCTGGCCCGGCGCGGTGCCGAGCGGCTGTG

B)  G  Q  C  P  *  G  A  H  P  G  P  A  R  C  R  A  A  V
C)   G  S  V  R  K  E  H  T  L  A  R  R  G  A  E  R  L
D)    A  V  S  V  R  S  T  P  W  P  W  P  G  A  V  P  S
```

Figure 5.4 Sequencing of the PCR product amplified using primers ICLNEW1 and ICLNEW2

A) This represents the DNA sequence obtained from sequencing of single stranded product, generated using primers ICLNEW1 and ICLNEW2. ICLNEW1 was used as the limiting primer and as the labelled primer to obtain sequence.

The X's represent a region of the gel for which sequence was not able to be read.

B) This is a translation of the DNA sequence from the 1st frame, into the respective amino acid sequence. The amino acids are represented by the one letter code. The ?'s are unknown amino acids, due to the DNA sequence which could not be read. The * , represents a termination codon.

C) This represents the translation from the second reading frame.

D) This represents the translation from the third reading frame.

1.

```
A)      R W Q G I E R T Y ? ? Q D V V R L S G S V R
        | | | | | | | | |     | | | |   |
B)    P R W Q G I E R T Y S A Q D V V P L L
```

2.

```
C)      R W Q G I E R T Y ? ? Q D V V R L S G S V R K E
        | |   | |   |   |       | | |   |   | | |     |
D)      R W E G I T R P Y S A E D V V K L R G S V N P E

        H T L A R R G A E R L
          | | |     | |
        C T L A Q L G A A K M
```

Figure 5.5 Comparisons with the peptide sequence deduced from the amplified PCR product with purified ICL from *S.coelicolo* and deduced sequence of ICL of *E.coli*

1. Comparison between the deduced peptide sequence of the PCR product with that from purified ICL of *S.coelicolor.*

A) The DNA sequence shown in Fig. 5.3 was translated from the second frame and part of that amino acid sequence is shown.

B) Part of the amino acid sequence obtained from Edman degradation sequencing of the 35kDa polypeptide (Fig. 4.10).

2. Similarity between the deduced peptide sequence of part of the ICL of *S.coelicolor* with the ICL of *E.coli.*

C) The DNA sequence shown in Fig. 5.3 was translated from the second frame and the amino acid sequence is shown.

D) Part of the *E. coli* ICL protein sequence is shown (Matsuoka and McFadden, 1988).

The major problem encountered was the high background, which made it difficult to identify positives. This seemed to be a problem caused by the DIG detection procedure. Since finishing laboratory work, many of the problems associated with the first experiments using DIG detection have been overcome.

Concurrently with the screening of the λ library, genomic DNA was screened and a partial plasmid library was constructed. This proved to be successful, so screening of the λ library was abandoned.

5.2.5 Screening of genomic DNA

Genomic DNA was digested to completion, using various restriction enzymes. The resulting fragments were separated on a 1% (w/v) agarose gel and visualised by ethidium bromide staining (Fig. 5.6a).

The DNA was blotted onto Hybond-N and the DNA fixed by UV crosslinking, as described in 2.3.17.2. The filter was then subjected to prehybridisaton. After prehybridisation, DIG labelled ICL/PCR product was added (10ng/ml of hybridisation solution) to the bag containing the hybridisation solution and hybridisation was allowed to proceed for 16 hours, before the filter was washed three times (2.3.18.2). Hybridisation was carried out at 65°C, in 1xSSC, 0.1% SDS and the filter washed with 0.1xSSC, 0.1% SDS, once at room temperature and twice at 65°C.

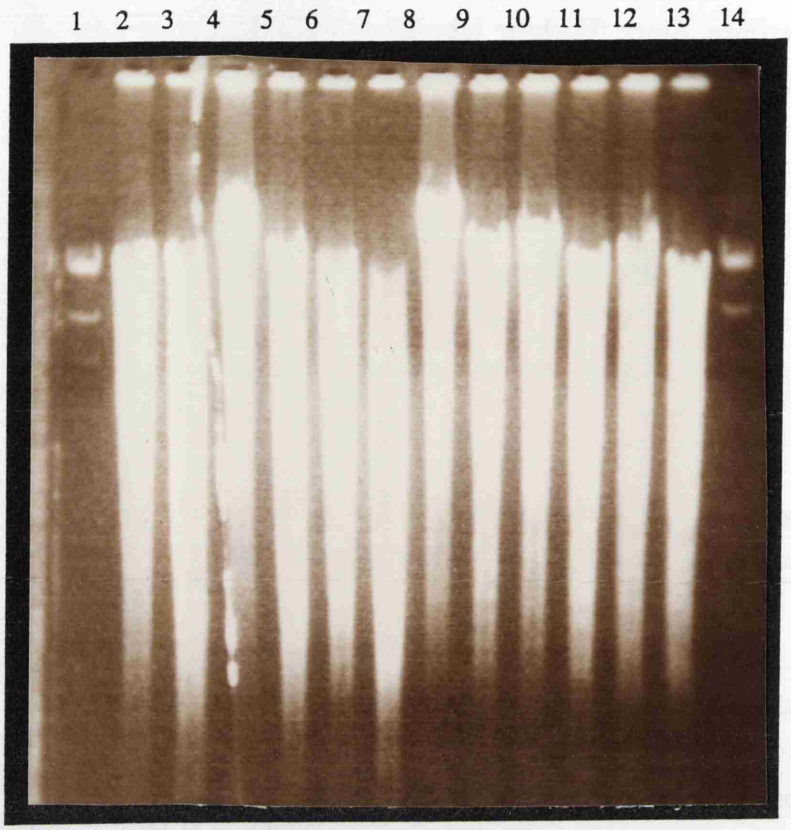
Chemiluminescent detection with DIG was then carried out according to the manufacturer's instructions. The filter was subjected to autoradiography for various lengths of time, and the films developed (Fig. 5.6b).

As the estimated molecular weight of ICL from *S.coelicolor* was 52kDa, the maximum size of the *icl* gene would be 1.56kb, assuming an average molecular weight for an amino acid of 100Da. A DNA fragment much larger than 1.56kb, which hybridised to the PCR probe was required to ensure that the entire *icl* gene was contained within it.

Nine digests yielded hybridising bands which were potentially large enough to contain the whole *icl* gene. These were *Bam* HI (9.3kb), *Bgl* II (11.5 kb), *Bam* HI/*Bgl* II (7.9kb), *Pst* I (5.75 kb), *Sst* I (11.5 kb), *Xho* I (9.3 kb), *Pst* I/*Sst* I (5.5kb), *Pst* I/*Xho* I (3.0 kb) and *Sst* I/*Xho* I (5.5kb) (Fig. 5.6).

Four of the digests were chosen for further investigation. These were the *Bam* HI (9.3 kb), *Bam* HI/*Bgl* II (7.9 kb), *Pst* I/*Sst* I (5.5 kb) and *Sst* I/*Xho* I (5.5 kb) digests.

a)



b)

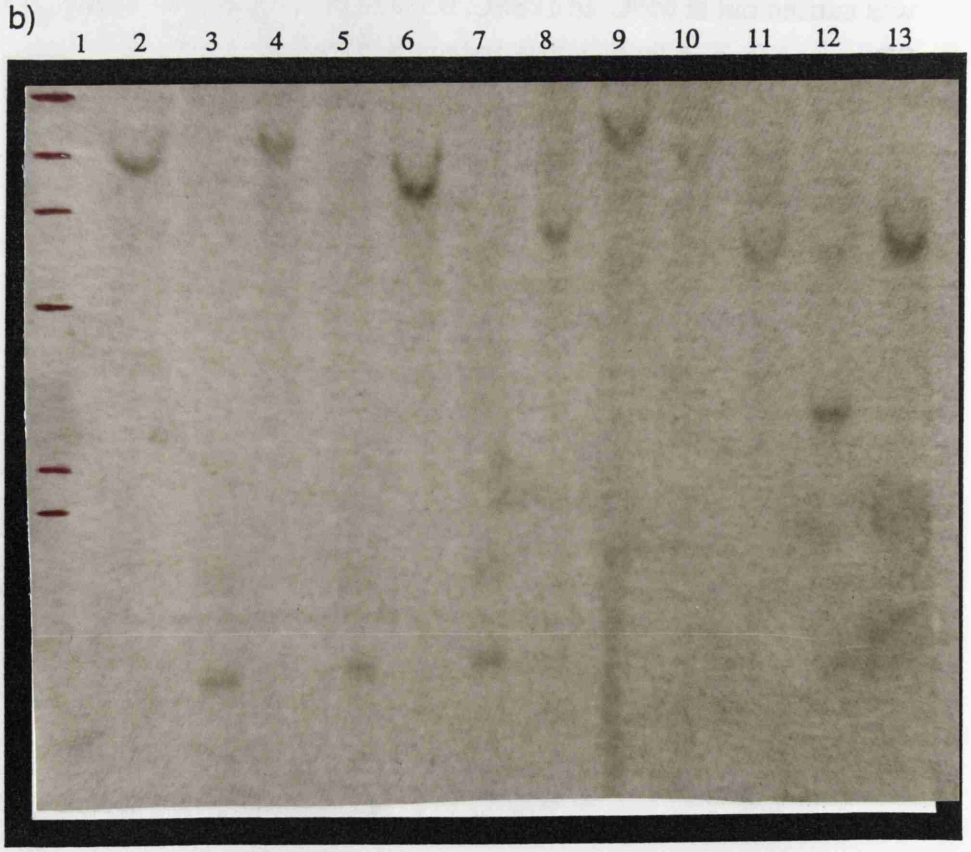


Figure 5.6 Restriction digests and Southern hybridisation of genomic DNA of *S. coelicolor* 1147

a) Genomic DNA of *S. coelicolor* 1147 was digested for 2 hours at 37°C, using 20 units of each of the restriction enzymes shown below. Fragments were separated by electrophoresis through a 1% (w/v) agarose gel (100V constant), for 4 hours. The gel was then photographed under UV illumination (2.3.11)

b) The DNA was transferred onto Hybond-N, by Southern blotting in alkali and the DNA fixed by UV crosslinking (2.3.17). DIG-labelled PCR product was used as the probe (5.2.4). Hybridisation was carried out at 65°C (with 1xSSC buffer), overnight. Washing was at 65°C and in 0.1xSSC. Chemiluminescent detection with DIG was carried out and autoradiography was at room temperature for 4 hours.

The size of each hybridising band is shown below. The size was calculated by comparison with the mobilities of the λ *Hind* III DNA markers (2.3.1), using a calibration graph.

1. λ *Hind* III markers (2.3.1)		8. *Pst* I	(5.75 kb)
2. *Bam* HI	(9.3 kb)	9. *Sst* I	(11.5 kb)
3. *Sal* I	(1.3 kb)	10. *Xho* I	(9.3 kb)
4. *Bgl* II	(11.5 kb)	11. *Pst* I/ *Sst* I	(5.5 kb)
5. *Sal* I/ *Bam* HI	(1.3 kb)	12. *Pst* I/ *Xho* I	(3.0 kb)
6. *Bam* HI/ *Bgl* II	(7.9 kb)	13. *Sst* I/ *Xho* I	(5.5 kb)
7. *Bgl* II/ *Sal* I	(1.3 kb)	14. λ *Hind* III markers	

The double digests were picked, because the non-compatible ends would prevent self-ligation of the vector. The *Bam* HI digest was also chosen, as dephosphorylated pUC18/19, previously digested with *Bam* HI was available in the laboratory.

5.2.6 Purification of size selected hybridising DNA

10µg of genomic DNA was again digested with the enzymes mentioned above, and the fragments separated on a 1% (w/v) agarose gel. A slice of gel, which contained DNA of the size which had previously given a hybridisation signal (Fig. 5.7), was excised and the DNA extracted by centrifugation through a Spin-X tube, as described in 2.3.13.

5.2.7 PCR of size selected genomic S.coelicolor DNA as template

A PCR was carried out on the purified DNA (5.2.6), to see which of the size-selected DNA fragments was most likely to contain the *icl* gene. The PCR reaction was carried out as in section 5.2.2. Only the 5.5kb *Sst* I/*Xho* I fragment yielded a PCR product of the correct size (Fig. 5.8). This DNA was used to construct a plasmid sub library.

5.2.8 Construction of a plasmid sub library

The plasmid pBluescript II (ks+) was digested with *Sst* I and *Xho* I. The DNA was subjected to electrophoresis and the linearised plasmid band excised from the gel. The DNA was then extracted by centrifugation through a Spin-X tube (2.3.13)

A ligation was set up containing approximately 100ng of the size-selected *Sst* I/*Xho* I DNA and 30ng of linearised pBluescript II (ks+) vector. There was little self -ligation of the vector because it had been digested with two restriction enzymes yielding non-compatible ends.

E.coli DS941 was made competent by the method of Hanahan (2.8.1.1) and different amounts of the ligation mix were then used to transform the cells. Colonies that contained recombinant plasmids (detected by the white colour of colonies plated on media containing X-gal, 2.2.8.3), were then picked and screened by hybridisation with the DIG-labelled PCR product.

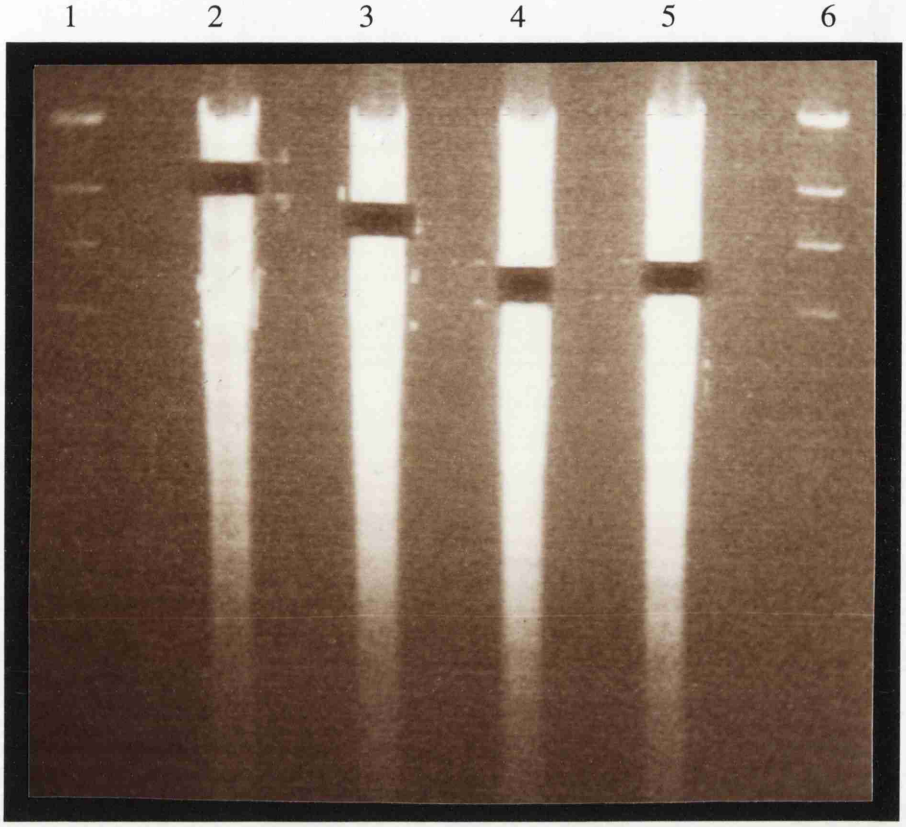


Figure 5.7 Size selection of DNA for analysis by PCR

10µg of genomic DNA of *S. coelicolor* 1147 was restricted with 20 units of each enzyme, for 2 hours at 37°C. The resulting digested DNA was separated on a 1% (w/v) agarose gel (100V constant), for 4 hours. The area of each digest (appearing as a dark box) containing the DNA which corresponded in size to the hybridising band (Fig. 5.6), was excised and the DNA purified. Molecular weight markers were run alongside (lanes 1 and 6), to enable sizing of the DNA (2.3.1).

Lane	Digest	Size of DNA excised
2	*Bam* HI	(~9.2-12 kb)
3	*Bam* HI/ *Bgl* II	(~6.6-9.0 kb)
4	*Pst* I/ *Sst* I	(~4.3-6.0 kb)
5	*Xho* I/ *Sst* I	(~4.4-6.1 kb)

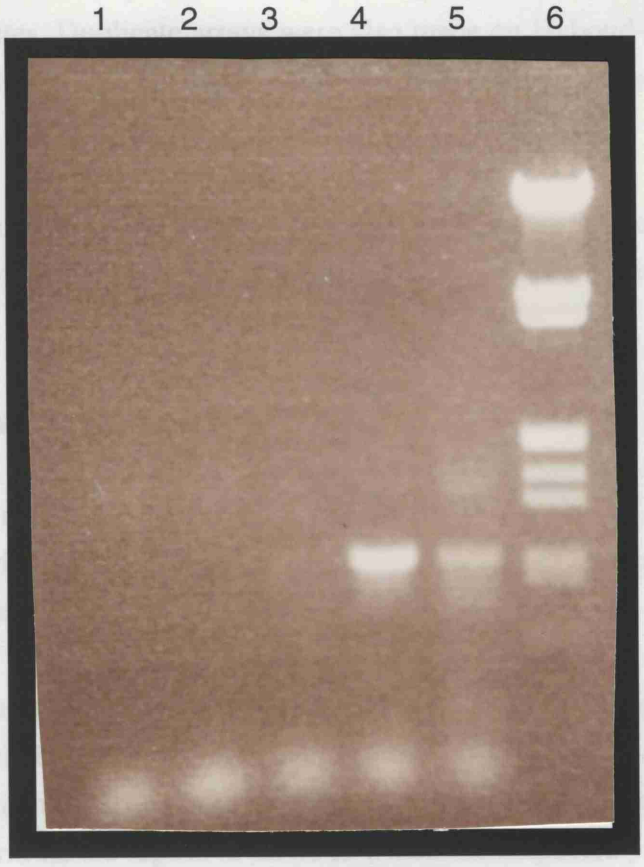


Figure 5.8 Analysis of size-selected genomic DNA of *S. coelicolor* 1147 by PCR

Purified size-selected DNA (Fig. 5.7) was subjected to PCR, to determine which of the fragments would amplify the PCR product, corresponding to part of the *icl* gene.

It was found that only DNA digested with *Xho* I and *Sst* I (~4.4-6.1 kb), would amplify the correct fragment. A PCR of genomic DNA was carried out as a control. Conditions were as described in 5.2.2, with the exception that only 100ng of DNA was used per reaction.

Lane	Digest	Size of the digested DNA used
1	*Bam* HI	(~9.2-12 kb)
2	*Bam* HI/ *Bgl* II	(~6.6-9.0 kb)
3	*Pst* I/ *Sst* I	(~4.3-6.0 kb)
4	*Xho* I/ *Sst* I	(~4.4-6.1 kb)
5	undigested control genomic DNA	
6	λ *Hind* III/ *Eco* RI markers	

400 colonies were picked and plated onto square L-agar plates containing ampicillin, as a 10x10 array on 4 plates. Duplicate arrays were also made on Hybond-N and the colonies grown overnight, before being treated for hybridisation, which was done under conditions descibed in 2.3.18.

Only one colony of the 400 gave a positive signal from both sets of filters (Figure 5.9). Plasmid (DS941/pICL92) was isolated from this recombinant and subjected to restriction mapping.

5.2.9 Restriction mapping of pICL92

pICL92 was digested with various restriction enzymes, which were known to cut within the pBluescript II (ks+) polylinker. The DNA fragments were separated on an agarose gel and a restriction map determined (Fig. 5.10).

The DNA was then blotted according to the method of Southern onto Hybond-N and subjected to prehybridisation and hybridisation as described previously in 5.2.5. By comparing the fragments which gave hybridisation signals, with the restriction map, it was possible to determine the region of DNA to which the amplified product hybridised.

It appeared that the *icl* PCR product hybridised most strongly to the 2.5kb *Pst* I/*Xho* I fragment. This 2.5 kb *Pst* I/*Xho* I fragment was then subcloned into pBluescript II (ks+), to create the plasmid pICL92B.

It was important to determine whether the whole of the *icl* gene had been cloned. This was achieved, by determining where the homologous PCR product hybridised within pICL92B. Southern analysis of restricted fragments of pICL92B (Fig. 5.11) revealed that the probe bound strongly to a 0.5 kb *Pst* I/*Sal* I fragment.

As the predicted size of the *S.coelicolor icl* gene was 1.5 kb and the cloned DNA (pICL92) either side of the 0.5 kb *Pst* I/*Sal* I fragment was greater than this size, it was clear that the whole of the gene had been cloned. However the orientation of the gene within the insert was not known at this time.

Figure 5.9 Colony hybridisation

400 colonies of *E.coli* were grown on four separate plates. DNA from the lysed colones was transferred onto Hybond-N (in duplicate) and the filters probed with DIG-labelled ICL/PCR product (5.2.5).

Hybridisation was carried out at 65°C in 1xSSC. Washing was at 70°C in 0.1xSSC. DIG detection was carried out and autoradiography was for 2 hours at room temperature. One set of the duplicate filters is shown and only one positively hybridising colony could be identified.

This positively-hybridising recombinant was named pICL92.

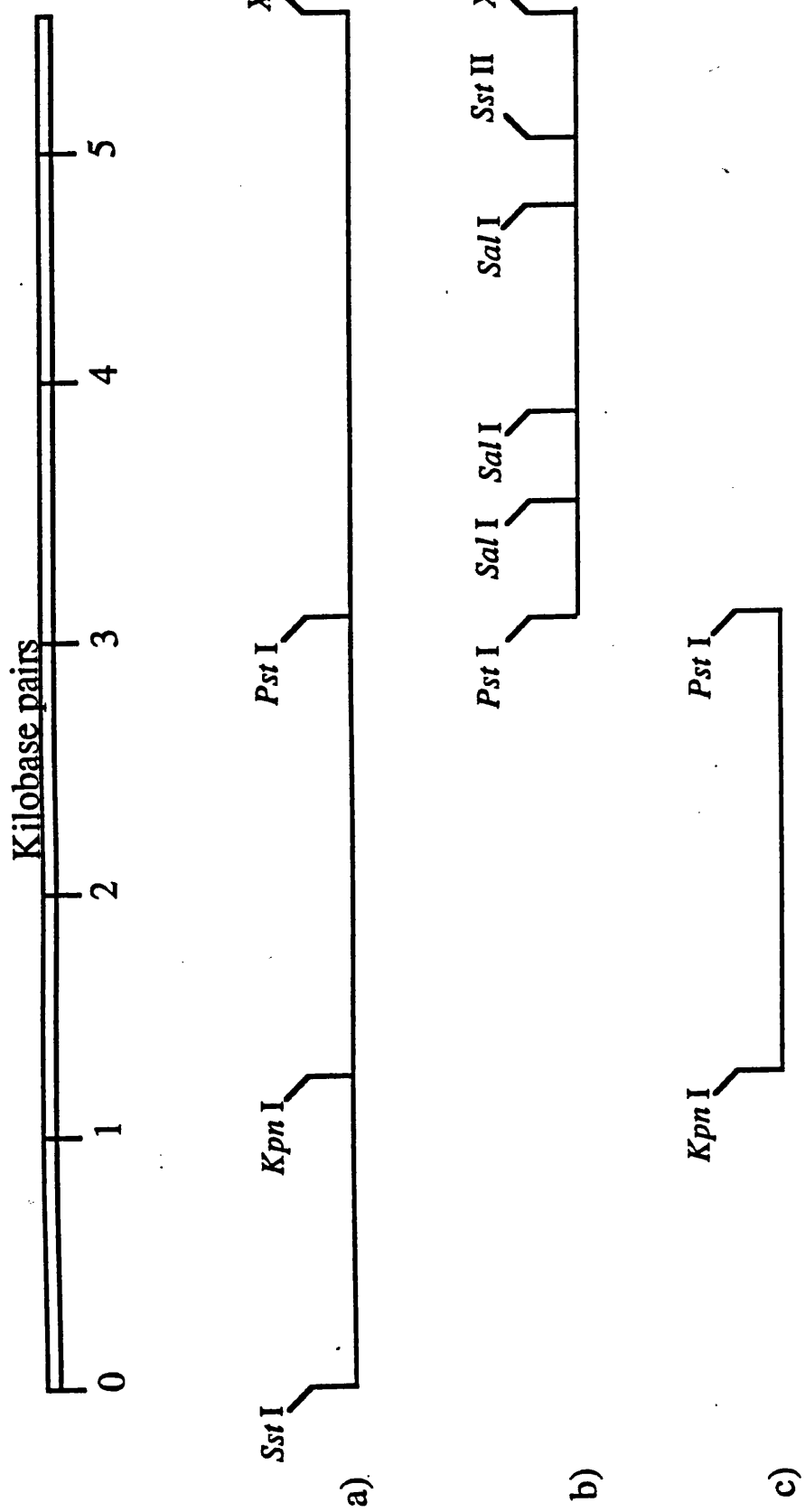


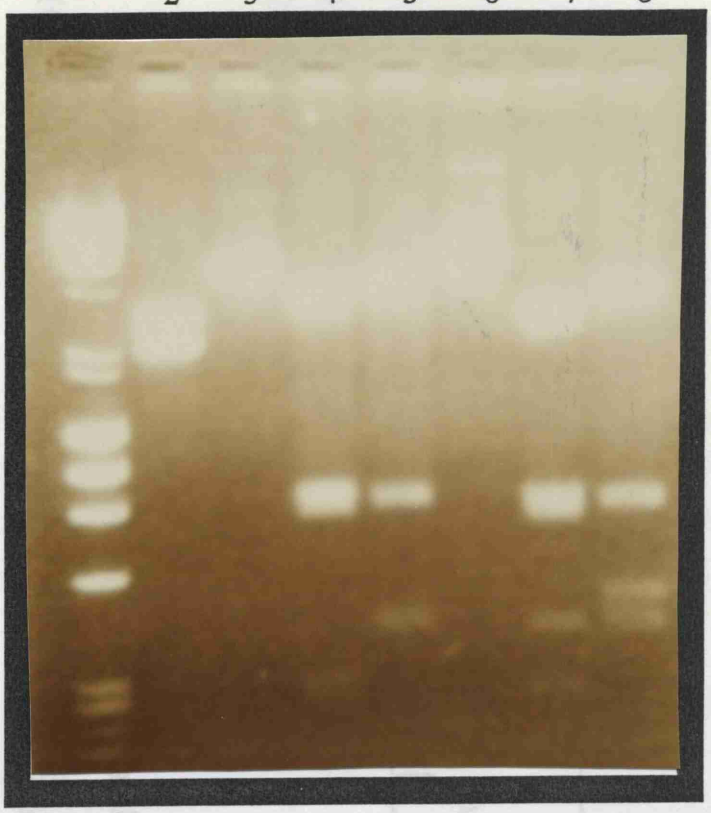
Figure 5.10 Restriction maps of plasmid clones and subclones used in this study

a) A clone containing the plasmid pICL92 was obtained by screening a plasmid sublibrary created in pBluescript II (ks+) (5.2.7) by colony hybridisation. A partial plasmid map was determined and relevant restriction sites are shown (5.2.8)

b) The 2.4 kb *Pst I/Xho* I fragment from pICL92 was cloned into pBluescript II (ks+), creating pICL92B. Futher restriction sites were determined by plamid mapping (5.2.8). More than one *Sst* II site was present. However only the site shown could be mapped with certainty.

c) The 1.8 kb *Kpn I/Pst* I fragment from pICL92 was cloned into pBluescript II (ks+), creating pICL92C.

a)



b)

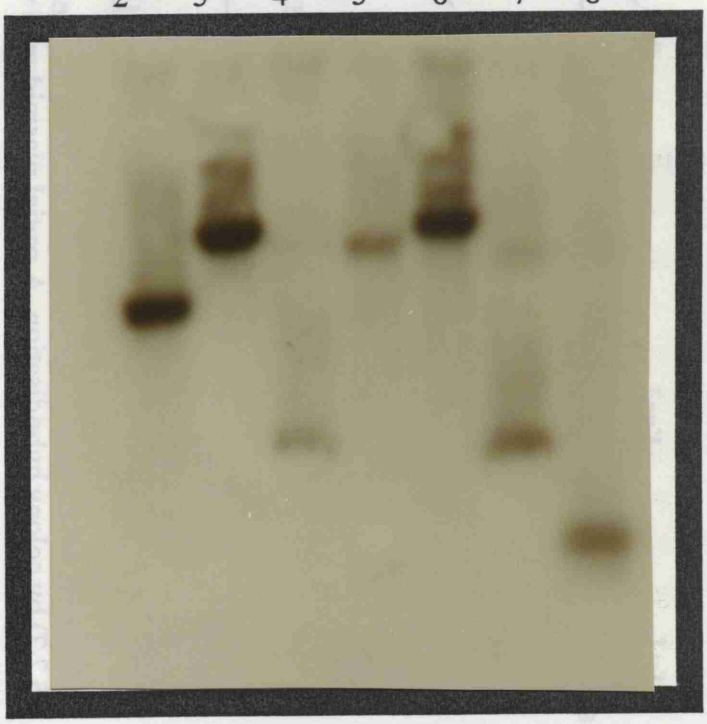


Figure 5.11 Southern analysis of pICL92B

a) Plasmid pICL92B was digested with various restriction enzymes, which were known to cut within the pBluescript II (ks+) polylinker. The resulting fragments were separated on a 1% (w/v) agarose gel and a restriction map determined (5.10).

b) The DNA was transferred to Hybond-N membrane by Southern blotting. Hybridisation was carried out using the homologous PCR product (5.2.2), in order to determine where the PCR product bound within the plasmid. Hybridising fragments are shown on the autoradiograph.

Lane	Digsets	Hybridising fragment (kb)
2	*Pst* I/*Xho* I	2.5
3	*Kpn* I	5.5
4	*Sst* II	0.8
5	*Sal* I	4.2
6	*Bam* HI	5.5
7	*Kpn* I/*Sst* II	0.8
8	*Sal* I/ *Bam* HI	0.5

CHAPTER 6

DNA Sequencing and analysis of *icl*

6.1 Introduction

It was of interest to sequence the *icl* gene and the surrounding regions, for a number of reasons. For example, it was of great importance to find out if *icl* was part of an operon containing other metabolic genes, as it is in *E.coli*. Few operons containing genes for primary metabolism have been studied in *Streptomyces* (Smith and Chater, 1988; Fornwald *et al,* 1987), so it would be of interest to study a new operon.

It would be important to study the control of expression of *icl*, because it has been shown (4.2) that the enzyme is not expressed (or at least extremely poorly) when cells are grown on glucose as a sole carbon source. However, when the cells were grown on substrates based on Tween, extremely high levels of enzyme were observed. It would be of interest to study if the expression is controlled at the level of transcription, or translation (or both). If *icl* were part of an operon, then there could also be a repressor or activator protein, which would modulate the level of transcription.

Other questions which could be addressed by sequencing around the *icl* locus would be whether there is an IDH kinase/phosphatase gene in close proximity. The work of Taylor (1992), had suggested that the *S.coelicolor idh* gene is not phosphorylated in the manner found in *E.coli*, so the likelihood is that the kinase/phosphatase gene was not present in *S.coelicolor*.

It was also important to find out why the attempts at cloning the *S.coelicolor icl* gene using oligonucleotides designed against conserved regions of other ICL's were unsuccessful. It would be possible to deduce from the DNA sequence if *S.coelicolor* ICL was atypical within these regions conserved in other species and therefore that the oligonucleotides had been designed based on an incorrect premise, resulting in poor hybridisation to genomic DNA.

This chapter describes the strategy employed to obtain the sequence of the *icl* gene of *S.coelicolor* and the sequence up and downstream from it. The analysis of the DNA, such as open reading frames, putative terminators and analysis of the protein sequence obtained from translation of the DNA sequence are also presented.

6.2 Results and Discussion

6.2.1 Construction of sub-clones used for DNA sequencing

To simplify the task of sequencing, sub-clones were contructed in M13.

The plasmid pICL92B was digested with the resctriction enzymes *Pst* I and *Xho* I, in order to excise the insert. The insert was then separated from the vector by electrophoresis and the band excised and purified using a Spin-X column. The purified DNA was then digested with *Sal* I, to yield 4 fragments.

Two *Sal* I/*Sal* I fragments of 0.4kb and 0.8kb, one *Pst* I/*Sal* I fragment of 0.5kb and one *Xho* I/*Sal* I fragment of 0.6kb were obtained. The DNA was separated on a 1.5% agarose gel, the bands excised and the DNA purified, as before.

The M13 vectors, mp18 and mp19 were digested with the enzymes mentioned above, to enable cloning of the inserts in both orientations, with respect to the -40 primer.

The plasmid, pICL92C was also digested with *Kpn* I and *Pst* I and the insert subcloned into similarly-digested M13mp18 and M13mp19, in the same manner as described above.

The *Sal* I/*Sal* I inserts were only subcloned into M13mp18, as both orientations were obtained.

The sub-cloning yielded 10 M13 clones, which were then used for sequencing (Fig. 6.1)

6.2.2 Sequencing strategy

The method chosen for sequencing of the DNA, was the di-deoxy chain termination method of Sanger *et al* (1977). Many commercial kits are available and a number of different DNA polymerases may be employed depending on the template being sequenced.

The most commonly used DNA polymerase was the T7 polymerase, which has been modified to remove its 5' to 3' exonuclease activity. This is available from USB as "Sequenase" or from Pharmacia as modified T7 polymerase.

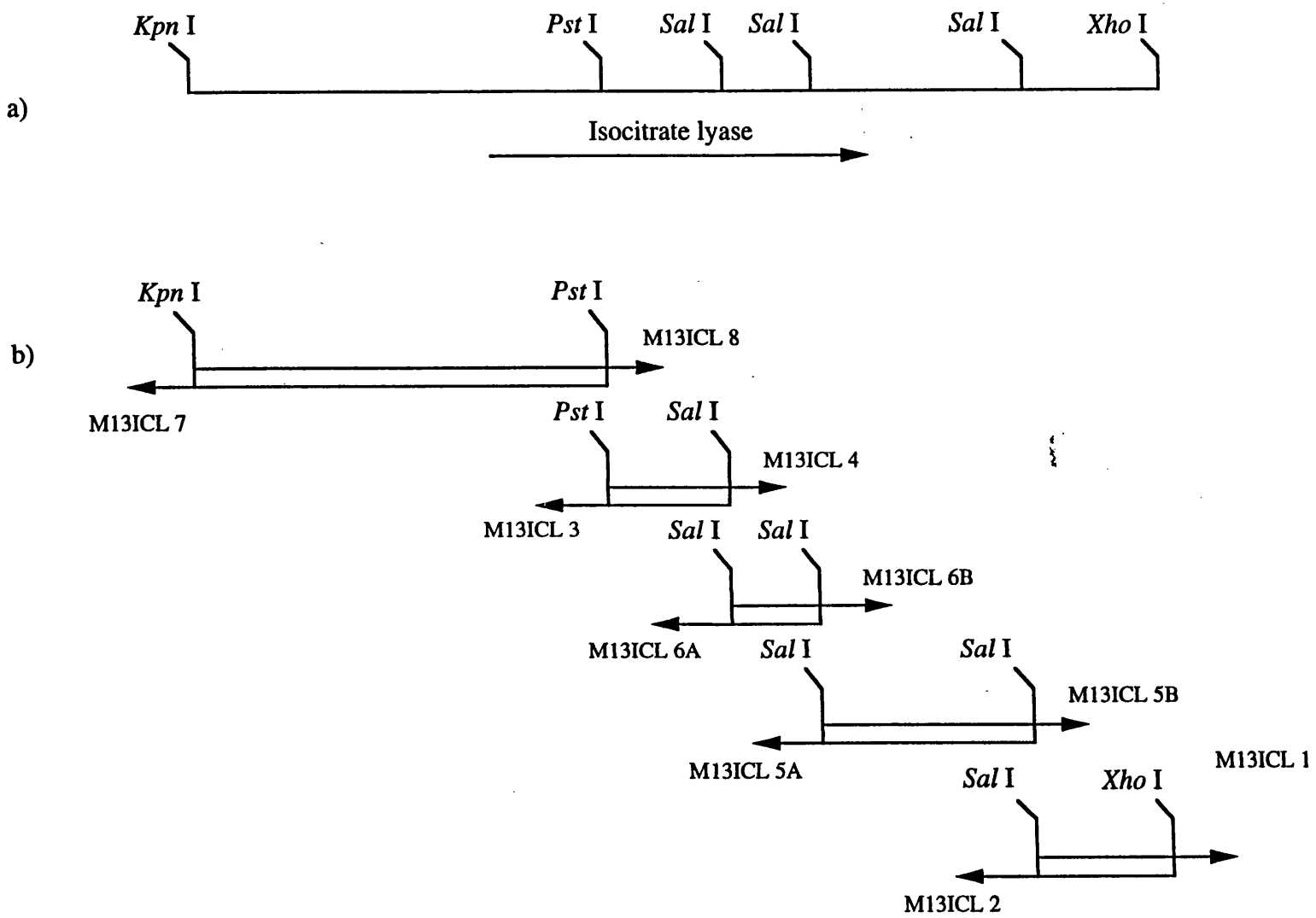

a)
Kpn I
Pst I
Sal I
Sal I
Sal I
Xho I
Isocitrate lyase
b)
Kpn I
Pst I
M13ICL 8
M13ICL 7
Pst I
Sal I
M13ICL 4
M13ICL 3
Sal I
Sal I
M13ICL 6B
M13ICL 6A
Sal I
Sal I
M13ICL 5B
M13ICL 5A
M13ICL 1
Sal I
Xho I
M13ICL 2

Figure 6.1 - Summary of the construction of M13 sub-clones used for sequencing of the *S. coelicolor icl*

a) This shows the entire region from which subclones were derived. Only the relevant restriction sites used to construct the subclones are shown. The arrow represents the open reading frame relative to the restriction sites.

b) Fragments of DNA containing part of the *icl* gene and the surrounding regions were introduced into M13mp18 and 19, for sequencing. Sequencing was carried out using the -40 primer and oligonucleotides which had been designed against the sequence obtained. The position of the clones relative to the *icl* gene is shown. The labelled arrows indicate the name of the sub-clone and the direction in which the DNA was sequenced (e.g. M13ICL4 was used to sequence the coding strand of *icl* and M13ICL3 used to sequence the non-coding strand).

The protocols were followed as per the manufacturers instructions, with the exception of the extension reaction, which was carried out at 42°C instead of 37°C. This was done to enable the polymerase to read through regions of secondary structure, which were formed due to the high G+C content of streptomycete DNA. When the polymerase is unable to read through secondary structure, it causes a pausing of the polymerase, which leads to a "four track stop". This manifests itself as a strong band across all four lanes of the autoradiograph, making it difficult to determine the sequence around this area of the gel. Raising the temperature of the extension reaction can help to reduce this problem.

The high G+C content of streptomycete DNA also results in a number of compressions. "Four track stops" are caused by the formation of secondary structures during the sequencing reactions. Compressions are due to secondary structures not being resolved while the gel is running. This causes the DNA to run aberrantly, with on occasions two bands from different lanes appearing to have the same mobility. This makes the task of reading the sequence difficult. There are however deaza-nucleotides, which help prevent the G-C base pairing, which is the cause of much of the secondary structure. These nucleotide analogues were used routinely to help resolve formation of compressions (Mizosawa *et al,* 1986).

The two most common types of templates used for sequencing were single-stranded templates generated from M13, or denatured-double stranded templates obtained from plasmid clones. Sequencing of the single-stranded template, would routinely give 250-350 nucleotides of sequence, whereas the double-stranded template would give 200-250 bases.

Certain regions of DNA contained such a high degree of secondary structure that T7-polymerase could not effectively read through it. When this was the case, *Taq* polymerase was used to alleviate this problem.

Taq polymerase is a thermostable DNA polymerase from *Thermus aquaticus*, which allows sequencing reactions to be carried out at elevated temperatures, of up to 70°C. The elevated temperature reduces the formation of secondary structures and enables the polymerase to read through the DNA.

Less sequence (200-250 bases) could be read from a Taq sequencing reaction, as the bands are thicker and more diffuse. This meant that it was not the DNA polymerase of choice and was only used when secondary structure formation became a problem.

All 10 M13 suclones were sequenced using the -40 primer, which aneals just outside of the multiple cloning site. Oligonucleotides were then designed against the sequence obtained, to enable further sequencing of the sub-clones. The whole of the *icl* gene was sequenced on both strands (Fig. 6.2). Some of the surrounding regions were only sequenced on one strand, and would need to be validated by sequencing of the other strand.

The DNA was sequenced on average 2-3 times, in order to correct any mistakes that may have arisen. The sequence obtained in shown in Fig. 6.3.

6.2.3 Identification of Open Reading Frames ORF's

The program "CODONPREFERENCE" from the GCG package (Devereaux *et al* 1984), was used to identify ORF's . The program finds regions of each reading frame, in a DNA sequence, that show either a strong codon preference (Gribskov *et al,* 1984), or a strong G+C% bias in the third position of each codon (Bibb *et al,* 1984).

The program is used to help identify protein coding regions, the frame which encodes the proteins and the direction of translation. It is also useful in predicting the location of DNA sequencing errors.

Using a previously-derived codon usage table, a statistic is calculated for a given window of sequence, which then shows the similarity to the usage table. The window is then moved along by one base and the statistic calculated again. The procedure is continued in this manner for the whole sequence, so that a continuous function is created. For an ORF to be predicted, the continuous statistic rises significantly above the level which would be generated from a random sequence.

The codon usage table was compiled from 67 previously-sequenced streptomycete genes (Fig. 6.8) and this was used in the calculation of the similarity statistic.

The third position G+C% bias is a very useful way to identify potential coding regions within streptomycete genes. In a streptomycete ORF a G or C is found in the third position of a codon in greater than 90% of cases. In non-coding DNA, the distribution is more random. Thus an ORF can be recognised from the third position G+C bias.

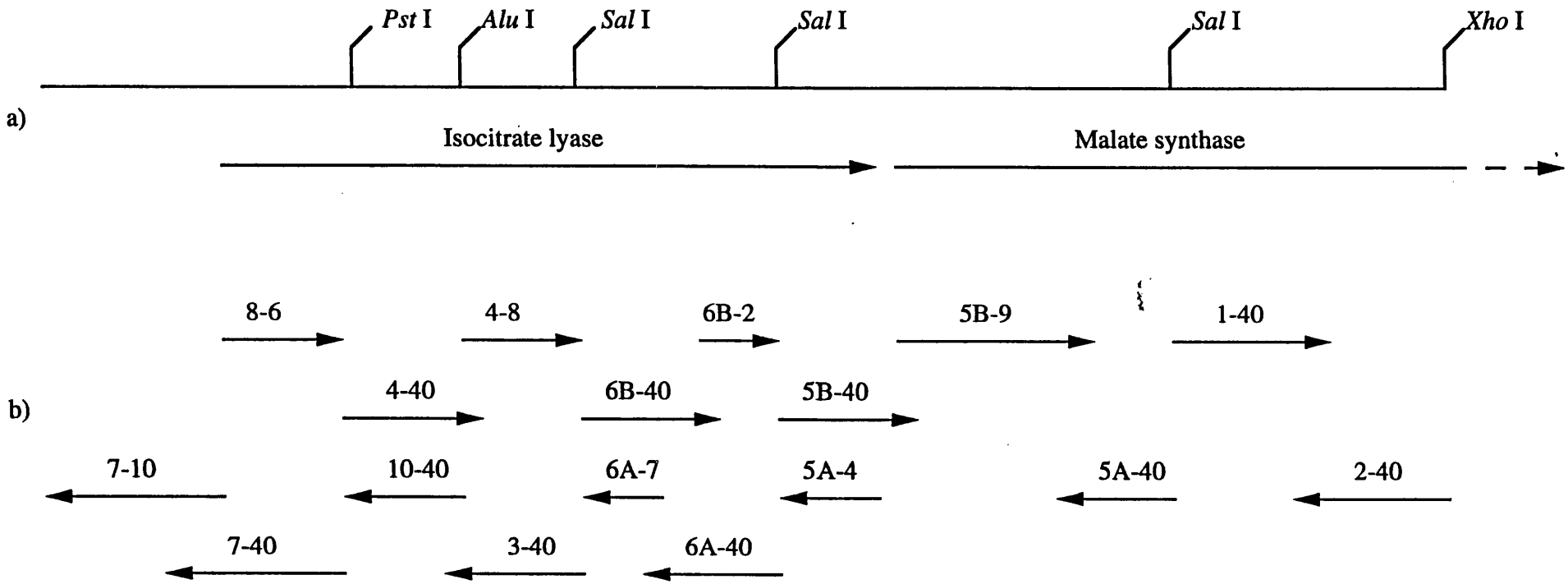

Pst I
Alu I
Sal I
Sal I
Sal I
Xho I
a)
Isocitrate lyase
Malate synthase
8-6
4-8
6B-2
5B-9
1-40
b)
4-40
6B-40
5B-40
7-10
10-40
6A-7
5A-4
5A-40
2-40
7-40
3-40
6A-40

Figure 6.2 Regions of sequence generated from sequencing of the various M13 subclones

a) The entire region from which sequence was obtained is shown. The two arrows represent the open reading frames of *icl* and malate synthase (*ms*).
The dashed arrow 3' to *ms* shows that some of the gene is still to be found 3' of the *Xho* I site.

b) Each arrow represents the direction and the relative amount of DNA sequenced from each primer.
The figures above the arrows represent the clone and the oligonucleotide used in sequencing (e.g. 5B-40, represents the clone M13ICL5B sequenced using the -40 primer. 5A-4, represents the clone M13ICL5A sequenced using oligonucleotide number 4).

```
                -35 I                    -10 I                          -35
                _____                    _____                          ___
(A) 1    CGCAAAGTTGGCAAAGTACGGCAACAAGATTCACAGAAGTTGGCATATCGCCAAGGTTGA
                                         -35 II    -35 III      -10 II
         IV                 -10 IV
         __                 _____
(A) 61   TGCACTGAGTGCCAATGCCAGAGTCTCAATGCGGCCCACCGCACCCGGCAGGCACCGCTC
         -10 III

(A) 121  GGCACGGGGGCACGATTCATCGCCCTGACGTCGGAGACCCCGGCATCCGTAGCGGTCTGC

                                                        -10 VI
                                                        _____
(A) 181  GCCCGGCAGCATTGCTCCCTTCGCCATCGGGTGCGGCGGGCCGCACCCCAGTAAGTCGAC

                -35 VI                          -35 V
                _____                           _____                  _
(A) 241  ACGCAGTGCCAACTTTGCTCGTTCCGGCGGCGGGGTTGGCAATCGGTACAGACCGAGGAG
                                                                  RBS
         -10 V
         _____
(A) 301  ACGGTGACAGTCATGGCAGAAGCGAGGACGCAGGCGGCCGAGGACCTGGCCCGGCGCTGG
(B)                  M  A  E  A  R  T  Q  A  A  E  D  L  A  R  R  W

(A) 361  GACACCGAGCCGAGGTGGCAGGGCATCGAGCGCACCTACAGTGCCCAGGACGTCGTCCGG
(B)      D  T  E  P  R  W  Q  G  I  E  R  T  Y  S  A  Q  D  V  V  R

(A) 421  CTCTCGGGCAGTGTCCGTGAGGAGCACACCCTGGCCCGGCGCGGTGCCGAGCGGCTGTGG
(B)      L  S  G  S  V  R  E  E  H  T  L  A  R  R  G  A  E  R  L  W

(A) 481  GCGGTCGACGTGCTCGCCCTGATGCAGGTGCGCGAGCCGCGCGAGTGGCCGCCAGTCCGG
(B)      R  Q  L  H  E  R  D  Y  V  H  A  L  G  A  L  T  G  G  Q  A

(A) 541  GTGCAGCAGATCAAGGCGGGCCTGCAGGCGATCTACCTGTCCGGCTGGCAGGTCGCCGCC
(B)      V  Q  Q  I  K  A  G  L  Q  A  I  Y  L  S  G  W  Q  V  A  A

(A) 601  GACGCCAACCAGGCCGGGCACACCTATCCCGACCAAAGCCTCTACCCGGCCAACTCCGTG
(B)      D  A  N  Q  A  G  H  T  Y  P  D  Q  S  L  Y  P  A  N  S  V

(A) 661  CCGCAGGTGGTGCGTCGGATCAACAACGCGCTGCTGCGCGCCGACCAGATCGCCACCAGT
(B)      P  Q  V  V  R  R  I  N  N  A  L  L  R  A  D  Q  I  A  T  S

(A) 721  GAGGGCGACAACTCCACCGACTGGCTGGCACCGATCGTCGCGGACGCGGAGGCCGGCTTC
(B)      E  G  D  N  S  T  D  W  L  A  P  I  V  A  D  A  E  A  G  F

(A) 781  GGCGGCCCGCTCAACGCCTTCGAGCTGGCCAAGGCGATGATCGCGGCGGGCGCGGCCGGC
(B)      G  G  P  L  N  A  F  E  L  A  K  A  M  I  A  A  G  A  A  G
```

```
(A) 841 ATCCACTACGAGGACCAGCTCGCCTCGGAGAAGAAGTGCGGGCACCTCGGCGGCAAGGTC
(B)     I  H  Y  E  D  Q  L  A  S  E  K  K  C  G  H  L  G  G  K  V

(A) 901 CTCGTCCCCACCTCGCAGCACATCCGCACCCTGAACGCAGCCCGTCTCGCCGCCGACATC
(B)     L  V  P  T  S  Q  H  I  R  T  L  N  A  A  R  L  A  A  D  I

(A) 961 GCTGACACCCCGACCGTGATCATCGCCCGCACCGACGCCCTCGCCGCCAATCTGCTGACG
(B)     A  D  T  P  T  V  I  I  A  R  T  D  A  L  A  A  N  L  L  T

(A)1021 TCCGACGTCGACGAGCGGGACGCTCAGTTCGTCACCGGTGAGCGCACGCCGGAGGGCTTC
(B)     S  D  V  D  E  R  D  A  Q  F  V  T  G  E  R  T  P  E  G  F

(A)1081 TACCGGGTCCGCAACGGCATGGCACCCGTGATCGCCCGTGGCCTGGCCTACGCCCCGTAC
(B)     Y  R  V  R  N  G  M  A  P  V  I  A  R  G  L  A  Y  A  P  Y

(A)1141 GCCGATCTGATCTGGGTCGAGACCGGCACTCCGGACCTGGAGCAGGCGCGCGAGTTCGCC
(B)     A  D  L  I  W  V  E  T  G  T  P  D  L  E  Q  A  R  E  F  A

(A)1201 GAGGCGATCCACGCCGAGCACCCGGACCAGATGCTGGCGTACAACTGCTCGCCGTCCTTC
(B)     E  A  I  H  A  E  H  P  D  Q  M  L  A  Y  N  C  S  P  S  F

(A)1261 AACTGGAAGGCCGCGCTGGACGACGACCAAATCGCCAAGTTCCAGCGCGAGCTGGGCGCG
(B)     N  W  K  A  A  L  D  D  D  Q  I  A  K  F  Q  R  E  L  G  A

(A)1321 ATGGGCTACAAGTTCCAGTTCATCACCCTGGCCGGCTTCCACTCCCTGAACCACGGCATG
(B)     M  G  Y  K  F  Q  F  I  T  L  A  G  F  H  S  L  N  H  G  M

(A)1381 TTCGACCTGGCCCGCGGCTATGCCGAGCACGGCATGACCGCCTACGTCGACCTCCAGGAG
(B)     F  D  L  A  R  G  Y  A  E  H  G  M  T  A  Y  V  D  L  Q  E

(A)1441 AGGGAGTTCGCCGCGCAGGCCCACGGCTTCACCGCCGTCAAGCACCAGCGCGAGGTCGGC
(B)     R  E  F  A  A  Q  A  H  G  F  T  A  V  K  H  Q  R  E  V  G

(A)1501 ACCGGCTACTTCGACCTGGTCTCCACCGCCGTCAACCCAGCCTCCTCCACGACCGCTCTC
(B)     T  G  Y  F  D  L  V  S  T  A  V  N  P  A  S  S  T  T  A  L

(A)1561 GCCGGGTCCACGGAGGAGGAGCAGTTCCACTAGGCCGCCGGACCGGGCGCCCCGCGTCCG
(B)     A  G  S  T  E  E  E  Q  F  H  *
                                      A

                ____________________          ____________________
(A)1621 GCCGAGCCCTCGGTCGAGCGGAGGGGGACGCTCCGTCCCCGTCCCCCTCCGCTCCCGTCC


(A)1681 CTCAGGAGAGCCGATGTCCACCACCTCACTCACCCACCACGTCGTGTCCTCGCCGCGCCC
(C)          RBS      M  S  T  T  S  L  T  H  H  V  V  S  S  P  R  P

(A)1741 GGTCCCCGGCACGACGAGATCCTCACCCCCGCCGCCCTGGACTTCGTCGGCGGGCTCGTC
(C)      V  P  G  T  T  R  S  S  P  P  P  P  W  T  S  S  A  G  S  S
```

```
(A)1801 GAGGCGTTCCAGCCGCGCCGACGGGATCTGGATGAAGGAGCGCCGCAGGTCGGCCCTGCG
(C)       R  R  S  S  R  A  D  G  I  W  M  K  E  R  R  R  S  A  L  R

(A)1861 GCTGGCGTCGGGCTCACCCCTCGACTTCCCCACGGTCACGTCCGCCGTGCGCAAGTACCC
(C)       L  A  S  G  S  P  L  D  F  P  T  V  T  S  A  V  R  K  Y  P

(A)1921 CACGTGGCGCGTGGCACCGCCCGCGCCCGGCCTGACCGACCGGCGGGTGGAGATGACCGG
(C)       T  W  R  V  A  P  P  A  P  G  L  T  D  R  R  V  E  M  T  G

(A)1981 GCCGCCCGACCGCCGCATGGCCGTCAACGCGCTCAACTCCGGCGCCCAGGTGTGGATGGC
(C)       P  P  D  R  R  M  A  V  N  A  L  N  S  G  A  Q  V  W  M  A

(A)2041 CGACTTCGAGGACGCCACTGCACCCCTGTGGGACAACATCATCGGCGGCCAGCTCACCCT
(C)       D  F  E  D  A  T  A  P  L  W  D  N  I  I  G  G  Q  L  T  L

(A)2101 CCTCGACGTCATCGAGCGGCGCATCGACTTCACGACTCCGGAGGGCAAGGAGTACCGGCT
(C)       L  D  V  I  E  R  R  I  D  F  T  T  P  E  G  K  E  Y  R  L

(A)2161 CGGTGACCGGCCGGCCACCATCATGGTCCGCCCCCGCGGCTGGCACCTCGACGAGGAGCA
(C)       G  D  R  P  A  T  I  M  V  R  P  R  G  W  H  L  D  E  E  H

(A)2221 CCTGGAGTACGACGGCAGGCCCGTACCCGCCACGCTCGTCGACTTCGGCCTGTACTTCTT
(C)       L  E  Y  D  G  R  P  V  P  A  T  L  V  D  F  G  L  Y  F  F

(A)2281 CCACTGTGCCCAGCGGCAGATCGACGCCGGCCACGGCCCGTACTTCTATCTGCCGAAGCT
(C)       H  C  A  Q  R  Q  I  D  A  G  H  G  P  Y  F  Y  L  P  K  L

(A)2341 GGAGAACCGCTACGAGGCCCGGCTGTGGAACGACGTCTTCCTGACGGCGCAGGAACTGCT
(C)       E  N  R  Y  E  A  R  L  W  N  D  V  F  L  T  A  Q  E  L  L

(A)2401 CGGCATCCCGCGCGGCACGGTCCGGGCCACCGTCCTCATCGAGACGATCACCGCGGCGTT
(C)       G  I  P  R  G  T  V  R  A  T  V  L  I  E  T  I  T  A  A  F

(A)2461 CGAGATGGAGGAGATCCTCCACGCGCTGCGCGAACACAGCGCCGGCCTCAACGCCGGCCG
(C)       E  M  E  E  I  L  H  A  L  R  E  H  S  A  G  L  N  A  G  R

(A)2521 CTGGGACTACCTCTTCAGCATCATCAAGACCTTCGGCCACCGCACCGACTTCCTCCTCCC
(C)       W  D  Y  L  F  S  I  I  K  T  F  G  H  R  T  D  F  L  L  P

(A)2581 CGACCGGGCGAAGGTCACGATGACCGCCCCCTTCATGCGCGCGTACACCGAACTCCTCGT
(C)       D  R  A  K  V  T  M  T  A  P  F  M  R  A  Y  T  E  L  L  V

(A)2641 CCGCACATGCCACAAGCGCGGCGTCCCACGCCATCGCGGCATGGCCGCCCAGGTCCCCGG
(C)       R  T  C  H  K  R  G  V  P  R  H  R  G  M  A  A  Q  V  P  G
```

```
 A)2701 CAAGGACCCGGCCGCGAACGAGGCCGCGTTCGCCAAGGTCCGCCTGGACAAGGAGCGCGA
(C)      K  D  P  A  A  N  E  A  A  F  A  K  V  R  L  D  K  E  R  E

(A)2761 GGCCGAGGACGGCTTCGACGGCTCCTGGGTCGCCCATCCCGGACTCGTCCCCGTCTGCCG
(C)      A  E  D  G  F  D  G  S  W  V  A  H  P  G  L  V  P  V  C  R

(A)2821 CGAGTCCTTCGACGCCGTCCTCGAG
(C)      E  S  F  D  A  V  L  E
```

Figure 6.3 DNA and amino acid sequence for the *icl* gene and surrounding regions

A) The entire region of DNA sequence obtained from sequencing M13 (Fig. 6.1) and plasmid clones (Fig. 5.10) is shown.

B) Translation of the ORF representing the *icl* gene is shown from nucleotide 313 to 1593.

C) Translation of a second ORF is shown from nucleotide 1695 to 2845.

I-V Represent 5 potential promoter regions found upstream from *icl* (Fig. 6.5)
VI Represents a potential promoter region which would result in transcription occuring in the opposite direction.

A potential terminator structure for the *icl* gene is marked **A** (Fig. 6.6)

All amino acid sequence was derived from the translation of the DNA sequences using the program "TRANSLATE", with each single letter amino acid code being written below the first base of each codon. Possible start and stop codons for *icl* and the second ORF are shown in bold, and putative ribosome binding sites are marked **RBS**.

Analysis of the "CODONPREFERENCE" plot (Fig. 6.4), revealed two identifiable ORF's. These are distinguished by firstly showing a high degree of similarity to the calculated codon frequency table and also showing a high G+C% bias in the third position. Secondly, there are far fewer rare codons in ORF's than in non-coding regions and thirdly, the program indicates possible ORF's by displaying a box, which is formed between any ATG codons and the next stop codon in that reading frame. When these parameters were considered together, the two ORF's were identified.

The plot shows there to be an ORF in the first frame starting at about 300 and ending at about base 1600. The second ORF is translated in the second frame, beginning around base 1700 and continuing through the end of the sequence. This means that this second gene has not been cloned in its entirety.

6.2.4 Translation of the *icl* gene from *S.coelicolor*

Streptomycete genes begin most frequently with an ATG, encoding formyl-methionine. They are however also capable of beginning with a GTG, which occurs 18% of the time (Seno and Baltz, 1989) and also very rarely a TTG.

The N-terminal amino acid sequence of the protein did not begin with a methionine and must therefore have been cleaved post-translationally from the protein. This is not uncommon for streptomycete proteins (Taylor, 1992 and P. White, personal communication). It was of interest to find out what the starting codon was. Analysis of the DNA sequence found there to be an ATG start codon at positions 313-315, which is consistent with the predicted start as estimated from the ORF depicted using "CODONPREFERENCE". The second and subsequent codon of the ORF corresponded to the N-terminal amino acid found from the sequence of the protein.

6.2.5 Database searching

The DNA sequence confirmed that the gene cloned indeed encoded the protein that had been purified - the translated DNA sequence matched that obtained from protein sequencing.

To investigate the similarity of the *S.coelicolor* ICL with other ICL's or indeed any other proteins, databases were searched by using the program "TFASTA", from the GCG package. "TFASTA" uses the Pearson and Lipman algorithm to search for similarity between a peptide sequence and any group of translated nucleotide sequences.

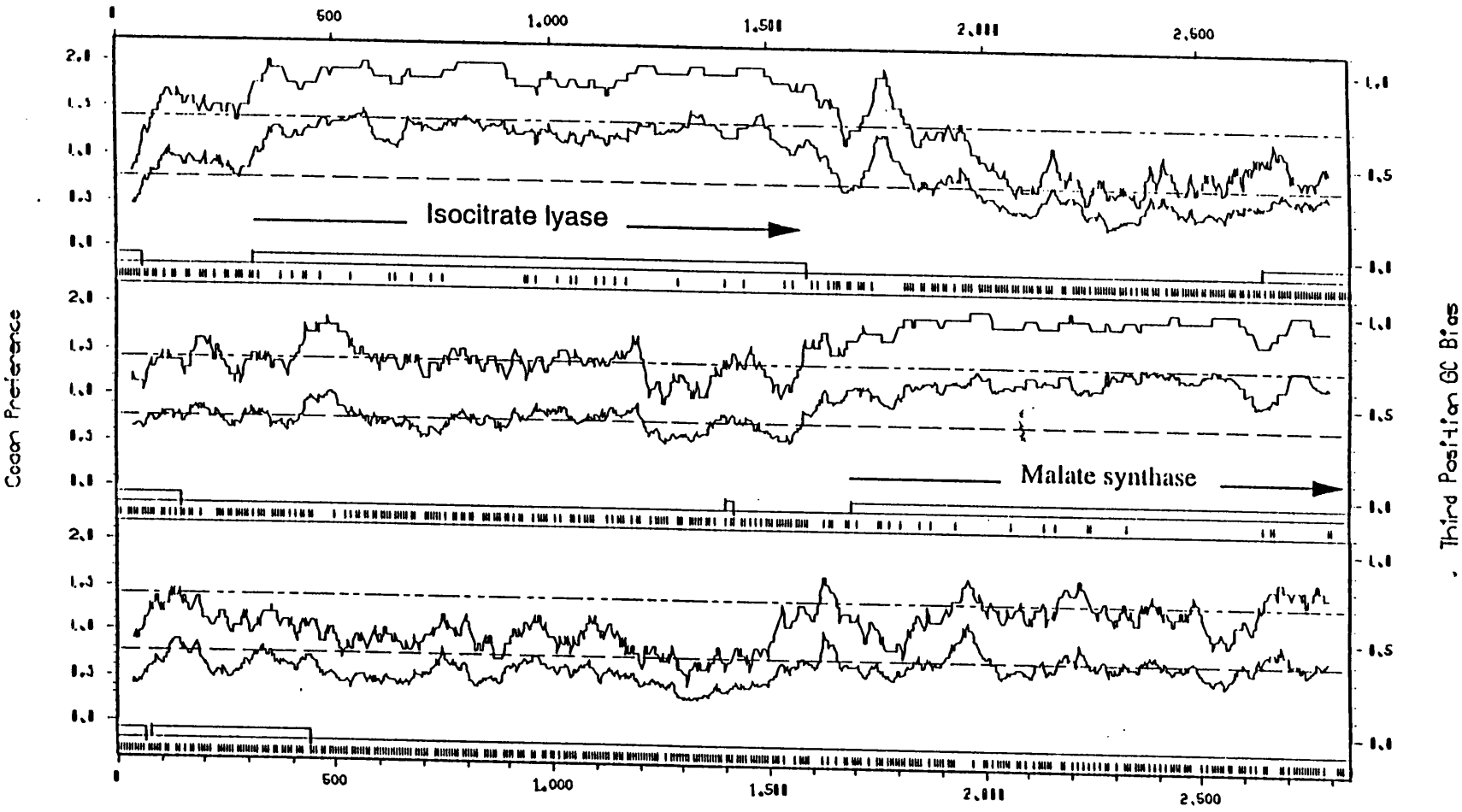

Codon Preference
Third Position GC Bias
Isocitrate lyase
Malate synthase
500
1,000
1,500
2,000
2,500

Figure 6.4 Analysis of the sequenced DNA for open reading frames - ORF's

The computer program "CODONPREFERENCE" was used to analyse the DNA sequence for ORF's.

The output of the program is a graphical plot, representing the possible ORF's from a given sequence.

The plot is separated into three sections, each section representing a frame from which the DNA would be translated. The top section represents frame 1, the second, frame 2 and the third, frame 3.

Each section then has a graphical trace representing the two methods depicting possible coding sequence. The top trace shows the G+C% for the third nucleotide of the codon in that frame. In streptomycete coding regions, the value is usually >90%.

The lower trace is the codon bias. This is calculated by comparing the codons of a given window (in this case 25), against a codon frequency table (Fig. 6.8), which has been compiled from previously sequenced streptomycete genes. This comparison is then given a value (P value), which if above 1 is indicative that the given codons are used more frequently than would be expected from a random sequence. The window is then moved along one base and the value calculated again. For the trace to represent a coding region, the trace needs to be consistently above 1.

Rare codons are also shown on the trace, by a mark on the horizontal axis of each trace. Rare codons were defined in this case, as codons with a P value of less than 1.

The plot shows two identifiable open reading frames, one from frame 1 starting around base 300 and a second ORF is found in frame 2, starting around base 1700. The two ORF's are only separated by a small region.

The first stop codon in the deduced reading frame, is a TAG, at position 1591-1593. This position is also in agreement with the predicted end of the ORF depicted from the "CODONPREFERENCE" analysis.

Translation of the ORF between these start and stop codons yields a coding region of 1281 nucleotides, resulting in a protein of 426 amino acids. The size of this protein would be 54 148 Da, which is in good agreement with the predicted size of ICL, as estimated by SDS-PAGE analysis.

The *icl* gene was translated into its peptide sequence and the GENEMBL database searched. The top five results were all ICL proteins from various organisms, with exceptional homology being found to the *E.coli* ICL (63.4% identity). No other proteins showed any significant sequence alignment.

The second ORF was unidentified, so it was of interest to investigate if the peptide sequence was similar to any other known protein sequence. The DNA was translated from the correct frame and the resulting peptide sequence used to search against the GENEMBL database. The top seven results were all malate synthase (MS) proteins. This was unexpected, because in *E.coli* the *ms* gene is upstream from *icl*, whereas in this case it is found downstream.

From this result, it can be seen that there is the possibility that the *S.coelicolor icl* and *ms* are part of an operon. However, further experimentation needs to be carried out to confirm this.

6.2.6 Identification of putative ribosome binding sites (RBS's) upstream from *icl* and *ms*

In order for the ribosomal subunit to recognise a translational start AUG, GUG or UUG, from any that may be found internally, there is usually a sequence upstream from the start which shows complementarity to the 3' end of the 16S rRNA, known as the Shine-Dalgarno sequence, or ribosome binding sequence (Shine and Dalgarno, 1974).

Although the Shine-Dalgarno sequence is usually found, it is not found in all cases. The *aph* gene from *S.fradiae*, when transcribed from the P1 promoter, lacks "leader" mRNA (Janssen 1989). In this case the transcriptional and translational start sites are the same and therefore no ribosome binding sequence exists.

The deduced Shine-Dalgarno sequences of 44 streptomycete genes were compiled (Strohl, 1992), to see how they differ across the species. The sequences were found to range from 5-12 nucleotides upstream of the initiation codon (average, 8.5 nucleotides). The binding strength of these sequences also varied from -2.2Kcal/mol to -22.2Kcal/mol, using the rules of Tinoco (1973), with an average of -11.3Kcal/mol. This brings into doubt the claims by McLaughlin *et al* (1981), that Shine-Dalgarno sequences of Gram-positive bacterial mRNA's, are typically able to form strong complexes with the 3' end of the 16S rRNA.

Strohl's analysis suggested a conserved Shine-Dalgarno sequence of (a/g)-G-G-A-G-G for streptomycetes. The sequences upstream from *icl* and *ms* were analysed for putative Shine-Dalgarno sequences. A RBS was found for each of the genes (Fig. 6.5).

Both RBS's had calculated free energies of binding close to the average for streptomycete RBS's and were spaced correctly from the initiation codon.

6.2.7 Identification of possible promoter sequences

Of the 139 streptomycete promoter sequences thus far observed, 29 of these are similar to the previously-described streptomycete "*E.coli*-like" promoter sequences (Hopwood *et al*, 1986). Many of these fall into the class of so-called "housekeeping" genes. *icl* could be considered as a "housekeeping" gene, so it would seem possible that this also would have a recognisable streptomycete "*E.coli*-like" promoter sequence.

There are two requirements for a promoter to be recognised as a streptomycete *E.coli*- like promoter. Firstly, it should appear to be similar to the -35 and -10 consensus sequences generated from the compilation of the 29 streptomycete promoter sequences, which appear to be similar to those recognised by RNA polymerase containing the σ^{70} subunit (Strohl, 1992). Secondly, there should be a spacer region between the -35 and -10 sequences of 16, 17 or 18 nucleotides. Over 92% of the *E.coli* promoters recognised by RNA polymerase holoenzyme with the σ^{70} attached, have spacers of 16-18 nucleotides.

The DNA sequence upstream of the *icl* gene was examined by eye, for similarity to streptomycete "*E.coli*-like" promoters (Fig. 6.6). Five possible promoters were observed (I-V). All showed a 5 out of 6 match to the -35 consensus and have between 2 and 5 matches out of 6 to the -10 consensus.

a) "16s rRNA" 5'-AGAAAGGAGGTGATC-3'

b) *icl* **A**CCG**AGGAG**AC**G**G**T**GACAGTCATG

c) *ms* CCTC**AGGAG**AGCCGATG

Figure 6.5 Putative ribosome binding sites for *ICL* and *MS*

a) This represents the sequence which is exactly complementary to the 16S ribosomal RNA from *S. lividans* and represents an ideal ribosome binding sequence for streptomycetes (Bibb and Cohen, 1982).

b) Shows the region upstream from the coding rgion of *icl*, which is most likely to be the ribosome binding site.

c) Shows the region upstream from the coding region of *ms*, which is most likely to be the ribosome binding site.

The letters shown in bold represent matches to the complementary sequence of the16S rRNA. The putative start ATG for each gene is underlined.

	-35		-10		
Consensus	TTGAC(G/A)	-- 16-18nt --	TAG(AA/GG)T		
I	**TTG**GC**A**	---- 16nt ----	**T**TC**A**CA	--- 278nt ---	ATG
II	**TT**CAC**A**	---- 17nt ----	A**AGGT**T	--- 255nt ---	ATG
III	**TTG**GC**A**	---- 15nt ----	**T**GC**A**C**T**	--- 247nt ---	ATG
IV	**TTGATG**	---- 16nt ----	G**ACGGT**	--- 229nt ---	ATG
V	**TTGATG**	---- 18nt ----	G**ACGGT**	--- 7nt ---	ATG
VI	**TTG**GC**A**	---- 16nt ----	**TGGGGT**	--- ????? ---	ATG

Figure 6.5 Possible promoter sequences found upstream of *icl*

The consensus -35 and -10 sequences for streptomycete *E. coli* -like promoters is shown, with the optimal spacing between the two regions.

The potential promoters for *ICL* are I - V (Figure 6.3). These potential -35 and -10 sequences have been identified by eye and the distances between them, and the distance to the translation start are shown. The letters in bold represent matches to the consensus sequence, shown above.

The putative promoter VI was identified from the same region of DNA, upstream of *icl*. However this promoter sequence is divergent to the rest and would cause transcription to occur in the opposite direction, for an as yet unidentified gene.

A sixth possible promoter (VI) was also identified. It is divergent to the others and would result in transcription occuring in the other direction. All of the potential promoters have spacer regions of between 15-18 nucleotides.

More than one of these sequences may act as a promoter, as there are several examples of genes with multiple promoters. Of 87 genes studied by Strohl (1992), 27 had more than one promoter. It may be that the promoter for the *icl* gene is not a streptomycete *E. coli* -like promoter, which is the case for the majority of streptomycete promoter sequences. Only further experimentation would reveal if this were true.

Strohl (1992) observed there to be between 9 to 345 nucleotides of leader mRNA for the 48 genes analysed. Since 312 bases upstream from the translation start have been sequenced, there is a high probability that the promoter sequence/s has been sequenced. S1 mapping or reverse transcriptase run-off experiments could be carried out to identify the possible promoter region/s.

In vitro site-directed mutagenesis of the putative promoter sequences could be carried out. The mutagenised promoters could then be used in *in vitro* transcription assays, to see if transcription has been decreased due to the alteration of the promoter sequence.

6.2.8 Identification of a stem-loop structure downstream of *icl*

A potential stable stem-loop structure was found in the sequence between *icl* and *ms* (Fig 6.7). The ΔG_O of formation for this stem-loop is -44.6 Kcal/mol (estimated using the rules of Tinoco *et al* (1973), which is similar to the free energy of one of the *gyl*CABX terminators (-44.4 Kcal/mol). The program "TERMINATOR" from the GCG package was used to evaluate this stem-loop structure as a potential rho-independent terminator.

The program identifies rho-independent terminator structures using the method of Brandel and Trifonov (1984 a,b). It is reported that when the program is run with a stringency of 3.5, 95% of all rho-independent terminators are identified. The program did not recognise this stem-loop structure as a terminator. However, this does not mean that it is not a terminator, simply that it does not fit the strucure of a rho-independent terminator.

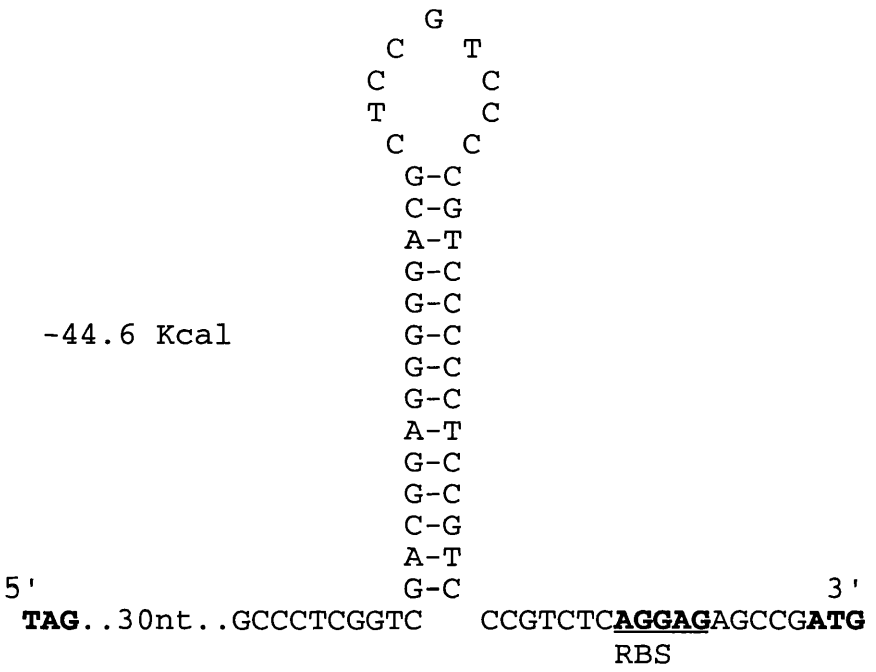


Figure 6.7 Putative terminator of the *icl* gene.

This stem-loop structure was identified by eye in the region downstream from *icl.* (Figure 6.3). The sequence was analysed using the "TERMINATOR" program from GCG, but no similarity to rho-independent terminators was observed.

The binding strength of the stem-loop stucture is shown, which is similar to other streptomycete terminators.

The stop codon of *icl* is shown in bold and the deduced ribosome binding site and start codon of *ms* are shown 3' to the stem-loop.

Transcriptional analysis would need to be carried out to determine if the stem-loop does indeed function as a terminator. If the structure functioned efficiently as a terminator there would have to be a promoter downstream to enable transcription of malate synthase. There are however only 18bp between the end of the stem loop structure and the start of malate synthase and no obvious promoter sequence is observed.

It may be that gene expression of malate synthase is controlled in part by this stem-loop structure, affecting transcription (originating from a promoter found upstream from *icl*), by causing a pausing of the RNA polymerase and a release of some of the RNA polymerase, resulting in a downshift in transcription through *ms*.

6.2.9 Codon usage

It has been suggested that if the abundance of a particular tRNA species is limiting, then genes with codons relating to that species could be limited in their expression. Thus genes with these codons would only be expected to occur where there is a need for control of translation in this manner. These codons would therefore occur less frequently than the synonomous codons used for the incorporation of the same amino acids.

This situation was reported by McCarthy and Gualerzi (1990) and has been found in *S.coelicolor*. The TTA codon (leucine) is extremely rare in *Streptomyces* and has only been found in genes that are involved in antibiotic biosynthesis or resistance, and genes that are regulated during differentiation (Leskiw *et al*, 1991b). The gene that encodes the tRNA species that recognises the TTA codon has been identified (*bld* A) and Lawlor *et al* (1987) found that mutation of this gene resulted in an organism that appeared to grow normally in a vegetative manner, however did not sporulate or produce antibiotics. Further work by Leskiw (1991a), showed the dependence of translation of the *car* B gene on the *bld* A gene. Leskiw converted the TTA codon found within the *car* B gene to CTC (which also encodes leucine), which resulted in the expression of *car* B in *bld* A mutants and no developmental regulation of *car* B.

Other rare codons such as TCT or TTT could be expected to play a similar role to TTA. However these codons are found to occur in genes involved in vegetative growth and secondary metabolism. It may be that these codons do not affect general translation, but can limit the relative amount of translation of a particular gene.

It was of interest to check the *S.coelicolor icl* gene for the use of such rare codons. It was expected that there would be relatively few, as the ICL protein was expressed at a high level when grown on a suitable substrate. Indeed it seems from inspection of the sequence, that very few rare codons are employed. A similar pattern of codon usage was found for the *icl* gene of *S.coelicolor*, when compared with the codon usage of streptomycete genes in general (Fig. 6.8). The only deviation from the usual codon usage was for TAT (encoding tyrosine). Its abundance in the *S.coelicolor icl* was 14%, whereas the codon was found to occur 5% of the time in previously sequenced streptomycete genes.

In summary the codon usage for the *S.coelicolor icl* , was similar to that compiled for the other streptomycete genes which had been sequenced to date.

6.2.10 Protein analysis

When the DNA sequence was obtained, it was possible to carry out the conceptual translation of the *icl* and *ms* genes, to study the amino acid sequences of the ICL and MS proteins.

It was already known that the eukaryotic ICL proteins were similar to each other. The *E.coli* protein was the only prokaryotic ICL to be sequenced to date, and it lacked an internal segment of 102 amino acids compared to the eukaryotic ICL's. It was known from work presented in this thesis (4.2.2.2) that the *S.coelicolor* ICL was similar in molecular weight to the *E.coli* enzyme, thus it was of interest to find out if this segment was also lacking from the *S.coelicolor* enzyme.

It would also be possible to examine the peptide sequences of the regions to which oligonucleotides were designed, in order to establish if they were different in the *S.coelicolor* enzyme. It might thus explain why the original cloning was unsuccessful.

6.2.11 Search for consensus sequences

The program "Macpattern" was used to search for consensus sequences within the conceptual translations of the *S.coelicolor icl* and *ms* sequences. The two enzyme sequences were compared against their respective amino acid consensus sequences, which are found in "MacPattern".

	A	C	G	T	
A	LYS 100- **95**	THR 19 - **30**	ARG 11 - **6**	MET 100 -**100**	G
	LYS 0 - **5**	THR 0 - **3**	ARG 0 - **1**	ILE 0 - **3**	A
	ASN 8 - **4**	THR 4 - **2**	SER 17 - **3**	ILE 0 - **5**	T
	ASN 92 - **96**	THR 77 - **65**	SER 6 - **28**	ILE 100 - **92**	C
C	GLN 92 - **93**	PRO 73 - **52**	ARG 29 - **36**	LEU 66 - **55**	G
	GLN 8 - **7**	PRO 7 - **2**	ARG 0 - **5**	LEU 0 - **0**	A
	HIS 0 - **6**	PRO 0 - **3**	ARG 14 - **7**	LEU 0 - **2**	T
	HIS 100- **94**	PRO 20 - **43**	ARG 46 - **46**	LEU 34 - **39**	C
G	GLU 97 - **81**	ALA 28 - **33**	GLY 9 - **19**	VAL 26 - **39**	G
	GLU 3 - **19**	ALA 6 - **4**	GLY 0 - **10**	VAL 0 - **2**	A
	ASP 4 - **4**	ALA 4 - **4**	GLY 9 - **8**	VAL 0 - **4**	T
	ASP 96 - **96**	ALA 62 - **59**	GLY 82 - **64**	VAL 74 - **56**	C
T	END 100 - **14**	SER 22 - **26**	TRP 100-**100**	LEU 0 - **3**	G
	END - - **3**	SER 0 - **3**	END - - **83**	LEU 0 - **0**	A
	TYR 14 - **5**	SER 0 - **1**	CYS 0 - **13**	PHE 0 - **1**	T
	TYR 86 - **95**	SER 56 - **39**	CYS 100- **87**	PHE 100- **99**	C

Figure 6.8 Comparison of codon usage in the *icl* gene to that used by streptomycete genes in general

This table shows how the codons used in the *S. coelicolor icl* compare to those used by streptomycete genes in general. The values represent the percentage occurance of a given codon for its amino acid.
The first nucleotide for each codon is shown in the left hand column, the second nucleotide, across the top, and the third down the right hand side.
The bold figures represent the average codon usage of 67 streptomycete coding regions, which were provided by Dr. M.Bibb and compiled using the program "Codonfrequency", provided in the GCG package. The table is more representative than that published by Seno and Baltz (1989), which was only compiled from 27 genes.
The figures in normal type show the codon usage of the *icl* gene.

Isocitrate lyase (ICL) consensus

The consensus shown below is based around a cysteine that is postulated to be part of the active site (Nimmo *et al*, 1989). The cysteine residue is found in the middle of a hexapeptide, which to date had been found to be conserved across all previously sequenced *icl*'s.

K-K-C-G-H-M

The search revealed that this pattern was not found in the *S.coelicolor* ICL enzyme sequence. It was found that the methionine residue was substituted for a leucine in the *S.coelicolor* enzyme. This can be considered as a conservative change, as both amino acids have non-polar side groups. It would now appear however, that the consensus sequence should be changed, to the one shown below (where the residues in brackets, show the variation at an individual site).

K-K-C-G-H-[ML]

Malate synthase (MS) consensus

A consensus pattern has also been recognised within the central region of the malate synthase sequences published to date. Although the *S.coelicolor ms* sequence was incomplete, a search was carried out to look for the pattern shown below, which had been previously found in all *ms* sequences.

R-[DN]-H-x-x-G-L-N-x-G-x-W-D-Y-[LIVM]-F

- where an x represents any amino acid.

The search did not reveal this pattern either, but on inspection of the amino acid sequence, a region which only missed being an exact match by one, was found (shown below). The change being shown in bold.

R - **E** - H-S-A-G-L-N-A-G-R-W-D-Y - L - F

Again the consensus pattern should be amended, to include the observations found from the sequence of the *S.coelicolor* enzyme. The amended consensus is shown below.

L-R-[DNQE]-H-[STI]-[ASV]-G-L-N-[CRA]-G-[RG]-W-D-Y-[MIL]-F

This amended consensus is different from the "MacPattern" consensus, in a number of ways. Firstly a leucine is added to the start of the pattern, as it was found to be conserved across all sequences. Secondly the residues which were labelled as x's (in the "MacPattern consensus), are shown as the amino acids actually found. It also includes two amendments which are the inclusion of a possible glutamine residue at the third position and the deletion of the valine from the penultimate residue.

6.2.12 Alignment of the available ICL and MS enzyme sequences

It was of interest to find out if the consensus regions used in the "MacPattern" program, were indeed the best regions to be used as signature patterns for these enzymes. The program "PILEUP" from the GCG package was used to line up all the available ICL and MS enzyme sequences.

"PILEUP" creates a multiple sequence alignment using a simplification of the progressive alignment method of Feng and Doolittle (1987). The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster can then be aligned to the next most related sequence or cluster of aligned sequences.

Two clusters of sequences can be aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments that include increasingly dissimilar sequences and clusters, until all sequences have been included in the final pairwise alignment.

The results from the two pileup's can be seen in Figs. 6.9 and 6.10. It can clearly be seen by eye that there is a great deal of similarity between all the ICL and MS sequences.

It can be noted from the ICL pileup that the bacterial sequences have been clustered together, as have the plant and the fungal sequences. It also appears that the bacterial sequences are more similar to the plant sequences than the fungal ones.

The *S.coelicolor* sequence is similar to the *E.coli* sequence in that it has a central region missing, which is present in all of the eukaryotic enzyme sequences.

```
      1                                                  50
Sco   .......... ....MAEART QAAEDLARRW DTEPRWQGIE RTYSAQDVVR
Eco   .......... .....MKTRT QQIEELQKEW .TQPRWEGIT RPYSAEDVVK
Ath   .....MIDKP NQIMEEEGRF EAEVAEVQTW WSSERFKLTR RPYTARDVVA
Bna   ..MAASFSVP SMIMEEEGRF EAEVAEVQTW WSSERFKLTR RPYTARDVVA
Ghi   ..MAASFSVP SMIMEEEGRF ETEVAEVQAW WNSERFKLTR RPYSARDVVA
Rco   ..MAASFSGP SMIMEEEGRF EAEVAEVQAW WNSERFKLTR RPYTARDVVA
Ani   .........M SYIEEEDQRY WDEVA.VKNW WKDSRWRYTK RPFTAEQIVA
Ncr   MAANNMVNPA VDPALEDELF AKEVEEVKKW WSDSRWRQTK RPFTAEQIVS
Ctr   .....MAYTK IDINQEEADF QKEVAEIKKW WSEPRWRKTK RIYSAEDIAK
Sce   .MPIPVGNTK NDFAALQAKL DADAAEIEKW WSDSRWSKTK RNYSARDIAV
                                        *     *      *   *
      51                                                 100
Sco   LSGSVREEHT LARRGAERLW RQLHE...RD YV.HALGALT GGQAVQQIKA
Eco   LRGSVNPECT LAQLGAAKMW RLLHGESKKG YI.NSLGALT GGQALQQAKA
Ath   LRGHL.KQGY ASNEMAKKLW RTLKSHQANG TASRTFGALD PVQVTMMAKH
Rna   LRGHL.KQGY ASNEMAKKLW RTLKSHQANG TASRTFGALD PVQVTMMAKH
Ghi   LRGSL.KQSY GSNEMAKKLW TTLKTHQANG TASRTFGALD PVQVTMMAKH
Rco   LRGNL.KQSY ASNELAKKLW RTLKTHQANG TASRTFGALD PVQVTMMAKH
Ani   KRGNL.KIEY PSNVQAKKLW GILERNFNK. EASFTYGCLD PTMVTQMAKY
Ncr   KRGNL.KIEY ASNAQAKKLW KILEDRFAKR DASYTYGCLE PTMVTQMAKY
Ctr   KRGTL.KIAY PSSQQSDKLF KLLEKHDAEK SVSFTFGALD PIHVAQMAKY
Sce   RRGTFPPIEY PSSVMARKLF KVLEKHHNEG TVSKTFGALD PVQISQMAKY
        *                     *              * *            *
      101                                                150
Sco   GLQAIYLSGW QVAADANQAG HTYPDQSLYP ANSVPQVVRR INNALLRADQ
Eco   GIEAVYLSGW QVAADANLAA SMYPDQSLYP ANSVPAVVER INNTFRRADQ
Ath   .LDTIYVSGW QCSSTHTSTN EPGPDLADYP YDTVPNKVEH LFFAQQYHDR
Bna   .LDTIYVSGW QCSSTHTSTN EPGPDLADYP YDTVPNKVEH LFFAQQYHDR
Ghi   .LDSIYVSGW QCSSTHTTTN EPGPDLADYP YDTVPNKVEH LFFAQQYHDR
Rco   .LDSIYVSGW QCSSTHTTTN EPGPDLADYP YDTVPNKVEH LFFAQQYHDR
Ani   .LDTVYVSGW QSSSTASSTD EPSPDLADYP MNTVPNKVNH LWMAQLFHDR
Ncr   .LDTVYVSGW QSSSTASSSD EPGPDLADYP YTTCPNKVGH LFMAQLFHDR
Ctr   .LDSIYVSGW QCSSTASTSN EPSPDLADYP MDTVPNKVEH LWFAQLFHDR
Sce   .LDTIYISGW QCSSTASTSN EPGPDLADYP MDTVPNKVEH LFKAQLFHDR
          * *** *              **   **     *  *           *
      151                                                200
Sco   IAT.....  SEGDNST...  ..DWLAPIVA DAEAGFGGPL NAFELAKAMI
Eco   IQW.....  SAGIEPGDPR YVDYFLPIVA DAEAGFGGVL NAFELMKAMI
Ath   KQREARMSMS REERT..KTP FVDYLKPIIA DGDTGFGGTT ATVKLCKLFV
Bna   KQREARMSMS REERA..KTP FVDYLKPIIA DGDTGFGGTT ATVKLCKLFV
Ghi   KQREARMSMS REERA..RTP YVDYLKPIIA DGDTGFGGTT ATVKLCKLFV
Rco   KQREARMSMS REERA..RTP YVDYLKPIIA DGDTGFGGTT ATVKLCKLFV
Ani   KQREERMTTP KDQR..HKVT NVDYLRPIIA DADTGHGGLT AVMKLTKLFV
Ncr   KQRQERLSVP KDQR..EKLA NIDYLRPIVA DADTGHGGLT AVMKLTKLFI
Ctr   KQREERLNMT KEERAN..TP YIDFLRPIIA DADTGHGGIT AIIKLTKLFI
Sce   KQLEARSKAK SQEELDEMGA PIDYLTPIVA DADAGHGGLT AVFKLTKMFI
                              *   ** * *   * **       * *
      201                                                250
Sco   AAGAAGIHYE DQLASEKKCG HLGGKVLVPT SQHIRTLNAA RLAADIADTP
Eco   EAGAAAVHFE DQLASVKKCG HMGGKVLVPT QEAIQKLVAA RLAADVTGVP
Ath   ERGAAGVHIE DQSSVTKKCG HMAGKVLVAV SEHINRLVAA RLQFDVMGTE
Bna   ERGAAGVHIE DQSSVTKKCG HMAGKVLVAV SEHINRLVAA RLQFDVMGTE
Ghi   ERGAAGVHIE DQSSVTKKCG HMAGKVLVAV SEHINRLVAA RLQFDVMGVE
Rco   ERGAAGVHIE DQSSVTKKCG HMAGKVLVAI SEHINRLVAA RLQFDVMGVE
Ani   ERGAAGIHIE DQAPGTKKCG HMAGKVLVPI SEHINRLVAI RAQADIMGTD
Ncr   EKGAAGIHIE DQAPGTKKCG HMAGKVLVPI QEHINRLVAI RAQADIMGSD
Ctr   ERGAAGIHIE DQAPGTKKCG HMAGKVLVPV QEHINRLVAI RASADIFGSN
Sce   ERGAAGIHME DQTSTNKKCG HMAGRCVIPV QEHVNRLVTI RMCADIMHSD
        ***  * * **    **** *  *            *    *   *
```

```
      251                                                        300
Sco   TVIIARTDAL AANLLTSDVD ERDAQFVTG. .......... ..........
Eco   TLLVARTDAD AADLITSDCD PYDSEFITG. .......... ..........
Ath   TVLVARTDAV AATLIQSNID ARDHQFILGA TNPSLRGKSL SSLLAEGMTV
Bna   TVLVARTDAV AATLIQSNID SRDHQFILGV TNPSLRGKSL SSLLAEGMAV
Ghi   TVLVARTDAV AATLIQTNVD TRDHQFILGA TNPNLRGKSL ANMLAEGMAA
Rco   TLLVARTDAE AANLIQSNVD TRDHQFILGV TNPNLRGKSL ATLLATGMAN
Ani   LLAIARTDSE AATLITSTID HRDHPFIIGS TNPDI..QPL NDLMVMAEQA
Ncr   LLCIARTDAE AATLITTTID PRDHAFILGC TNPDL..EPL AHLMMKAEAE
Ctr   LLAVARTDSE AATLITSTID HRDHYFIIGA TNPES..GDL AALMAEAEAK
Sce   LIVVARTDSE AATLISSTID TRDHYFIVGA TNPNI..EPF AEVLNDAIMS
          ****   ** *     *   *  *

      301                                                        350
Sco   .......... .......... .......... .......... ..........
Eco   .......... .......... .......... .......... ..........
Ath   GKNGPALQSI EDQWLGSAGL MTFSEAVVQA IKRMNLNENE KNQRLSEWLT
Bna   GNNGPALQAI EDQWLSSARL MTFSDAVVEA LKRMNLSENE KSRRVNEWLN
Ghi   GKNGPQLQAI EDNWLAIAQL KTFSECVMDA IKSMNITEDE KRRRMNEWMN
Rco   GKTGAELQAT EDNWLAMAQL KTFPECVMDA IKNMNAGEDE KRRRMNEWMN
Ani   GKNGAELQAI EDEWLAKAGL KLFNDAVVDA INNSPLP..N KKAAIEKYLT
Ncr   GKTGAQLQAI EDDWLAKADL KRFDEAVLDV IAKGKFS..N AKDLAAKYQA
Ctr   GIYGDELARI ETEWTKKAGL KLFHEAVIDE IKAGNYS..N KEALIKKFTD
Sce   GASGQELADI EQKWCRDAGL KLFHEAVIDE IERSALS..N KQELIKKFTS

      351                                                        400
Sco   .......... .......... .......... .ERTPEGFYR VRNGMAPVIA
Eco   .......... .......... .......... .ERTSEGFFR THAGIEQAIS
Ath   HARYENCLSN EQGRVLAAKL GVTDLFWDWD LPRTREGFYR FQGSVAAAVV
Bna   HARYENCLSN EQGRELAAKL GVTDLFWDWD LPRTREGFYR FQGSVTAAVV
Ghi   HSSYDKCLSN EQAREIAERL GLQNLFWDWD LPRTREGFYR FRGSVMAAIV
Rco   HTSYDKCLSY EQGREIADRM GLKNLFWDWD LPRTREGFYR FKGSVMAAVV
Ani   QSKG.K..SN LEARAIAKEI AGTDIYFDWE APRTREGYYR YQGGTQCAIN
Ncr   AVKG.KQISN REARAIARQL LGQEIFFDWE SPRTREGYYR LKGGCDCSIN
Ctr   KVNPLSHTSH KEAKKLAKEL TGKDIYFNWD VARAREGYYR YQGGTQCAVM
Sce   KVGPLTETSH REAKKLAKEI LGHEIFFDWE LPRVREGLYR YRGGTQCSIM
                                        *  **  *

      401                                                        450
Sco   RGLAYAPYAD LIWVETGTPD LEQAREFAEA IHAEHPDQML AYNCSPSFNW
Eco   RGLAYAPYAD LVWCETSTPD LELARRFAQA IHAKYPGKLL AYNCSPSFNW
Ath   RGWAFAQIAD IIWMETASPD LNECTQFAEG IKSKTPEVML AYNLSPSFNW
Bna   RGWAFAQIAD LIWMETASPD LNECTQFAEG VKSKTPEVML AYNLSPSFNW
Ghi   RGWAFAPHAD LIWMETSSPD MVECTRFAEG VKSMHPEIML AYNLSPSFNW
Rco   RGRAFAPHAD IIWMETAKPD FAECTAFAEG VKSMHPEIML AYNLSPSFNW
Ani   RAVAYAPFAD LIWMESKLPD YKQAKEFADG VHAVWPEQKL AYNLSPSFNW
Ncr   RAISYAPYCD AIWMESKLPD YAQAEEFAKG PR.VWPEQKL AYNLSPSFNW
Ctr   RGRAFAPYAD LIWMESALPD YNQAKEFADG VKAAVPDQWL AYNLSPSFNW
Sce   RARAFAPYAD LVWMESNYPD FQQAKEFAEG VKEKFPDQWL AYNLSPSFNW
      *    *   *   * *   **       **        *   * *** ******

      451                                                        500
Sco   K.AALDDDQI AKFQRELGAM GYKFQFITLA GFHSLNHGMF DLARGYAE.H
Eco   Q.KNLDDKTI ASFQQQLSDM GYKFQFITLA GIHSMWFNMF DLANAYAQGE
Ath   DASGMTDPLT C......... .......... .......... ..........
Bna   DASGMTDQQM MEFIPRIARL GYCWQFITLA GFHADALVVD TFAKDYAR.R
Ghi   DASGMTDEHM RDFIPRIAKL GFCWQFITLA GFHADALVTD TFARDFAR.R
Rco   DASGMTDEQM RDFIPRIARL GFCWQFITLG GFHADALVID TFAKDYAR.R
Ani   K.KAMPRDEQ ETYIKRLGAL GYAWQFITLA GLHTTALISD TFAKAYAK.Q
Ncr   K.TAMGRDDQ ETYIRRLAKL GYCWQFITLA GLHTTALISD QFAKAYSK.I
Ctr   N.KAMPADEQ ETYIKRLGQL GYVWQFITLA GLHTTALAVD DFANQYSQ.I
Sce   P.KAMSVDEQ HTFIQRLGDL GYIWQFITLA GLHTNALAVH NFSRDFAK.D
                            *   *****  * *
```

```
      501                                                        550
Sco   GMTAYVD.LQ EREFAAQAHG FTAVKHQREV GTGYFDLVST AVNPA.SSTT
Eco   GMKHYVEKVQ QPEFAAAKDG YTFVSHQQEV GTGYFDKVTT IIQGG.TSSV
Ath   .......... .......... .......... .......... ..........
Bna   GMLAYVERIQ REER...SNG VDTLAHQKWS GANYYDRYLK TVQGGISSTA
Ghi   GMLAYVEKIQ REER...NNG VDTLAHQKWS GANFYDRYLK TVQGGISSTA
Rco   GMLAYVERIQ REER...KNG VDTLAHQKWS GANYYDRYLK TVQGGISSTA
Ani   GMRAYGELVQ EPEM...ANG VDVVTHQKWS GANYVDNMLK MITGGVSSTA
Ncr   GMRAYGELVQ EPEI...DNG VDVVKHQKWS GATYVDELQK MVTGGVSSTA
Ctr   GMRAYGQTVQ QPEI...EKG VEVVKHQKWS GANYIDGLLR MVSGGVTSTA
Sce   GMKAYAQNVQ QREM...DDG VDVLKHQKWS GAEYIDGLLK LAQGGVSATA
      **  *    *   *      *       **    *    *
      551                                  587
Sco   ALAGSTEEEQ FH+..... .......... .......
Eco   ALTGSTEESQ F+..... .......... .......
Ath   .......... .......... .......... .......
Bna   AMGKGVTEEQ FKETWTRPGA AGMGEGTSLV VAKSRM+
Ghi   AMGKGVTEEQ FKETWTRPGA GNIGSEGNLV VAKARM+
Rco   AMGKGVTEEQ FKETWTRPGA MEMGSAGSEV VAKARM+
Ani   AMGKGVTEDQ FKS+..... .......... .......
Ncr   AMGKGVTEDQ FH+..... .......... .......
Ctr   AMGAGVTEDQ FKETKAKV+. .......... .......
Sce   AMGTGVTEDQ FKENGVKK+. .......... .......
      *      * * *
```

Figure 6.9 Alignment of the known ICL sequences

The program "PILEUP" from the GCG package was used to align the 10 available ICL sequences.

Sco = *S.coelicolor*, Eco = *E.coli*, Ath = *A.thaliana*, Bna = *B.nappus* (rapeseed), Ghi = *G.hirsutum* (cotton), Rco = *R.communis* (castor bean), Ani = *A.nidulans*, Ncr = *N.crassa*, Ctr = *C.tropicalis*, Sce = *S.cerevisiae*

Asterisks shown below the pileup represent amino acids which are conserved across all sequences. An incomplete *A.thaliana* sequence is shown, as the full DNA sequence has not yet been reportted.

```
      1                                                  50
Hpo   ......MAQI NPTTLDNVKI LGEVSDKPLL ..SKATPRDI LTKDALKFIV
Sce   .........M VKVSLDNVKL LVDVDKEPFF KPSSTTVGDI LTKDALEFIV
Ani   .......MSQ VDAQLKDVAI LGSVSNEA.. .......RKI LTKEACAFLA
Ncr   .......MAS VETLLQGVTI SGPIEEHQ.. .......RKI LTPQALSFVA
Ctr   ...MSTPFPK TADKVKGVQI LGPIPDEA.. .......KHI FNQETLAFVA
Sco   ........MS TTSLTHHVRV LAAPGPRH.. .......DEI LTPAALDFVG
Csa   MGSLGMYSES GLTKKGSSRG YDVPEGVDIR GRYDEEFAKI LNKEALLFIA
Cma   MGSLGMYSES AVRKK.SSRG YDVPEGVDIR GRYDEEFARI LNKEALLFVA
Ghi   MIGLGSYGYT APSSKKIN.A YDVPQGVDIR GRFDEEFAKI LTKDALQFVA
Rco   .MRYDTYGDS APIKKTGA.G YDVPEGVDIR GRYDGEFAKI LTRDALQFVA
Bna   ......MELE TSVYRPNVAV YDSPDGVEVR GRYDQVFAKI LTRDALGFVA
Eco   .......... ......MTEQ ATTTDELAFT RPYGEQEKQI LTAEAVEFLT
                                                    *
      51                                                 100
Hpo   LLHRSFNETR KQLLENRQKV QERLDAGE.S LHFLEETKYI REDPNWKCL.
Sce   LLHRTFNNKR KQLLENRQVV QKKLDSGSYH LDFLPETANI RNDPTWQGP.
Ani   ILHRTFNPTR KALLQRRVDR QAEIDKGHLP .DFLPETKHI RDDPSWKG.A
Ncr   LLHRSFNQTR KNLLERRHVR QAEIDRGVLP .DFLPETKHI RENPTWKGAA
Ctr   TLHRGFEARR QELLNNRKEQ QKLRDQGFLP .DFLPETEYI RNDSTWTGPA
Sco   GLVEAFQPRR RDLMKERRRS ALRLASGS.P LDFPTVTSAV RKYPTWR.VA
Csa   DLQRTFRNHI KYSMECRREA KRRYNEGGLP GFD.PATKYI RDSE.WT.CA
Cma   DLQRTFRNHI RYSMECRREA KRRYNEGAVP GFD.PATKYI RESE.WT.CA
Ghi   DLQREFRNHI KYAMECRKEA KRRYNEGALP GFD.PATRYI REGK.WT.CV
Rco   DLQREFRNRI RYAIECRKEA KSRYNAGALP GFEHPATKYI REGE.WT.CA
Rna   ELQREFRGHV RYAMECRREV KRRYNSGAVP GFD.PSTKFI RDGE.WV.CA
Eco   ELVTHFTPQR NKLLAARIQQ QQDIDNGTLP DF.ISETASI RDAD.WK.IR
                        *                   *    *    *
      101                                                150
Hpo   PTHPKLQCRK IEITGPPDAK MIVNAFNTDV FTYMTDFEDS CSPTWNNMIY
Sce   ILAPGLINRS TEITGPPLRN MLINALNAPV NTYMTDFEDS ASPTWNNMVY
Ani   PPAPGLVDRR VEITGPTDRK MVVNALNSDV .......... .APTWDNMIN
Ncr   PAAPPLVDRR VEMTGPTDRK MVVNALNSDV YTYMADFEDS SAPTWANMVN
Ctr   .LAPGLIDRR CEITGPTDRK MVINALNSNV ATYMADFEDS LTPAWKNLVE
Sco   PPAPGLTDRR VEMTGPPDRR MAVNALNSGA QVWMADFEDA TAPLWDNIIG
Csa   PVPPAVADRR VEITGPVERK MIINALNSGA KVFMADFEDA LSPNWENLMR
Cma   SVPPAVADRR VEITGPVERK MIINALNSGA KVFMADFEDA LSPNWENLMR
Ghi   PFPPAVADRR VEITGPVERK MIINALNSGA KVFMADFEDA LSPSWENLMR
Rco   PVPPAVADRK VEITGPVERK MIINALNSGA KVFMADFEDA LSPSWENLMR
Rna   SVPPAVADRR VEITGPVERK MIINALNSGA KVFMADFEDA LSPSWENLMR
Eco   GIPADLEDRR VEITGPVERK MVINALNANV KVFMADFEDS LAPDWNKVID
                  * ***     *  ** *                    *
      151                                                200
Hpo   GQVNLYDAIR DRIDFTNEAT GKRYKINREG RRVPVMIVRP RGWHMVDKHI
Sce   GQVNLYDAIR NQIDF..DTP RKSYKLNGNV ANLPTIIVRP RGWHMVEKHL
Ani   GQINLYDAIR RQVDFKQGQK EY..KL.RTD RTLPTLIARA RGWHLDEKHF
Ncr   GQVNLYDAIR RQIDFKQGPK EY..KL.RTD RTLPTLIVRP RGWHLEEKHV
Ctr   GQVNLYDAVR RNLSATINGK QYSLNL.EKG RHIPTLIVRP RGWHLTEKHV
Sco   GQLTLLDVIE RRIDFTTPEG ....KEYRLG DRPATIMVRP RGWHLDEEHL
Csa   GQINLKDAVD GTISFHDRVR NRVYKLN... DRTAKLFVRP RGWHLPEAHI
Cma   GQINLKDAVD GTISFHDKAR NKVYKLN... DQTAKLFVRP RGWHFAEAHI
Ghi   GQINLKDAVE GTITFNDKAR NRVYKLN... NEIAKLFVRP RGWHLPEAHI
Rco   GQVNLRDAVN GTISFHDKAR NRVYKLN... DQIAKLFVRP RGWHLPEAHI
Rna   GQVNLKDAVD GSITFNDKAR NKVYKLN... DQVAKLFVRP RGWHLPEAHI
Eco   GQINLRDAVN GTISYTNEA. GKIYQLK... PNPAVLICRV RGLHLPEKHV
      **   * *                                *  ** *    *
```

```
      201                                                        250
Hpo   LVDGEPISAS IMDFGLFFFH NAKFL...LS QGLAPFFYLP KMEHYKEAQL
Sce   YVDDEPISAS IFDFGLYFYH NAKEL...IK LGKGPYFYLP KMEHHLEAKL
Ani   TVDGEPISGS LFDFGLYFFH NAKEL...VA RGFGPYFYLP KMESHLEARL
Ncr   TIDGEPVSGS LFDFGLYFFH NAKEL...VQ RGFGPYFYPP KMESHLEARL
Ctr   LVDGTPVSGG IFDFAVYFYN SAKEA...IA QGFGPYFYLP KMEHHLEAKL
Sco   EYDGRPVPAT LVDFGLYFFH CAQRQ...ID AGHGPYFYLP KLENRYEARL
Csa   FIDGEPATGC LVDFGLYFFH NHANFRRSQG QGYGPFFYLP KMEHSREAKI
Cma   FIDGEPATGC LVDFGLYFFH NHANFRRSQG QGSGPFFYLP KMEHSREAKI
Ghi   FIDGEPATGC LVDFGLYFYH NYATFRNTQG QGFGPFFYLP KMENSREAKI
Rco   LIDGEPATGC LVDFGLYFYH NYAAFRRNQG AGYGPFFYLP KMEHSREAKI
Rna   LIDGEPATGC LVDFGLYFFH NYAKFRQTQG SGFGPFFYLP KMEHSREAKI
Eco   TWRGEAIPGS LFDFALYFFH NYQAL...LA KGSGPYFYLP KTQSWQEAAW
                 **   *                 * **** *     **
      251                                                        300
Hpo   WNDIFCVAQD CLEIPRGTIK ATVLIETLPI SYQLDEVLYA LRDHSAGLNC
Sce   WNDVFCVAQD YIGIPRGTIR ATVLIETLPA AFQMEEIIYQ LRQHSSGLNC
Ani   WNDVFNLAQD YIGMPRGTIR GTVLIETITA AFEMEEIIYE LRDHSSGLNC
Ncr   WNDAFNLAQD YVGIPLSTIR GTVLIETITA AFEMDEIIFE LRNHTSGLNR
Ctr   WNDIFNYSQD YIGLKRGTIR ASVLIETIPA VFQMDEIIYQ LREHSAGLNC
Sco   WNDVFLTAQE LLGIPRGTVR ATVLIETITA AFEMEEILHA LREHSAGLNA
Csa   WNSVFERAEK MAGIERGSIR ATVLIETLPA VFQMNEILYE LRDHSVGLNC
Cma   WNSVFERAEK MAGIERGSIR ATVLIETLPA VFQMDEILYE LRDHSVGLNC
Ghi   WNSVFEKAEK MAGIEKGSIR ATVLIETLPA VFQMDEIFYE LRDHSVGLNC
Rco   WNCVFEKAEK MAGIERGSIR ATVLIETLPA VFQMNEILYE LRDHSVGLNC
Bna   WNSVFERAEK MAGIERGSIR ATVLIETLPA VFQMNEILYE LRDHSVGLNC
Eco   WSEVFSYAED RFNLPRGTIK ATLLIETLPA VFQMDEILHA LRDHIVGLNC
      *   *                   ****         *     ** *  ***
      301                                                        350
Hpo   GRWDYMFSTI KRLRNQKQHI LPDRHQVTMT VPFMTNYVKQ LIKICHKRGV
Sce   GRWDYIFSTI KRLRNDPNHI LPNRNQVTMT SPFMDAYVKR LINTCHRRGV
Ani   GRWDYIFSFI KKFRQHPNFV LPDRSDVTMT VPFMDAYVKL LIKTCHKRGV
Ncr   GGWDYIFPFI KEVRRFPNFV LPDRSDVTMT VPFMEAYVKL LIKTLHRLVV
Ctr   GRWDYIFSYI KCLRNHPDFI LPDRSQVTMA APFMSSYVKL LVHTTHKRKV
Sco   GRWDYLFSII KTFGHRTDFL LPDRAKVTMT APFMRAYTEL LVRTCHKRGV
Csa   GRWDYIFSYV KTFQAHPDRL LPDRVLVGMT QHFMRSYSDL LIRTCHRRGV
Cma   GRWDYIFSYV KTFQAHLDRL LPDRVQVGMA QHFMRSYSDL LIRTCHTVVC
Ghi   GRWDYIFSYV KTFQGHPDRL LPDRGQVGMT QHFMRSYSDL LIRTCHRRGV
Rco   GRWDYIFSYV KTFQAHPDRP LPDRVQVGMT QHFMKSYSDL LVWTCHRRGV
Bna   GRWDYIFSYV KTFQAHPDRL LPDRVLVGMG QHFMRSYSDL LIRTCHKRGV
Eco   GRWDYIFSYI KTLKNYPDRV LPDRQAVTMD KPFLNAYSRL LIKTCHKRGA
      * *** *    *          **    * *    *   *    *  * *
      351                                                        400
Hpo   HAMGGMAATI PIKDDPEKNA AAMEAVRQDK LREVLAGHDG TWIAHPGLLP
Sce   HAMGGMAAQI PIKDDPAANE KAMTKVRNDK IRELTNGHDG SWVAHPALAP
Ani   HAMGGMAAQI PIKDNAEAND KAMEGVRADK LREVRAGHDG TWVAHPALAS
Ncr   HAM.GMAAQI PIKDDKAAND KAMEGVRADK LREARAGHDG TWVAHPALAS
Ctr   HALGGMAAQI PIKDDEARNR AALENVTKDK LREVTLGCDS CWVAHPALVP
Sco   PRHRGMAAQV PGKD.PAANE AAFAKVRLDK EREAEDGFDG SWVAHPGLVP
Csa   HAMGGMAAQI PIRDDPKANE VALELVRKDK LREVKAGHDG TWAAHPGLIP
Cma   H.VGGMAAQI PIRDDPKANE MALELVRKDK LREAKAGHDG TWAAHPGLIP
Ghi   HAMGGMAAQI PIRDDPTANE AAFELVRKDK QREVKAGHDG TWAAHPGLIK
Rco   HAMGGMAAQI PIRDDPAANK AALELVRKDK LREVKAGHDG TWAAHPGLIP
Bna   HAMGGMAAQI PIRDDPKANE MALDLVKKDK LREVRAGHDG TWAAHPGLIP
Eco   FAMGGMAAFI PSKDEEHNNQ V.LNKVKADK SLEANNGHDG TWIAHPGLAD
          ****   *  *           *  **   *   *     *  **
```

```
      401                                                        450
Hpo   TALSVFQEHM P.TPNQIHVQ KNVE...ITE ADLVDTNIPD GKITMKGVSA
Sce   ICNEVF.INM G.TPNQIYFI PENV...VTA ANLLETKIPN GEITTEGIVQ
Ani   IASEVFNKYM P.TPNQMHVR REDV..NITA NDLLNTNVPG .KITEDGIRK
Ncr   IALEVFNKHM P.TPNQLFNR REDV..KIGQ QDLLNMNVPG .SSTEDGIRK
Ctr   VVLKVFNEHM K.GPNQISLP PKEPFKPITQ RDLLSPFVPG AKITEQGIRA
Sco   VCRESFDAVL E......... .......... .......... ..........
Csa   ACMEVFTNNM GNAPNQIRSM RRDDAANLTE EDLLQQ..PR GVRTMEGLRL
Cma   ACMEVFTNSM GNAPNQIRSA RRDDAANLTE DDLLQQ..PR GVRTLEGLRL
Ghi   TCMEVFTNNM GNTPNQIETV KRDYASNLTE DDLLQR..PR GVRTMEGLRL
Rco   ACMEVFANNM GNTPHQIQAM KREDAANITE EDLIQR..PR GVRTLEGLRL
Bna   ICMDAFSH.M GNNPNQIKSM KRDDASAITE EDLLQI..PR GVRTLEGLRL
Eco   TAMAVFNDIL GSRKNQLEVM REQDAP.ITA DQLLAP..CD GERTEEGMRA
            *           *                 *           *
      451                                                        500
Hpo   NIYIGLNYME SWLRGLGCVP INNLMEDAAT AEVSRLQLYS WCKHAVKMDD
Sce   NLDIGLQYME AWLRGSGCVP INNLMEDAAT AEVSRCQLYQ WVKHGVTLKD
Ani   NLNIGLSYME GWLRGVGCIP INYLMEDAAT AEVSRSQLWQ WARHGVT.TS
Ncr   NLNTGLGYTE PWIRGVGCVP IKHPQEDAAT AEVSRSQLWQ WVKHRVT.TA
Ctr   NIVIGISYIE AWLRNVGCVP INYLMEDAAT AEVSRTQIWQ WVTHGAK.TD
Sco   .......... .......... .......... .......... ..........
Csa   NTRVGIQYLA AWLTGAGSVP LYNLAEDAAT AEISRVQNWQ WLKYGVELDG
Cma   NTRVGIQYLA AWLTGTGSVP LYNLMEDAAT AEISRVQNWQ WLKYGVELDG
Ghi   NTRVGIQYLA AWLTGSGSVP LYNLMEDAAT AEISRVQIWQ WLKYGVELDG
Rco   NTRVGIQYLA AWLTGTGSVP LYNLMEDAAT AEISRVQNWQ WLKYGVELDG
Bna   NTRVGIQYLA AWLTGSGSVP LYNLMEDAAT AEISRVQNWQ WIRYGVELDG
Eco   NIRVAVQYIE AWISGNGCVP IYGLMEDAAT AEISRTSIWQ WIHHQKTLSN
      *     *    *    *  *      ***** ** **      *
      501                                                        550
Hpo   T..GKTITPE FISKLIDEEA ERCAANKP.. .NNK..FKIA ADCLKKEING
Sce   T..GEKVTPE LTEKILKEQV ERLSKASPLG DKNK..FALA AKYFLPEIRG
Ani   E..GKKVDKA YALRLLKEQA DALAAKGPKG ..NK..FQLA GRYFSGQVTG
Ncr   E..GKHVDKR YPLKLLKEAD RQRLAKAPQG ..NK..FNLA AQYFASQVTG
Ctr   T..GKVITKE YVKQLLDEEY AKLTKNAKPG ..NK..FKRA FEYFAPEALG
Sco   .......... .......... .......... .......... ..........
Csa   DGLGVRVNKE LFGRVVEEEM ERIEREVGKE RFKKGMYKEA CKMFTRQCTA
Cma   DGLGVRVNKE LFARVVEEEM ERIEREVGKE KFRKGMYKEA CKMFTRQCTA
Ghi   DGLGVRVN.H VFGRVVEEEM ARIEREVGKE KFKKGMYKEA CKIFTRQCTA
Rco   DGLGVKVTFD LLGRVVEDEM ARIEREVGKE KFKKGMYKEA CKMFVRQCAA
Bna   DGLGVRVSKE LFGRVVEEEM ERIEKEVGKD KFKRGMYKEA CKMFTKQCTA
Eco   ...GKPVTKA LFRQMLGEEM KVIASELGEE RFSQGRFDDA ARLMEQITTS
         *                                          *
      551                                 581
Hpo   HTEVAEFLTT LLYDDIV.TI GPEVDISSLK*
Sce   E.KFSEFLTT LLYDEIVSTK ATPTDLSKL*
Ani   E.DYADFLTS LLYNEI.SSP G...TASKL*
Ncr   E.DYADFLTC LLYNEI.TSA GNSLPASKL*
Ctr   E.KYSDFVTT LIYDDV.TTI GRALPGERL*
Sco   .......... .......... ..........
Csa   P.NLDDFLTL DAYNYIVIHH PRELSKL*
Cma   P.TLDDFLTL DAYNHIVIHH PRELSRL*
Ghi   S.TLDDFLTL DAYNYIVIHH PKDVSSKL*
Rco   P.TLDDFLTL DAYNNIVIHY PKG.SSRL*
Bna   A.ELDDFLTL AVYDHIVAHY PINASRL*
Eco   D.ELIDFLTL PGYRLLA*
```

Figure 6.10. Alignment of the known MS sequences

The program "PILEUP" from the GCG package was used to align the 12 available MS sequences.

Hpo = *H.polymorpha*, Sce = *S.cerevisiae*, Ani = *A.nidulans*, Ncr = *N.crassa*, Ctr = *C.tropicalis*, Sco = *S.coelicolor*, Csa = *C.sativus* (cucumber), Cma = *C.maxima* (pumpkin), Ghi = *G.hirsutum* (cotton), Rco = *R.communis* (cator bean), Bna = *B.nappus* (rapeseed), Eco = *E.coli*.

An incomplete *S.coelicolor* sequence is shown, as the full DNA sequence has still to be completed.

Asterisks shown below the pileup represent amino acids which are conserved across all sequences.

The results of the MS pileup are slightly different from the ICL pileup, in that the *S.coelicolor* MS sequence is shown to be more similar to the fungal sequences than the bacterial ones. It should be borne in mind that the similarity clustering does not represent any phylogenetic relationship between the sequences. Also since the *S.coelicolor*MS sequence is incomplete, the final result may turn out to be different to that shown.

6.2.13 Proposal of new signature motifs for ICL and MS

The results of the pileup's, enabled the regions of greatest similarity to be observed and new signature patterns to be proposed.

ICL signature patterns

L-A-Y-N-[LC]-S-P-S-F-N-W

or

Q-I-F-T-L-[AG]-G (*A.thaliana* sequence not included)

MS signature patterns

P-[FY]-F-Y-L-P-K

or

G-L-N-[CRA]-G-[RG]-W-D-Y

or

E-D-A-A-T-A-E-[IV]-S-R (*S.coelicolor* sequence not included)

The two sequences mentioned as not included, have simply been omitted as the sequences were incomplete.

It is proposed that these new signature motifs should not replace the existing motifs, but augment the information which can be used. They could be used by other workers, who might want to clone the genes by using oligonucleotides designed against these sequences. Alternatively they can be used as extra signatures when searches are being carried out for these proteins.

6.2.14 Comparison of the *S.coelicolor ICL* to other ICL's

The program "PILEUP" helped establish that the *S.coelicolor* ICL is similar to the other ICL enzymes. However, visually it is difficult to see the extent of similarity and which regions of the protein are most similar to each other.

For this reason the programs "COMPARE" and "DOTPLOT" from the GCG package were utilised to obtain a graphical representation of the similarity between the *S.coelicolor* ICL and the other ICL's.

"COMPARE" compares two sequences and writes a file of the points where matches of a certain quality are found. The points in the output file can be plotted with the "DOTPLOT" program.

"COMPARE" was used according to the method of Maizel and Lenk (1981). In this case a comparison of two proteins was made within a specified window. When the number of residues equaled or was greater than a specified stringency, then a match was scored. The window of one protein was compared to every register of the second protein and scored respectively. The window of the first protein was then moved along by one and the analysis carried out again. In this way every region of the first protein was compared with every region of the second protein, which shows the similarity of the two proteins, even if there are insertions or deletions in one protein with respect to the other.

The program "DOTPLOT" was then used to visualise all of the structures in common between the two sequences. A dot was placed in a position corresponding to the middle of a matching window. Lines were then built up where there were regions of similarity between the two proteins.

"COMPARE" was run with a window of 30 and a stringeny of 18, i.e., there had to be 18 matches out of 30 in a given window, for a point to be scored.

The output from several plots can be seen in Fig. 6.11. This shows that the *S.coelicolor* ICL is very similar to the other ICL's. The comparison between the *S.coelicolor* ICL and the *E.coli* ICL shows that these two ICL's are extremely similar to each other across the entire protein sequence.

The *S.coelicolor* ICL sequence also shows a great deal of similarity to the other ICL's. The plots obtained show a similar pattern. However the *S.coelicolor* ICL shows differences to the other ICL's at the N-terminus, it also lacks a central portion which is found in the eukaryotic ICL's.

Fig. 6.11, also shows plots of some of the other ICL's compared against each other. It can be seen that the eukaryotic ICL's (rape seed, castor bean, *A.nidulans* and *C.tropicalis*), are all very similar to each other.

The plots of the *E.coli* ICL against the eukaryotic ICL's show a similar pattern to that found when the *S.coelicolor* ICL was compared against them.

Thus, it appears that the two prokaryotic ICL's are extremely similar to each other and that they both show the same differences to the eukaryotic ICL's. That is, they differ from eukaryotic ICL's in the N-terminal region and they both lack a 102 amino acid stretch found in the central region of eukaryotic ICL's.

It is not known why these differences should come about, but it may be that the eukaryotic sequences contain signal sequences for compartmentalisation of the protein. These sequences would not be expected to be found in the prokaryotic ICL's.

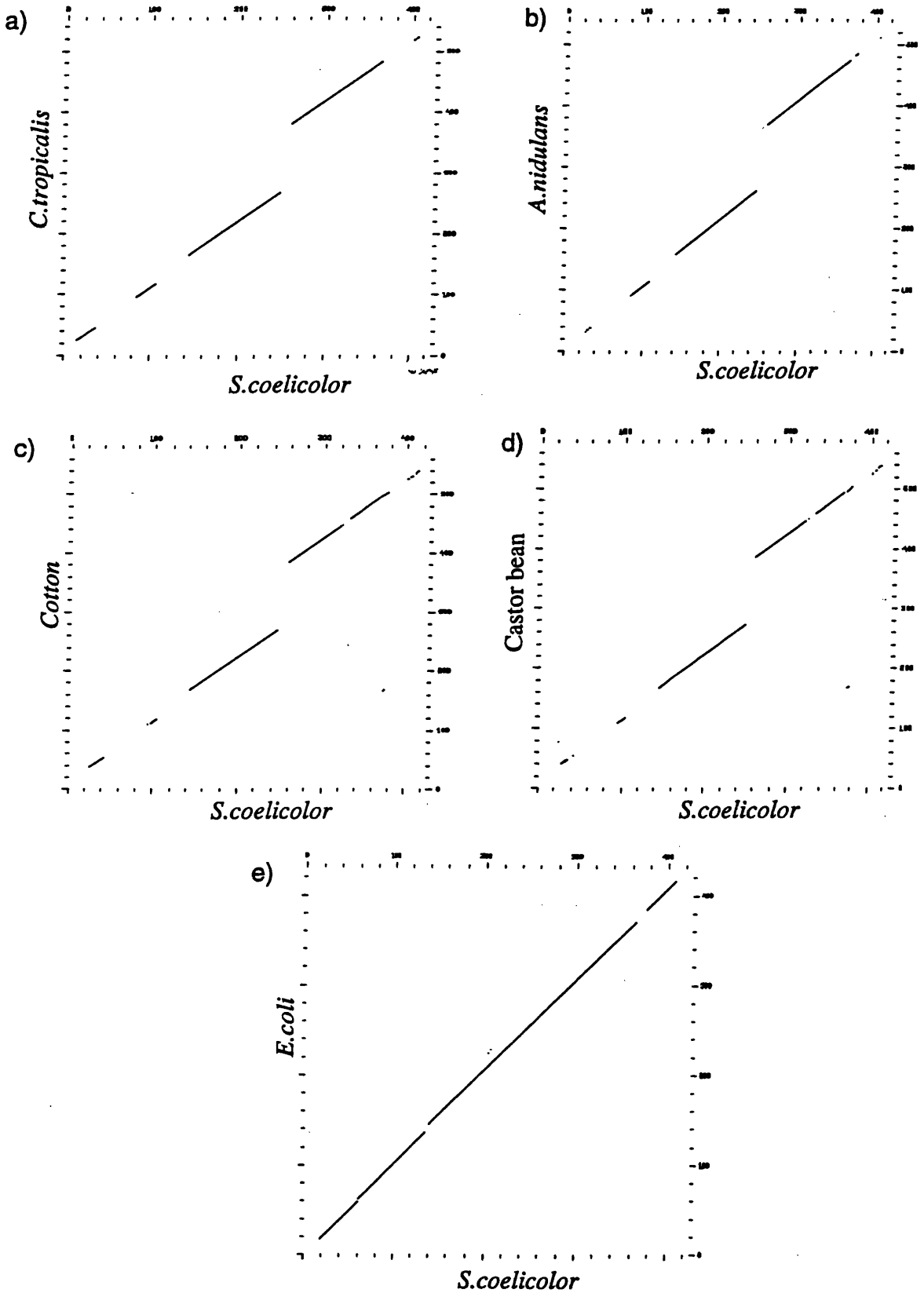
Eukaryotic ICL's are found within glyoxysomes, which are a form of peroxisomes. There is a recognition sequence for peroxisomes, which has been implicated in peroxisomal targetting.

(SAC) - (KHR) - L

When the eukaryotic sequences were searched for this sequence, no match to this motif was found.

Interestingly, when the malate synthase enzymes are scanned for this motif, matches were found at the C-terminus of the enzyme. 8 out of 10 eukaryotic MS enzymes have a match to this motif. The only prokaryotic MS is that from *E.coli*, which is slightly smaller than the eukaryotic enzymes and does not contain this motif.

"COMPARE" and "DOTPLOT" were also used to look at the MS sequences (see Figure 6.12). It appears that all of the MS enzymes, both eukaryotic and prokaryotic are extremely similar to each other.


a)
C.tropicalis
S.coelicolor
b)
A.nidulans
S.coelicolor
c)
Cotton
S.coelicolor
d)
Castor bean
S.coelicolor
e)
E.coli
S.coelicolor

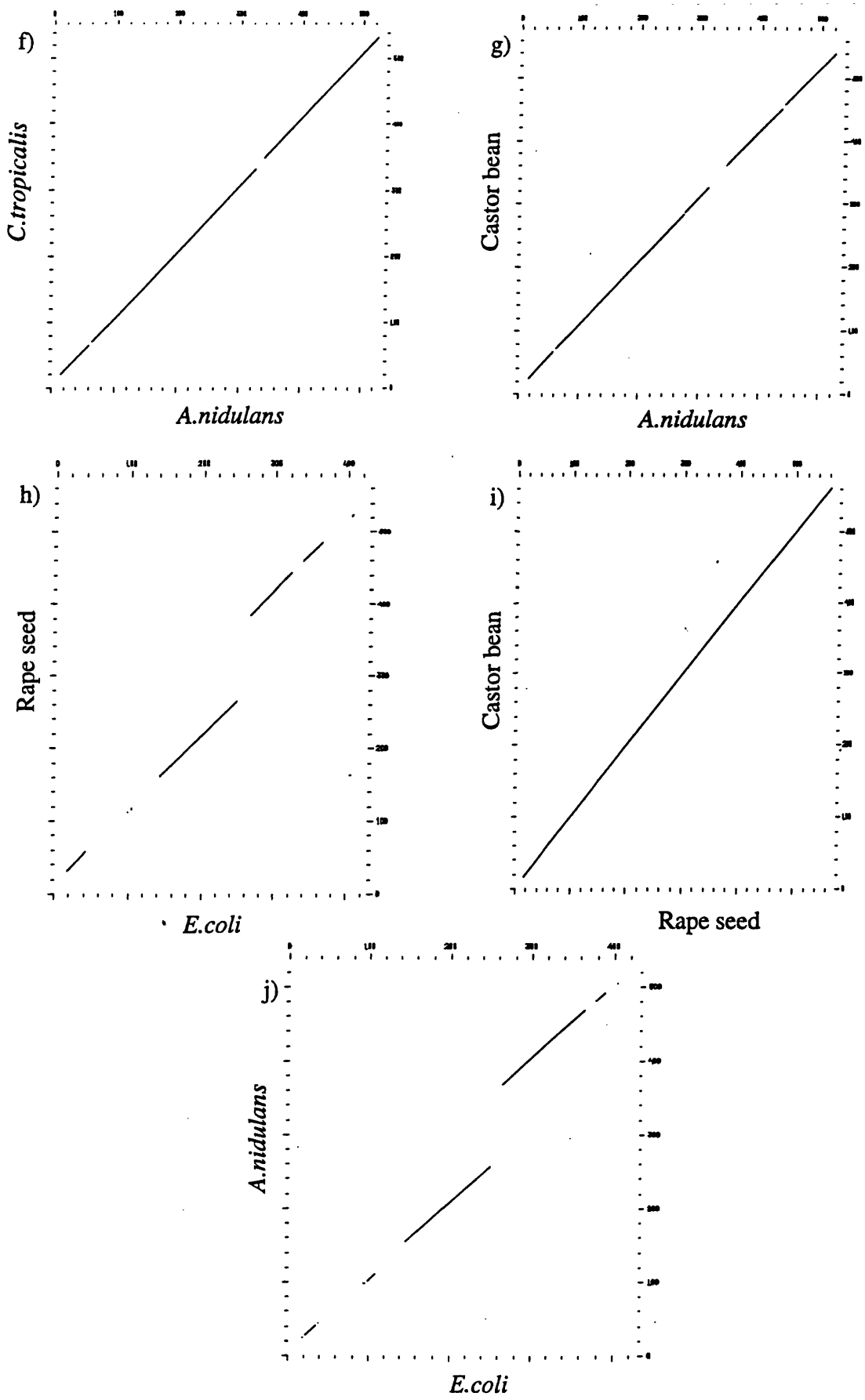
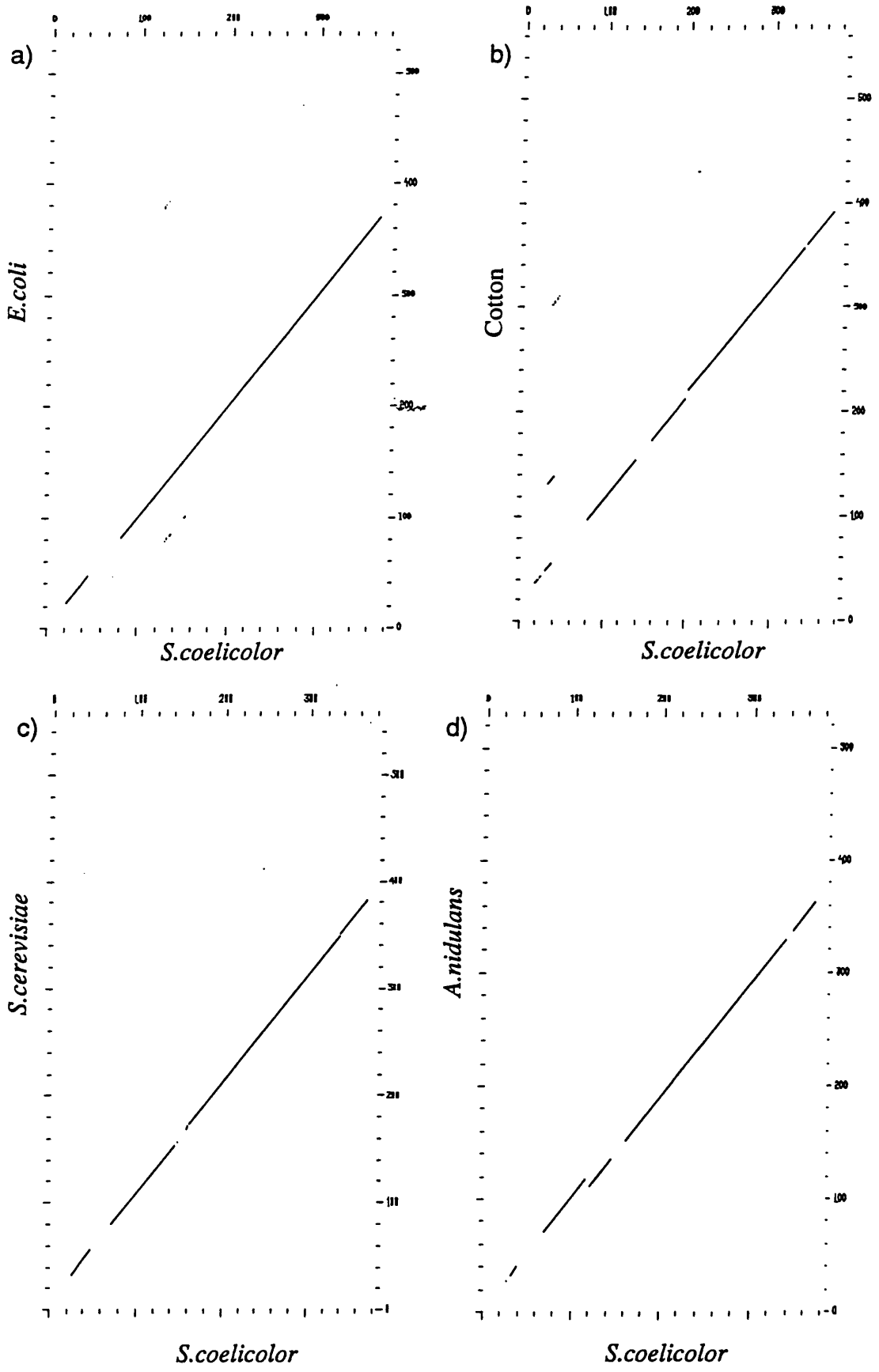

f)
C.tropicalis
A.nidulans
g)
Castor bean
A.nidulans
h)
Rape seed
E.coli
i)
Castor bean
Rape seed
j)
A.nidulans
E.coli

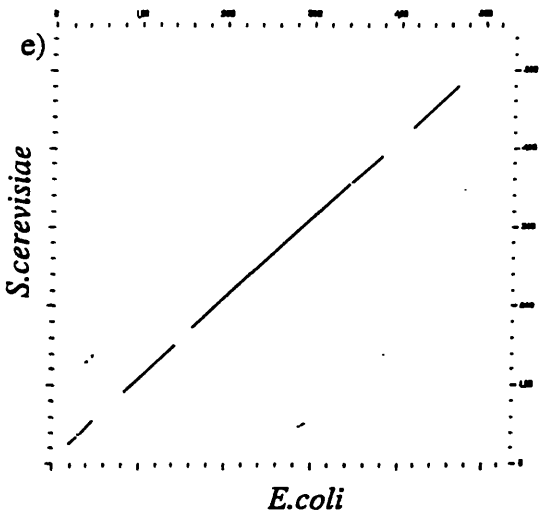
Figure 6.11 Comparison of some ICL sequences using "DOTPLOTS"

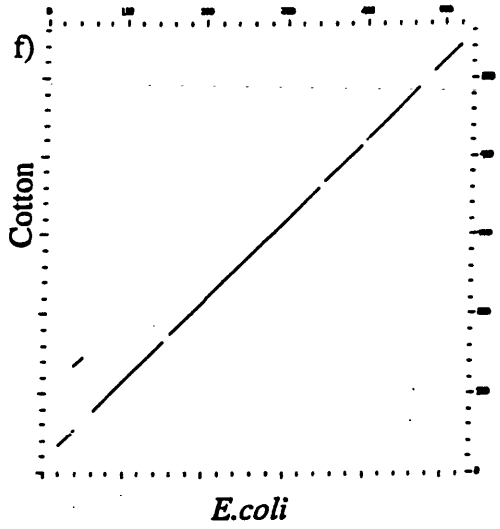
The amino acid sequence of *S.coelicolor* ICL was compared to the amino acid sequences of other ICL's. These other ICL's were also compared to each other to determine their similarities.

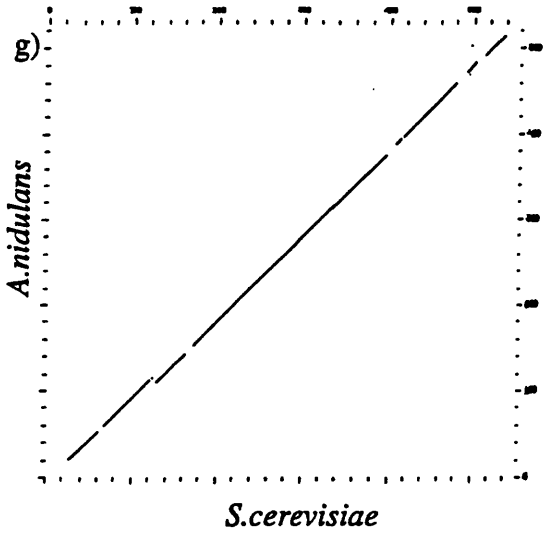
The comparisons were carried out using the GCG program "COMPARE", using a window of 30 and a strigency of 18. Results were then displayed using the program "DOTPLOT".

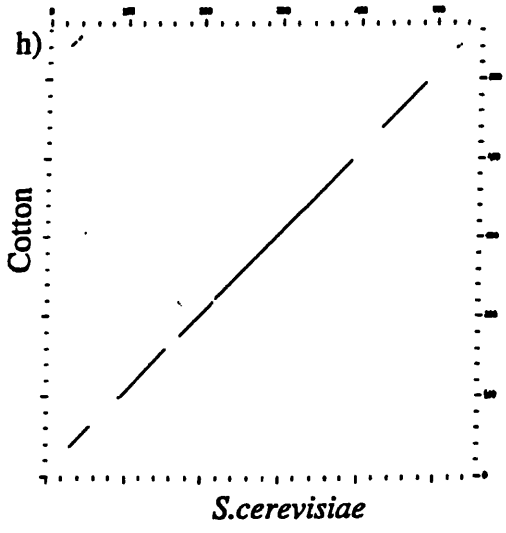
a) *C.tropicalis* ICL vs *S.coelicolor* ICL
b) *A.nidulans* ICL vs *S.coelicolor* ICL
c) Cotton ICL vs *S.coelicolor* ICL
d) Castor bean ICL vs *S.coelicolor* ICL
e) *E.coli* ICL vs *S.coelicolor* ICL
f) *C.tropicalis* ICL vs *A.nidulans* ICL
g) Castor bean ICL vs *A.nidulans* ICL
h) Rape seed ICL vs *E.coli* ICL
i) Castor bean ICL vs Rape seed ICL
j) *A.nidulans* ICL vs *E.coli* ICL


a)
E.coli
S.coelicolor
b)
Cotton
S.coelicolor
c)
S.cerevisiae
S.coelicolor
d)
A.nidulans
S.coelicolor


e)
S.cerevisiae
E.coli


f)
Cotton
E.coli


g)
A.nidulans
S.cerevisiae


h)
Cotton
S.cerevisiae

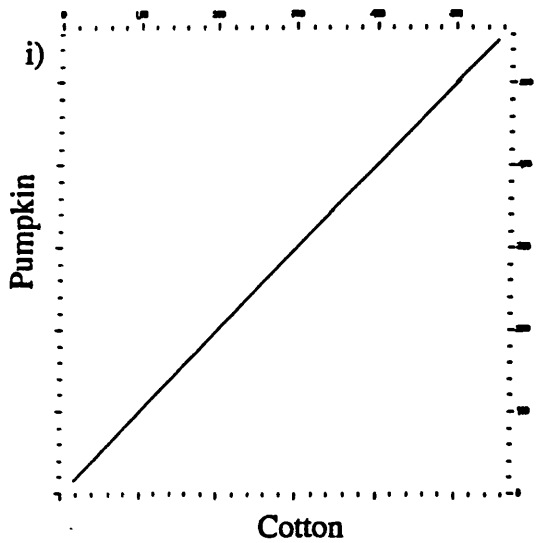

i)
Pumpkin
Cotton

Figure 6.12 Comparison of some MS sequences using "DOTPLOTS"

The amino acid sequence of partially sequenced*S.coelicolor* MS was compared to the amino acid sequences of other MS's. These other MS's were also compared to each other to determine their similarities.

The comparisons were carried out using the GCG program "COMPARE", using a window of 30 and a strigency of 18. Results were then displayed using the program "DOTPLOT".

a) *E.coli* MS vs *S.coelicolor* MS
b) Cotton MS vs *S.coelicolor* MS
b) *S.cerevisiae* MS vs *S.coelicolor* MS
d) *A.nidulans* MS vs *S.coelicolor* MS
e) *S.cerevisiae* MS vs *E.coli* MS
f) Cotton MS vs *E.coli* MS
g) *A.nidulans* MS vs *S.cerevisiae* MS
h) Cotton MS vs *S.cerevisiae* MS
i) Pumpkin MS vs Cotton MS

Chapter 7

Overexpression and attempted disruption of the *S.coelicolor* IDH

Chapter 7 Overexpression and attempted disruption of the *S.coelicolor* IDH

7.1 Introduction

Previous work carried out by Taylor (1992) had led to the purification of IDH and cloning of the *idh* gene from *S.coelicolor*. This gene product turned out to be particularly interesting as it was the first type II IDH to be cloned and sequenced.

Fukanaga *et al* (1988) had suggested the classification of bacterial IDH's into two groups. The first group (type I) consists of dimeric enzymes with subunit molecular weights of c. 45kDa. The second group (type II) consist of enzymes which are monomeric with molecular weights around c. 80kDa

The *S.coelicolor* IDH was found to be c. 80kDa and monomeric in nature. It was also found not to cross-react with anti-*E.coli* IDH antiserum (*E.coli* IDH is a type I enzyme). The *S.coelicolor* IDH was also found to have a N-terminus similar to *R.vannielli* IDH and *Vibrio spp.* IDH-II, which were both classified as type II IDH's.

It was thus of great interest and importance to continue this work further. Overexpression was to be attempted first, in order to prove that the cloned gene did in fact encode IDH As the DNA sequence of the *S.coelicolor idh* was different to the other *idh*'s sequenced previously, there was still a slim possibility that the cloned DNA was not in fact *idh*. The enzyme had to be overexpressed to confirm that the cloned DNA encoded IDH from *S.coelicolor*.

It was also of interest to express the *S.coelicolor idh* gene in *E.coli idh* mutants, firstly with cells grown on glucose, to find out if the *S.coelicolor* IDH could complement an *E.coli idh* mutant, and secondly for cells grown on acetate, to find out if the *E.coli* phosphorylation system will be able to covalently modify the *S.coelicolor* IDH, as it would its own IDH.

If overexpression of *S.coelicolor* IDH could be obtained in *S.coelicolor*, it would be of interest to study the effect of the elevated level of the enzyme on growth. Indeed, overexpression of the enzyme might have an effect on antibiotic biosynthesis.

A second important experiment was to attempt to disrupt the *idh* gene in *S.coelicolor*. This was important for a number of reasons. Firstly, it would test if a second enzyme with IDH activity was present. Secondly, once an *idh* mutant was constructed it might be possible to study the effects on antibiotic synthesis of varying the level of the IDH enzyme.

This could be achieved by intoducing the cloned *idh* gene into an *S.coelicolor idh* mutant, on a plasmid that allows controllable expression of the gene. By controlling the levels of IDH and also the stages of growth when it was produced, it might be possible to engineer metabolism, and ultimately the levels of antibiotics produced. This would be of great importance to industry, where optimum production of antibiotics is sought.

This chapter deals with the overexpression of *S.coelicolor* IDH in *S.lividans* and *S.coelicolor*. It also gives details of the preliminary experiments designed to lead to the overexpression in *E.coli* of the IDH from *S.coelicolor*. Finally attempts to disrupt the *idh* gene in *S.coelicolor* are described.

7.2 Results

7.2.1 Overexpression of *S.coelicolor* IDH in *S.lividans*

S.lividans was used as an intermediate host before overexpression was attempted in *S.coelicolor*, because it was easier to obtain *S.lividans* protoplasts which gave a sufficiently high transformation frequency. *S.lividans* protoplasts could be made which transformed at c.100-fold higher efficiency than *S.coelicolor*.

The strategy was to transform *S.lividans* protoplasts with a ligation mix, containing *idh* DNA and the appropriate vector. The transformants were then checked for overexpression of IDH. When an overexpressing transformant was obtained, plasmid DNA was prepared and this was then used to transform *S.coelicolor* 213.

Two different ways of overexpressing *S.coelicolor* IDH were attempted. The first method was to use a plasmid containing the strong *veg* promoter, which had been obtained from *B.subtilis* (Moran *et al,* 1982). Cloning *idh* downstream of this promoter would hopefully result in transcription of *idh* and ultimately translation of the gene. The plasmid to be used for this purpose was pCBS7 which contains the *veg* promoter cloned into the *Eco* RI/ *Bam* HI sites of the pIJ487 (Ward *et al*, 1986) polylinker region.

The second overexpression experiment relied on the possibilty of *idh* being expressed from its own promoter. Although no promoter studies had been carried out, a number a putative promoters had been observed by inspection of the DNA sequence upstream *idh* (Taylor, 1992). Thus *idh* was to be cloned with approximately 600bp of upstream sequence, in which it was hoped that the promoter or promoters would be present.

7.2.1.1 Construction of the sub-clone pCBS*idh*1

An intermediate step had to be carried out before the DNA encoding *idh* could be cloned into pCBS7. pRTB1 (Taylor, 1992) was digested with *Eco* RV and *Sma* I which generated a 2.5kb fragment which contained only 45 bp of sequence upstream from the predicted start of the *S.coelicolor IDH*. The DNA was separated on a 1% TAE agarose gel and the fragment purified using a spin-X tube (2.3.13).

The resulting fragment was ligated with pBluescript II KS+, which had been digested previously with *Sma* I. This DNA was then used to transform *E.coli* DS941 which had been made competent by the method of Hanahan (2.2.8.1b).

This resulted in transformants which had the *S.coelicolor IDH* cloned in both orientations with respect to the T7 promoter. Plasmid DNA was then prepared from 8 transformants (2.3.2). This DNA was digested with *Bam* HI and the resulting fragments separated on a 1% agarose gel. Depending on the orientation of the fragment within the plasmid, the sizes of the separated fragments were either 3.6 and 1.75kb or 0.7 and 4.65kb. A transformant (p*idh*100) that yielded fragments of 3.6 and 1.75kb was used for further sub-cloning, as this had the *idh* gene in the required orientation.

p*idh*100 DNA was digested with *Xba* I and *Hind* III and the resulting 2.5kb *idh*-containing fragment purified in the manner described above. A ligation was set up with pCBS7 DNA, which had been digested and purified in a similar manner.

This ligation mix was used to transform protoplasts of *S.lividans* TK64 (2.2.9.2) and the protoplasts plated onto R2 regeneration plates. After 16 hours at 30°C, the plates were overlayed with thiostrepton, in order that only recombinant colonies would continue to grow.

After 7 days a number of sporulating colonies were obtained. The spores from these colonies were inoculated into YEME medium, which had been supplemented with thiostrepton. The cultures were then left to grow at 30°C for 3 days, until thick growth was observed.

Plasmid DNA was prepared from the cultures (2.3.2.). The resulting DNA was separated along with control pCBS7, on a 1% TAE agarose gel. (Fig. 7.1). 3 out of the 4 transformants (B,C, and D) checked appeared to contain plasmids which were greater in size than pCBS7. These clones were taken to contain the *idh* insert. One clone (tranformant B), *S.lividans*/ pCBS*idh* 1, was picked for further experimentation.

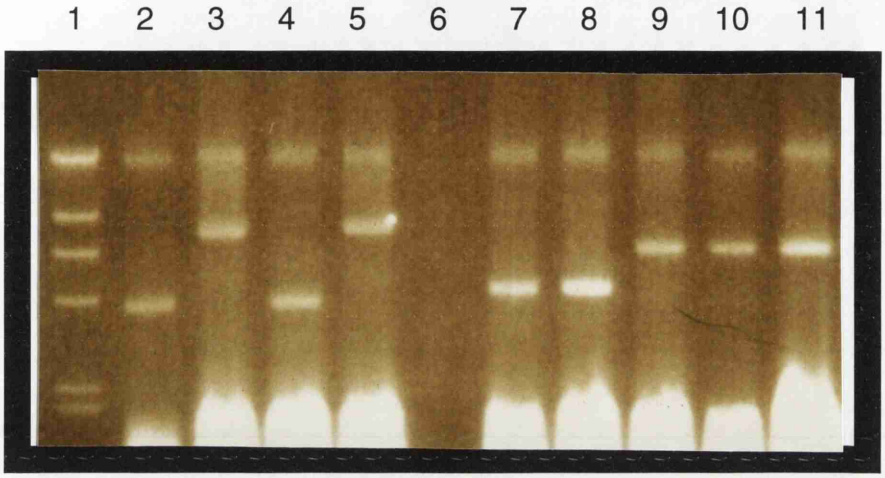


Figure 7.1 Agarose gel to check which transformants contained inserts

Transformants were obtained and allowed to sporulate. Spores were then used to inoculate 10ml cultures of YEME, supplemented with thiostrepton. 5ml of the resulting cultures was used to prepare plasmid DNA (2.3.2). Half of the resulting preparation was then run on a 0.8% agarose gel to determine which of the transformants contained inserts.

Lane	DNA	Lane	DNA
1	λ*Hind* III markers	7	pCBS7 (control)
2	pIJ486	8	transformant A
3	transformant 1	9	transformant B
4	transformant 2	10	transformant C
5	transformant 3	11	transformant D
6	---		

7.2.1.2 Construction of plasmid p*idh*486

pRTB1 was digested with *Xba* I and *Hind* III to generate a 5kb fragment, which contained *idh* with 0.6kb of upstream sequence. This fragment was then ligated with pIJ486 which had been previously digested with *Xba* I and *Hind* III.

This ligation mix was then used to transform *S.lividans* TK64 protoplasts in exactly the same manner as described above. The protoplasts were then regenarated and overlayed, as described above.

Spores from the transformants obtained were used to inoculate YEME, as described above and plasmid DNA obtained from the cultures. This DNA was separated on a 1% TAE agarose gel (Fig. 7.1). 2 out of the 3 transformants (1 and 3) appeared to contain inserts, by comparison with control DNA of pIJ486. One clone (transformant 1), *S.lividans*/p*idh*486 was picked for further experimentation.

7.2.2 SDS-PAGE of crude extracts from *S.lividans* containing plasmids p*idh*486 and pCBS*idh* 1

The two *S.lividans* clones which contained the plasmids p*idh*486 and pCBS*idh*1 were inoculated into YEME supplemented with thiostrepton, plus two control cultures (*S.lividans* TK64 containing pIJ486 or pCBS7). The cultures were grown for 3 days until thick growth was observed. 1ml of each culture was then spun down and a crude protein extract obtained (2.4.1). 20µl of each of the extracts was then run on a 10% (w/v) SDS PAGE gel and the protein visualised by staining with Coomassie (2.4.4e; Fig. 7.2). It was clearly observed that IDH was overexpressed from plasmid p*idh*486. This plasmid contained *idh* including upstream sequences. Consequently this upstream sequence most likely contains a promoter or promoters. No over-expression was observed from the *veg* promoter in plasmid pCBS*idh*1.

7.2.3 Construction and assay of S.coelicolor 213 overexpressing IDH

Plasmid p*idh*486 from *S.lividans* /p*idh*486 which overexpressed IDH, was prepared. 20ng of p*idh*486 was then used to transform *S.coelicolor* 213 protoplasts. The transformants were selected as described in 7.2.1.1.

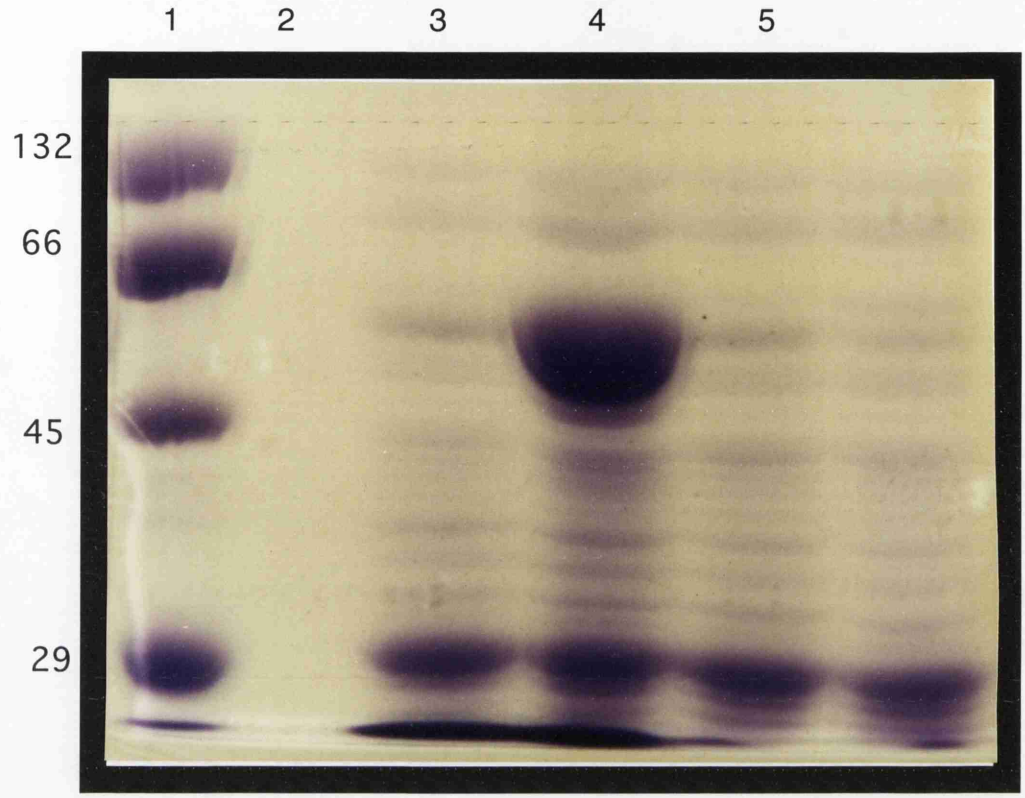


Figure 7.2 SDS-PAGE of crude extracts of *S.lividans* containing plasmids p*idh* 486 and pCBS *idh*1

Two different plasmids containing *S.coelicolor idh* were constructed and cloned into *S.lividans* TK64. Crude extracts were prepared as described in section 2.2.4 and 20µl run on a 10% SDS gel. The bands were visualised by coomassie staining.

Lane	Construct
1	Markers (2.4.5)
2	pIJ486 (Control)
3	p*idh*486
4	pCBS7 (Control)
5	pCBS*idh* 1

Spores from one of the transformants were used to inoculate 50ml of YEME culture medium, which was supplemented with thiostrepton. A control culture of plasmid-free *S.coelicolor* 213 was also grown. The cultures were grown until thick growth was observed. The mycelia from the 50ml of culture medium was spun down and crude extracts prepared as described in 2.4.1. The crude extracts from the control culture and the overexpressing strain were assayed for IDH activity (2.4.4i)). The control culture had a specific activity of 57mU/mg and the overexpressing strain 2.25U/mg. This represents a 39.5 fold overexpression. It does appear that the cloned DNA encodes *S.coelicolor idh.*

7.2.4 Attempts at overexpression of *S.coelciolor* IDH in *E.coli*

Previously Taylor (1992) had tried to complement an *E.coli idh* mutant, EB106, using the plasmids pRTB1, pRTB2 and pRTB3 (Taylor, 1992). No complementation was observed for any of the constructs.

The work decribed in the previous section showed that a promoter or promoters were present in the region upstream from *idh.* Sequence analysis of this upstream region showed that the appropriate sequences normally required for expression of genes in *E.coli* were absent (Taylor, 1992). It was probable that the *idh* promoter sequence was not recognised by *E.coli.* It is well known that only a fraction of streptomycete genes are capable of being expressed in *E.coli*, due to the lack of "*E.coli*-like" promoter sequences (Jaurin and Cohen 1985).

This laboratory had previously expressed streptomycete genes in *E.coli* utilising constructs based on the pT7-7 plasmid (K.Linton and G. Young, personal communications). An attempt was made to express *S.coelicolor* IDH using this plasmid.

7.2.4.1 The pT7-7 vector

The pT7-7 plasmid comes from a group of vectors designated pET vectors (**p**lasmid **e**xpression by **T**7 polymerase). The pT7-7 plasmid contains a polylinker, a promoter recognised by T7 RNA polymerase and a translation start site (containing the ribosome binding site). The presence of a unique *Nde* I restriction site (recognition sequence CATATG) allows for cloning of the ATG start codon of a gene at the optimum distance from the ribosome binding site.

The vector is placed under the control of the T7 RNA polymerase rather than overexpressing the gene by *E.coli* RNA polymerase for several reasons:

1) T7 RNA polymerase transcribes at a faster rate than that of *E.coli* RNA polymerase.
2) T7 RNA polymerase is highly selective and initiates transcription only from its own promoter sequences.
3) T7 RNA polymerase is resistant to rifampicin. Rifampicin can be added to cells which results in exclusive expression of genes directed by T7 RNA polymerase, as *E.coli* RNA polymerase is sensitive to rifampicin.
4) It is also a useful vector to use when trying to express Streptomycete DNA, as Streptomycete DNA is G+C rich; *E.coli* RNA polymerase often stalls when trying to transcribe Streptomycete DNA; T7 RNA polymerase does not.

7.2.4.2 Mutagenesis of the DNA encoding the N-terminus of IDH from *S.ceolicolor*

S.coelicolor idh did not have an *Nde* I site in the DNA encoding its N-terminus, nor did it start with an ATG. Mutagenesis had to be carried out in order to create an ATG start codon, which is also part of an *Nde* I site.

S.coelicolor idh was first subcloned from pRTB1 into pT7-7, by excision from pRTB1 using *Eco* RV and *Sma* I ligated into pT7-7, which had been digested with *Sma*I. This created the construct as shown in Fig. 7.3.

Mutagenesis was facilitated using PCR. The first primer was designed against the sequence of the N-terminus of the *idh*, with the exception of the first 6 nucleotides, which coded for an *Nde* I restriction site. This also had the effect of altering the start GTG to ATG. The second primer (oligo 14) had been designed previously for sequencing of *idh* (Taylor, 1992). Fig. 7.4 shows a representation of the PCR carried out.

The PCR was set up as described in 2.3.24. 100pmol of each primer was used and 10ng of pRTB1 (Taylor, 1992) was used as template. The cycle was, 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 1minute. This cycle was repeated 30 times.

The expected size of the PCR product was ~1.3kb. A product of this size was observed when an aliqout of the reaction was run on an 1.5% agarose gel.

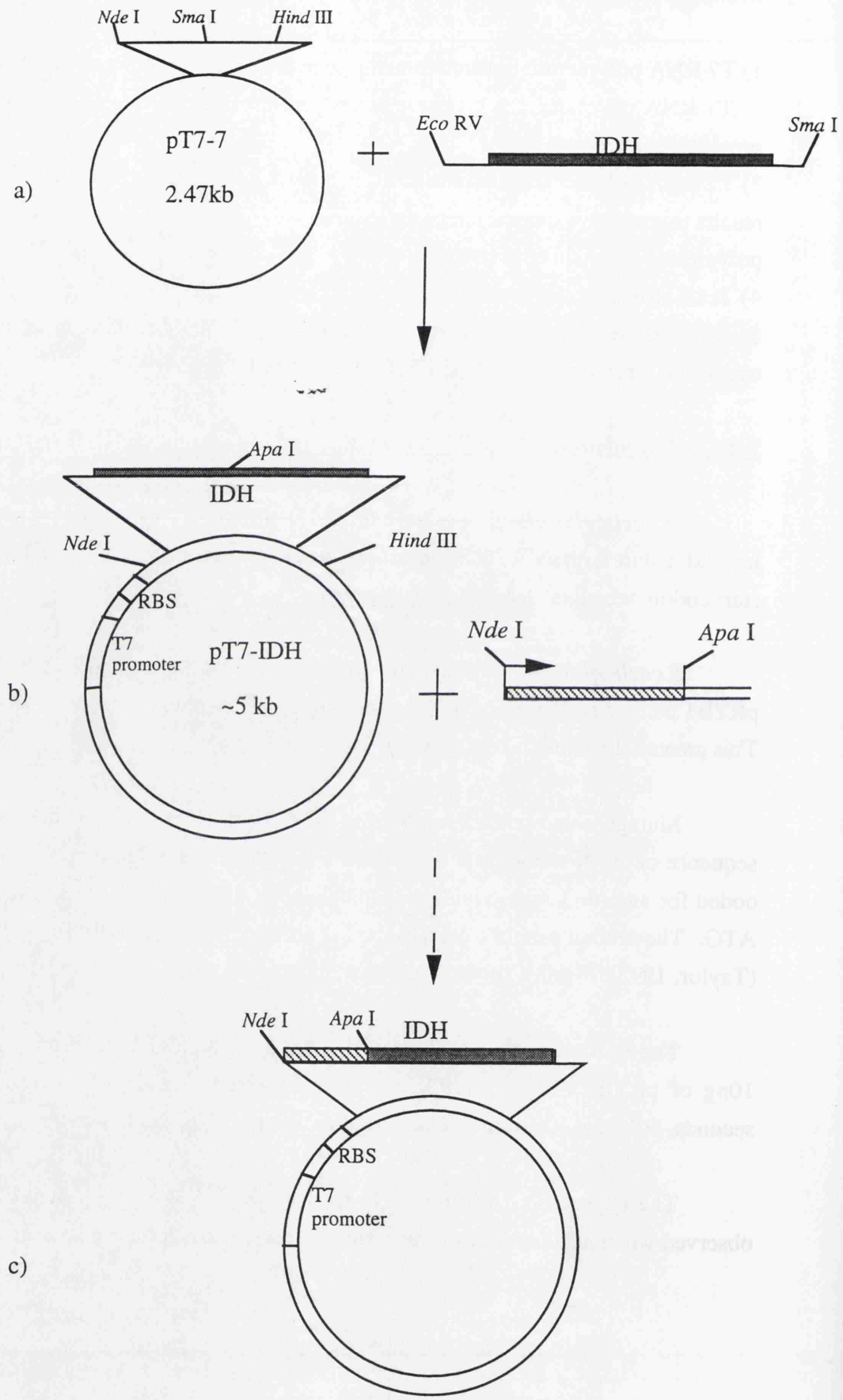

a)
Nde I
Sma I
Hind III
pT7-7
2.47kb
Eco RV
IDH
Sma I
b)
Apa I
IDH
Nde I
Hind III
RBS
T7
promoter
pT7-IDH
~5 kb
Nde I
Apa I
c)
Nde I
Apa I
IDH
RBS
T7
promoter

Figure 7.3 Representation of the stages in the creation of an overexpressing copy of *S.coelicolor idh* in *E.coli*

a) In order to first clone the *S.coelicolor idh* into pT7-7, the *idh* gene was excised from pRTB3 (Taylor, 1992), on *Eco* RV/*Sma* I ends and cloned into *Sma* I digested pT7-7.

b) This created pT7-IDH. Expression would not be acheived from this construct as the start of the gene was not optimally spaced from the RBS and did not initiate with an ATG.

Mutagenesis of the start site was to be acheived by cloning of a PCR product (see figure 7.4), which allowed for the start of *idh* to be an ATG, which would also be optimally spaced from the RBS.

c) This is the conceptual representation of the construct that would have resulted from cloning the PCR product into the *Nde* I/*Apa* I sites of pT7-IDH

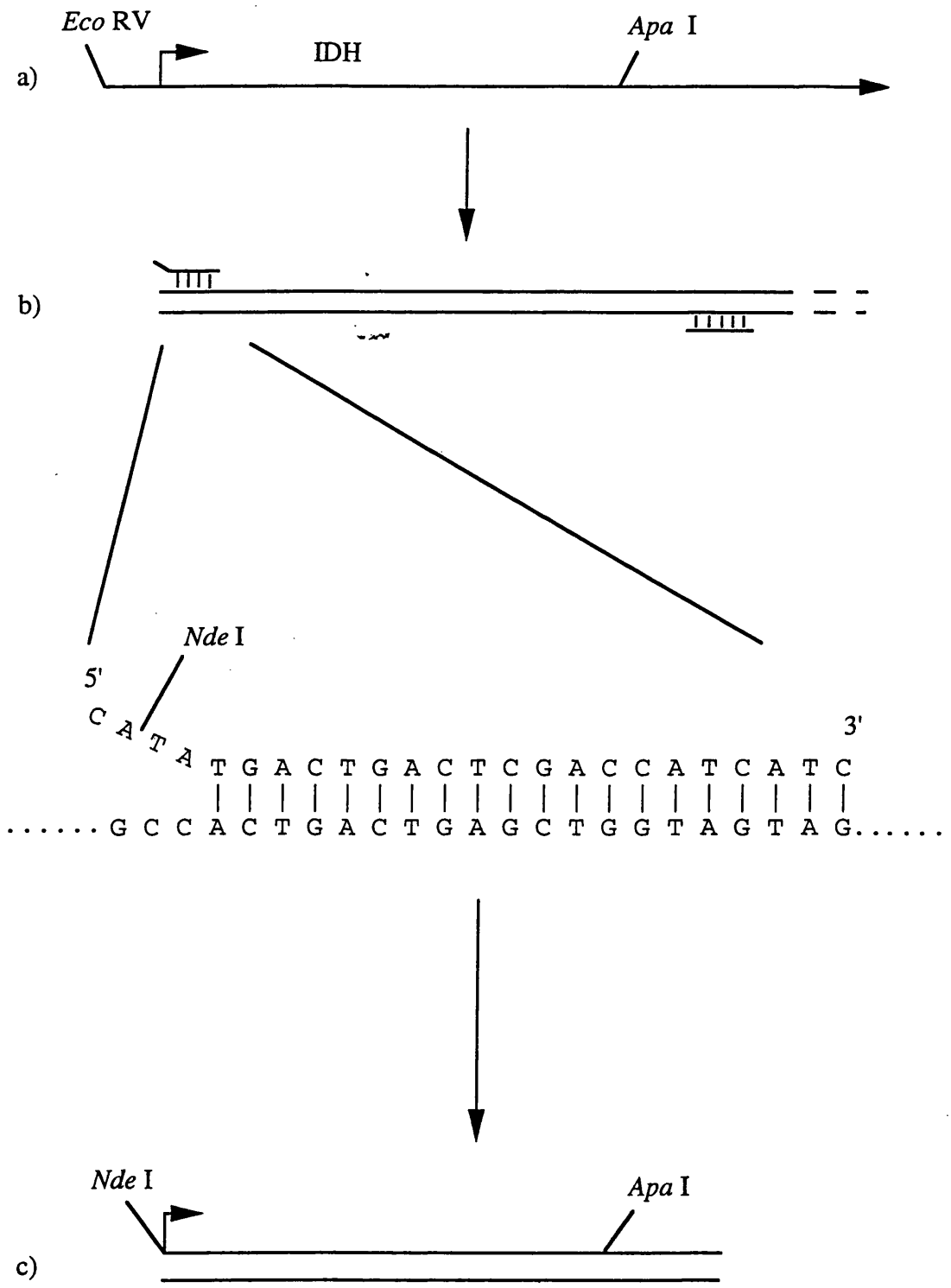

a)
Eco RV
IDH
Apa I
b)
Nde I
5'
C A T A T G A C T G A C T C G A C C A T C A T C
3'
. . . . . G C C A C T G A C T G A G C T G G T A G T A G. . . . . .
c)
Nde I
Apa I

Figure 7.4 PCR amplification of part of the *S.coelicolor idh* gene

a) This shows a representation of part of the *S.coelicolor idh* gene. The arrow represents the start of the gene and the *Apa* I site is a unique site found internal to the gene.

b) This is a representation of how the PCR primers anneal to the *idh* gene. In particular it shows the primer which creates the *Nde* I site at the N-terminus and mutates the start to an ATG from a GTG. The second primer (oligo 14) was designed previously for sequencing of *idh* (Taylor, 1992)

c) This shows the PCR product which may be cloned into *Nde* I/*Apa* I digested pT7-IDH (Fig. 7.3)

Attempts were made to clone the PCR product. Unfortunately this was not achieved even though the cloning experiment was tried many times, and different cloning strategies were carried out. It was hoped that when the PCR product was cloned into a suitable cloning vector (such as pBluescript), an Nde I/Apa I insert could be excised. This would then be ligated into similarly-digested pT7-IDH. This would have generated a construct in which the *idh* initiation codon was an ATG which was spaced optimally from the ribosome binding site. (Fig. 7.3)

7.2.5 Attempts to disrupt the genomic copy of the *S.coelicolor idh*

Disruption of the *S.coelicolor idh* was planned to study the effects of a mutation in *idh* on growth and antibiotic production. It would also enable growth studies to be carried out using a controllable exogenous copy of *idh.*

The method to be tried involved cloning of an antibiotic marker into the middle of the *idh* gene and then incorporation of this disrupted copy into the genome. This would result in the non-functional expression of the *idh* gene and an *idh*$^-$ phenotype. It was expected that this would result in mutants which were auxotrophic for proline as is the case in *idh* mutants of *E.coli*.

A plasmid copy of *idh* had to be first altered by cloning the *erm*-E gene (which encodes resistance to the antibiotic erythromycin) into the middle of *idh*. The *erm*-E gene was to be cloned with its upstream promoter sequences. This would allow for expression of the *erm*-E gene, which would aid in the screening process for gene disruption.

7.2.5.1 Construction of p*IDH-erm* E1

pIJ4026 contained the *erm*-E copy which was excised using *Kpn* I. The *erm*-E containing DNA was then gel purified. p*IDH*100 (7.2.1.1) was restricted with *Mlu* I. This plasmid contained only one *Mlu* I site which was located centrally within *idh*. The *idh* containing DNA was also gel-purified. The fragments contained overhangs which were incompatible for ligation. In order that the *erm*-E DNA could be cloned into the middle of the *idh* gene, the sticky ends of DNA had to be converted to blunt ends. The ends of the *erm*-E DNA were made blunt by using Mung bean nuclease. The *idh* DNA was made blunt by using the Klenow fragment of DNA polymerase I. A ligation was then set up and left at 16°C, overnight. The ligation mix was used to transform competent *E.coli* DS941.

Transformants were obtained. Plasmid was prepared from 12 transformants and the plasmid size determined to check that *erm*-E was cloned into the middle of *idh*. Only 1 out of 12 transformants checked appeared to contain the *erm*-E insert. The plasmid was called p*IDH-erm* E1 (Fig. 7.5).

7.2.6 Conjugation of *Ecoli* S17-1 with *S.coelicolor* 1147

Firstly plasmid p*IDH-erm* E1 was prepared and used to transform competen S17-1. Transformants were selected for their ability to grow on ampicillin-containing plates. A single colony was used to inoculate a 5ml culture of 2xYT, containing ampicillin as a selectable marker. The culture was then grown overnight.

Spores of *S.coelicolor* 1147 were filtered and resupended in 5ml of dH_2O. The procedure as described in 2.2.10 was then followed and the cells overlayed with erythromycin. The plasmid p*IDH-erm* E1 would not be maintained in *S.coelicolor* as it did not contain a streptomycete origin of replication. Any erythromycin-resistant exconjugants that were obtained, had to be as a result of homologous recombination between the wild-type copy and the disrupted copy . Unfortunately no exconjugants were obtained.

7.3 Discussion

It has been shown that the DNA cloned within pRTB1 does indeed contain the *S.coelicolor idh* gene. Overexpression of this gene was achieved in both *S.lividans* and *S.coelicolor*.

Attempts to overexpress *S.coelicolor* IDH in *E.coli* still have to be completed. The gene has been cloned into pT7-7, however the N-terminus still needs to be altered so that it is optimally spaced from the RBS and also initiates from an ATG. It was hoped that this could be achieved by mutating the N-terminus by way of PCR. Indeed a product of the correct size was obtained. However attempts to clone this fragment proved unsuccessful.

It was hoped that when the PCR product had been cloned, it could be excised by digestion with *Nde* I and *Apa* I and ligated into similarly digested pT7-*IDH*. This would create an *idh* gene which was optimally spaced from the RBS and initiated from an ATG start codon.

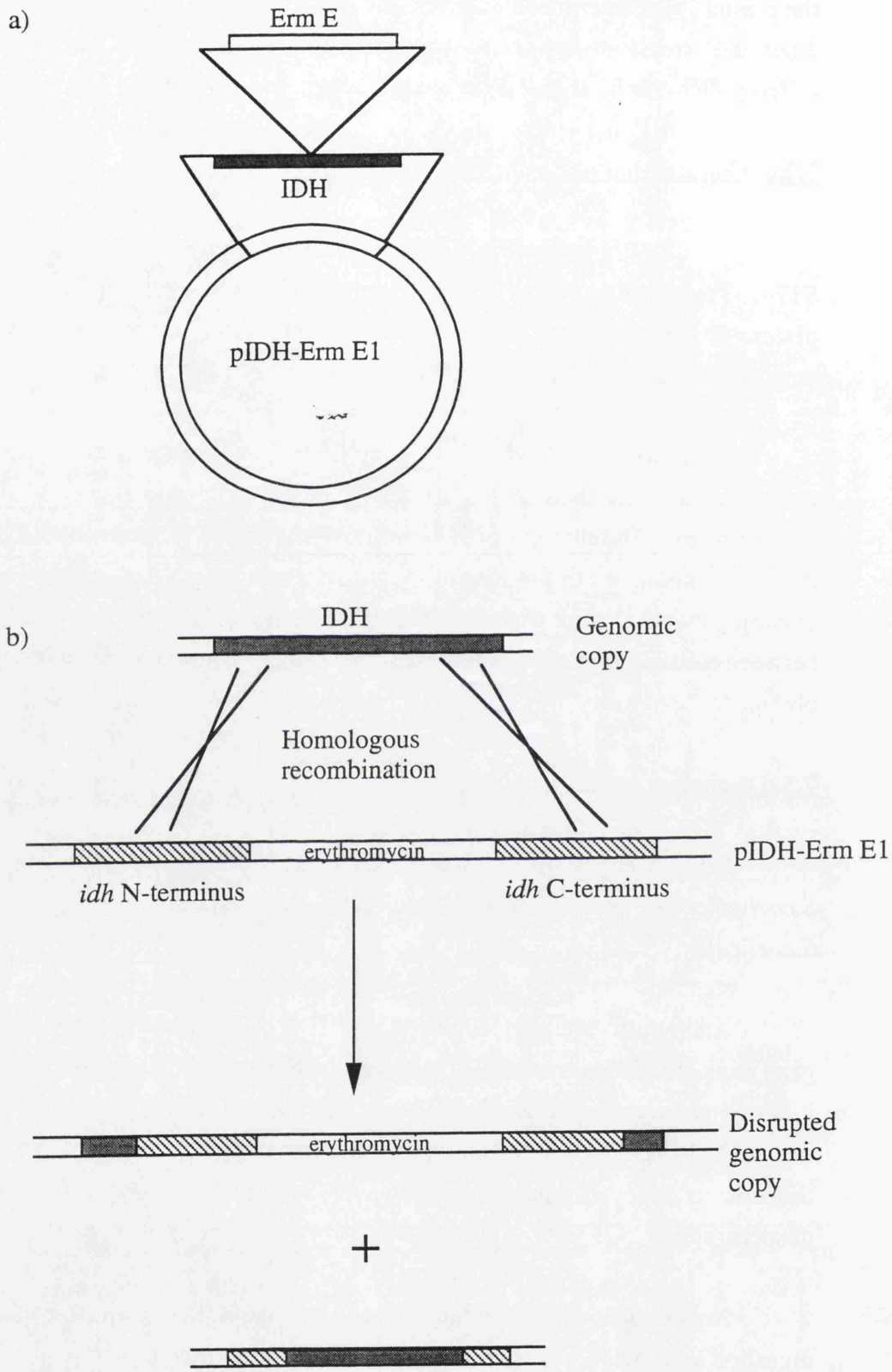

a)
Erm E
IDH
pIDH-Erm E1
b)
IDH
Genomic copy
Homologous recombination
erythromycin
pIDH-Erm E1
idh N-terminus
idh C-terminus
erythromycin
Disrupted genomic copy
+

Figure 7.5 Conceptual representation of a method to disrupt the *S.coelicolor idh* gene

a) This shows the plasmid pIDH-*erm* E1 which is a pBluescript derived plasmid with the IDH gene cloned into the *Eco* RV/*Sma* I sites of the polylinker region and the *erm*-E gene cloned into a unique *Mlu* I restriction site, found internal to the *idh* gene.

b) This represents the double cross-over event required to result in a disrupted copy of the *S.coelicolor idh* gene.

A single cross-over event does not result in a disrupted copy and no cross-over events result in loss of the plasmid, as there is no streptomycete origin of replication on the plasmid.

It was not known why cloning of the PCR product could not be achieved. Cloning of other PCR products had been previously been achieved (3.2.7), using the vector pDK101 (Kovalic *et al*, 1991) restricted with *Xcm* I, to create T overhangs. This method was tried, as was the method of using vectors engineered by adding dTTP to blunt-ended restriction sites, using Taq polymerase. Another method involved concatemerisation of the PCR product before digestion with restriction enzymes which then allows ligation into similarly- digested vectors. None of the above methods proved successful even though they were attempted a number of times. Due to lack of time this work had to be abandoned. However it would still be of interest to try to overexpress *S.coelicolor idh* in *E.coli.*

The attempts to disrupt the *S.coelicolor idh* proved unsuccessful. However, other members of this laboratory have identified a number of problems with this strategy. It appears that the use of erythromycin as a selectable marker for integration should not be used, as often, spontaneous resistant mutants have been found to propagate. This makes the screening of integrants a difficult process (P. McKernan personal communication).

It has also been observed that the frequency of obtaining a double cross-over event is extremely low. Thus the vast majority of integrants are due to single cross-over events, which result in an intact undisrupted copy of *idh* remaining. However, singles can be resolved to double cross-overs by non-selective growth.

Given that the gene has been completely sequenced, another method of obtaining a disrupted *idh* copy could be tried. This method is based on cloning a fragment of *idh* lacking both the N-terminus and the C-terminus. If this fragment is introduced into *S.coelicolor*, only a single cross-over event is required in order to obtain disruption of the *idh* gene. This is because, the cross-over event would result in two mutated copies of *idh*, one lacking an N-terminus and the other lacking a C-terminus. Both copies would be unable to generate native proteins.

Again, due to the lack of time, this work had to be cut short. However, currently this work is being carried on by J.Dowman in this laboratory.

CHAPTER 8

General Discussion

8.1 Introduction

The TCA cycle has an important role in the production of precursors for both primary biosynthesis and secondary metabolism. Streptomycetes produce a wide variety of secondary metabolites (in particular, antibiotics) which are of commercial interest. It is thus of interst to study the possible control of the TCA cycle and how flux through the cycle is altered between primary and secondary metabolism.

A particulary important group of antibiotics produced by some *Streptomyces* species, are the polyketide antibiotics. It has been elucidated that these antibiotics are synthesised from acetyl-CoA units.

When *E.coli* is grown on acetate, or fatty acids as a sole carbon source, the glyoxylate bypass is induced. This pathway allows for the metabolism of acetyl-CoA units (produced from acetate or fatty acids) by the TCA cycle. It was envisaged that this pathway may also be present in *Streptomyces*. If so, the pathway must be tightly controlled to allow metabolism of acetyl-CoA by the TCA cycle during primary metabolism and for the biosynthesis of polyketide antibiotics during secondary metabolism.

In *E.coli* the enzyme at the branch point of the TCA cycle and the glyoxylate bypass is IDH. This enzyme is strictly controlled by phosphorylation/ dephosphorylation. Cloning and characterisation of the IDH from *S.coelicolor* was carried out by Taylor (1992), but it is still unclear if this enzyme is also controlled by phosphorylation/ dephosphorylation.

In *E.coli*, the enzymes of the glyoxylate bypass (ICL and MS) are part of an operon, namely the *ace* operon. It was of interest to study these enzymes in *S.coelicolor* and to elucidate if they were part of an operon.

8.2 Attempts to clone *S.coelicolor icl* using primers designed against consensus sequences.

At the start of this project, three other *icl* gene sequences had been determined It was envisaged that an oligonucleotide could be designed against a region of high similarity (after aligning the amino acid sequences) and that this oligonucleotide could be used in hybridisation studies to clone the *S.coelicolor icl* gene.

A 1kb hybridising region of DNA was cloned and sequenced, but no similarity to any of the previously sequenced *icl's* was observed. To date this region of DNA shows no significant similarity to any proteins in the sequence databases.

It was unclear why this attempt was unsuccessful. Later work which enabled the gene sequence of the *S.coelicolor icl* to be determined, showed that the oligonucleotide contained only four mismatches to the actual gene sequence (Fig. 8.1a).

Knowing the percentage homology of the oligonucleotide, it was possible to calculate the theoretical Tm using the equation below (Albertson, 1988).

$$Tm = 81.5 + 16.6(\log M) + 0.41(G+C\%) - (820/l) - 1.2(100-h)$$

Where;
Tm=temperature at which half of the DNA hybrids melt. G+C=percent G+C content.
h=percent homology between probe and target sequence. l=length of probe
m=molarity of monovalent cation in the hybridisation buffer.

Using the values, M=0.66 (4xSSC), G+C%=62.5, l=32 and h=87.5, the equation gives an approximate Tm=65°C for the washing conditions employed using oligo*icl*. Washing of the filters was in fact only carried out at 60°C, so it would be expected that the oligonucleotide should remain bound.

There are a number of possible reasons why attempts to clone the *S.coelicolor icl* gene using this oligonucleotide were unsuccessful. The first and most obvious explanation is poor technique. The initial hybridisation studies may not have been stringent enough, resulting in the cloning of DNA which bound non-specifically to the genomic DNA. An experiment which has not been attempted, but would be of interest, would be to carry out another hybridisation experiment, but this time using the Tm information shown above.

Another possible explanation is that the Tm may not be as calculated. This could be due to the fact the the mismatches are clustered in the centre of the oligonucleotide. This is believed to have a greater effect on the melting temperature, than mismatches spread throughout an oligonucleotide.

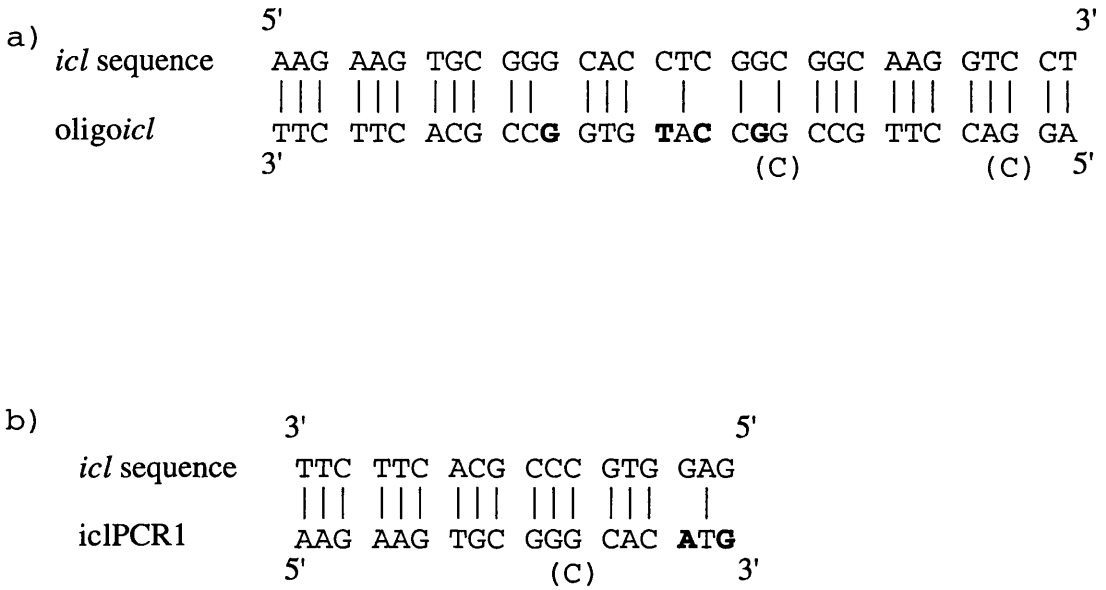


Figure 8.1 Binding of designed oligonucleotides to the *S.coelicolor icl* sequence

a) This shows the binding of oligo*icl* to the sequence of the *S.coelicolor icl.* Four mismatches were observed and these are shown in bold.

b) This shows the binding of iclPCR1 to the sequence of the *S.coelicolor icl.* Two mismatches were observed and these are shown in bold.

The attempts to clone a region of the *S.coelicolor icl* using primers designed against similar amino acid regions by PCR were also unsuccessful. The reason for this became evident once the *S.coelicolor icl* sequence was obtained. It was observed that although iclPCR2 was correct and showed no mismatches to the actual gene sequence, iclPCR1 was not. It contained 2 mismatches, one being the 3' base. This mismatch would not allow the binding of the 3' base of the oligonucleotide to the DNA, and subsequently would not allow priming to occur from this oligonucleotide (Fig. 8.1b).

8.3 Purification and characterisation of ICL from *S.coelicolor*

It is unknown why *S.coelicolor* cannot be grown successfully on acetate as a sole carbon source. It may be that an active transport system is required, or depolarisation of the membrane may be occuring as the acetate is transported. However, once a suitable carbon source was found (Tween), which caused induction of ICL, it was possible to effect purification of ICL. The purification procedure employed steps which had been previously used to purify other ICL's.

The protein had a molecular weight similar to the other purified prokaryotic ICL's and displayed some similar characteristics. In particular, the poor instability of the enzyme posed a problem during purification. The enzyme also showed an absolute requirement for magnesium and the apparent requirement of a cysteine group for its activity.

8.4 Cloning and sequencing of the *icl* gene from *S.coelicolor*

Oligonucleotides were designed against the amino acid sequence obtained from sequencing of the N-terminal regions of the ICL enzyme and chymotryptic-digested peptides. These oligonucleotides were used to amplify an internal region of the *S.coelicolor icl*, which was subsequently used as a probe, to clone the entire gene.

The entire *S.coelicolor icl* gene and regions upstream and downstream have been sequenced. A translation of the *icl* gene has been shown to be similar to the other sequenced *icl's* and in particular the *E.coli icl*. Both the *E.coli* and *S.coelicolor* ICL's lack a central region, which is present in the eukaryotic ICL's. It is unknown at this time what purpose this region serves in the eukaryotic enzymes. It may contain targetting sequences for directing the protein to glyoxysomes, but this has yet to be determined.

Upstream from the *S.coelicolor icl* gene a number of potential promoter sequences have been identified. These have been observed by showing similarity to the consensus for the *E.coli*-like, *Streptomyces* promoters. It has still to be determined if any of these are in fact promoter sequences.

A region downstream from *icl* has been identified as containing a potential terminator structure. Further downstream the gene encoding *ms* has been identified. Approximately two thirds of this gene has been sequenced and when translated, shows a high level of similarity to the other sequenced MS's. No putative promoters were observed upstream from the *ms* gene.

There is the possibilty that the *icl* and *ms* genes form part of an operon, as in *E.coli.* This operon would however be clearly different from the *E.coli* operon as the gene order is different. In *E.coli*, *ms* is found upstream from *icl.*

8.5 Overexpression and attempted disruption of the *S.coelicolor* idh.

A 40-fold overexpression of the *S.coelicolor idh* was achieved in *S.coelicolor* by cloning the *idh* gene into the multicopy streptomycete vector, pIJ486.

Expression was effected from a promoter/s found upstream from *idh.*

Attempts to overexpress the *S.coelicolor idh* in *E.coli* have still to be completed.

An attempt to disrupt the *S.coelicolor idh* by replacing the genomic copy of *idh* with an altered copy containing an antibiotic resistance marker in the center of the gene has been tried. However no exconjugants which showed resistance to the antibiotic were obtained. It should be stressed that this was only carried out once and so does not reflect the possiblity of whether this disruption will work or not. Improvements to the technique have also been recently developed.

8.6 Future work

Now that ICL has been successfully purified and the gene cloned and sequenced, there are many further experiments which need to be carried out.

At the DNA level it is important to study the expression of this gene. S1 mapping experiments could be carried out to identify the promoter/s and the possibility of whether a terminator is found downstream of *icl*.

S1 studies will help to elucidate whether the *icl* and *ms* genes form part of an operon and also if the stem-loop structure affects the level of *ms* transcription.

Growth studies may be carried out to determine what the effects of different carbon sources and growth stages have on the expression of the *icl* and *ms* genes. Ultimately, it may be possible by disrupting genomic copies of the genes and having controllable copies on plasmid vectors, to study the effects of altering the levels of the enzymes at different stages of the life cycle. The goal of this would be to control the growth of the culture to produce greater levels of antibiotic.

Of course the first task would be to disrupt the genomic copies of these genes and also to clone copies into suitable controllable expression vectors. At the same time disruption of the *idh* gene should be carried out, so that studies involving the interaction of the IDH and ICL enzymes can be carried out.

It will also be necessary to clone the rest of the *ms* gene. Further sequencing should then be carried out to determine if there are any other genes downstream of *ms*. It would be of particular interest to determine if an *idh* kinase/phosphatase gene is present, or if a repressor gene can be found.

S.coelicolor is used as a model system, but if the level of antibiotic production can be increased, then it may be possible to use the *S.coelicolor icl, idh* and *ms* genes as probes, to clone the same genes from commercially-producing strains. These genes could then be controlled in the manner determined using *S.coelicolor* to try to effect an increase in antibiotic production in these strains.

Physiologically, it will be interesting to study the growth of *Streptomyces* on various fatty acids and acetate. Further work with acetate is required to determine the problems of growth on this carbon source.

Biochemically it may be possible to mutagenise residues thought to be involved in the activity of the enzyme. In particular the cysteine residue (amino acid 188 of the *S.coelicolor* ICL enzyme) which is conserved across all the proteins so far sequenced. This would help to elucidate the possible mechanism of action of the enzyme.

REFERENCES

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

Albertson, C., Haukanes, B-I., Aasland, R., and Kleppe, K. 1988. Optimal conditions for hybridisation with oligonucleotides: a study with myc-oncogene DNA probes. Anal. Biochem. **170**:193-202

Baronofsky, J.J., Schreurs, W.J.A., and Kashket, E.R. 1984. Uncoupling by acetic-acid limits growth of and acetogenesis by *Clostridium-thermoaceticum.* App and Env Micro **48**:1134-1139.

Beeching, J.R, and Northcote, D.H. 1987. Nucleic acid (cDNA) and amino acid sequences of isocitrate lyase from castor bean. Plant Mol. Biol. **8**:471-475.

Bennett, P.M, and Holms, W.H. 1975. Reversible inactivation of the isocitrate dehydrogenase of *Escherichia coli* ML308 during growth on acetate. J. Gen. Microbiol. **87**:37-51.

Berdy, J. 1980. Recent advances in and prospects of antibiotic research. Process Biochem. **15**:28-35.

Berg, P. 1956. Acyl adenylates: an enzymic mechanism of acetate activation. J.Biol.Chem. **222**:991-1013.

Bibb, M.J., and Cohen, S.N. 1982. Gene expression in *Streptomyces* construction and application of promoter probe plasmid vectors in *Streptomyces lividans.* Mol. Gen. Genet. **187**, p265-277.

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

Binstock, J.F., Pramanik, A., and Schultz, H. 1977. Isolation of a multienzyme complex of fatty acid oxidation from *Escherichia coli* Proc. Natl. Acad. Sci. USA **74**:492-495.

Birnboim, H.C., and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513-1523.

Biro, S., and Chater, K.F. 1987. Cloning of *Streptomyces-griseus* and *Streptomyces-lividans* genes for glycerol dissimilation. Gene **56:**79-86.

Black, P.N., Said, B., Ghosn, C.R., Beach, J.V., and Nunn, W.D. 1987. Purification and characterisation of an outer membrane protein involved in fatty acid transport in *Escherichia coli.* . J. Biol. Chem. **262**:1412-1419.

Borthwick, A.C., Holms, W.H., and Nimmo, H.G. 1984. The phosphorylation of *Escherichia-coli* isocitrate dehydrogenase in intact-cells. Biochem J. **222**:797-804.

Borthwick, A.C., Holms, W.H., and Nimmo, H.G. 1984. Amino-acid sequence round the site of phosphorylation in isocitrate dehydrogenase from *Escherichia-coli* ML308. FEBS Letters **174**:112-115.

Borthwick, A.C., Holms, W.H., and Nimmo, H.G. 1984. Isolation of active and inactive forms of isocitrate dehydrogenase from *Escherichia coli* ML308. Eur. J. Biochem. **141**:393-400.

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

Brandel, V., and Trifonov, E.N. 1984. A computer programme for testing potential prokaryotic terminators. Nucleic Acids Res. **12**:4411-4427.

Brice, C.B., and Kornberg, H.L. 1968. Genetic control of isocitrate lyase activity in *Escherichia coli*.. J. Bacteriol. **96**:2185-2186.

Byrne, C., Stokes, H.W., and Ward, K.A. 1988. Nucleotide sequence of the *aceB* gene encoding malate synthaseA in *E.coli*. Nucleic Acids Res. **16**:9342.

Campbell, J.J.R., Smith, R.A., and Eagles, B.A. 1983. A deviation from the conventional tricarboxylic acid cycle in *Pseudomonas aeruginosa*. Biochim. Biophys. Acta **11**:594.

Celenza, J.L., and Carlson, M. 1986. A yeast gene that is essential for release from glucose repression encodes a protein-kinase. Science **233**:1175-1180.

Chater, K.F., and Merrick, M.J. 1979. *Streptomyces* In: Developmental Biology of Prokaryotes, ed J.H. Parish pp93-114 Oxford: Blackwell.

Chell, R.M., Sundaram, T.K., and Wilkinson, A.E. 1978. Isolation and characterisation of isocitrate lyase from a thermophilic *Bacillus* sp. Biochem. J. **173**:165-177.

Chin, A.M., Feldheim, D.A., and Saier, M.H.Jr. 1989. Altered transcriptional patterns affecting several metabolic pathways in strains of *Salmonella typhimurium* which overexpress the fructose regulon. J. Bacteriol. **171**:2424-2434.

Chung, T., Klumpp, D.J., and LaPorte, D.C. 1988. Glyoxylate bypass operon of *Escherichia coli* : cloning and determination of the functional map. J. Bacteriol. **170**:386-392.

Chung, Tw.W., Resnik, E., Stueland, C., and Laporte, D.C. 1993. Relative expression of the products of glyoxylate bypass operon - contributions of transcription and translation. J. Bacteriol. **175**:4572-4575.

Cioni, M., Pinzauti, G., and Vanni, P. 1981. Comparative biochemistry of the glyoxylate cycle. Comp.Biochem Biophys. B. **70**:1-26.

Cleveland, D.W., Fischer, S.G., Kirschner. M.W., and Laemmli, U.K. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J. Biol. Chem. **252**:1102-1106.

Conai, L., Dietrich, R.A., Maslyar, D.J., Baden, C.S., and Harda, J.J. 1989. Coordinate expression of the transcriptionally regulated isocitrate lyase and malate synthase in *Brassica napus* L. Plant Cell **1**:293-300.

Cortay, J.C., Bleicher, F., Rieul, C., Reeves, H.C., and Cozzone, A.J. 1988. Nucleotide-sequence and expression of the *aceK* gene coding for isocitrate dehydrogenase kinase phosphatase in *Escherichia-coli.* J. Bacteriol. **170**:89-97.

Cortay, J.C., Bleicher, F., Duclos, B., Cenatiempo, Y., Gautier, C., Prato, J.L. and Cozzone, A.J. 1989. Utilisation of acetate in *Escherichia coli*: structural organisation and differential expression of the *ace* operon. Biochimie **71**:1043-1049.

Cortay, J.C., Negre, D., Galinier, A., Duclos, B., Perriere, G., and Cozzone, A.J. 1991. Regulation of the acetate operon in *Escherichia-coli* - purification and functional-characterization of the IclR Repressor. Embo Journal. **10**:675-679.

Dean, A.M., and Koshland, D.E. Jr. 1990. Electrostatic and steric contributions to regulation at the active site of isocitrate dehydrogenase. Science **249**:1044-1046.

Dekleva, M.L., and Strohl, W.R. 1988. Biosynthesis of ε-rhodomycinone from glucose by *Streptomyces* C5 and comparison with intermediary metabolism of other polyketide-producing streptomycetes. Can. J. Microbiol. **34**:1235-1240.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids res. **12**:387-395.

Dirusso, C.C., and Nunn, W.D. 1985. Cloning and characterization of a gene (*fadR*) involved in regulation of fatty-acid metabolism in *Escherichia-coli.* J. Bacteriol. **161**:583-588.

Ebbighausen, H., Weil, B., and Kramer, R. 1991. Carrier-mediated acetate uptake in *Corynebacterium glutamicum*.. Archives of Microbiology. **155**:505-510.

Ebright, R. 1986. Proposed amino acid-base pair contacts for 13 sequence-specific DNA binding proteins. p.207-219 *In* D.L. Oxender (ed.), Protein structure, folding and design. Alan, R. Liss. Inc. New York.

Elmansi, E.M.T., Nimmo, H.G., and Holms, W.H. 1985.The role of isocitrate in control of the phosphorylation of isocitrate dehydrogenase in *Escherichia-coli* ML308. FEBS letters. **183:**251-255.

El-Mansi, E.M.T., Mackintosh, C., Duncan, K., Holms, W.H. and Nimmo, H.G. 1987. Molecular cloning and over-expression of the glyoxylate by pass operon from *Escherichia coli* ML308. Biochem. J. **242**:661-665.

Feng, D-F., and Doolittle, R.F. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Molecular Evolution. **25**:351-360.

Fornwald, J.A., Schmidt, F.J., Adams, C.W., Rosenberg, M., and Brawner, M.E. 1987. Two promoters, one inducible and one constitutive, control transcription of the *Streptomyces lividans* galactose operon. Proc. Natl. Acad. Sci. USA **84**:2130-2134.

Frerman, F.E. 1973. The role of acetyl-coenzyme A in the transferase uptake of butyrate by isolated membrane vesicles of *Escherichisacoli.* Arch. Biochem. Biophys. **159**:444-452.

Gainey, L.D.S., Connerton, I.F., Lewis, E.H., Turner, G., and Ballance, D.J. 1992. Characterisation of the glyoxysomal isocitrate nlyase genes of *Aspergillus nidulans* (*acuD*) and *Neurospora crassa (acu3).* Curr. Genet. **21**:43-47.

Galassi, C., Giachetti, E., Pinzauti, G., and Vanni, P. 1988. Ptroteolytic activity in pure preparartions of isocitrate lyase from *Pinus pinea.* Phytochemistry **27**:1303-1307.

Galinier, A., Bleicher, F., Negre, D., Perriere, G., Duclos, B., Cozzone, A.J., and Cortay, J.C. 1991. Primary structure of the intergenic region between *ace*K and *icl*R in the *Escherichia-coli* chromosome. Gene. **97**:149-150.

Garnak, M. and Reeves, H.C. 1979. Phosphorylation of isocitrate dehydrogenase of *Escherichia coli.* Science **203**:1111-1112.

Garnak, M. and Reeves, H.C. 1979. Purification and properties of phosphorylated isocitrate dehydrogenase of *Escherichia coli.* J. Biol. Chem. **254** (16):7915-7920.

Gemmrich, A.R. 1979. Isocitrate lyase in germinating spores of the fern *Anemia phyllitidis.* Phytochemistry **18**:1143-1146.

Giachetti, E., Pinzauti, G., Bonaccorsi, R., and Vanni, P. 1988. Isocitrate lyase from *Pinus-pinea* - characterization of its true substrate and the action of magnesium-ions European J. Biochem. **172**:85-91.

Gibson, T.J. 1984. Studies on the Epstein-Barr virus genome. PhD Thesis, Cambridge, England.

Gribskov, M., Devereux, J., and Burgess, R.R. 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**:539-549.

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

Harley, D.K., and McClure, W.R. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237-2255.

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

Hodgson, D.A. 1982. Glucose repression of carbon source uptake and metabolism in *Streptomyces coelicolor* A(3)2 and its perturbation in mutants resistant to 2-deoxyglucose. J. Gen. Microbiol. **128**:2417-2430.

Holms, W.H. 1986. The central metabolic pathways of *Escherichia-coli* - relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate. Current Topics in Cellular Regulation. **28**:69-105.

Holms, W.H. 1987. Control of flux through the citric acid cycle and the glyoxylate bypass in *Escherichia coli.* Biochem. Soc. Symp. **54**:17-31.

Holms, W.H., and Bennett, P.M. 1971. Regulation of isocitrate dehydrogenase activity in *Escherichia coli* on adaptation to acetate. J. Gen. Microbiol. **65**:57-68.

Hopwood, D.A., and Wright, H.M. 1983. CDA is a new chromosomally determined antibiotic from *S.coelicolor* A3(2). J. Gen. Microbiol. **129**:3575-3579.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate.D.J., Smith, C.P., Ward, J.M., and Schrempf, H. 1985. Genetic manipulation of *Streptomyces* a laboratory manual. The John Innes Foundation, Norwich, ngland.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Janssen, G.R., Malpartida, F., and Smith, C.P. 1986. Regulation of gene expression in antibiotic producing *Streptomyces*, p.251-276. *In* I. Booth and C. Higgins (ed.) Regulation of gene expression. Cambridge University Press. Cambridge.

Hopwood, D.A., and Kieser, T. 1990.The *Streptomyces* genome.In: "The Bacterial Chromosome." Drlica, K. and Riley, M. (ed). American Society for Microbiology, Washington DC. pp. 147-162.

Hoyt, J.C., Robertson, E.F., Berlyn, K.A., and Reeves, H.C. 1988. *Escherichia coli* isocitrate lyase: properties and comparison. Biochim. Biophys. Acta. **966**:30-36.

Hoyt, J.C., and Reeves, H.C. 1992. Phosphorylation of *acinetobacter* isocitrate lyase. Biochemical and Biophysical Research Communications. **182**:367-371.

Hunter, I.S., and Baumberg, S. 1989. Molecular genetics of antibiotic formation. In S.Baumberg, I.S.Hunter and P.M.Rhodes (eds.) SGM Symposium 44. Microbial products: New approaches. Cambridge University Press. p122-162.

Hurley, J.H., Dean, A.M., Sohl, J.L., Koshland, D.E., and Stroud, R.M. 1990. Regulation of an enzyme by phosphorylation at the active-site. Science. **249**:1012-1016.

Hurley, J.H., Dean, A.M., Thorsness, P.E., Koshland, D.E.Jr. and Stroud, R.M. 1990. Regulation of isocitrate dehydrogenase by phosphorylation involves no long-range conformational change in the free enzyme. J. Biol. Chem. **265** (7):3599-3602.

Ikeda, T., and LaPorte, D.C. 1991. Isocitrate dehydrogenase kinas/phosphatase: *ace*K alleles that express kinase, but not phosphatase activity. J. Bacteriol. **173**:1801-1806.

Jammel, S., El-Gut, T., McFadden, B.A. 1984. Isolation and properties of watermelon isocitrate lyase. Phytochemistry **23**:2753-2759.

Janssen, G., Ward, J.M., and Bibb, M.J. 1989. Unusual transcriptional and translational features of the aminoglycoside phosphotransferase gene (*aph*) from *Streptomyces fradiae*. Genes aqnd Development. 3:415-429.

Jaurin, B., and Cohen, S.N. 1985. Streptomyces contain *Escherichia coli*-type A+T rich promoters having novel structural features. Gene **39**:191.

Jenkins, L.S., and Nunn, W.D. 1987. Regulation of the *ato* operon by the *atoC* gene in *Escherichia-coli*. J. Bacteriol. **169**:2096-2102.

Jenkins, L.S., and Nunn, W.D. 1987. Genetic and molecular characterization of the genes involved in short-chain fatty-acid degradation in *Escherichia-coli* - the *ato* system. J. Bacteriol. **169**:42-52.

Johanson, R.A., Hill, J.M., and McFadden, B.A. 1974. Isocitrate lyase from *Neurospora crassa* I. Purification, kinetic mechanism and interaction with inhibitors. Biochim. biophys. Acta **364**:341-352.

John, P.C.L., and Syrett, P.J. 1967. The purification and properties of isocitrate lyase from *Chlorella.* Biochem. J. **105**:409-416.

Kahn, F.R., and McFadden, B.A. 1982. Isocitrate lyase from flax. Terminal residues, composition, active site and catalysis. Plant physiol. **70**:943-948.

Kameda, K., and Nunn, W.D. 1981. Purification and characterisation of acyl conenzyme A synthetase from *Escherichia coli.* J. Biol. Chem. **256**:5702-5707.

Kell, D.B., Peck, M.W., Rodger, G., Morris, J.G. 1981. On the permeability to weak acids and bases of the cytoplasmic membrane of *Clostridium-pasteurianum.* Biochem. Biophys. Res. Comm. **99**:81-88.

Klumpp, D.J., Plank, D.W., Bowdin, L.J., Stueland, C.S., Chung, T., and Laporte, D.C. 1988. Nucleotide-sequence of *aceK*, the gene encoding isocitrate dehydrogenase kinase phosphatase. J. Bacteriol. **170**:2763-2769.

Kornberg, H.L. 1966. The role and control of the glyoxylate cycle in *Escherichia coli.* Biochem. J. **99**:1-11.

Kovalic, D., Kwak, J.H., and Weisblum, B. 1991. General method for direct cloning of DNA fragments generated by the polymerase chain reaction. Nucleic Acids Res. **19**:4560.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680-685.

Levine, S.M., Ardeshir, F., and Ames, G.F. 1980. Isolation and characterisation of acetate mutants of *Escherichia coli* and *Salmonella typhimurium.* J. Bacteriol. **143**:1081-1085.

Laporte, D.C., and Chung, T. 1985. A single gene codes for the kinase and phosphatase which regulate isocitrate dehydrogenase. J. Biol. Chem. **260**:15291-152971

LaPorte, D.C., and Koshland, D.E.Jr. 1982. A protein with kinase and phosphatase activities invoved in the regulation of tricarboxylic acid cycle. Nature **300**:458-460.

Leskiw, B.K., Bibb, M.J., and Chater, K.F. (1991a). The use of a rare codon specifically during development. Mol. Microbiol.

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

Mackintosh, C., and Nimmo, H.G. 1988. Purification and regulatory properties of isocitrate lyase from *E.coli* ML308. Biochem. J. **250**:25-31.

Maizel and Lenk 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. (USA) **78** (12):7665-7669.

Malhorta, O.P., and Srivastava, P.K. 1982. Isolation and characterisation of isocitrate lyase of castor endosperm. Arch. Biochem. Biophys. **214**:164-171.

Mallinder, P.R., and Moir, A. 1991. Mutants of *Bacillus-stearothermophilus* defective in the uptake and metabolism of acetate. J. Gen. Microbio. **137**:779-785.

Mahorta, O.P., and Singh, J. 1992. Role of Histidine-residues and mechanism of action of isocitrate lyase of castor seadling endosperm. Plant Science **81**:155-162.

Malhorta, O.P., Singh, J., and Srivatava, P.K. 1991. Significance of sulphydryl-groups in the activity of isocitrate lyase of castor endosperm. Plant Science **75**:47-53.

Maloy, S.R., Bohlander, M., and Nunn, W.D. 1980. Elevated levels of glyoxylate shunt enzymes in *Escherichia coli* strains constitutive for fatty acid degradation. J. Bacteriol. **143**:720-725.

Maloy, S.R, Ginsburgh, C.L, Simons, R.W., and Nunn, W.D. 1981. Transport of long and medium chain fatty-acids by *Escherichia-coli*-K1. J. Biol. Chem. **256**:3735-3742.

Maloy, S.R., and Nunn, W.D. 1981. Role of gene *fadR* in *Escherichia-coli* acetate metabolism. J.Bacteriol. **148**:83-90.

Maloy, S.R., and Nunn, W.D. 1982. Genetic-regulation of the glyoxylate shunt in *Escherichia-coli* K-12. J. Bacteriol. **149**:173-180.

Martin, J.F., and Demain, A.L. 1980. Control of antibiotic biosynthesis. Microbiol. Reviews **44**:230

Matsuoka, M., and McFadden, B.A. 1988. Isolation, hyperexpression, and sequencing of the *aceA* gene encoding isocitrate lyase in *Escherichia-coli*. J. Bacteriol. **170**:4528-4536.

Mattern, S.D., Brawner, M.E., and Westpheling, J. 1993. Identification of a complex operator for *galP1*, the glucose sensitive, galactose-dependent promoter of the *streptomyces* galactose operon. J. Bacteriol. **175**:1213-1220.

Mazodier, P., Petter, R., and Thomson, C. 1989. Intergeneric conjugation between *E.coli* and *Streptomyces* species. J. Bacteriol. **171**:3583-3585.

McCarthy, J.E.G., and Gualerzi, C. 1990. Translational control of prokaryotic gene expression. Trends in Genetics **6**:78-85.

McFadden, B.A., Rao, G.R., Cohen, A.L., and Roche, T.E. 1968. Isocitrate lyase from *Pseudomonas indigofera*. V. Subunits and terminal residues and the relation to catalytic activity. Biochemistry **7**:3574-3582.

Miller, J.R., and Ingolia, T.D. 1989a. Cloning β-lactam genes for streptomyces spp. and fungi. In "Genetics and molecular biology of industrial microorganisms". Edited by Herschberger, C., Queener, S., Hegermann, G.; Based on the proceedings of the fourth ASM conference on the genetics and molecular biology of industrial microorganisms, Bloomington, 1988.

Miller, J.R., and Ingolia, T.D. 1989b. Cloning and characterisation of β-lactam biosynthetic genes. Molecular Microbiol. **3**:689-695.

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

Moran, C.P., Lang, N., Lehrice, S.F.C., Lec, G., Stephens, M., Sonenshein, A.L., Pero, J., and Losick, R. 1982. Nucleotide-sequences that signal the initiation of transcription and translation in *Bacillus subtilis* Mol. Gen. Genet. **186**:339-346.

Negre, D., Cortay, J.C., Old, I.G., Galinier, A., Richaud, C., Saintgirons, I., and Cozzone, A.J. 1991. Overproduction and characterization of the *iclR* gene-product of *Escherichia-coli*-K-12 and comparison with that of *Salmonella-typhimurium*-LT2. Gene. **97**:29-37.

Negre, D., Cortay, J.C., Galinier, A., Sauve, P., Cozzone, A.J. 1992. Specific interactions between *iclR* repressor of the acetate operon of *E.coli* and its operator. J. Mol. Biol. **228**:23-29.

Nikaido, H., and Vaara, M. 1985. Molecular basis of bacterial outer membrane permeability. Microbiol. reviews. **49**:1-32.

Nimmo, G.A., Borthwick, A.C., Holms, W.H., and Nimmo, H.G. 1984. Partial purification and properties of isocitrate dehydrogenase kinase/phosphatase from *Escherichia coli* ML308. Eur. J. Biochem. **141**:401-408.

Nimmo, H.G., Borthwick, A.C., El-Mansi, E.M.T., Holms, W.H., Mackintosh, C., and Nimmo, G.A. 1987. Regulation of enzymes at the branchpoint between the citric acid cycle and the glyoxylate bypass in *Escherichia coli.* Biochem. Soc. Symp. **54**:93-101.

Nimmo, H.G., Douglas, F., Kleanthos, C., Campbell, D.G., and Mackintosh, C. 1989. Identification of a cysteine residue at the active site of *Escherichia coli* isocitrate lyase. Biochem. J. **261**:431-435.

Nunn, W.D. 1986. A molecular view of fatty acid catabolism in *Escherichia coli.* Microbiol. Reviews **50**:179-192.

Olson, J.A. 1954. The D-isocitric lyase system: the formation of glyoxylic and succinic acids from D-isocitric acid. Nature **174**:695-696.

Overath, P., Pauli.G., and Schairer,H.U. 1969. Fatty acid degradation in*Escherichia coli.* An inducible acyl-CoA synthetase, the mapping of *old*-mutations, and the isolation of regulatory mutants, and the isolation of regulatory mutants. Eur. J. Biochem. **7**:559-574.

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

Pawar, S., and Schultz, H. 1981. The structure of the multienzyme complex of fatty acid oxidation from *Escherichia coli.* J. Biol. Chem. **256**:3894-3899.

Pinzauti, G., Giachetti, E., and Vanni, P. 1982. Isocitrate lyase of conifers (*Pinus pinea*). Int. J. Biochem. **14**:267-275.

Pinzauti, G., Giachetti, E., and Vanni, P. 1983. Isocitrate lyase - artifacts and multiple enzyme forms. Arch. Biochem. and Biophys. **225**:137-142.

Pramanik, A., Pawar, S., Antonian, E., and Schultz, H. 1979. Five different enzymic activities are associated with the multienzyme complex of fatty acid oxidation from *Escherichia coli*.. J. Bacteriol. **137**:469-473.

Ramseier, T.M., Negre, D., Cortay, J.C., Scarabel, M., Cozzone, A.J., and Saier, M.H. 1993 *In-vitro* binding of the pleiotropic transcriptional regulatory protein, FruR to the *fru, pps, ace, pts* and *icd* operons of *Escherichia-coli* and*Salmonella-typhimurium*.. J. Mol. Biol. **234**:28-44.

Ramseier, T.M., and Saier, M.H. (1993) Defining a FruR binding consensus sequence based on 6 operator sites in *Salmonella-typhimurium* and *Escherichia-coli.* Faseb Journal. **7**:1215.

Reiss, U., and Rothstein, M. 1974. Isocitrate lyase from free living nematode *Turbatrix aceti*: purification and properties. Boichemistry **13**:1796-1800.

Rieul, C., Bleicher, F., Duclos, B., Cortay, J.C., and Cozzone, A.J. 1988. Nucleotide sequence of the *aceA* gene coding for isocitrate lyase in *Escherichia coli.* Nucleic Acids Res. **16**:5689-5690.

Roberts, L.M., and Lord, J.M. 1981. Synthesis and posttranslational segregation of glyoxysomal isocitrate lyase from castor bean endosperm. Eur. J. Biochem. **119**:43-49.

Robertson, E.F., Hoyt, J.C., and Reeves, H.C. 1987. Evidence of histidine phosphorylation in isocitrate lyase from *Escherichia-coli.* Federation Proceedings. **46**:1952.

Robertson, E.F., Hoyt, J.C., Reeves, H.C. 1987. Invitro phosphorylation of *Escherichia-coli* isocitrate lyase. Current Microbiology **15**:103-105

Robertson, E.F., and Reeves, H.C. 1987. Purification and characterization of isocitrate lyase from *Escherichia-coli.* Current Microbiology. **14**:347-350.

Rogers, J.E., and McFadden, B.A. 1976. Isocitrate lyase from *Pseudomonas indigofera*: pH dependence of catalysis and binding of substrates and inhibitors. Arch. Biochem. Biophys. **174**:695-704.

Ruchti, M., and Widmer, F. 1984. Isocitrate lyase in germinating soybean. Experientia **40**:612.

Ruchti, M., and Widmer, F. 1986. Isocitrate lyase in germinating soybean cotyledons: effects of metabolites *in vitro* J. Plant physiol. **123**:507-511.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning, 2nd edition. Cold Spring Harbor Press.

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

Saz, H.J. 1954. The enzymic formation of glyoxylate and succinate from tricarboxylic acids. Biochem. J. **58:**20.

Schumperli, D, Mckenney, K., Sobieski, D.A., and Rosenberg, M. 1982. Translational coupling at an intercistronic boundary of the *Escherichia-coli* galactose operon. **30**:865-871.

Seno, E.T., and Baltz, R.H. 1989. Structural organization and regulation of antibiotic biosynthesis and resistance genes in actinomycetes. In: "Regulation of Secondary Metabolism in Actinomycetes". Shapiro, S. (ed.). CRC Press. pp 1-48.

Seno, E.T., and Chater, K.F. 1983. Glycerol catabolic enzymes and their regulation in wild-type and mutant strains of *Streptomyces coelciolor* A(3)2. J. Gen. Microbiol. **129**:1403-1413.

Seno, E.T., Bruton,C.J., and Chater, K.F. 1984. The glycerol utilisation operon of *Streptomyces coelicolor*: genetic mapping of *gyl* mutations and the analysis of cloned *gyl* DNA. Mol. Gen. Genet. **193**:119-218.

Shapiro, J.A., and Adhya, S.L. 1969. The galactose operon of *E.coli* K-12 II. A deletion analysis of operon structure and polarity. Genetics **62**:249-264.

Sheu, C.W., Konings, W.N., and Freese, E. 1972. Effects of acetate and other short-chain amino-acids on sugar and amino acids uptake of *Bacillus subtilis* J. Bacteriol. **111**:525-530.

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

Simons, R.W., Egan, P.A., Chute, H.T., and Nunn, W.D. 1980. Regulation of fatty acid degradation in *Escherichia coli*: isolation and characterisation of strains bearing insertion and temperature-sensitive mutations in gene *fadR*. J. Bacteriol. **142:**621-632.

Smith, C.P., and Chater, K.F. 1988. Structure and regulation of controlling sequences for the *Streptomyces-coelicolor* glycerol operon. J. Mol. Biol. **204**:569-580.

Smith, C.P., and Chater, K.F. 1988. Cloning and transcription analysis of the entire glycerol utilization (*gylabx*) operon of *Streptomyces-coelicolor* A3(2) and identification of a closely associated transcription unit. Mol. Gen. Genet. **211**:129-137.

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

Spratt, S.K., Black, P.N., Ragozzino, M.M., and Nunn, W.D. (1984). Cloning, mapping, and expression of genes involved in the fatty acid degradative multienzyme complex of *Escherichia-coli*. J. Bac.. **158**, p535-542.

Stadman, E.R., and Barker, H.A. 1950. Fatty acid synthesis by enzyme preparations of *Clostidium kluyveri*. J. Biol. Chem. **184**:769-793.

Strohl, W.R. 1992. Compilation and analysis of DNA-sequences associated with apparent streptomycete promoters. Nucleic Acids Res. **20**:961-974.

Sunnarborg, A., Klumpp, D., Chung, T., and Laporte, D.C. 1990. Regulation of the glyoxylate bypass operon - cloning and characterization of *iclR*. J. Bacteriol. **172**:2642-2649.

Taylor, R.D. 1992. Purification and characterisation of the isocitrate dehydrogenase from *Streptomyces coelicolor* and cloning of its gene. Ph.D. Thesis. University of Glasgow.

Thorsness, P.E. and Koshland, D.E. Jr. 1987. Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. J. Biol. Chem. **262** (22):10422-10425.

Tinocco, I.Jr., Borer, P.N., Dengler, B., Levin, M.D., Uhlenbeck, O.C., Crothers, D.M., Gralla, J. 1973. Improved estimation of secondary structure in ribonucleic acids. Nature new biology **246**:40-41.

Tsukamoto, C., Ejiri, S., and Katsumatra,T. 1986. Purification and some properties of isocitrate lyase from the pollen of *Pinus densiflora* Sieb et Zucc. Agric. Biol. Chem. **50**:127-134.

Turley, R.S., Choe, S.M. and Trelease, R.N. (1990). Characterisation of a cDNA clone encoding the complete amino acid sequence of cotton isocitrate lyase. Biocim. Biophys. Acta. **1049**:223-226.

Uchida, M., Ueda, M., Matsuki, T., Okada, H., Tanaka, A., and Fukui, S. 1986. Properties of isocitrate lyase from an alkane utilisable yeast, *Candida tropicalis* Agric. Biol. Chem. **50**:127-134.

Vanderwinkel, E., and DeVliegher, M. 1968. Physiologie et genetique de L'isocitrase et des malate synthases chez *Escherichia coli*. Eur. J. Biochem. **5**:81-90.

Vanni, P., Giachetti, E., Pinzauti, G. and McFadden, B.A. 1990. Comparative structure, function and regulation of isocitrate lyase, an important assimilatory enzyme. Comp. Boichem. Physiol. **95B**:431-458.

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

Westpheling, J. and Brawner, M. 1989. Two transcribing activities are invoved in expression of the *streptomyces* galactose operon. J. Bacteriol. **171:**1355-1361.

White, P.J., Young, J., Hunter, I.S., Nimmo, H.G. and Coggins, J.R. 1990. The purification and characterisation of 3-dehydroquinase from *Streptomyces coelicolor.* Biochem. J. **265**:735-738.

Woese, C.R. 1987. Bacterial Evolution. Microbiological Reviews. **51:**221-271.

Wright, F., and Bibb, M.J. 1992. Codon usage in the G+C rich *Streptomyces* genome. Gene **113**:55-65