Purification and Characterisation of the Isocitrate Dehydrogenase From *Streptomyces coelicolor* and Cloning of It s Gene.

by

Richard David Taylor.

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

Departments of Genetics and Biochemistry.

June 1992.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

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

Richard David Taylor, June 1992.

Dedicated to my Parents

and Grandparents,

with thanks for their

love and support.

CONTENTS.

CHAPTER 2: Materials and methods.

CHAPTER 4: Cloning of the *icd* B gene that encodes IDH in *S.coelicolor.*

CHAPTER 5: DNA Sequencing and analysis.

CHAPTER 7: General discussion.

List of figures.

Chapter 1.

Chapter 2.

Chapter 4.

Chapter 5.

Chapter 6.

Abbreviations.

AMP.	ampicillin
ATP	adenosine triphosphate
BSA	bovine serum albumin
(k)Da	(kilo) daltons
DMSO	dimethylsulphoxide
DNAse	deoxyribonuclease
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)
IMDH	3-Isopropylmalate dehydrogenase (EC 4.2.1.33)
IPTG	isopropyl-b-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 phospate
NADPH	reduced nicotinamide adenine dinucleotide phospate
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-b-galactoside

Acknowledgements.

Many thanks to my supervisors, Hugh Nimmo and Iain Hunter, for wading through this thesis and to the lab for putting up with periodic bouts of moans, groans and despondency and for providing dancing partners for the odd Ceilidh.

A special thank-you to Marion Stone for keeping the lab in order and Alison (Wild thing of Borneo) Wylie for the preparation of *S.coelicolor* DNA that actually cut. The ladies on the third floor require a special mention for putting up with the steamy sessions with the autoclave room and the ensuing foul smells.

Thanks also to Graeme Walker and Kenny McDowell for their excellent materials and methods section, which made my task so much easier and to Linda for typing all those awkward references.

My gratitude (and apologies) to the entire 5th floor of Genetics and to John Coggins and his group for allowing me to hog the Apple Macs for hours on end.

Thanks to my mate, John Anders for his friendship and navigation skills and for keeping me supplied with the water of life and also to Lesley for keeping me fed!

Finally I would like to extend a special thank you to the "Foreigners" (especially the Great Dane) for keeping me entertained, and providing numerous distractions from the writing of this thesis!

This work was supported by the SERC (antibiotics initiative) with contributions from SKBeecham, Glaxo, ICI and Celltech.

Summary.

Summary.

As part of an effort to increase our understanding of central metabolic pathways in streptomycetes, the TCA cycle enzyme, isocitrate dehydrogenase (IDH), was studied from *Streptomyces coelicolor*. In *Eschericia coli* grown on acetate as a sole carbon source, IDH plays an important role in controlling the flux of carbon between the glyoxylate bypass and the TCA cycle. The study of any such control in streptomycetes is of interest because these pathways are important in providing energy and precursors for both primary and secondary metabolism.

A purification procedure was developed that enabled the isolation of the IDH from *S.coelicolor*. This enzyme was found to be a monomeric protein with a molecular weight of around 80,000, quite different to the IDH from *E.coli* . Based on this information and on the immunological properties and N-terminal sequence of the enzyme, it was proposed that the *S.coelicolor* IDH could be grouped with the IDH-II from *Vibrio spp.* ABE-1, into a separate class of enzymes that were named the type II IDHs. Recent characterisation of the IDH from *Rhodomicrobium vannielii* indicates that these properties may be common to all type II IDHs.

Determination of the N-terminal sequence of IDH enabled the design and construction of an oligonucleotide that was used as a probe specific for the gene encoding IDH (*icd* B). It was used to screen two types of *S.coelicolor* libraries. One was a plasmid sub-library constructed from genomic DNA that had been pre-selected to contain a high proportion of fragments that contained the *icd* B gene. The other was a λ library (constructed in λ-GEM11) that represented sequences from the entire *S.coelicolor* genome. Screening of the plasmid library allowed the isolation of a recombinant plasmid (called pRT60) that contained a 5kb insert. Partial sequencing and restriction mapping of this clone showed that only the first 800bp of *icd* B had been isolated on this fragment. The rest of the DNA fragment contained DNA upstream of *icd* B. Screening of the λ library produced a clone (λ1) that contained the whole of the *icd* B gene (*c*.2.5 kb) plus *c*.6.4 kb of the region downstream of *icd* B plus *c*.2 kb of the region upstream of the gene.

A 5kb *Pst* I fragment was subcloned from λ1 into pBluescript II SK$^+$ to produce pRTB1. The insert is likely to contain all of the up and downstream signals required for expression in *S. coelicolor*, as well as the *icd* B coding region. Further subcloning enabled the gene to be isolated on a 2.7 kb *E.co* RV/*Xma* I fragment (pRTB 2 and 3) that contained very little up and downstream sequences. None of these plasmids were able to complement a mutant of *E.coli* that was devoid of IDH activity. This was probably due to lack of expression as many streptomycete promoters are not recognised by the *E. coli*

DNA-dependent RNA polymerase. Analysis of the DNA sequence upstream of *icd* B revealed no obvious *E.coli* -like promoters supporting this hypothesis.

The whole of the *icd* B gene (plus another 197 bp upstream and 137 bp downstream) was sequenced using a series of oligonucleotides. Analysis of the predicted protein sequence showed that the gene encoded a protein with an N-terminal sequence identical to that obtained by Edman degradation of the purified protein. This information, coupled with the similarity of the sequence to the N-terminal sequence of the IDH from *R. vannielii,* proves that the gene encoding IDH has been cloned. The codon bias and G+C content of *icd* B is typical of those found for other streptomycete genes. Analysis of the codon bias and G+C content also suggested that there is a possible open reading frame 103 bp upstream and 66 bp downstream of *icd* B.

The IDH from *E.coli* is known to have an amino acid sequence similar to that of the isopropylmalate dehydrogenases (IMDHs) from various organisms. Comparison of several IMDHs and the IDH from *E.coli,* with the IDH from *S.coelicolor,* showed that the streptomycete enzyme is significantly different. However closer inspection of the sequences revealed a small region of similarity that is present in all of the IMDH and IDH sequences available, including the NAD dependent IDH from yeast. This region coincides with the region known to bind NADP in the IDH from *E.coli,* suggesting that both NAD and NADP binding sites in IDH and IMDH type enzymes are related. A novel consensus sequence is presented that describes some of the conserved residues found between the aligned sequences at the proposed NADP binding site.

During this sequence analysis it was noticed that the all of the IDH and IMDH sequences, with the exception of the *S.coelicolor* IDH sequence, show significant similarity along their whole length. This is despite the fact that the different enzymes differ in subunit organisation and coenzyme specificity. It has therefore been proposed that all of the IMDH/IDH type enzymes (except the IDH from *S. coelicolor*) have subunits that are evolutionarily related and that some of these subunits have developed specialist functions. Sequence alignments of the IDH from *S. coelicolor* with other IDHs and IMDHs may indicate a distant relationship between this type II enzyme and the IMDH/IDH enzymes, although the alignments could be coincidental.

CHAPTER 1.

General Introduction.

1.0 Introduction.

The aim of this chapter is to provide a background to some of the subjects discussed in this thesis. Some relevant aspects of streptomycetes are discussed first (section 1.1). This is followed by a discussion of the role and control of IDHs from various organisms.

During the course of the work described in this thesis, a putative $NAD(P)^+$ binding site was discovered in the IDH sequence from *S. coelicolor*. Similarities to other enzyme sequences were identified and used to create a novel $NAD(P)^+$ consensus sequence. Section 1.3 is therefore devoted to a dicussion of some of the information known about $NAD(P)^+$ binding sites in various enzymes.

The final section in this chapter describes the general aims of this study and the strategies employed to achieve them.

1.1 An introduction to streptomycetes.

This section aims to provide the reader with an insight into some of the more unusual features of the genus. It is not meant to be a comprehensive review of the information known about streptomycetes but is meant to provide information relevant to the work carried out during this study.

1.1.1 Some general aspects of streptomycetes.

Streptomyces belongs to the group of bacteria called the actinomycetes. These are a group of filamentous soil bacteria that obtain nutrients by the degradation of organic material. To do this, they excrete a wide variety of enzymes such as cellulases, xylanases, chitinases and lipases (Tomich, 1988a). They also provide an important source of other enzymes of commercial interest such as restriction endonucleases and amylases.

Starvation causes the substrate mycelia to branch further (see figure 1.1) and produce aerial hyphae. These eventually produce chains of spores, with the concomitant lysis of the substrate mycelia usually accompanied by the production of antibiotics (Chater, 1984).

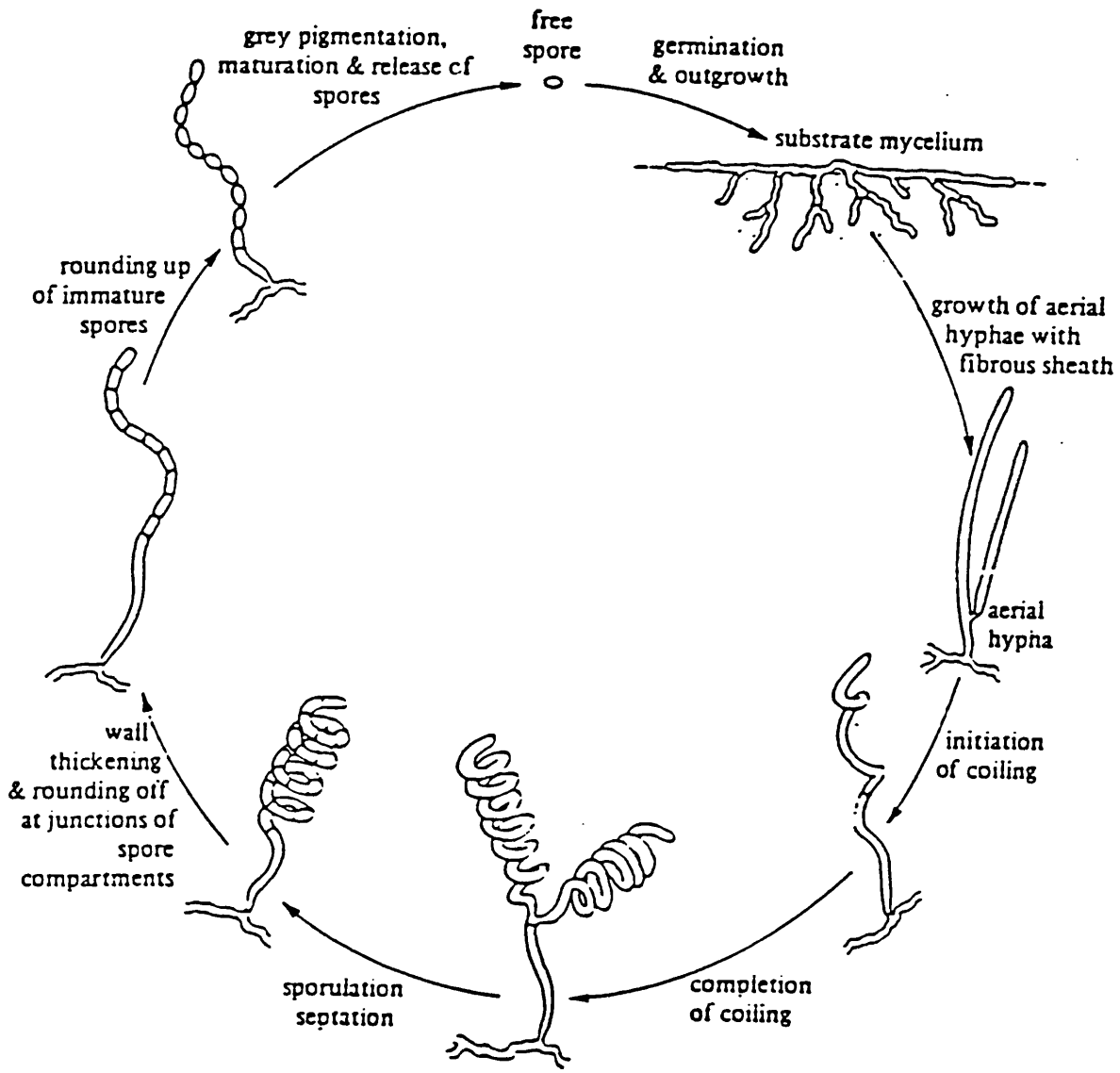


Figure 1.1 Life cycle of streptomycetes.

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

This ability to undergo both biochemical and morphological changes has stimulated interest in streptomycetes as a model for the study of differentiation. Perhaps the most extensive research on streptomycetes though is due to their ability to produce a wide range of antibiotics, accounting for over 70% of the naturally occurring antibiotics discovered so far (Tomich, 1988a).

The advances that have been made in the molecular biology of streptomycetes (Hopwood *et al.*, 1985a and Tomich, 1988a and b) have allowed analysis of the biology of these organisms at the genetic level. Again, much of the attention has been directed toward the genes that control the production of antibiotic synthesis and resistance. For many antibiotics all of the genes required for production and resistance are clustered together on the chromosome, making it comparatively easy to clone whole biosynthetic pathways within a single plasmid construct, by selecting for fragments of DNA that confer antibiotic resistance to the host. For instance Malpartida and Hopwood (1984) cloned the whole cluster of genes that allow for the production of the polyketide antibiotic, actinorhodin, into a naive host which then produced the antibiotic. Introduction of only part of the cluster into hosts (*eg. S.violaceoruber*) that normally produce other polyketides led to to the production of the first genetically engineered hybrid antibiotics (Hopwood *et al.*, 1985b).

S. coelicolor is the best studied streptomycete producing several antibiotics including actinorhodin, undecylprodigiosin and methylenomycin. Actinorhodin is an isochromanequinone antibiotic that is produced by a polyketide type pathway (see figure 1.2). Polyketide synthesis shows certain analogies to the synthesis of long chain fatty acids which are synthesised by the sequential addition of two-carbon units. The first molecule in fatty acid biosynthesis (the starter unit) is always acetyl-CoA whilst the units used to extend the chain are always malonyl-CoA (for unbranched fatty acids). The condensation of the extender units onto the growing chain results in the loss of carbon dioxide resulting in the net incorporation of a two carbon acetyl group. Different polyketides can however use different starter and extender units, sometimes using a mixture, allowing a greater variety of structures (Sherman *et al.*, 1988).

Another major difference between the two types of synthesis is that the keto groups of fatty acids are always fully reduced after the addition of each extender unit. In polyketide synthesis these are sometimes only partially reduced or not reduced at all, dependent on the polyketide being made. These modifications may occur immediately after addition of the extender unit or may be left until chain growth is complete (Sherman *et al.*, 1988).

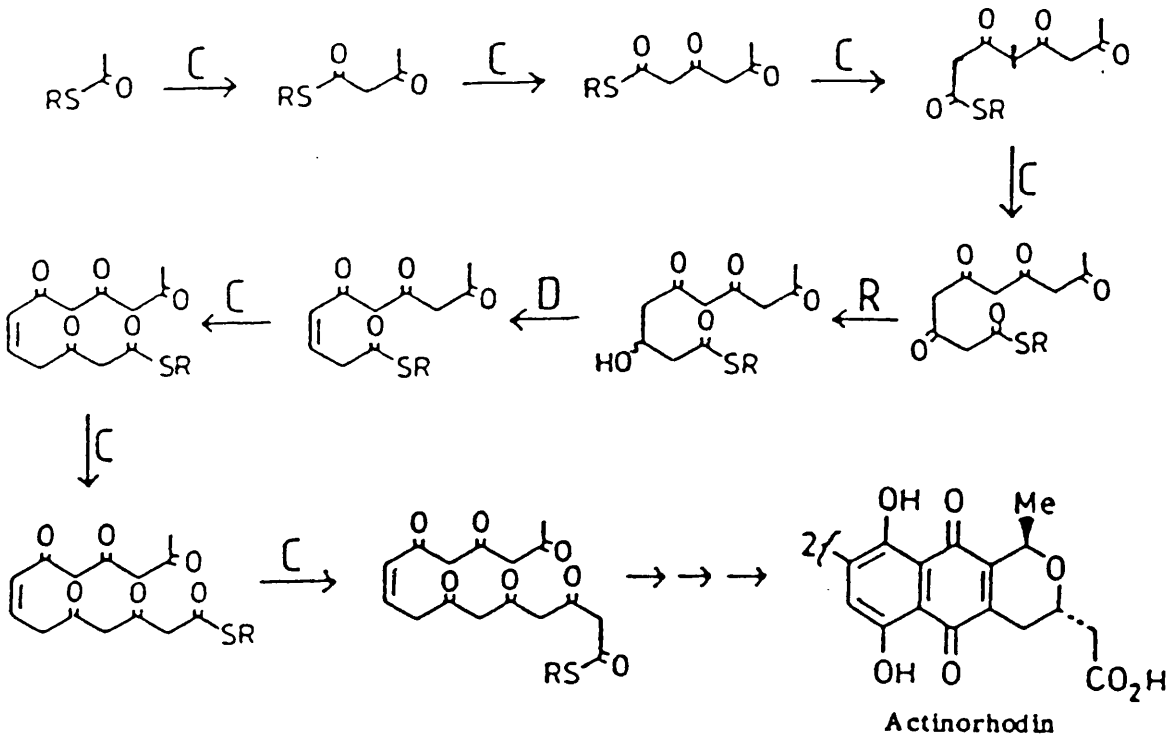


Figure 1.2 Proposed pathway for the assembly of actinorhodin.

This diagram shows the sequential condensation of acetate units (indicated by the letter as C) with a single reduction and dehydrogenation (indicated by the letters R and D respectively). From Sherman *et al.*, 1988.

The synthesis of actinorhodin (figure 1.2) provides a good example of polyketide synthesis with each individual acetate unit (derived from malonyl-CoA) being added sequentially to the growing chain with a total of only one reduction during the synthesis. Two of these cyclised "mature chains" are joined together to give the final actinorhodin molecule.

1.1.2 Why study IDH in *S. coelicolor* ?

As described above streptomycetes have become the object of extensive research. Understandably, the commercial use of these organisms to produce antibiotics has encouraged research into the systems which control and produce antibiotics. However little is known about the pathways and systems that produce the basic precursors of antibiotic synthesis which will be provided by primary metabolic pathways and may become a limiting step in the overproduction of antibiotics.

As described above, polyketides such as actinorhodin are synthesised from the head-to-tail condensation of acetate units which are derived from acetyl-CoA. The TCA cycle also consumes acetyl-CoA and may therefore compete with antibiotic synthesis for this common precursor. Consequently, it is of interest to see if the TCA cycle is regulated in some way to control this flux. Indeed, the same factor (A-factor) that stimulates antibiotic synthesis (streptomycin) in *S.griseus* also causes the indirect inhibition of IDH and glycolytic enzymes (see section 1.1.4).

Control over the TCA cycle is likely to be even more strict when one considers that many industrial fermentations are carried out on media based on oils, such as oil seed rape. These oils are likely to metabolised directly to acetyl-CoA for entry into the TCA cycle and will therefore require the induction of the glyoxylate bypass as an anaplerotic pathway. By analogy to *E. coli* , the competition for isocitrate is likely to be controlled *via* the control of IDH, adding further complexity to the control of the cycle both before and during antibiotic synthesis. This enzyme was therefore an obvious target for study.

S. coelicolor was chosen for these studies because it is genetically the best characterised streptomycete allowing for a variety of classical and recombinant DNA techniques (Hopwood *et al.*, 1985a) to be used in the study of this enzyme.

1.1.3 RNA polymerase and promoter heterogeneity.

For many genes, from many organisms, promoters can be identified as recognisable sequence motifs that conform to a consensus. The sequences are required for the RNA polymerase (RNApol) to bind to DNA at the correct position to initiate transcription. In

E. coli these consensus sequences lie 10 and 35 nucleotides upstream of the 5' end of the mRNA and have therefore been called the -10 and -35 sequences respectively (Hawley and McClure, 1983).

Initial studies on the promoters from streptomycetes used promoter probe vectors to show the presence of promoter activity. Only a few randomly cloned DNA fragments from streptomycete DNA could direct transcription (*ie.* show promoter activity) in *E. coli* whereas many of the DNA fragments from *E. coli* could direct transcription in streptomycetes (Bibb and Cohen, 1982). This suggested that streptomycetes have the ability to transcribe genes with *E. coli* -like promoters, but that most of streptomycete promoters are significantly different and cannot be transcribed by *E. coli.*

This situation is reminiscent of the situation in *B.subtilis* where genes involved in different stages of the life-cycle possess different classes of promoters. These are recognised by specific RNA-polymerase (RNApol) holoenzymes whose specificity is imposed by the incorporation of one of a number of different sigma (σ) factors (Losick and Pero, 1981). Varying levels of the different σ factors at different times of the life cycle therefore allows preferential transcription of different sets of genes at appropriate points in the life-cycle. For instance four of the eight σ factors produced by *B.subtilis* recognise promoters for sporulation specific genes and are produced only at the onset of sporulation (Errington, 1991, cited in Strohl 1992).

Westpheling *et al.* (1985) were the first to show directly that there was more than one RNApol holoenzyme in *S. coelicolor.* They found one RNApol that was associated with a 35 kDa σ factor (σ^{35}) and could transcribe the *E. coli* -like *veg* promoters of *B.subtilis.* The second RNApol had either a σ^{37} or σ^{49} associated with it and was found to direct transcription from the non-*E. coli*-like *ctc* promoter from *B.subtilis*, which is involved in sporulation and motility. The specificity of certain holoenzymes for certain types of streptomycete promoters was shown by Buttner *et al.* (1988). This group found that the *dag*A gene from *S. coelicolor* possessed four different promoters and identified three different holoenzymes that would transcribe three of them. This suggested that there are at least four σ factors in *S. coelicolor* that can transcribe this one gene. Two of these σ factors probably correspond to the σ^{35} and σ^{49} identified by Westpheling *et al.*, (1985). The third was a novel 38 kDa σ factor (σ^{38}).

Tanaka *et al.* (1988) have cloned four σ factor genes called *hrd*A, B, C and D, all of which show remarkable sequence similarity to the *veg* σ factor gene from *B.subtilis* and the σ^{70} gene of *E. coli* (the "housekeeping" σ factors). The close similarity suggests that these factors are different to the σ^{49} and σ^{37} factors isolated by Wespheling *et al.* (which recognise *ctc* type promoters) although one of these genes could encode σ^{35} which

recognises the *veg* promoters. Disruption of the genomic copies of the genes *hrd*C and D had no apparent effect on the viability of the organism and no obvious phenotype (Buttner *et al.*, 1990). However attempts to disrupt the *hrd* B mutants were unsuccessful indicating that this mutation is lethal and that the σ factor encoded by this gene is therefore an absolute requirement for growth.

This shows that there are at least six different factors present in *S. coelicolor* (*hrd*A, B, C and D and the *ctc* transcribing factors σ^{37} and σ^{49}; factor σ^{28} and σ^{35} could possibly be encoded by one of the *hrd* genes). A seventh σ factor called σ^{whiG} has also been identified. Mutants that lack a functional σ^{whiG} factor are unable to produce mature pigmented spores which results in the colonies having a white colour (hence the mutation was called *whi* te) and were unable to produce antibiotics. This sigma factor is therefore unnecessary for vegetative growth but is required for the transcription of genes involved in both morphological and biochemical differentiation. Cloning and sequencing of a piece of DNA that complements this mutation has shown that the gene resembles the motility factor (σ^{28}) of *B.subtilis* (Chater *et al.*, 1989).

Unfortunately little is known about which σ factors recognise which promoters in streptomycetes. Attempts to classify the different promoters that have been sequenced by similarity to each other have only really succeeded in highlighting the diversity of promoter sequences present in streptomycetes (Seno and Baltz, 1989). No obvious distinctions were found between those promoters from "housekeeping" and primary metabolic genes and those involved in differentiation and secondary metabolism.

A more recent review by Strohl (1992) was able to take advantage of the fact that 139 promoters have now been sequenced from 87 genes. From this study it was shown that only 29 of these could be classified as having "*E. coli* -like promoters" that might be recognised by the *E. coli* housekeeping sigma factor, σ^{70}. This classification required that the promoter showed similarity to the consensus sequences TTGACA and TATAAT (for the -35 and -10 regions respectively) and that the two sequences were between 16 and 18 nucleotides apart. Some of these promoters were in fact known to be active in *E. coli* demonstrating the functional significance of this sequence classification. The rest of the promoters studied showed a wide diversity of sequences. Some had *E. coli* like -10 sequences or -35 sequences but not both, whilst others had both consensus sequences but a distance between the two of greater than the 16 and 18 nucleotides. Other streptomycete promoters show no obvious sequence similarity to the *E. coli* -like consensus at all. Some promoters show similarities to each other but the functional significance of this is not yet known. It was suggested that some of these sequences could be involved in providing recognition sites for positive or negative regulators as well

as providing a recognition site for the RNApol. In fact some of these promoters could form secondary structures which could be involved in this sort of regulation.

Another unusual feature of transcription initiation in streptomycetes is the widespread use of multiple promoters. Out of the 87 genes studied by Strohl (1992), 27 had more than one promoter. Multiple promoters presumably allow transcription of genes by more than one RNApol like the situation for *dag* A (see above). Where multiple promoters were found, 13 cases were identified in which a single region of DNA contains overlapping divergent promoters. This must form part of some complex regulatory system which allows for the coordinated control of more than one gene from the same DNA region.

The main conclusion that can be drawn from these studies is that the organisation of streptomycete promoters is more complex than in other prokaryotes and that only further study will reveal the working of these mechanisms.

1.1.4 Control of metabolism by A-factor like molecules.

Another unusual feature of streptomycetes is the control of antibiotic biosynthesis by a microbial "pheromone-type" molecule called A-factor (activating factor). In *S.griseus* the *afs*A$^+$ gene is required for A-factor synthesis. Mutations in this gene render the strain unable to sporulate or to produce the antibiotic streptomycin. However, addition of purified A-factor to the media in which the mutants were grown restored these functions, showing that A-factor is diffusible and indicated the "pheromone-like" action of the factor (Kokhlov, 1973).

It has been reported that *S.griseus* possesses a protein that has the ability to bind A-factor (Miyake *et al.*, 1989). In strains incapable of synthesising A-factor, secondary mutations were isolated that restored antibiotic synthesis and sporulation despite the lack of A-factor (Miyake *et al.*, 1990). These secondary mutations were found to lack a functional A-factor binding protein leading to the conclusion that the binding protein acts as a repressor and that A-factor binds to the protein to prevent repression (*ie.* acts as an anti-repressor).

A-factor mutants of *S.griseus* were also found to show abnormally high levels of the NADP$^+$ dependent enzymes, IDH, glucose 6-phosphate dehydrogenase (G6PDH) and 6-phosphogluconate dehydrogenase (6PGDH) (Grafe *et al.*, 1980). Normally the amount of enzyme activity for these enzymes during antibiotic synthesis was barely detectable indicating that the inhibition of these enzymes and stimulation of antibiotic synthesis are in some way linked. The inhibition of IDH and G6PDH was shown to be due to the

action of the potent inhibitor adenosine di-phosphoribose phosphate, which was shown to be produced from $NADP^+$ by the action of the enzyme $NADP^+$ glycohydrolase (NGH). The levels of NGH in wild-type *S.griseus* were found to increase during the onset of antibiotic synthesis but were very low in *afs*A mutants. It therefore seems that IDH and the glycolytic enzymes are controlled by A-factor *via* the action of NGH; antibiotic genes are stimulated directly by A-factor. NGH has been shown not to play a role in the stimulation of antibiotic synthesis because *afs*A mutants that lack the A-factor binding protein lack NGH but still produce streptomycin. Grafe *et al.* (1980) also reported the results of some *in vitro* inhibition studies of IDH by adenosine di-phosphoribose phosphate, although the relevance is somewhat doubtful as they were performed solely on IDH purified from pig heart.

Other streptomycete species including *S. coelicolor* have also been found to produce A-factor type molecules. The mode of action of A-factor in *S. coelicolor* however seems to be different to the situation in *S.griseus*. In *S. coelicolor*, *afs*A mutants were found to sporulate and produce antibiotics as normal. They are also different in requiring the presence of a second gene, called *afs*B, for the expression of *afs*A (reviewed by Horinouchi and Beppu, 1990).

S. coelicolor has also been found to require a third gene, called *afs*R, for the expression of A-factor, acting as a postive regulator (Hong *et al.*, 1991). The action of this gene product was found to be stimulated when phosphorylated by a protein kinase. More recent work by Horinouchi's group (ISBA meeting, 1991) has shown that this protein kinase (encoded by *afs*K) is a membrane protein that is controlled by auto-phosphorylation.

The actual role of A-factor in controlling secondary metabolism in *S. coelicolor* is unclear and there are no reports on the effects of A-factor on primary metabolism in this strain. It should therefore be borne in mind that A-factor has been shown to have an effect on primary metabolic genes in *S.griseus* and may play a similar role in *S. coelicolor*.

1.2 Role of TCA cycle in *E. coli* grown on glucose and acetate.

The function of the tricarboxylic acid (TCA) cycle is to produce energy and biosynthetic precursors and is therefore an example of an amphibolic pathway (figure 1.3). Most fuel molecules such as fatty acids, amino acids and carbohydrates will eventually be converted to acetyl-CoA which then condenses with OAA to form citic acid in the first reaction of the cycle. Complete oxidation of an acetyl group to CO_2 and H_2O produces three molecules of NADPH and one molecule of FADH as well as one molecule of ATP. Reduced nucleotides are used to provide the cell with reducing equivalents for use in biosynthesis as well as for producing more ATP via the electron transport chain. However as well as losing carbon from the cycle as CO_2, carbon is also lost as the intermediates, oxaloacetate (OAA), 2-oxoglutarate (2OG) and succinyl-CoA (Succ-CoA) are withdrawn from the cycle for the synthesis of other cell constituents (figure 1.3). If acetyl groups are not replaced at some point in the cycle the TCA cycle will stop due to the loss of intermediates.

During growth on glucose, oxaloacetate (OAA) is replaced by the anaplerotic enzyme phosphoenolpyruvate carboxylase (PEPc) which directly produces OAA by the carboxylation of PEP. Growth on acetate however requires a different pathway, called the glyoxylate bypass, as acetate is converted directly to acetyl-CoA for metabolism by the TCA cycle, by-passing the production of PEP and pyruvate completely (Kornberg, 1966).

As the name suggests the glyoxylate bypass by-passes those reactions of the TCA cycle that release CO_2, by converting one molecule of acetyl-CoA plus one molecule of isocitrate to one molecule of of succinate and malate (see figure 1.3). This requires the presence of the two enzymes isocitrate lyase (ICL) and malate synthase (MS), which catalyse the first and second reactions of the pathway respectively (see figure 1.3).

1.2.1 Control of IDH activity by IDH kinase/phosphatase in *E. coli*.

Under conditions in which both the glyoxylate bypass and TCA cycles are functioning, there is competition between ICL and IDH for available isocitrate. The flux between these pathways therefore needs to be coordinated to ensure the optimum ratio of flux through each pathway. There is no evidence for allosteric control of either of these enzymes though raising the question of how the two enzymes are controlled. Work performed by several groups indicated that the control was mediated *via* the regulation of IDH activity and is discussed below. The mechanism by which IDH activity can control flux is discussed in section 1.2.2.

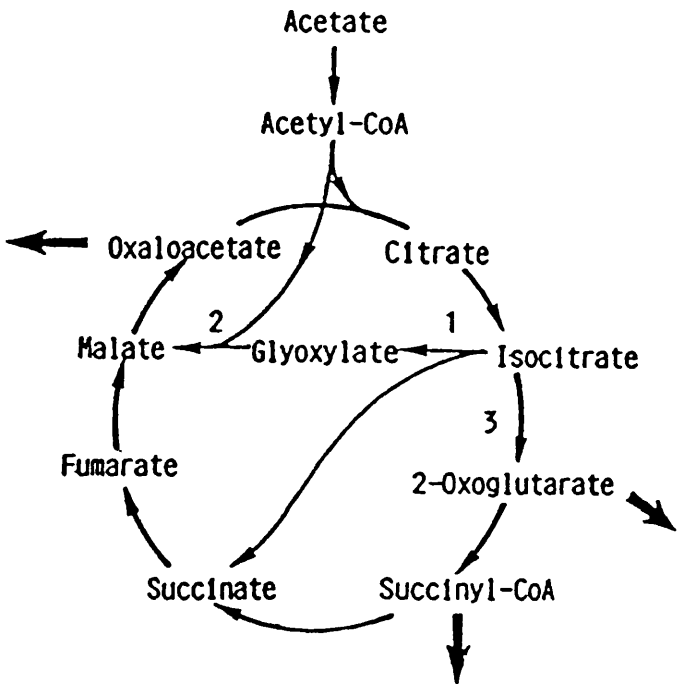


Figure 1.3 The TCA cycle and glyoxylate bypass.

The TCA cycle is represented as a cycle of reactions. The glyoxylate bypass (reactions 1 and 2) is drawn with arrows passing through the middle of the TCA cycle. Large arrows indicate flux to biosynthesis.

1) Isocitrate lyase (ICL).
2) Malate synthase (MS).
3) Isocitrate dehydrogenase (IDH).

Holms and Bennett (1971) found that *E. coli* grown on limiting amounts of glucose exhibited diauxic growth. As cells metabolised glucose, acetate was excreted into the medium. The acetate was utilised by the cells once the preferred carbon source (in this case glucose) had been exhausted. Concomitant with the change from growth on glucose to acetate, the total level of IDH activity was found to decrease. Once all of the acetate was exhausted IDH activity increased again showing that the inactivation of IDH was fully reversible.

When *E. coli* was grown on acetate as a sole carbon source, specific activity of IDH was found to be only 30% of the value obtained during growth on glucose or pyruvate (Bennett and Holms, 1975). However, addition of pyruvate to such cultures increased IDH activity to levels similar to those found in glucose-grown cells. Once pyruvate was exhausted, the IDH activity returned to the original lower activity.

Dialysis of IDH and mixing experiments showed that this effect was not mediated by the action of low molecular weight effectors. Addition of chloramphenicol to cultures grown on acetate, just before the addition of pyruvate, did not prevent the increase in IDH activity either, showing that *de novo* protein synthesis was not responsible for this increase. It was therefore suggested that a reversible covalent modification of the enzyme might be responsible (Bennett and Holms, 1975).

Garnak and Reeves (1979a and b) showed that during the period in which IDH activity decreased, IDH became phosphorylated at a serine residue. LaPorte and Koshland (1982) demonstrated that both the phosphorylation an de-phosphorylation activities were catalysed by a bifunctional protein which was called the isocitrate dehydrogenase kinase/phosphatase (IKP). The phosphorylation catalysed by the IKP caused the complete inactivation of IDH. This is unusual as most other phosphorlation mechanisms alter the affinity of an enzyme for substrate (see section 1.2.3 and Barford, 1991). The IKP protein was found to be a homodimer meaning that the two activities must in fact be present on a single polypeptide. This was confirmed by the cloning of a single gene (*ace* K) that was able to produce IKP activity when expressed (LaPorte *et al.*, 1985). However, it is still unclear if these activities are catalysed by the same or distinct active sites (Ikeda *et al.*, 1991).

Purification of the IKP allowed further studies of the properties of this enzyme *in vitro* (Nimmo *et al.*, 1984). Both the kinase and phosphatase activities of the IKP were found to be active simultaneously but subject to allosteric regulation by a number of effectors. Such effectors will stimulate phosphatase activity and inactivate kinase activity when the cell requires high IDH activity.

Purification and analysis of both the inactive and active forms of IDH from cells grown on acetate showed that the only difference between the two forms was the presence of a phosphorylated serine on the inactive IDH. Subsequent proteolysis of the phosphorylated, inactive, enzyme allowed Borthwick *et al.* (1984c) to identify a peptide which contained the phosphorylated serine. Comparison of this to the entire gene sequence, later published by Thorsness and Koshland (1987), showed that this serine corresponds to SER-113.

1.2.2 How do changes in IDH activity control the flux between the TCA cycle and glyoxylate bypass ?

It has been mentioned in section 1.2.1 that IDH activity in cells grown on acetate is lower than cells grown on glucose and that this control is mediated by the bifunctional kinase/phosphatase enzyme IKP. However the role of this activation/inactivation cycle has not yet been addressed. Holms and Bennett first proposed that the inactivation of IDH during growth of *E. coli* on acetate might allow ICL to compete more effectively against IDH for available isocitrate. This suggestion was supported by the following data.

Although different values have been reported for the K_m of ICL for isocitrate, it is clear that the K_m is higher than the K_m of IDH for isocitrate (Nimmo *et al.*, 1987). A consequence of this is that ICL cannot compete with IDH for isocitrate when intracellular concentrations of the substrate are low (Nimmo *et al.*, 1984). This situation is found in cells grown on glucose, where the isocitrate concentration is below the levels of detection (ElMansi *et al.*, 1985). During growth on acetate, IDH is partially inactivated, reducing the flux of isocitrate through the TCA cycle and leading to a build up of isocitrate to an intracellular concentration of 570 μM (ElMansi *et al.*, 1985 and Holms, 1987).

It was described in section 1.2.1 that IDH activity is controlled by a phosphorylation/dephosphorylation mechanism in which both activities are catalysed by IKP. The role of this system is essentially to maintain the isocitrate concentration at a value that will sustain the correct balance of flux between the glyoxylate bypass and TCA cycle. To control this system, the IKP is sensitive to various metabolites that can either stimulate phosphatase activity or inhibit kinase activity, depending on the requirement of the cells for precursors and energy. For instance, increasing levels of the TCA cycle intermediates, PEP, pyruvate, 3-phosphoglycerate, OAA and 2-oxoglutarate signal a reduced requirement for the glyoxylate bypass. To reduce the flux of isocitrate through the glyoxylate bypass, activity of IDH is increased, *via* stimulation of the phosphatase activity and inhibition of the kinase activity. Isocitrate has a similar affect on the kinase

and phosphatase activity and could be the most important effector, as a low concentration of this metabolite directly prevents flux through the bypass.

AMP and ADP also stimulate phosphatase and inhibit kinase activity. These metabolites probably indicate reduced cellular energy levels, a situation where flux through the TCA cycle would be favoured (Nimmo *et al.*, 1987).

1.2.3 Mechanism by which IDH is inactivated.

Phosphorylation usually affects enzyme activity by altering the affinity for substrate or by modifying the response to effectors. Glycogen phosphorylase (GP) provides a good example of an enzyme controlled in this sort of way. In its de-phosphorylated form, GP is dependent on AMP for activity. Phosphorylation of a serine residue on the enzyme induces a large conformational change, which relieves the enzyme from AMP dependence. This contrasts with the situation found for the IDH from *E. coli* where phosphorylation (at SER-113) causes the almost complete inactivation of the enzyme. The differences between the roles of these two types of phosphorylation are discussed in a review by Barford (1991).

To investigate how the phosphorylation of SER-113 causes the inactivation of the *E. coli* IDH, mutants were created in which the SER-113 residue was changed by site-directed mutagenesis to amino acids with different properties. Substitution of SER-113 by the negatively charged amino acid aspartate (S113D), which has the same charge as phosphorylated serine but not the same shape, caused virtually complete inactivation of the enzyme (Thorsness and Koshland, 1987). Substitution of the positively charged amino acid, lysine (S113K), or introduction of the bulky amino acid tyrosine (S113Y) only affected enzyme activity by a comparatively small amount (Dean *et al.*, 1989). These experiments therefore suggested that it is the negative charge of the phosphorylated serine that is the main factor contributing to the inactivation of IDH, whereas steric interference contributes little to the inactivation.

Crystal structures were determined for both the phosphorylated and native IDH in order to determine if phosphorylation induces conformational changes in IDH that may contribute to inactivation of the enzyme (Hurley *et al.*, 1990a). The only conformational changes observed were minimal and localised to around the site of phosphorylation. Comparison of the mutant enzymes that had aspartate or glutamate residues replacing SER-113 showed that this small change plays no part in inactivation as the inactive mutants had conformations that were more similar to the native IDH than to the phosphorylated form.

It has therefore been shown that most of the inactivation was caused by the electrostatic effect of having a negative charge at position 113. No mention has yet been made of how this charge causes the inactivation. Before this crystal structure was available, Nimmo and colleagues suggested that this charge acted by preventing the binding of $NADP^+$ to the phosphorylated enzyme. Fluorometric titration experiments indicated that NADPH would bind to active, native enzyme but not to the phosphorylated form, whilst phosphorylation of the active enzyme seemed to protect it from limited proteolysis in a similar way to the protection afforded by $NADP^+$ bound to the enzyme (Garland and Nimmo, 1984). Inactive enzyme was also found to be incapable of binding to Procion Red columns, unlike the native form which would bind and could then be eluted with $NADP^+$ (Borthwick *et al.*, 1984 b). Yet further evidence was presented by McKee and Nimmo (1989) when they found that inactivation of the enzyme by phenylglyoxal could be prevented either by phosphorylation of the enzyme or by the addition of $NADP^+$.

However, equilibrium dialysis experiments performed by Dean *et al.* (1989) indicated that it was isocitrate that could not bind to the phosphorylated enzyme and that $NADP^+$ could bind equally well to both the phosphorylated and non-phosphorylated forms. Similar experiments were also performed with the mutant IDH in which the phosphorylatable serine had been changed to an aspartate residue causing virtually complete inactivation of the enzyme. Again, the enzyme was found to bind $NADP^+$ but not isocitrate.

Once it had been established that enzyme activity was prevented by the exclusion of the substrate, experiments were performed to determine how phosphorylation of the enzyme caused such a dramatic reduction in enzyme activity (Dean and Koshland, 1990). The amino acid in IDH that is phosphorylated by the IKP (SER-113) was replaced with amino acids with different properties and the effect on IDH activity measured. The introduction of negatively charged amino acids (*eg.* S113D as mentioned above), showed that the negative charge caused a 10^5 fold increase in the K_m for isocitrate, whereas steric interference (as shown by S113Y) caused a 10^4 fold increase. Loss of a hydrogen bond also caused a 2 fold decrease in the K_ms (shown by S113A).

Measurements of K_ms were also made for the native and mutant IDHs using the substrate analogue, 2R-malate. Although this analogue lacks the γ-carboxyl group of isocitrate it does act as a poor substrate for the IDH. The K_ms of the native IDH, the phosphorylated IDH and each of the mutant enzymes were found to be very similar. Since the analogue differs from isocitrate only in the fact that it does not possess a γ-carboxyl group, it was concluded that the phosphorylation of IDH prevented isocitrate binding by repulsion of the isocitrate at this group. When the crystal structure of the IDH-

isocitrate complex was determined it was shown that the SER-113 does indeed make contact with the carboxyl group of isocitrate (Hurley *et al.* 1990b).

1.2.4 IDHs from some other sources.

It was first noticed by Fukanaga *et al.* (1988) that most bacterial $NADP^+$ dependent IDHs can be classified into one of two groups, based on their quaternary structure. Most of the enzymes studied so far belong to the class which is typified by the enzyme from *E. coli* and the IDH-I from *Vibrio spp.* ABE-1. They are homodimers with a subunit size of c.45 kDa. The second class of enzymes are all monomers of molecular weights around 80,000 which are typified by the IDH from *Azotobacter vinelandii* and the IDH-II from *Vibrio spp.* These these two classes of IDHs have been called type I and type II respectively, after the two enzymes found in *Vibrio spp.* ABE-1.

Although most bacteria have been found to have either one or the other of these enzyme types (see figure 1.4 for summary), *Vibrio spp.* was found to contain one of each, whilst *A.calcoaceticus* was found to contain a type I enzyme and another, as yet unclassified IDH with a native molecular weight of 300 kDa (Self and Weitzman, 1972).

1.2.4.1 Type I IDHs.

Based on the subunit organisation alone (homodimers of *c.*45kDa subunits), most bacterial IDHs studied to date can be classified as type I IDHs. Fukanaga *et al.* (1988) showed that they could also be classified according to their cross-reaction with antibodies. Antibodies raised against the IDH-I from *Vibrio spp.* were found to cross-react with the *E. coli* and *Salmonella typhimurium* IDHs, both of which are dimeric enzymes. Moreover the antibody raised against the type II enzymes did not cross-react with these type I IDHs, showing that the monomeric IDHs are substantially different from the type-I enzymes.

Fukanaga *et al.* (1988) also showed that the N-terminal sequence of the *Vibrio spp.* IDH-I enzyme was similar to the N-terminal sequence from *E. coli* . It is now becoming clear that this is a general property of the type I IDHs since it was reported that the N-termini of the IDHs from *Synechocystis spp.* (Muro-Pastor and Florencio, 1992 and *Thermus thermophilus* (Eguchi *et al.*, 1988) were also similar to the *Vibrio spp.* IDH-I and *E. coli* enzymes. It has since been found that that the entire protein sequence of the *T. thermophilus* enzyme is highly homologous to the *E. coli* sequence (Miyazaki *et al.*, 1992) as is the sequence from another dimeric IDH from *Bacillus caldotenax* (C.Wild, personal communication) indicating that the group-I IDHs are evolutionarily related.

IDH Class	Organism	Native Mr (kda)	Reference.
I	*Vibrio spp.* IDH-I	88 (D)	Ochiai *et al.*, (1979)
I	*A.calcoaceticus* IDH-I	100 (D)	Self and Weitzman (1972)
I	*T.thermophilus*	115 (D)	Eguchi *et al.* (1988)
I	*P.variotii*	99 (D)	Takao *et al.*, (1986)
I	*R.spheorodes*	105 (D)	Chung and Braginski (1972)
I	*B.stearothermophilus*	92.5 (D)	Howard and Becker (1970)
I	*E. coli*	80 (D)	A.Borthwick (1984c)
I	*Synechocystis spp.*	112 (D)	Muro-Pastor and Florencio (1992)
II	*Vibrio spp.* IDH-II	81 (M)	Ochiai *et al.*, (1979)
II	*A.vinelandii*	78 (M)	Barrera and Jurtshuck (1979)
II	*Rhodomicrobium vannielii*	75-80 (M)	Leyland and Kelly (1991)
ND	*A.niger*	60 (M)	Meixner-Monori *et al.* (1986)
ND	*A.calcoaceticus* IDH-II	300 (M)	Self and Weitzman (1972)

Figure 1.4 Summary of IDHs from some other organisms.

The table above classifies several IDHs by virtue of subunit size and quaternary structure.

The letters M and D in brackets indicate whether the enzyme is a monomer or a dimer respectively.

ND indicates that the enzyme belongs to a class of IDH that has not yet been determined and may therefore be unique.

In common with the *E. coli* enzyme none of these enzymes seem to be allosterically controlled by any of the possible effectors such as AMP, ADP, ATP, $NAD(P)^+$, NAD(P)H or intermediates of central metabolism. Also like the *E. coli* enzyme they are inhibited by the concerted effect of glyoxylate and oxaloacetate, although the physiological significance of this is doubtful (Nimmo, 1986). It was reported that the enzyme from *Synechocystis spp.* was inhibited by ADP and 2-oxoglutarate but the high concentrations required again make this effect seem insignificant in relation to control of the enzyme *in vivo*.

1.2.4.2 Type II IDHs.

At the time that Fukanaga *et al.* (1988) had proposed the grouping together of the dimeric enzymes, there were only two examples of the 80 kDa monomeric IDHs. One was the IDH-II from *Vibrio spp.* and the other was from *A.vinelandii,* for which there is little further information available. The recent publication of a study of another monomeric form, from *Rhodomicrobium vannielii* (Leyland and Kelly, 1991), allows a more meaningful comparison of the monomeric IDHs to be made. It was reported that the N-terminal amino acid sequence of the *R.vannielii* enzyme was very similar to the N-terminal sequence of the IDH-II from *Vibrio spp.*, but dissimilar to the*Vibrio spp.* IDH-I, or any of the eukaryotic IDHs for which N-terminal data was available. Leyland and Kelly (1991) also reported that antisera raised against the *R.vannielii* IDH did not cross

react with the IDHs from *T. thermophilus* or *E. coli* either, further emphasising the differences between the type I and type II IDHs.

Unlike the IDHs from *Vibrio spp.* and other organisms, the *R.vannielii* enzyme has been reported to have both NAD^+ and NADP-linked IDH activity (Leyland and Kelly, 1991). The NAD-linked activity was reported to be inhibited by ATP. However the specificity (defined as k_{cat}/K_m) of the enzyme for NAD^+ is 400 times lower than for $NADP^+$, indicating that this activity and therefore the metabolic effect of ATP may be insignificant. It was also reported that the $NADP^+$ dependent IDH activity was inhibited by OAA plus glyoxylate in the concerted manner reported for all IDHs so far studied (Nimmo, 1986). However, the $NADP^+$ linked activity was also inhibited by OAA alone, a situation only previously reported for *P.variotti* which is a dimeric enzyme.

The classification of IDHs into type I and II does not take account of the properties of the IDH-II from *A.calcoaceticus*, which does not seem to belong to either class. The *A.calcoaceticus* enzyme not only has a native molecular weight (300 kDa) much larger than either of these groups, but it was also found to be stimulated by the addition of only 20 μM pyruvate or glyoxylate. It was also stimulated 5-fold by the addition of 500 μM

ATP, perhaps suggesting that each of these effectors regulates the IDH allosterically (Self and Weitzman, 1972; Self *et al.*, 1973; Kleber and Aurich, 1975).

1.2.4.3 Possible role of IDH isoenzymes.

The existence of more than one form of IDH in *Vibrio spp.* raises questions as to the function and control of each enzyme in cell metabolism. During growth on succinate it was found that the ratio of IDH-I activity to IDH-II was 1:5.6 (Ishii *et al.*,1987). However, during growth on acetate the level of IDH-I activity increased to 3 times its level on glucose, whilst the IDH-II activity remained virtually constant, altering the ratio of activities to 1:1.5. This change in activity was found to be due to increased levels of synthesis of IDH-I, as addition of chloramphenicol to the culture prevented this increase. This was further confirmed by monitoring the relative amounts of each enzyme, using antisera that had been raised against each.

It was assumed that this may somehow play a role in controlling the metabolism of isocitrate during growth on acetate. To elucidate the different roles that the enzymes may play, Fukanaga *et al.* (1988) isolated a mutant of *Vibrio spp.* (YG-83) that was completely devoid of any IDH-I activity yet had levels of IDH-II similar to the wild type strain (ABE-1). This strain grew at rates similar to the wild type on both acetate and succinate. The authors therefore concluded that IDH-I was not essential for growth on either substrate. However it is possible that this result is an artefact of the experiment since strain YG-83 was selected after two rounds of mutagenesis. The original mutagenesis produced an IDH-I mutant called YG-30 that was unable to grow on acetate without a cysteine supplement in the media. Its growth rate on glucose was also 10 fold lower than the growth rate of the wild-type strain. This cysteine auxotrophy was assumed to have occured by the coincidental introduction of a second mutation during mutagenesis. A second round of mutagenesis was therefore performed on the mutant YG-30 to select for a revertant to cysteine prototrophy which still lacked IDH-I activity. It was this strain (YG-83) that was used in the subsequent growth experiments. It is therefore possible that the successive rounds of mutagenesis could have modified the strain in some unexpected way that has caused the apparent loss of the requirement for IDH-I.

The situation found in *A.calcoaceticus* was found to be quite different to that in *Vibrio spp.* Although growth on acetate caused an increase in IDH-I activity by 1.5 fold compared to growth on succinate, IDH-II activity also increased, altering the ratio of IDH-I to IDH-II from 1:1.1, to 1:1.8 (Reeves *et al.*, 1986). Experiments were not performed that would determine if this alteration was due to altered levels of protein synthesis, or by alteration of the enzyme activities. Once again, it was hypothesized that

this has an effect on the control of the TCA cycle and glyoxylate bypass during growth on acetate. It was pointed out in section 1.2.4.2 that the *A.calcoaceticus* IDH-II enzyme is probably subject to allosteric control. It could therefore be that the activity of the allosterically controlled IDH is increased during growth on acetate to allow control of flux between the glyoxylate bypass and TCA cycle.

The role of isoenzymes and how essential they are for growth on different substrates in these two organisms is therefore still far from clear. Isolation of mutants for each of these organisms deficient in each of the enzymes would help to clarify the situation.

1.2.4.4 IDH isoenzymes from pig heart and *Saccharomyces cerevisiae.*

Many eukaryotes contain several forms of IDH, some of which are specific to certain organelles. The situation in *S.cerevisiae* provides a good example. This organism contains two mitochondrial IDHs, one of which is NADP-dependent and one of which is NAD-dependent, as well as another cytosolic NAD-linked IDH. The function of each enzyme is not really known, although it is assumed that the NAD-linked enzymes produce NADH for the production of energy *via* the electron transport chain. The $NADP^+$ form probably produces NADPH and intermediates for cellular biosynthesis.

There is also a difference in the way that the enzymes are controlled. The mitochondrial NAD-linked IDH is controlled allosterically by effectors such as NAD^+, AMP and citrate, whereas the NADP-linked enzyme activity is not affected by metabolites (Keys and McAlister-Henn, 1990). The NAD-IDH is different to the NADP-IDH in quarternary structure too. NAD-IDH was reported to be an octameric enzyme made from equal proportions of two subunits (IDH1 and IDH2) that are immunologically distinct and have different N-terminal amino acid sequences. The $NADP^+$ isoenzyme is a dimer made from identical 45 kDa subunits (called IDP1) and is therefore similar to the NADP-IDH from *E. coli* .

The NAD-IDH from pig-heart has a subunit arrangement different from any of the IDHs mentioned so far. It was first reported that this was an oligomeric enzyme made from three different subunits in the ratio $4\alpha:2\beta:2\gamma$ (Ramachandran and Colman, 1980). However, Huang and Colman (1990) have since reported that the enzyme is actually a tetramer made up of three different subunits in the arrangement $\alpha\beta+\alpha\gamma$.

1.2.5 Some sequence and structural similarities between the IDH from *E. coli* and other proteins.

The first IDH encoding gene was cloned from *E. coli* by LaPorte *et al.* (1985) which led to its sequencing by Thorsness and Koshland (1987). The inferred amino acid sequence was found to have no recognisable $NADP^+$ binding site of the type originally proposed by Rossmann *et al.* (1974). What is more, it showed no sequence similarity to any other known dehydrogenase except for the NAD^+ dependent enzyme, isopropylmalate dehydrogenase (IMDH) from *Thermus aquaticus*, which was also found to lack the Rossmann, type nucleotide binding sequence (see section 1.3.1).

The relevance of this similarity was not fully recognised however until the crystal structure of the *E. coli* IDH was elucidated by Hurley *et al.* (1989). It was then reported that sequence alignments could be made between the *E. coli* IDH and the six IMDHs for which sequence information was available, with between 25% and 29% identity. Since the IMDHs catalyse a very similar reaction to the IDHs and with a very similar substrate, it is perhaps not surprising to find that there is such a significant overall sequence conservation (Hurley *et al.*, 1991).

A cluster of positively charged residues were found to be absolutely conserved in most of the sequences analysed, corresponding to ARG-119, ARG-129, ARG-153, TYR-160 and LYS-230 of the *E. coli* sequence. All of these residues were found to be contained within a pocket that perhaps allows them to interact with the negatively charged isocitrate (Hurley *et al.*, 1989). The residues that correspond to SER-113 of the IDH sequence are not conserved in the IMDHs, though. One would expect this, because SER-113 in IDH interacts with the γ-COOH group of isocitrate, a functional group which is not present in isopropylmalate.

Elucidation of the three-dimensional structure of the isocitrate-IDH complex provided proof that the amino acids mentioned above, and SER-113, do in fact interact directly with the isocitrate (Hurley *et al.*, 1990b). It was expected from the amount of sequence conservation observed between the *E. coli* IDH and the IMDHs that the topology of the enzymes would be very similar. This was found to be the case when the crystal structure of the IMDH from *Thermus thermophilus* (a close relative of *T.aquaticus*) was solved (Imada *et al.*, 1991). The arrangement of secondary structure in both enzymes is identical except for the absence in the IMDH of two β-sheets and two α-helices that are present in the IDH.

It is therefore apparent from similarities in both the primary and tertiary structures, that both types of enzymes have diverged from a common ancestral protein to perform these two types of reaction which are similar. This evolution has led not only to the difference in substrate specificity but also to the difference in cofactor requirement.

As the work described in this thesis was nearing completion, three more IDH sequences were published. The IDH from *T. thermophilus* could be classified as a class I, *E. coli* -type IDH, by virtue of its subunit organisation and molecular weight (see section 1.2.4.1). It was therefore perhaps of little surprise that the recently published amino acid sequence for the *T. thermophilus* enzyme suggested that it was related evolutionarily to the IMDH/IDH group of enzymes (Miyazaki *et al.*, 1992).

The *T. thermophilus* IDH was found to be 33% identical to the IMDH from the same organism and 37% identical to the *E. coli* IDH over most of its length. However the *T. thermophilus* IDH possesses an extra 140 amino acids at its C-terminus. Also, in common with the IMDHS, it lacks two of the β-sheets and two of the α-helices that are found in the *E. coli* IDH (see above). All the residues expected to bind substrate are conserved with respect to the *E. coli* sequence, including the amino acid corresponding to SER-113.

The other two IDH sequences published are both from *S.cerevisiae* mitochondria. One of the sequences is of IDH2; one of the two subunits of the NAD^+ dependent enzyme (see section 1.2.4.4 and Cupp and McAlister-Henn, 1991). The other sequence is of the subunit that makes up the homodimeric, NADP-dependent IDH, IDP1 (see section 1.2.4.4 and Haselbeck and McAlister-Henn, 1991). Cupp and McAlister-Henn (1991) reported that the NAD-IDH showed 33% identity to the IDH from *E. coli* , despite the difference in cofactor specificity and quarternary structure. The same group also reported that there was no "significant homology" between the yeast and *E. coli* NADP-dependent IDHs, even though the yeast IDH could be classified as a type I IDH, like the enzyme from *E. coli* (Haselbeck and McAlister-Henn, 1991).

As well as the entire sequence for those proteins mentioned so far, peptide sequence is also available from each of the subunits that make up the NAD-IDH from pig-heart (Huang and Colman, 1990). It is described in section 1.2.4.4 that this enzyme has a different quarternary structure to both the $NADP^+$ and NAD^+ dependent enzymes mentioned so far being made up of three different subunits in the arrangement $\alpha\beta+\alpha\gamma$. Three of the 5 peptides sequenced from the α subunit were aligned with the sequence of the IDH2 subunit from yeast, as were four out of five peptides from the β subunit and all five of the peptides from the γ subunit (Cupp and McAlister-Henn, 1991).This suggests

that all of the pig heart NAD-IDH subunits are related to each other and also to the NAD^+ enzyme from yeast.

1.3 $NADP^+$ binding sites.

From the information discussed in the previous sections it would now seem that there are several IDHs and IMDHs that show relationships to each other (see section 1.2.5) based on sequence similarities. However none of these proteins were found to show any sequence or structural similarity to any other dehydrogenase and did not have a classical LDH-type $NAD(P)^+$ binding site (section 1.3.1). In fact it was not until the crystal structure of the *E. coli* IDH-NADP complex was determined, that residues involved in nucleotide binding were identified.

It is now becoming clear that there are a number of different types of nucleotide binding sequences and structures, some of which are discussed here.

1.3.1 "Rossmann fold" (LDH) type nucleotide binding structures.

It had been noticed that the NAD^+ and FAD^+ binding domains of many proteins share a common structural motif that was called the "LDH fold", after the first enzyme in which this was identified. This structure consisted of two βαβαβ motifs which form a β-sheet that is responsible for the binding of the ADP moiety of the nucleotide, giving it the alternative name of the "βαβ-ADP binding motif" (Wierenga *et al.*, 1986). The first motif of the pair is involved in the binding of the adenine moiety whilst the second seems to be involved in the binding of the nicotinamide moiety (Rossmann *et al.*, 1974).

Analysis of other nucleotide binding proteins by Wierenga *et al.*, (1986) allowed the identification of eleven key amino acids that were conserved within this fold which, in turn, allowed the definition of a fingerprint that could be used to test other sequences for possible nucleotide binding sites. Some of these positions in the fingerprint required absolute conservation, whilst others described the type of side chain that an amino acid must possess. Most noticeable was the conservation of the sequence GXGXXG, which forms the loop between the first β-strand (called βA in the nomenclature of Rossmann *et al.*, 1974) and the first α-helix (called αB; see figure 1.5). The first glycine (GLY-27 for spiny dogfish LDH; see figure 1.5) allows for the tight turn required of the loop whilst the second glycine (GLY-29, in figure 1.5) allows for the close approach of the 5'-pyrophosphate from NAD^+ to the N-terminus of the α-helix. The third glycine (GLY-32, in figure 1.5) allows for the intercalation of the second β-strand, βB, and the helix αB at the point at which they are closest. All of the amino acids in the βB strand that come close to the helix αB were found to be hydrophobic residues that enable the two structures to

come close to each other with the minimum of unfavourable interactions (see amino acids labeled in boxes (□) figure 1.5). An acidic amino acid was also always found at the end of the second β-strand which allowed the formation of a hydrogen bond between the side chain and the 2'-hydroxyl of the NAD^+ (ASP-52 for the spiny dogfish LDH).

Wierenga *et al.* (1986) noted that the consensus sequence proposed by them excluded many of the $NADP^+$ binding enzymes including those, such as GR, that are known to contain βαβ-type nucleotide binding folds. This was due to two differences. Firstly, they noticed that the core consensus sequence of nucleotide binding residues found in $NADP^+$-linked enzymes is GXGXXA and not GXGXXG. Secondly, they realised that the acid amino acid found in NAD-linked enzymes, that is required to interact with the 2'-hydroxyl of the NAD^+, would actually hinder the binding of $NADP^+$ because of the negatively charged 2' pyrophosphate group. NADP-dependent enzymes were therefore frequently found to contain hydrophobic residues at the equivilent position. It was therefore proposed that these two main differences between NAD and NADP-linked enzymes could account for the different coenzyme specificities.

This simple explanation seemed to be supported by the results of a series of elegant experiments by Scrutton *et al.* (1990), which are discussed in section 1.3.2. However, recent work by Lilley *et al.* (1991) on the glutamate dehydrogenase (GDH) from *Clostridium symbiosum* has shown that this model cannot be said to be applicable to every NAD(P)-dependent enzyme.

1.3.2 Mutagenesis of glutathione reductase converts the enzyme from being $NADP^+$ dependent to NAD^+ dependent.

Glutathione reductases (GRs) are known to contain LDH type βαβ-folds because the crystal structure had been elucidated for the enzyme from human erythrocytes (Karplus *et al.*, 1987). This enzyme belongs to a family of proteins called the "flavoprotein disulphide oxidoreductases", most of which are NADP(H)-dependent. One exception to this is the dihydrolipoamide dehydrogenase (DLDH) class of enzymes which are NAD-dependent, yet show significant sequence similarity to other members of the GR family. Scrutton *et al.* (1990) therefore compared the amino acid sequences from seven of the NADP-dependent enzymes (including GRs) with the sequences from seven of the NAD^+ dependent enzymes (which included DLDHs) in an attempt to identify residues that might be involved in imposing coenzyme specificity.

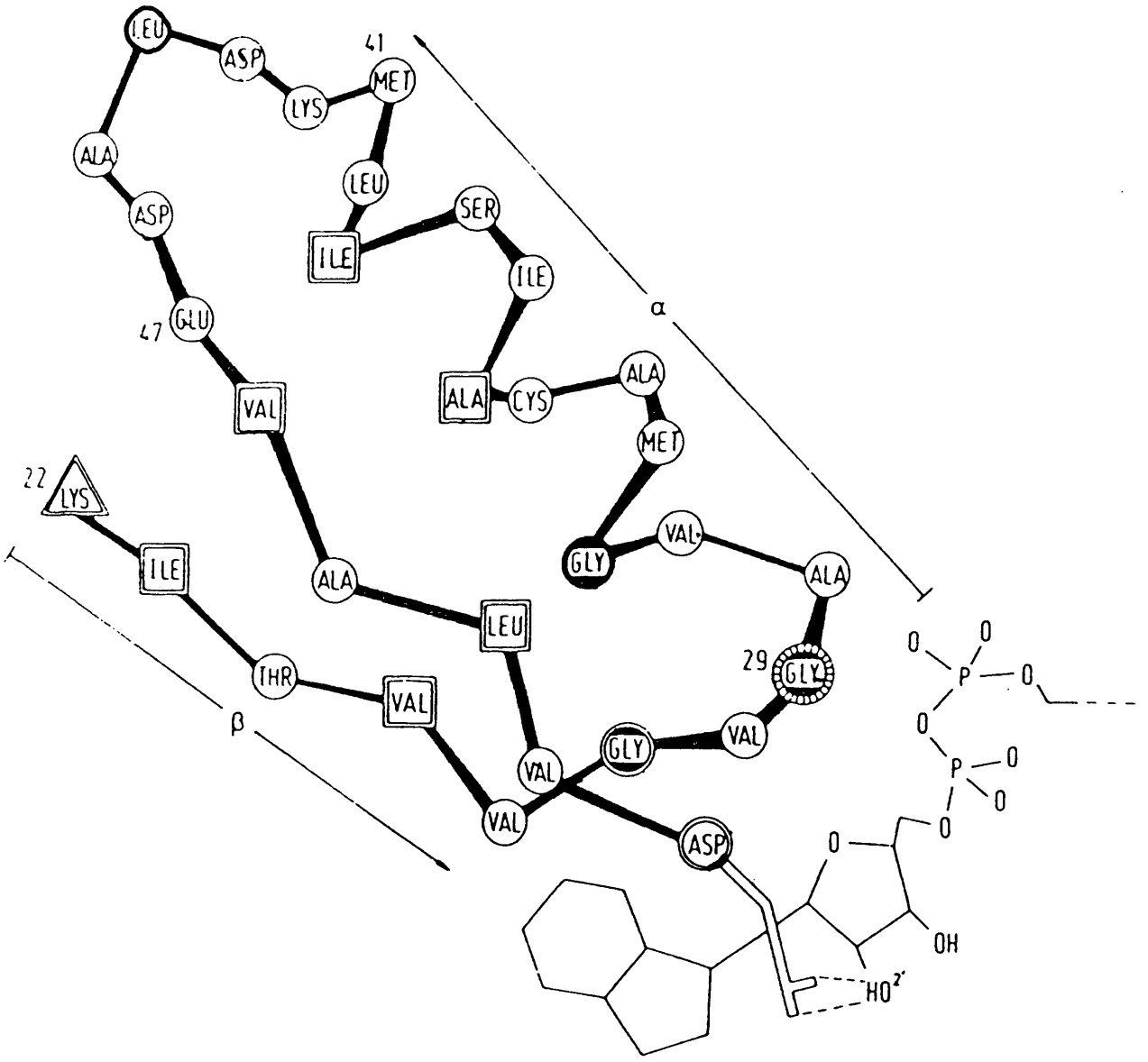


Figure 1.5 Schematic diagram of the Wierenga βαβ fold of LDH.

An NAD^+ molecule is shown in the position in which it is bound to the spiny dogfish M-lactate dehydrogenase (LDH). From Wierenga *et al.*, 1986.

□= basic or hydrophobic amino acids.

Δ= small and hydrophobic amino acids.

Residues were identified that were conserved in the $NADP^+$ dependent enzyme but not in the NAD-dependent enzymes, and *vice versa*. Site-directed mutagenesis was used to alter amino acids in the GR from *E. coli* to those used by the DLDH from the same organism, in an attempt to alter the coenzyme specificity of GR. These changes are discussed below and summarised in figure 1.6.

One of the most striking differences between the two types of enzymes was the complete conservation of arginine residues in two positions in all of the $NADP^+$ dependent enzymes (shown as positions 4 and 7 of the sequence shown in figure 1.6) whereas there was no significant conservation in the DLDHs. Conversion of either of these arginines in the *E. coli* GR to the corresponding residues found in the *E. coli* DLDH sequence caused the K_m for $NADP^+$ to increase by 25 fold, whilst the k_{cat} decreased by 3 fold. Conversion of both residues at once still gave a 25 fold increase in the K_m for $NADP^+$ but decreased the final k_{cat} further, to only 5% of the wild type level. Although the conversion of these residues obviously led to a decreased affinity and reactivity for the $NADP^+$ dependent activity of the mutant GR, it was found to only have a small effect on the ability of the enzyme to use NAD^+ (there was a 3 fold drop in the K_m for NAD^+). This is consistent with the use of these residues to bind to the 2'-phosphate of $NADP^+$, which is lacking in the NAD^+ molecule.

The other striking difference is the presence of alanines found in all of the $NADP^+$ dependent enzymes, (positions 1 and 2 figure 1.6) instead of the glycines found in most of the NAD^+ dependent enzymes. The alanine at position 1 corresponds to the third glycine in the Wierenga consensus sequence, GXGXXG, which was reported to be necessary for the close approach of helix αB to the β-sheet, βB (Wierenga *et al.*, 1986). Conversion of the alanine at position 1 (figure 1.6) in the *E. coli* GR to the corresponding glycine found in the NAD^+ dependent enzymes caused a 40 fold decrease in the K_m of the enzyme for NAD^+ although the K_m for $NADP^+$ and the k_{cat} for both co-enzymes changed little. Introduction of the mutation in position 2 of the sequence was only performed in the presence of the other three mutations already mentioned (A179G, R198M, R204L) but was found to have little effect on any of the parameters for either of the co-enzymes.

Wierenga *et al.* (1986) had already proposed that in LDH type enzymes, the acidic residue at the end of sheet βB was used in binding of the 2'-hydroxyl of NAD^+ and that this type of residue would hinder the binding of $NADP^+$ (see above). This residue corresponds to position 3 of figure 1.6, where all the DLDHs were found to contain a conserved glutamate (other NAD linked dehydrogenases *eg.* LDH usually have aspartate) whereas all of the $NADP^+$ dependent enzymes had non-acidic residues that would not interfere with the negatively charged 2'-phosphate found on $NADP^+$. Valine-197 was

```
                        1       2               3 4 5 6       7
GR    - 174 G-X-G-X-X-A-X-X-X-A-(6X)-G-(6X)-X-R-K-H-(3X)-R-(2X)-D 207
DLDH  - 180 G-X-G-X-X-G-X-X-X-G-(6X)-G-(6X)-E-X-M-D-(3X)-P-(2X)-D 213
```

Summary of changes made to GR.

1. A179G	3. V197E	5. K199F	7. R204L
2. A183G	4. R198M	6. H200D	

Figure 1.6 Glutathione reductase is converted from NADP$^+$ to NAD$^+$ dependency by site directed mutagenesis

Summary of the changes made to the *E. coli* NADP$^+$ dependant GR in order to alter its coenzyme specificity to NAD$^+$. The amino acids changed are indicated below and with the exception of change number 7, were changed to the equivalent amino acid found in the NAD$^+$ dependent DLDH from *E. coli* (Scrutton *et al.*, 1990). The changes were made on the basis of those amino acids that were conserved within the enzyme type (*ie.* NAD$^+$ or NADP$^+$) which are shown in the sequences above. The non-conserved amino acids are shown as Xs.

therefore substituted for a glutamate in the mutant GR that had already been modified at the four positions previously mentioned (A179G, A183G, R198M, R204L). At the same time two other mutations were introduced (K199F, H200D) in order to eliminate any residual positive charges in the surrounding area (positions 3, 5 and 6 in figure 1.6). This change increased the k_{cat} of the NAD^+ dependent enzyme activity by 3 fold. The k_{cat} for $NADP^+$ dependent activity was also decreased by 2 fold. The K_ms for the two coenzymes were also altered inversely, decreasing by 1.3 fold for NAD^+ and increasing by 3 fold for $NADP^+$.

With all of these mutations together, the effect was therefore to increase the specificity (k_{cat} /K_m) of GR for NAD^+ by 70 fold whist decreasing the specificity for $NADP^+$ by 250 fold. It is worth pointing out though that there are some combinations of mutations that produced unexpected and as yet unexplained results, perhaps indicating a more complex model for specificity than the one proposed. However the experiments did show that both the NAD^+ and $NADP^+$ nucleotide binding domains studied are in fact related and can be interconverted by appropriate mutations.

In summary then, it would appear that there are "three classes" of amino acids that impose coenzyme specificity in different ways. Amino acids in positions 4 and 7 allow for favourable interactions with the 2'-phosphate of $NADP^+$, which contributes to the ability of the enzyme to use $NADP^+$, but does not affect its ability to use NAD^+. Conversely the presence of a glycine at position 1 seems to increase the affinity of the enzyme for NAD^+ considerably, whilst not affecting the enzymes specificity for $NADP^+$ by very much. The third set of amino acids that are involved in imposing this specificity are those involved in binding of the 2'-hydroxyl of NAD^+. These amino acids not only contribute to the ability of the enzyme to bind NAD^+ but also contribute in the rejection of $NADP^+$ by creating unfavourable interactions with the 2'-phosphate of this molecule.

1.3.3 GDH from *C.symbiosum* does not conform to the Scrutton model for NAD(P)$^+$ binding.

The general aplicability of the simple model for coenzyme specificity to all proteins has recently been thrown into some doubt by studies of the glutamate dehydrogenases (GDHs) from *Clostridium symbiosm* (Lilley *et al.*, 1991).

This enzyme is an NAD^+ dependent enzyme and yet contains a nucleotide binding site (identified by the presence of an βαβ-nucleotide binding fold in the crystal structure) with the sequence GXGXXAXXXA. This sequence conforms to the consensus normally associated with the binding of $NADP^+$ in $NADP^+$ dependent enzymes (Scrutton *et al.*, 1990). NAD^+ dependent enzymes are expected to possess the sequence GXGXXG

(Wierenga *et al.*, 1986). The empirical correlation of nucleotide binding sites therefore needs to be reassessed since some other factor must determine the coenzyme specificity of the *C.symbiosum* enzyme.

1.3.4 Non-LDH type nucleotide binding domains.

1.3.4.1 Flavoprotein pyrimidine nucleotide cytochrome reductases.

Karplus *et al.* (1991) had shown that the enzymes ferredoxin-$NADP^+$ reductase, cytochrome P450 reductase, nitrate reductase and sulphite reductase were very similar at the amino acid sequence level and that they should be grouped together as a family of enzymes called the "Flavoprotein pyrimidine nucleotide cytochrome reductases". This family therefore contains both NAD^+ and $NADP^+$ dependent enzymes. Determination of the structure of ferredoxin-$NADP^+$ reductase (FNR) from spinach allowed the amino acids involved in binding of FAD^+ and $NADP^+$ to be elucidated (Karplus *et al.*, 1991). Sequence alignments of the FAD^+ and $NADP^+$ binding sites of FNR to the NAD^+ dependent enzymes, nitrate reductase and cytochrome b_5 reductase, and the $NADP^+$ dependent enzymes, cytochrome P450 reductase and sulphite reductase, showed that residues were conserved amongst enzymes with both coenzyme specificities.

The domain that contains $NADP^+$ binding residues looked topologically similar to the domain found in the LDH type nucleotide binding fold, consisting of a 5-stranded parallel β-sheet surrounded by 6 α-helices. $NADP^+$ was found to bind to the C-terminal edge of this sheet, like in the LDH-type folds. The negatively charged 5'-phosphate of the $NADP^+$ approaches the N-terminus of an α-helix, where interactions with the protein are probably stabilised by the positively charged dipole. This is again similar to the LDH-type fold. It was also reported by Hyde *et al.* (1991) that those amino acids in $NADP^+$ dependent members of the family that would bind the 2'-phosphate of $NADP^+$, were different from the amino acids used by NAD^+-dependent members of the family. This is again reminiscent of the structure in the LDH type folds. Two invariant glycines were also found within the conserved sequence, GXGXXP which in part coincides with the LDH type consensus sequence, GXGXXG/A.

There are however significant differences between the nucleotide binding fold of this family of enzymes and those found in the LDH family. For instance the loop containing the conserved GXGXXP sequence was found to be at the end of a β-sheet that would correspond to the βC-sheet of the LDH rather than the normal βB-sheet. It was also found to make a much tighter turn than the loop found in the LDH family of enzymes.

The complete conservation of the proline in the FNR type sequences, instead of the glycine normally expected for an NAD^+ binding fold, or the alanine expected in $NADP^+$ binding folds, is also a difference between this family of enzymes and the LDH-type folds. It also indicates that this amino acid cannot play a major role in determining the coenzyme specificity of the enzymes, as it is conserved in both NAD^+ and $NADP^+$ linked forms.

The differences found between the LDH and FNR type enzymes may indicate that they have both evolved separately the ability to bind NAD(P), although a divergent relationship cannot be ruled out.

1.3.4.2 Aldose reductase.

One of the conserved features in all of the $NAD(P)^+$ binding enzymes discussed so far is the binding of the cofactor at the C-terminal edge of a parallel β-sheet. Another is the use of the positive dipole found at the N-terminal end of an α-helix to stabilise protein interactions with the the negatively charged pyrophosphate moiety of NAD(P).

However neither of these features are used by aldose reductase from pig lens, which is found to be a single domain made from an eight stranded αβ–barrel that could not contain a typical βαβ-ADP binding fold (Rondeau *et al.*, 1992). This enzyme uses both the C and N-terminal ends of the β-sheets to bind $NAD(P)^+$ and relies solely on the use of positively charged amino acids to stabilise the pyrophosphate moiety.

It remains to be seen if this structure is used by any other enzymes for the binding of nucleotides but indicates another way in which nature has evolved a mechanism for binding of $NAD(P)^+$.

1.3.4.3 *E. coli* IDH nucleotide binding site.

It was mentioned in section 1.2.5 that the *E. coli* IDH shows little similarity to any other dehydrogenase, except for the IMDHs to which it is related. Analysis of the crystal structure of the *E. coli* IDH and of sequences from both types of enzyme revealed that there is no βαβ-ADP binding fold either (Hurley *et al.*, 1989). It was not until the structure of the $NADP^+$ bound to the IDH was determined that one could examine the structural features and the amino acids involved in the binding of $NADP^+$.

$NADP^+$ was actually found to bind within the same cleft as isocitrate which is formed between the large and small domains of the subunit and is different to any of the binding structures discussed so far. The negatively charged pyrophosphate moiety is found to

bind at a loop that is found at the N-terminus of an α-helix, taking advantage of the positive charge found there. This is therefore similar to both the LDH and FNR type folds (Wierenga *et al.*, 1985). The close approach of a loop to the N-terminus of an α-helix therefore seems to be a feature used in many $NADP^+$ binding sites.

In summary then it could be said that there are now several examples of $NAD(P)^+$ binding structures which seem to have evolved independently of each other. Some similarities in the structures employed (and in some cases the amino acids) would seem to indicate that the number of structures that can be used in nucleotide binding is limited. However there is still significant variability even between related binding structures.

1.4 Aims of the project.

As described in section 1.1.2, it is of interest to study IDH because of the control it is likely to have over the TCA cycle and glyoxylate bypass and the effect that it may have on antibiotic biosynthesis, especially of the polyketide, actinorhodin. Alteration of the levels of IDH activity in *S. coelicolor* will provide information as to what control (if any) IDH has over both primary and secondary metabolism and may provide an insight into how it could be altered to increase antibiotic production.

IDH activity could be increased by the introduction of multiple copies of the gene for IDH whilst mutations can be introduced into the chromosomal copy by insertional disruption studies. The ultimate aim of this project was therefore to clone the gene from *S. coelicolor* that encodes IDH so that these studies could be performed.

Perhaps the best way to clone this gene is by complementation of a mutant of *S. coelicolor* that has no IDH activity (and would therefore probably be auxotrophic for glutamate or proline). Randomly cloned DNA fragments could be introduced into the mutant and transformants selected for restoration of prototrophy. Unfortunatlely no *S. coelicolor* IDH mutants are available. Attempts to make them have proved difficult, probably due very poor growth rates or the lethality expected for such mutants (S.Fischer, personal communication).

Mutants of *E. coli* do exist which lack IDH activity which could be used in similar complementation studies. However this makes two important assumptions;

a) That the streptomycete IDH gene will be expressed in *E. coli*. From the comments in section 1.1.3. it is clear that this could not be relied upon.

b) That the streptomycete IDH enzyme will be able to replace the function of the *E. coli* IDH to an extent that will allow prototrophic growth.

The method of choice was therefore to clone the gene using the technique of so-called "reverse genetics". This involves obtaining part of the amino acid sequence of the protein of interest from which one can predict the most likely DNA sequence encoding the region (based on the codon usage bias of other streptomycete genes). Oligonucleotides can then be constructed that should hybridise to the coding sequence, providing a specific probe for selection of DNA fragments that contain the gene of interest.

As there was no protein sequence information available for the *S. coelicolor* IDH the first step was therefore to purify the protein to a level that would enable the determination of either N-terminal or internal peptide sequence (by Edman degradation). The aims of this project were therefore as follows.

1) To purify the IDH from *S. coelicolor* and study some of the relevant basic properties of the enzyme.
2) To obtain N-terminal and/or internal protein sequence.
3) To design and synthesise an oligonucleotide based on the protein sequence and use as a probe for screening libraries and/or digests of *S. coelicolor* genomic DNA.
4) To sequence the enzyme for comparison to other proteins, especially the *E. coli* IDH and to search for any interesting conserved features at either the DNA or amino acid level.
5) To initiate complementation studies in *E. coli* and overexpression and gene disruption experiments in streptomycetes.

CHAPTER 2

Materials and methods.

2.1 Chemicals and biochemicals.

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

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

ATP, DTT, NAD+, NADP+, NADPH, phosphoenolpyruvate, 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, U

Agarose, isopropyl-β-thiogalactoside (IPTG), phenol (ultrapure), and 5-bromo-4-chloro-3-indoyl-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 Honeywell and Stein Ltd., Wellington, UK.

Oligonucleotides were synthesised on an Applied Biosystems Model 280A DNA synthesiser, at the Institute of Genetics, University of 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.

2.2 Enzymes, proteins and kits.

A kit for molecular weight determination and the following enzymes and proteins were 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.

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

λ DNA packaging kits "GIGAPACK™ GOLD" and the plasmid vectors pBlusescript II (KS^+) and (SK^+) were supplied by Stratagene (La Jolla, USA).

2.3 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 amino acids, were sterilized by filtration through Nalgene 0.22µm pore membranes.

2.3.1 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.

4 g glucose, 4.5 g NaNo3, 5 g NaCl, 5 g Na_2SO_4, 1 g $MgSO_4.6H_2O$, 0.5 g $CaCl_2$, 0.0lg $ZnSO_4$, 1.2 g Tris buffer, 1 g Junlon, 20 g 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.

b) Acetate NMM.

Same as above except;- Glucose was replaced by 1 g/l of NaAcetate. NaCl was ommited. The pH of the mixture was adjusted to pH 7 with NaOH prior to autoclaving at 5 psi for 50 minutes.

2.3.2 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.

2.4 Buffers used in purification of IDH.

a) Phosphate buffer.

Stock phosphate comprises of 0.5 M KH_2PO_4 adjusted to pH 7.0 with KOH.

b) Buffer A.

50 mM phosphate buffer (pH 7.0), 1 mM EDTA and 0.4 M DTT. DTT was always added immediately prior to use from a freshly thawed frozen stock.

c) Buffer B.

As buffer A but 500mM phosphate buffer.

d) Extraction buffer.

As buffer A but with 3.5 mM benzamidine (made in dH_2O and kept as a frozen stock) and 1.2 mM PMSF (made up fresh in absolute alcohol) which were added to inhibit proteases. Benzamidine and PMSF were added to buffer A just before addition to cells.

e) Assay buffer

0.15 M Tris-HCl (pH 7.5) and 0.5 mM $MgCl_2$ was stored at 4°C.

f) Sample buffer (2X).

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

2.5 Chromatography media.

Pre-packed Mono Q HR5/5 (1 ml volume), HR10/10 (8 ml volume), phenyl-Superose, Superose 6 and Superose 12 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.

2.6 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 sample and boiled for 2 mins. The amount of sample loaded was usually based on the activity of the sample. 0.05 U of IDH activity was required to visualise a band of IDH protein by coomassie staining.

e) Protein staining.

Protein was localised on gels either by staining with Coomassie blue. Staining was with Coomassie reagent (0.1% Coomassie brilliant blue G250 in 10% (v/v) glacial acetic acid 50% (v/v) methanol) for 1 hour at 40°C. The gel was then destained in 10% acetic acid, 10% methanol at 40°C overnight. Gels were then dried down for storage or autoradiography.

f) Drying down destained polyacrylamide protein gels.

Slab gels were dried down for easy storage or autoradiography. 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 sandwhich dried down for 30-60 mins on a Biorad Slab Gel Dryer model 1125.

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 37°C. Each 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 sample.

i) Concentration with Centricon -30 micro concentrators.

Centricon™ 30 micro concentrators were obtained from Amicon Corporation Cherry Hill Dr., M.A., U.S.A and used essentialy as described in the manufactureres

instructions. Samples were loaded into a concentration device in a maximum volume of 2 mls. The apparatus were then subjected to centrifugation at 3000.g for the appropriate length to reduce the sample to 200-300µls (*e.g.* 20 mins will reduce 2mls to 200µls at 3000.g). If the original sample had a volume greater than 2mls the retentates were pooled and the volume reduced further by another round of concentraion.

2.7 Production of *S.coelicolor* spores.

a) Preparation suspensions and storage.

Concentrated spore suspensions were required for inoculating liquid cultures of *S. coelicolor*. 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.

2.8 Growth of *S.coelicolor* 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.

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 72 hours or just as the mycelia started to produce undecylprodigiosin (red) antibiotic. Mycelia harvested from NMM cannot be stored frozen because the Jonlon forms a sicky glue like consistency once thawed.

b) Complex media (YEME).

Most growth experiments were grown on this complex media 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 800 mls of media. Cells grown on YEME could be harvested after 48 hours and stored as cell pellet at -20°C indefinitely until required.

2.9 Harvesting of mycelia.

After growth mycelia were recovered from the growth media by centrifugation (20 minutes 7,500 g) and resuspended in extraction buffer (100 mM-potassium phosphate pH 7, 0.4 mM-DTT, 1.2 mM-PMSF, 3.5 mM-benzamidine) and the cells lysed using a French Press. Cells grown on YEME (section 2.8b) were resuspended and washed twice in 1/10th culture volume distilled dH_2O prior to resuspension in extraction buffer.

2.10 Preparation of crude 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 (l4 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 10,000.g for 1 hour at 4°C.

2.11 Methods for characterisation of IDH.

2.11.1 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.

a) Standard spectrophotmetric assays.

These were performed with assay buffer. To this was added 0.4 mM NADP, 2.5 mM DL-isocitrate which were kept as frozen stocks (-20°C) and added to assays just before use. Enzyme 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 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 disapearance 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.

b) Spectrofluorometric assays.

When the K_ms for isocitrate and NADP were determined for IDH, the accurate concentration of the respective substrates was needed. To determine the concentration of isocitrate, excess NADP (0.4 mM) was added to the assay mixture with an appropriate amount of active IDH enzyme. The spectrophotometeric reading was made to zero against this blank. Dilutions of the isocitrate stock (containing an estimated concentration of less than 0.3mM with respect to the D-form) were then added to the cuvette and the reaction allowed to proceed to completion. The final absorbance reading at 340 nm was then used to calculate the amount of NADPH produced and therefore the amount of isocitrate consumed) in the reaction, using the extinction co-efficient for NADP ($E_{340nm}=6.23x10^3$ M/cm). The concentration of NADP was calculated in a similar way using excess isocitrate (2.5 mM) and limiting amounts of NADP (less than 0.4 mM). Reactions were again started by the addition of isoctrate.

When determining K_ms for isocitrate and NADP for IDH, and when screening substances for activation/inhibition effects on IDH activity, the production of NADPH was followed on a spectrofluorometer. This method was more sensitive being able to detect lower concentrations of NADPH. A Perkin-Elmer LS-5 luminescence specrtometer was used with an excitation wavelength of 340 nm and a detection wavelength of 460 nm (slit width 10nm for both).

2.11.2 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-galactosidase	116,000
rabbit muscle phosphorylase b	97,400
bovine albumin	66,000
egg albumin	45,000
bovine erythrocyte carbonic anhydrase	29,000

2.11.3 Native M_r.

Gel permeation chromatography on a Superose 12 column was used to estimate the native M_r of the purified protein. This was carried out at room temperature using a Pharmacia f.p.l.c. apparatus. The column was eluted with buffer A (flow rate = 0.3 ml/min, fraction size 0.3 ml). The eluate was monitored at 280 nm and the column calibrated with the proteins listed below.

Protein	Native M_r.
horse apoferritin	443,000
rabbit muscle pyruvate kinase	232,000
rabbit muscle aldolase	158,000
pig muscle lactate dehydrogenase	140,000
hexokinase	100,000
ovalbumin	45,000

2.11.4 N-terminal sequence analysis by gas-phase sequencing.

N-terminal amino acid sequencing of intact *S.coelicolor* IDH was carried out using an Applied Biosystems model 470 protein sequencer by with on-line detection of amino acid thiohydantoins. All glassware for protein chemistry was heated for 1 hour in 6N HCl or 6N HNO_3 then left steeping overnight. It was then rinsed exhaustively with distilled water and dried in a hot oven.

2.11.5 Phosphorylation of IDH with *E.coli* IDH kinase/phosphatase.

IDH kinase/phosphotase was partially purified by Fiona Douglas (Department of Biochemistry, Glasgow University) essentially using the method of Nimmo *et al.* (1984). The reaction mixture was;
25 mM Mops-NaOH (pH 7.3), 0.5 mM EDTA, 0.5 mM DTT, 100 mM $MgCl_2$, 0.5 mM ATP and IDH kinase/phosphotase (c. 6 μU of activity). Approximately 700 mU of IDH activity was added to the assays. ATP was added last to start the reaction which was left for 1 hour at 37°C. IDH enzyme activity was measured over a time period.

2.11.6 Western blotting procedures.

2.11.6.1 Buffers.

a) Transfer Buffer.

20 mM Tris.Cl pH7.2, 150 mM NaCl, 0.5% Tween 20.

b) Substrate buffer.

50 mls 10 mM Tris.Cl (pH 7.4), 150 μl 4% H_2O_2, 30 mg of chloronapthol dissolved in 10 mls methanol. (H_2O_2 was added immediately prior to use.

c) Blotting buffer (x10 stock).

20 mM Tris.Cl pH7.2, 150 mM NaCl.

2.11.6.2 Transfer of protein to nitrocellulose.

Proteins in samples were seperated by SDS PAGE as described in section 2.6d, except that 5 μls of pyroniny was added to allow eventual visualisation of protein on the nitrocellulose. Proteins were then transferred to 0.45μ nitrocellulose papers (Adderman and Co., Kingston-upon-Thames, Surrey KT2 6NH) using a Bio-Rad trans-blot cell containing Buffer A with 16.6% methanol. The transfer was run at 0.65 mA/cm^2 of filter for 4 hours.

2.11.6.3 Probing nitrocellulose filters with antibodies.

The nitocellulose paper was placed in blotting buffer containing 5% normal donkey serum at 4°C for at least 16 hours. The nitrocellulose paper was then rinsed twice in blotting buffer. The paper was then soaked in transfer buffer containing 5% (v/v) normal donkey serum and 0.5% (v/v) rabbit anti-*E.coli* IDH antisera (gift from H.G.Nimmo, Gsagow University) at 25°C for 90 mins, after which the nitrocellulose paper was washed in transfer buffer 4 times for 12 mins each, followed by washing in blotting buffer for 12 mins. The nitrocellulose paper was then soaked in blotting buffer 5% containing 5% (v/v) normal donkey serum and 0.05% peroxidase conjugate donkey anti-rabbit IgG at 23°C for 90 mins. It was then washed five times (12 mins each time) in blotting buffer and developed by soaking in substrate buffer for 5 mins at 23°C. The final

product was then washed in dH_2O and left to dry. Stain fades with time so important filters should be photographed immediately.

Normal donkey serum and peroxidase donkey anti rabbit IgG were obtained from the Scottish Antibody Production Unit.

2.12 *E.coli* strains.

The *E.coli* strains used in this study are shown below.

STRAIN	GENOTYPE	REFERENCE/SOURCE
DS941	*rec* F143, *pro* A7, *str* 31, *thr* 1, *leu* 6, *tsx* 33, *mtr*12, *his* 4, *arg* E3, *lac* Y^+, *lac* ZΔM15, *lac* I^q, *gal* K2, *ara* 14, *sup* E44, *xyl* 5.	Gift from Dave Sherratt (Institute Genetics, Glasgow University.)
EB106	*icd* 11, *dad* R1, *trp* A62, *trp* E61, *tna* 5, λ^-	Apolostolakos *et al.*, 1982 (Genetic Stock centre, Yale University,.U.S.A)
TG1	*sup* E, *hsd* Δ5, *thi*, Δ(*lac-pro* AB), F'[*tra* D36, *pro* AB^+, *lac* I^q, *lac* ZΔM15].	Gibson (1984)
NM621	F^-,*sup* E44, *thy* $^+$, *rec* D, *thi* 1, *leu* B6, *lac* Y1, *ton* A21, *mcr* A, *mcr* B, *hsd* R, *rec* D1009, *pho* R::Tn10	Whittaker *et al.*, 1988

2.13 Plasmid and phage vectors.

The plasmid pUC18 was obtained from Pharmacia. Plasmid **pBluescript™ II (KS⁺ and SK⁺)** was obtained from Stratagene (La Jolla, CA, USA).

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

pCK505 is a pBR322 derived plasmid containing the wild type *icd* gene from *E. coli*. (LaPorte *et al.*, 1985) and was a gift from Prof. D.E.Koshland Jr., University of California, Berkerly, U.S.A.).

2.14 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) Top agar.

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

d) Top agarose.

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

e) Minimal agar

Agar was made to 17.5 g/l with water. To every 100mls 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 neccasary.

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

28g K_2HPO_4, 8g KH_2PO_4, 16g KH_2PO_4 1g trisodium citrate, $MgSO_4.7H_2O$, made up to 1 litre with distilled water.

g) Minimal top agar

6.5 g/l bacto-agar in dH_2O.

2.15 Growth of *E.coli* strains.

a) DS941 and NM621.

Normally propagated on L-agar and L-broth, depending on the application.

b) TG1

It is neccasary to maintain the strain on minimal media in order to maintain a selection for the F' episome which allows the production of pilli neccassary for M13 bacteriophage infection. Growth in liquid medua is always performed on YT for the same reasons.

c) EB106

For growth on liquid media L-broth was always used. The strain is mutated in the *icd* gene (encoding IDH in *E. coli*) that for unknown reasons makes the strain resistant to naladixic acid (Nal^R) as well as making the strain auxorophic for glutamate or proline. It is therefore important to always grow this strain with Naladixic acid (10 µg/ml) in order to prevent reversion of the gene to wild type. For growth on solid media, minimal media

was usually used which is supplemented with 40mM proline to allow growth of the *icd*⁻ mutant.

2.16 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 a isolate single colony.

2.17 General DNA methods.

2.17.1 Commonly used buffers.

a) TE buffer (10x).

100 mM Tris.Cl pH 8, 1 mM EDTA. Sterilised using an autoclave and stored at room temperature. Use as a 1x solution for most applications.

b) Phage buffer (1x).

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

c) TBE buffer (10x) pH8.3.

109g Tris, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$ made up to 1 litre in distilled water, pH should be *c*.8.3.

d) TAE Buffer (10x) pH 8.2.

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.

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 µg/µl in TE (final concentration 1x) with loading buffer added to 1x concentration. 10 µls 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 and 2,027 bp. and 560bp.

2.17.2 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.

2.17.3 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.

2.17.4 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.

2.17.5 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.

2.17.6 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 20µl of 1x ligation bufferprovided by BRL, containing 1U of T_4 ligase per µg of DNA. The reactions were incubated for 4 hours at room temperature or overnight at 16°C.

2.17.7 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 2.17.9). 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 (section 2.17.1f).

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 (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 (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, heated to 70°C for 2min and cooled on ice before loading onto the gel.

2.17.8 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).

2.17.9 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 endm 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 apllied 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.

2.17.10 Autoradiography.

Autoradiography was performed in metal cassettes (medical chest X-ray type) using Kodak X-OMATS film. Enhancement in the intensity of the autoradiographic images was obtained when the film was exposed at low temperature (-70°C) in close contact with a du Pont Cronex Lightening Plus intensifying screen. The X-ray films were developed using a Kodak X-OMAT automatic processor, Model ME-I.

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

2.18.1 Preparation of competent cells.

a) Hanahan method.

For high transformation efficiency cells, 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. Then, 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.

b) $CaCl_2$ 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.

2.18.2 Transformation procedure.

Transformations were carried out in sterile 1.5 ml microfuge tubes. Ligation mix refers to any DNA being used in the transforming process and usually was a ligation mixture. 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. 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.

2.18.3 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. This maintained selection for the plasmid which makes cells carrying the vector resistant to ampicillin.

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

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.

2.19 Preparation of plasmid DNA.

2.19.1 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) culture volumes of *E. coli.*

a) Birnboim Doly I (BDI)

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

b) Birnboim Doly II (BDII)

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

c) Birnboim Doly III (BDIII)

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.

2.19.2 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 (see sections 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. 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.

2.19.3 Large scale plasmid preparations

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 was 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 Bec$K_m$an 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 relaxed 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.

2.19.4 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 was 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.

2.20 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.*, (1985a).

2.21 Techniques with *E.coli* bacteriophage λ

2.21.1 Preparation of plating bacteria for infection with bacteriophage λ.

50 ml of L-broth with 0.2% (w/v) maltose, 20 mM-MgSO4 was inoculated with a single colony of the appropriate *E.coli* strain (*e.g.* NM621), and grown overnight 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.* 1.6×10^9 cells/ml). The cells were stored at 4°C and remained viable for at least 3 weeks.

2.21.2 Infection of bacteriophage λ, plating and titre.

Serial 10-fold dilutions of λ phage stock (or packaged λ DNA) were prepared in phage buffer (10 mM-Tris HCl pH 7.5, 10 mM-$MgSO_4$). Bacteriophage λ infection was achieved by adding 100 µl aliquots of each dilution to 100 µl (1.5 x 10^8 cells) of plating bacteria suspension. 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.

2.21.3 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 yields 10^6-10^7 infectious bacteriophage particles, which can be stored indefinitely at -70°C in phage buffer/DMSO without loss of viability.

2.21.4 Preparation of λ DNA.

a) Growth of λ and preparation of phage lysates.

Approximately 1.6 x 10^8 cells, of prepared bacterial suspension were infected with 10^5 plaque forming units (pfu) and plated out as described in 2.21.2. Plates were incubated overnight or until the plaques touched one another. 10 ml of phage buffer was added to the resulting lysates and plates agitated gently at room temperature for 2 hours. The phage buffer was transferred to microfuge tubes, centrifuged to remove bacterial debris and the supernatant retained. To the supernatants, 10 µl of 1 mg/ml DNaseI and 10 µl of 10 mg/ml RNase A was added and this was incubated at 37°C for 1 hour. 70 µl of DMSO per 1 ml phage buffer was added to the λ phage lysates and they were stored at 4°C.

b) DNA Preparation from λ phage lysates.

An equal volume of 20% (w/v) PEG 8000, 2M-NaCl in phage buffer was added to one volume of λ phage lysate (see 2.21.2) and incubated for 1 hour at 0°C. The precipitated phage particles were recovered by centrifugation in a microfuge (10 minutes, 4°C) and the supernatant discarded. The pellet was resuspended in 5 µl of 10% (w/v) SDS and incubated for 5 minutes at 68°C. 10 µl of 5M-NaCl was added to the mixture which was then extracted once with an equal volume of phenol/chloroform (see 2.17.2)

then extracted once with chloroform alone. An equal volume of isopropanol was added to the retained aquaeous phase and following incubation at -70°C for 15 minutes the λ phage DNA was recovered by centrifugation in a microfuge (15 minutes, 4°C). The DNA pellet was then washed with 70% (v/v) ethanol, allowed to air dry and dissolved in 50 µl of 1xTE buffer. This DNA was suitable for restriction analysis, Southern blotting, and sub-cloning.

2.22 Construction of a λ genomic library.

2.22.1 Partial digestion of genomic DNA with *Sau* 3A.

a) Small scale digestions.

These were performed eesentially as described in the Promega "Protocols and Applications Guide". Ten serial dilutions of *Sau* 3A enzyme were made in ice cold *Sau* 3A reaction buffer (1x) to span the concentration range 0.7 mU/µl to 200mU/µl. To 45 µl of a solution containing 1 µg of *S.coelicolor* genomic DNA (in 1x *Sau* 3A reaction buffer), 5 µls of the diluted enzyme was added. Digestion was then allowed to proceed for excactly 30 mins before being stopped by the addition of 1 µl of EDTA (0.5 M). DNA was analysed by electrophoresis on a 0.4% (w/v) agarose gel to determine which concentration(s) of *Sau* 3A produced the greatest flourescence in the 15-23 kb range. Electrophoresis was carried out slowly (*eg.* 50V for 2.5 hours in a minigel) at 4°C to prevent melting of the low percentage gel.

b) Large scale digestions.

A full scale digestion was performed with 200 µg of DNA using HALF the optimal *Sau* 3A concentration determined in the small scale digests (Sambrook *et al.*, 1989). All reactions were scaled up to keep the DNA, buffer and enzyme concentrations the same as in the small scale reactions.

2.22.2 Size selection of digested DNA.

Partially digeted DNA was precipited using ethanol and the pellet resuspended in 100 µls of TE (1x). Loading buffer was added and the sample loaded in to a well 5 cm long in a large 0.4% agarose gel (no ethidium bromide). DNA fragments were separated by electrophoresis overnight at 4°C. DNA was visualised using a long wave UV monitor, to reduce damage done to the DNA, and a gel slice excised that contains DNA in the 15-23 kb range. DNA was electroeluted from the gel purified and purified by phenol/chloroform extraction followed by ethanol extraction (section 2.17.9).

2.22.3 Ligation of DNA

The replacement vector (λ–GEM-11) used was supplied by Promega as *Bam* HI arms. The linker fragments had been removed and the arms de-phosphorylated with alkaline phosphatase by the supplier to prevent re-ligation of the arms to each other.

To 500 ng of vector was added 500 ng of the size selected DNA. The DNA mixture was then ethanol precipitated and the pellet resuspended in 5 µl of ligase buffer containing 1 µl of DNA ligase. The reaction was allowed to proceed for 4 hours at room temperature. A control was also performed using 100 ng of vector alone.

2.22.4 *In vitro* packaging of ligated λ DNA.

Stratagene GIGAPACK GOLD II was used to package the ligation mixture essentially as described in the manufacturers protocol guide supplied with the kit. However each reaction set was split into three in order to allow three packaging reactions from each set (instead of the reccommended one reaction per set). This had no apparent effect on the packaging effeciency. Packaging reactions were also performed on the control ligation performed with no insert to determine the number of background plaques produced (section 2.22.3). The phage particles produced by this *in vitro* packaging was used to infect a culture of NM621 ready for titration as described in section 2.21.2.

2.22.5 Amplification and storage of library.

Phage particles produced by *in vitro* packaging start to rapidly lose viability after 4 weeks storage at 4°C (Kaiser and Murray, 1985). An appropriate aliquot of DNA of the primary library was therefore used to infect a culture of NM621 (section 2.21.2) which was then plated at low density to give c.3000 plaques per 150 cm^2 plate and allowed to grow for 8-10 hours at 37°C. Plates were overlayed with 10 mls of phage buffer and and left overnight at 4°C to allow phage to diffuse into the buffer. Buffer was then decanted and the agar surface rinsed with another 2 mls which was added to the original 10 mls in a 50 ml Falcon™ tube.

Chloroform was added and the mixture left for 15 mins at room temperaure to kill any bacteria. Cell debris was removed by centrifugation at 2000g for 5 mins. DMSO was added to a final concentration of 7% and the supernatant divided into 1.2 ml aliquots for storage indefinitely at -70°C.

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

2.23.1 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 foriegn DNA into such vectors by ligation, they can be introduced into a suitable M13 host such as TG1 (*ie.* one that produces pilli).

E.coli TG1 was made competent by one of the method above 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 -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 as.

2.23.2 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 2.21.3. However M13 are sensitive to chloroform and DMSO and so these additives were ommited.

2.23.3 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 ethanol precipitated 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.

2.24 Labelling of DNA with ^{32}P.

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

Bacteriophage T4 polynucleotide kinase can catalyze 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, 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 using gel filtration (section 2.24b).

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.

2.25 Southern blotting.

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

2.25.1 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.

2.25.2 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 2X SSC to remove any adhering agarose.

2.26 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.

2.26.1 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); the volume of prehybridisation fluid was determined by filter surface area multiplied by 0.2. The bag was placed in an agitating waterbath and the filter prehybridised for at least 4 hours at 60°C.

2.26.2 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 cocentration 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 (section 2.24) was added to the bag before it was sealed. Hybridisation conditions were as described for each individual hybridisation.

Hybridised filters were washed then three times in large volumes of buffer at ionic strengths appropriate to the experimental conditions. The temperature of this buffer 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 (section 2.17.10). Films were developed by a Kodak X-OMAT processor.

2.27 Screening of plasmid clones by colony hybridisation

Recombinant pUC clones were screened using a modification of the method described by 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 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 air dry, colony side up. Finally, the filters were baked at 80°C for 2 hours. The filter was then hybridised with a nucleic acid probe (section 2.26).

2.28 Screening a phage λ library

a) First round screening

Cells from a prepared bacterial suspension were infected (section 2.21.2) with phage from the bacteriophage λ library at a ratio of 10^4 pfu/108 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 syring needle. Filters was removed and treated as described for colony filters (section 2.27).

b) Secondary screening

Desired plaque regions or single plaques were isolated as described in section 2.23.2. 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 (section 2.28a).

When identified by hybridisation and subsequent autoradiography a plug of positive phage was stabbed from the secondary plate and stored (section 2.23.2).

2.29 DNA Sequencing techniques.

2.29.1 Sequencing reactions.

Dideoxy sequencing (Sanger *et al.*, 1977) was carried out on single-stranded M13 templates using either a TAQuenase™ 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 manufacturaers insructions. However the extension and labelling reactions as well as termination reactions were carried out at 65°C to help further reduce artefacts caused by secondaty structure. ^{35}SATP was used for labeling.

b) Sequencing with T7 polymerase.

Extension and labeling 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 labeling.

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.

2.30 Denaturing PAGE for DNA sequencing.

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

2.30.1 Preparation of polyacrylamide gels.

6% (w/v) denaturing polyacrylamide gels were used for sequencing and single-stranded probe isolation. 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.

2.30.2 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.

2.30.3 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.

CHAPTER 3

Purification and Characterisation of IDH from *S.coelicolor.*

3.1 Introduction.

One of the aims of this project was to clone the gene encoding IDH in *S.coelicolor*. To this end it was decided to purify the protein in order to obtain N-terminal amino acid sequence; an oligonucleotide probe could then be constructed for use in identifying the gene by hybridisation. This approach had an added advantage over other cloning methods, allowing the determination of some of the physical and kinetic properties of the enzyme.

In this chapter I describe the successful procedures used for the purification of IDH from *S.coelicolor* and the characterisation of some of the enzyme's properties. A comparison of some of these properties to those of other IDHs characterised is also discussed.

3.2 Results.

3.2.1 Activities of ICL and IDH alter on different carbon sources .

An initial experiment was performed in conjunction with Dr H.Bramwell (Departments of Genetics and Biochemistry, Glasgow University) in which *S.coelicolor* was grown in NMM supplemented with either glucose or acetate. The results in figure 3.1 show that there is very little ICL activity in crude extracts prepared from mycelia grown with glucose as the sole carbon source. However significant ICL activity is present in crude extracts prepared from mycelia grown with acetate as the sole carbon source. There is also a reduction in the specific activity of IDH in acetate grown mycelia compared to glucose grown mycelia. This is similar to the situation found in *E.coli* where a reduction in IDH specific activity is required for the operation of the glyoxylate bypass (section 1.2.2).

3.2.2 Purification of IDH from *S.coelicolor*.

S.coelicolor was grown in the complex liquid medium, YEME, which produces approximately five times the biomass obtained from *S.coelicolor* grown on NMM (see section 2.8). The typical yield of biomass derived from 1.6 L of medium was 4-5 g (wet weight) from which crude extracts were obtained that contain around 500 mg of protein. The total IDH activity in crude extracts is usually 100-150 U.

	Acetate	Glucose
Total protein (mg)	120	106
Total IDH (U)	11.35	34.20
Specific activity IDH (mU/mg)	94.59	322.63
Total ICL (U)	1.88	0.176
Specific activity ICL (mU/mg)	15.67	1.66

Figure 3.1 IDH and ICL activity from *S.coelicolor* grown in glucose and acetate minimal media.

S.coelicolor was grown for 60 hours in NMM with either glucose or acetate as the sole carbon source (section 2.8). Crude extracts were prepared as described in section 2.9 and 2.10 and the protein concentration (section 2.6c) and ICL and IDH activities (sections 2.6h and 2.11.1a respectively).

Harvesting of mycelia and all stages of the purification were carried out at 4°C, unless otherwise stated, accept for the running of FPLC columns which was performed at room temperature. Buffers and media are described in sections 2.3-2.5.

Step 1; Production of a crude extract.

Harvested mycelia (5-6 g) was resuspended in 15-20 mls of extraction buffer to obtain a smooth paste. Cells were disrupted by three passages of this paste through a french pressure cell as described in section 2.10. Cell debris was then removed by centrifugation at 10,000 g for 1 hour. The clear yellow supernatant was decanted and subjected to further purification.

Step 2; Ammonium sulphate fractionation.

Powdered ammonium sulphate was gradually added to crude extract, with stirring, to bring the solution to 45% saturation by adding 278 mg of ammonium sulphate to every millilitre of crude extract. After stirring for a further 30 mins the precipitated protein was removed by centrifugation at 12,000 g. The supernatant was then brought to 80% saturation in a similar way by the addition of a further 245 mg/ml of ammonium sulphate and the protein again recovered by centrifugation.

The pellet obtained was then resuspended in 1-3 mls of buffer A (section 2.4) for the next stage of the purification.

Step 3; FPLC using anion exchange Mono-Q columns.

The resuspended pellet was dialysed overnight against 1 L of buffer A with at least two changes. To check that dialysis had gone to completion the conductivity of the solution was compared to that of buffer A alone.

For samples with a total protein content of up to 400 mg a large scale Mono-Q column (HR10/10) was used (section 2.5) with a flow rate of 4 ml/min. If the total protein content of the sample was less than 50 mg, then a small scale Mono-Q column (HR 5/5) was used in preference with a flow rate of 1 ml/min, since resolution and reproducibility seemed to be superior. Columns were equilibrated by washing with buffer A before use.

Samples were loaded onto a Mono-Q column and the column washed with buffer A until no further protein emerged. A linear gradient of 100% buffer A to 100% buffer B (creating a phosphate gradient of 50-500 mM at pH 7) was then applied to the column with a constant flow rate.

Most of the IDH activity was eluted between 185 mM and 240 mM phosphate (for an example see figure 3.2).

When fractions containing more than 10% of the total activity were pooled, 80-90% of the original enzyme activity was routinely obtained. An increase in the specific activity of IDH (purification factor) of approximately 10 fold was obtained.

Step 4; FPLC using a Superose 12 gel filtration column.

A Superose 12 column was equilibrated in Buffer A and 300 µl of the pooled fractions from step 3 loaded onto the column. Buffer A was passed through the column at a flow rate of 0.3 ml/min whilst collecting 0.3 ml fractions (see figure 3.3).

When scaling up the purification procedure the volume of the pooled fractions from the Mono Q was often 2-4 mls. This volume was therefore reduced by the use of Centricon-30 micro-concentrators (following the manufacturers recommended protocols), into a volume more appropriate for loading onto the Superose 12 column. It was found that up to 500 µl could be loaded without affecting the purity of the final samples, provided that less than 10 mgs of protein was loaded at a concentration of less than 30 mg/ml.

Fractions containing IDH activity from either side of the main peak were analysed by SDS-PAGE and stained with coomassie blue and those fractions containing only a single major band were pooled for further use. Although small amounts of contaminating proteins can be seen (see figure 3.4), it is estimated that greater than 95% of the total protein in the pooled fractions is IDH. Samples from the Mono-Q and Superose 12 purification stages were analysed by PAGE (figure 3.4). A purification table (figure 3.5) summarises the results obtained from a typical purification.

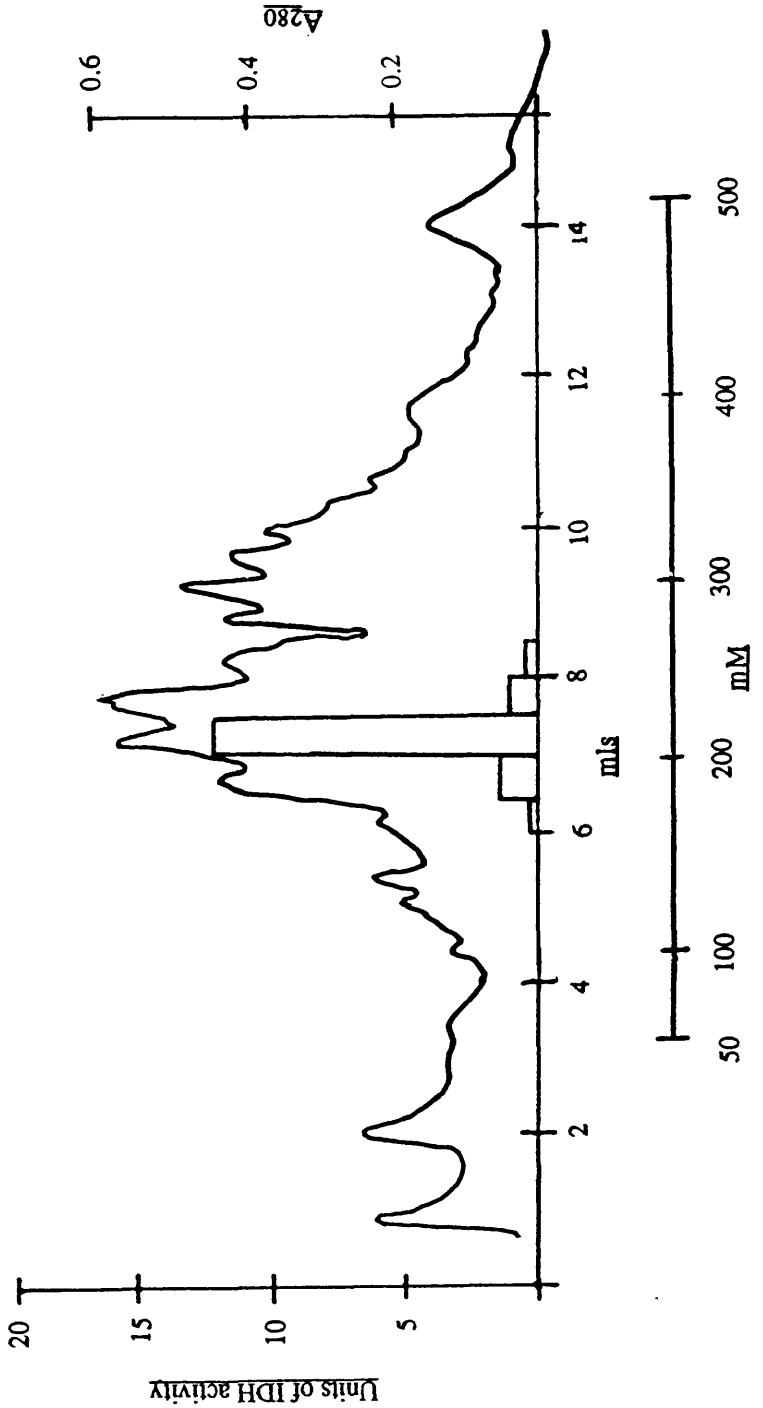


Figure 3.2 Mono Q ion exchange chromatography on the FPLC.

Protein from the 45-80% ammonium sulphate fraction that had been dialysed in Buffer A was applied to a Mono Q HR 5/5 column using a 10 ml loop and the column washed with Buffer A. The trace shows the A280 of protein eluting from the column against the elution volume (labeled as mls). A gradient was applied to the column at the point indicated by the lower horizontal axis of 100% Buffer A to 100% Buffer B (representing an increase in phosphate concentration of 50 mM to 500 mM). The fractions in which enzyme activity was found are shown in the histogram. The peak three fractions were pooled in this case.

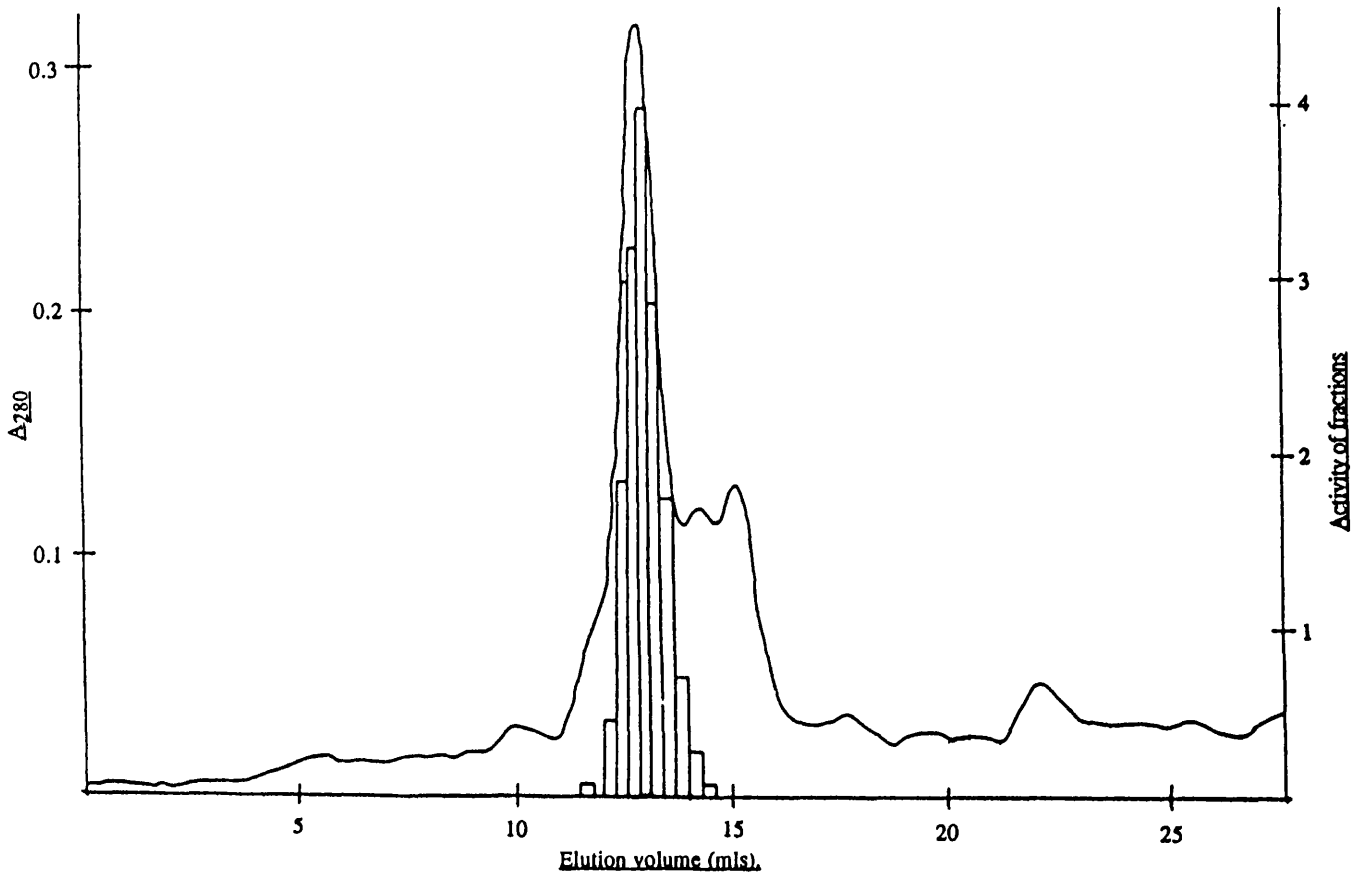


Figure 3.3 **Superose 12 gel filtration chromatography on the FPLC.**

Pooled fractions from a Mono Q column were applied to the Superose 12 column in a volume of 500 µl and 0.3 ml fractions collected.

The trace shows the A_{280} of the eluant. The histogram shows the IDH activity in the appropriate fractions. In this case the four fractions containing the most activity were pooled.

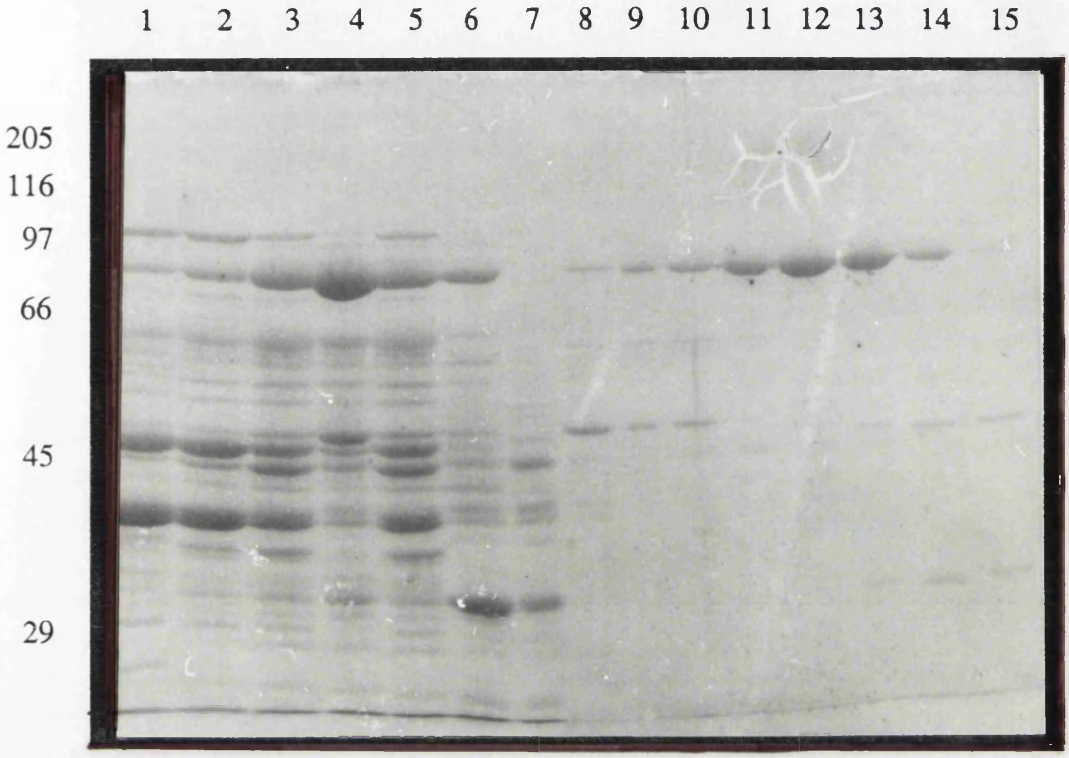


Figure 3.4. SDS-PAGE of samples from the Mono-Q and Superose-12 columns.

Fractions (*c.* 1.5 mls) eluting from a Mono-Q HR 10/10 column were assayed for activity. The three fractions with the highest activity were pooled. The pooled fractions (track 4) and some of the other fractions containing IDH activity, were analysed by SDS-PAGE (section 2.6d). Fractions from the Superose-12 column were also analysed.

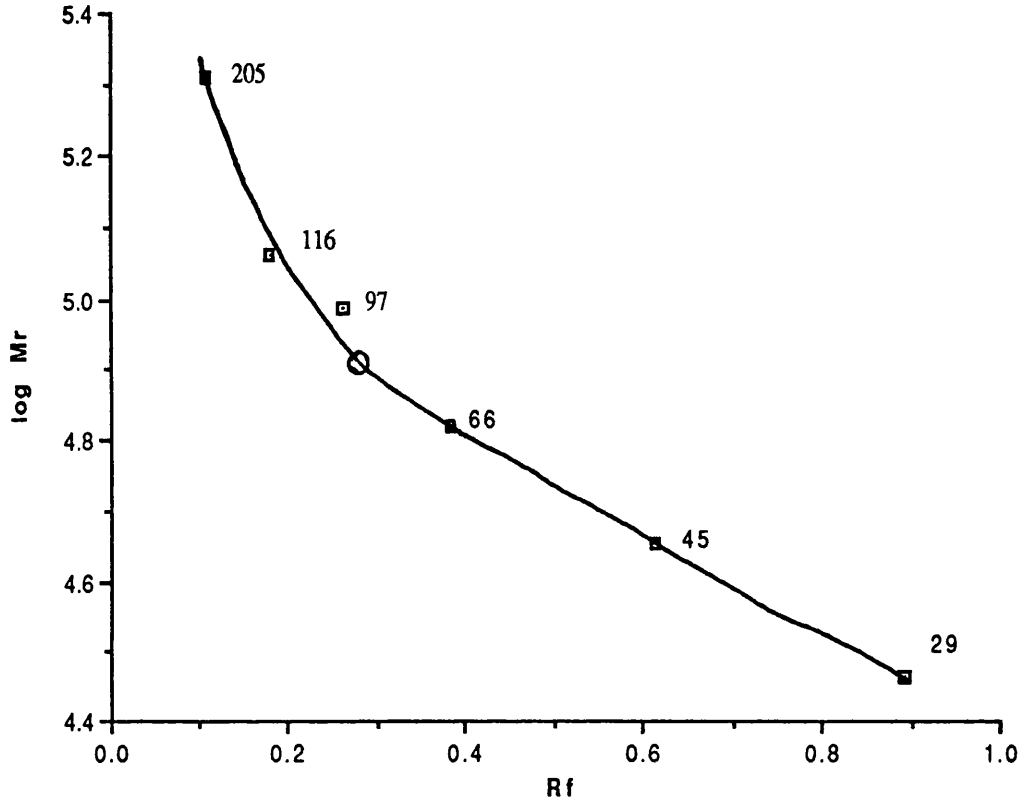
Approximately 8μg of protein was loaded into lanes 1-7 and 4μg into lanes 8-15.

Track	Fraction.	Elution volume.	IDH Activity (U/mg)	Track	Fraction.	Elution volume.	IDH Activity (U/mg)
1.	Mono Q	25.5	0.26	8.	Superose 12	12.8	0.01
2.	Mono Q	22.5	0.26	9.	Superose 12	12.5	0.04
3.	Mono Q	21	0.29	10.	Superose 12	12.2	0.04
4.	Mono Q	16.5+18+19.5	0.60	11.	Superose 12	11.9	0.14
5.	Mono Q	13.5	0.43	12.	Superose 12	11.6	0.24
6.	Mono Q	12	0.13	13.	Superose 12	11.3	0.12
7.	Mono Q	9	0.00	14.	Superose 12	11.0	0.05
				15.	Superose 12	10.7	0.01

Sample.	Total Activity (U)	Total Protein (mg)	Specific activity (U/mg)	Yield (%)	Purification Factor
Crude extract	367	1602	0.23	100	1
Resuspended 45-80% Ammonium Sulphate pellet.	308	269.8	1.14	84	5
Dialysed ammonium sulphate pellet	229	-	-	62	-
Pooled Mono-Q (HR10/10) fractions.	110	5.4	20.4	30	89
Concentrated Mono-Q fractions	84	-	-	23	-
Pooled Superose12 fractions	31	1.13	27.33	8	119

Table.3.5 Purification table showing main steps of purification.

Cells were grown for 60 hours on the complex media YEME and the purification described in section 3.2.2 carried out. The purification factor shows the increase in specific activity compared to the specific activity of the crude extract.



Key.
M_r is molecular weight in Daltons.
R_f is the mobilty of the markers defined as distance travelled by protein relative to dye front.
Numbers indicate the Mr of the standard molecular weight markers (section 2.11.3).

Figure 3.6 SDS PAGE calibration curve.

The gel in figure 3.4 was used to calculate the R_f values for the standard molecular weight markers (section 2.11.3) which had been seperated by SDS PAGE. The results were plotted against the logarithm of their molecular weight to produce the calibration curve above. The R_f of the *S.coelicolor* IDH was also calculated (shown as 'O' in the above graph) allowing calculation of its molecular weight (81 kDa). The curve was fitted by eye. Other experiments produced similar results (+/- 5kDa).

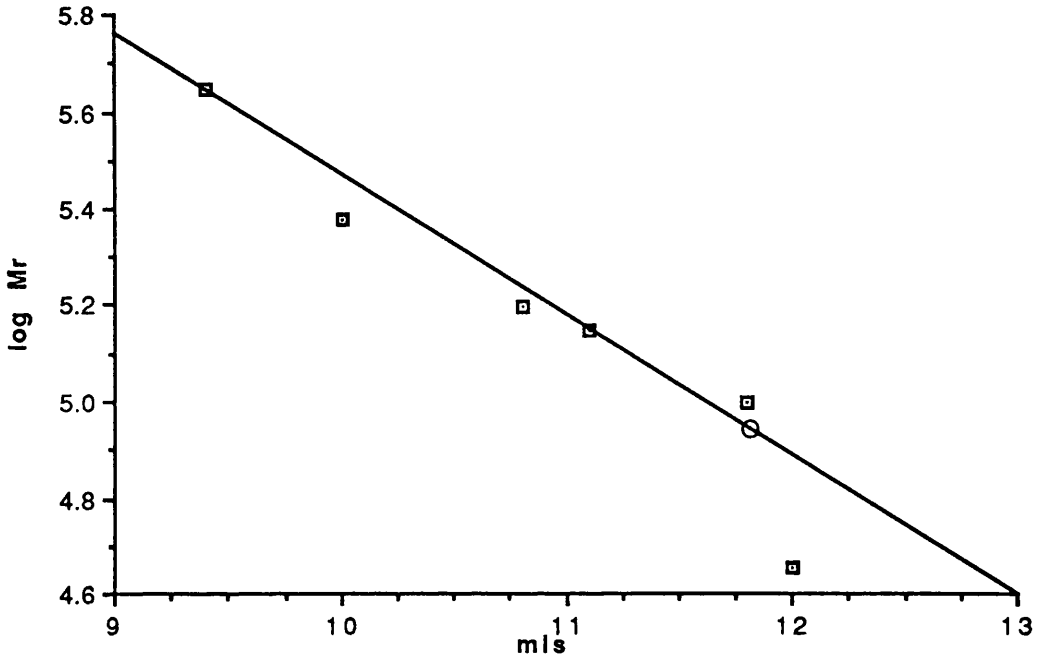


Figure 3.7 Standard curve for Superose 12 calibration.

Standard molecular weight markers (section 2.11.3) were loaded onto a Superose 12 column and the elution volume for each was determined. The average data from three runs are represented in the graph above. The elution volume of the *S.coelicolor* IDH protein was also determined in triplicate and the average value shown by the point 'O'. The average native molecular weight of IDH was calculated to be 89 kDa (+/- 5 kDa).

3.2.3 Native and subunit molecular weight determination.

The subunit molecular weight of the *S.coelicolor* IDH was estimated by comparing the mobility of the protein on SDS PAGE to the mobility of markers with known molecular weights (figure 3.6). The results of several such experiments showed that the *S.coelicolor* IDH has a subunit molecular weight of approximately 80,000 (+/- 5 kDa).

A Superose 12 column was used in order to determine the native molecular weight of the active form of IDH. Comparison of the elution volume of IDH to that of the markers (figure 3.7) shows that the *S.coelicolor* IDH has a native molecular weight of approximately 90,000 (+/- 5 kDa). This is similar to the value determined for the subunit molecular weight, indicating that the *S.coelicolor* IDH is a monomer. This is different to the situation in *E.coli*, where the enzyme is a dimer.

3.2.4 Western blotting.

Because the *S .coelicolor* and *E.coli* IDHs have such different subunit molecular weights it was of interest to determine if their was any similarity between them. Western blots of the *S.coelicolor* enzyme were probed with the *E.coli* anti-IDH antiserum (section 2.11.6). Samples from various stages of the purification were tested, as well as a sample of partially purified IDH from *Bacillus caldotenax* (a gift from C.Wild), whose subunit molecular weight is known to be similar to *E.coli*. The *E.coli* enzyme was also applied to the gel as a positive control. Figure 3.8 shows that the protein corresponding to the *S .coelicolor* IDH does not cross react with the antibody raised against the *E.coli* enzyme. However an *S.coelicolor* protein with a similar mobility to that of the *E.coli* and *B.caldotenax* IDHs does cross react. The intensity of the band bears no relation to the amount of IDH activity loaded This indicates that *S.coelicolor* contains a protein with some similarities to the *B.caldotenax* and *E.coli* IDHs, but it is not the IDH that has been the subject of purification in this work.

3.2.5 Attempt to phosphorylate the *S.coelicolor* IDH with the *E.coli* IDH kinase/phosphatase.

As already described in section 1.2.1, *E.coli* IDH activity is controlled by a reversible phosphorylation mechanism which is catalysed by the IDH kinase/phosphatase (IKP). We were interested to discover how the *S.coelicolor* enzyme may be controlled and whether it may be regulated by a similar mechanism to that occurring in *E.coli*. Attempts were therefore made to see if the *E.coli* IKP could inactivate the *S.coelicolor*

IDH *in vitro* . Conditions were used that were similar to those used for the inactivation of the *E.coli* enzyme *in vitro* (see section 2.11.5).

Figure 3.9 shows that the activity of the *S.coelicolor* enzyme was not affected by the kinase/phosphatase under the conditions used, whereas the control with the *E.coli* enzyme shows 82% inactivation after 12 mins. Neither enzyme showed any decrease in activity when incubated under identical conditions without the kinase/phosphatase.

This must indicate that there are significant differences between the IDHs from *S.coelicolor* and *E.coli* that prevents the IKP from recognising the *S.coelicolor* enzyme as a substrate. It is therefore unlikely that the *S.coelicolor* IDH is controlled by an IKP similar to that from *E.coli*. This does not however preclude the possibility that the *S.coelicolor* IDH may be controlled by a different phosphorylation mechanism or indeed some other mechanism of covalent modification.

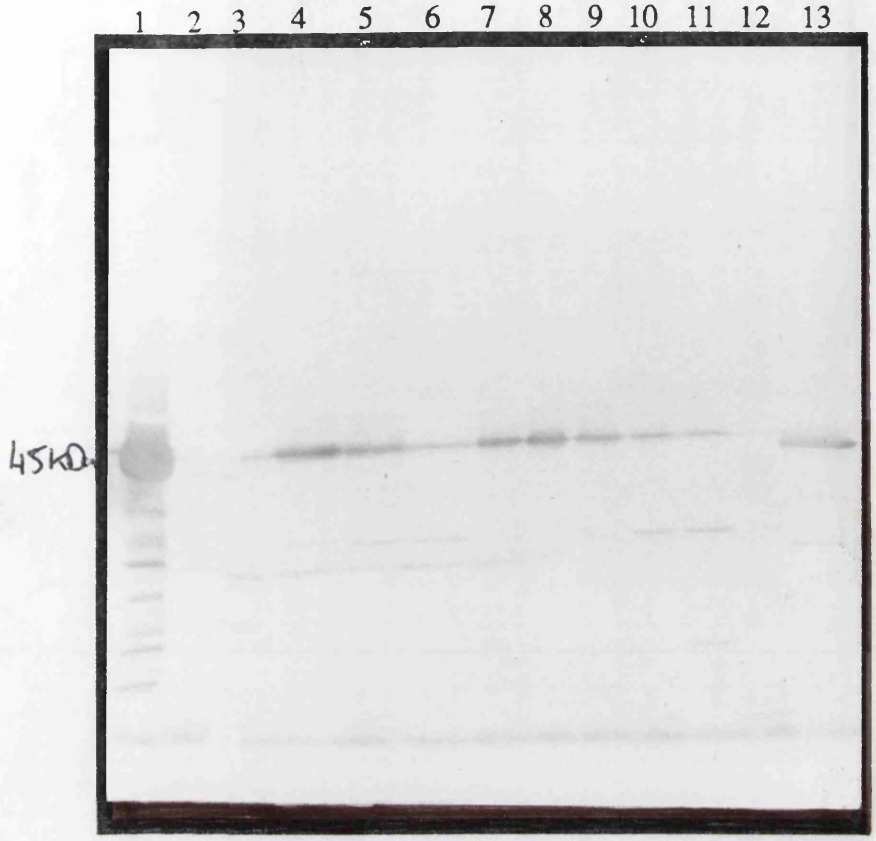


Figure 3.8 **Western blot of IDH enzymes probed with anti-*E.coli* IDH antiserum.**

Samples from various stages of the purification of the *S.coelicolor* IDH were analysed by SDS PAGE on a 10% gel. The *E.coli* enzyme was also loaded as a control as well as a partially pure extract from *Bacillus caldotenax* (gift from C.Wild). Protein was transferred onto nitrocellulose and probed with rabbit anti-*E.coli* IDH antiserum (section 2.11.6).

Lane 1	Partially purified IDH from *E.coli* ,	0.50 U.
Lane 2	Crude extract from *S.coelicolor*	0.07 U.
Lane 3	Ammonium Sulphate (45-80%)	0.07 U.
Lane 4-6	Fractions from across the Mono Q peak of activity.	0.07 U.
Lane 7-11	Fractions from across the Superose 12 peak of activity.	0.07 U.
Lane 12	High molecular weight markers.	-
Lane 13	Partially purified fraction from *B.caldotenax*.	*c*.0.10 U

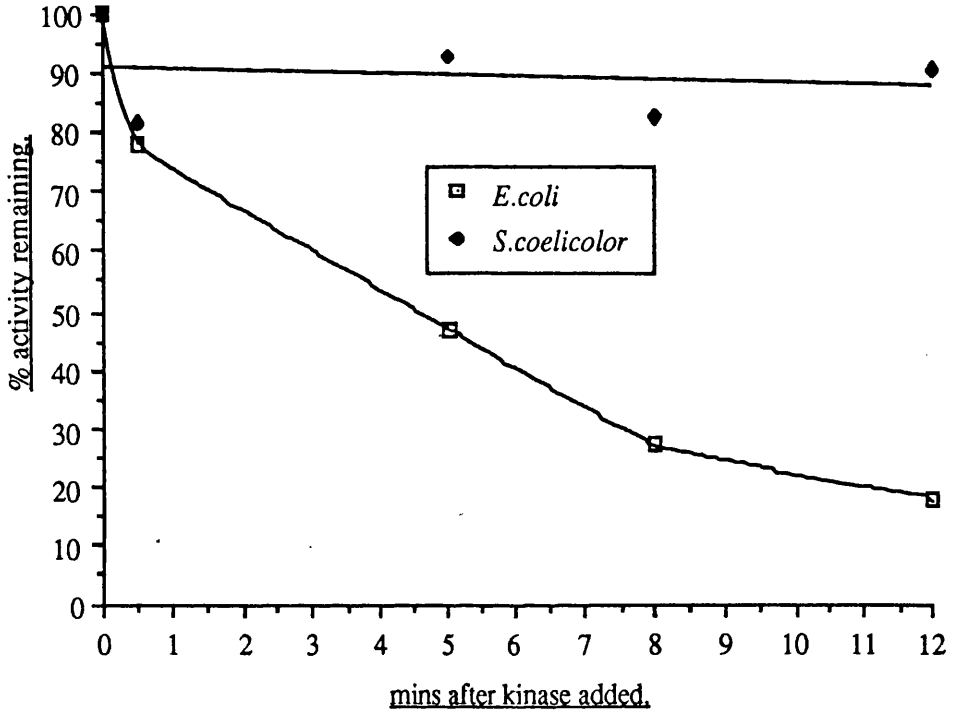


Figure 3.9 Effect on *S.coelicolor* IDH activity of Incubation with the *E.coli* IKP.

Active *S.coelicolor* IDH eluted from the final stage of purification was incubated for 60 mins with the *E.coli* IKP (IDH kinase/phosphatase) as described in section 2.11.5. The activity of the enzyme was measured at various time points. An identical experiment was performed with the *E.coli* IDH as a control. Approximately 700 mU of IDH was added to the reaction mixture in each case

3.2.6 K_ms of IDH for $NADP^+$, isocitrate and Mn^{2+}.

Initial attempts to measure K_ms of IDH were made using a spectrophotometer (section 2.11.1a). The values for both isocitrate and $NADP^+$ were in the range 1-2 μM. At these low K_m values it was neccessary to repeat experiments using a spectrofluorometer (section 2.11.1b) which has greater sensitivity for detecting NADPH. The K_m for Mn^{2+} was large enough to measure using the spectrophotometric measurements (see below). All reactions were initiated by the addition of isocitrate and the initial reactions measured at least in triplicate.

The K_m was then calculated by the method of Lineweaver and Burk (1934) (see figures 3.10-3.12). Mean K_m values for isocitrate, $NADP^+$ and Mn^{2+} of 1.30 μM , 2.06 μM and *c.* 13 μM respectively.

The specific activity of the purified IDH assayed with saturating concentrations of Mn^{2+}, isocitrate and $NADP^+$ gave a maximum rate of reaction (Vmax) of 36 U per milligram of protein. The turnover number for the enzyme has therefore been calculated to be 2880 min^{-1}.

3.2.7 Screening for inhibitors and activators of IDH activity.

IDH from *E.coli* is not allosterically controlled and seems to be entirely controlled by the effects of the kinase/phosphatase (see Chapter 1). It was of interest to find out if the *S.coelicolor* enzyme is controlled by a mechanism similar to that employed by *E.coli* or if it is controlled by the effects of metabolites on enzyme activity. Its response to various metabolites also provides further data for comparison to IDHs from other organisms.

Any of the metabolic intermediates from glycolysis and the TCA cycle could be potential inhibitors of IDH enzyme activity. Other candidates include those molecules that are known to act as signals of the energy state of the cell such as ADP, ATP, AMP and cAMP and those indicating the reducing capacity of the cell, namely NAD^+, $NADP^+$, NADH and NADPH.

Assays were performed in the presence of 1 mM of the metabolites to be tested under conditions similar to those used for the determination of the K_ms of the enzyme. Isocitrate and $NADP^+$ were present at K_m concentrations to ensure that the enzyme activity was less than V_{max} allowing changes in activity to be observed.

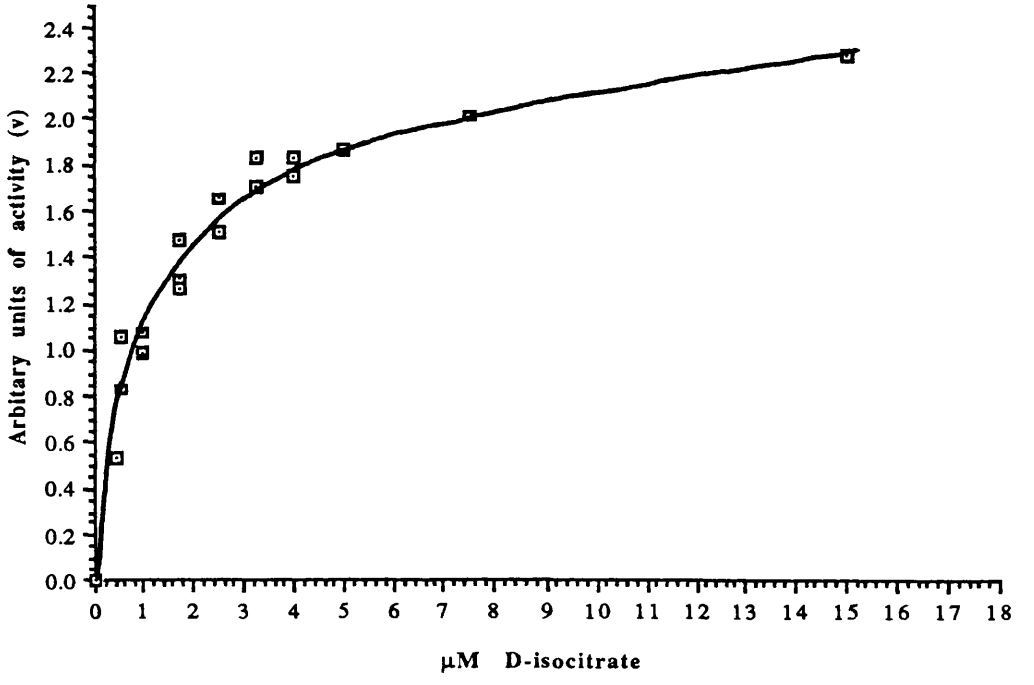
It can be seen from figure 3.13 that all of the metabolites tested had some sort of inhibitory effect on the enzyme activity, although it is doubtful that many of them would be significant at the concentrations to be found in the cell (see section 3.3.4).

3.2.8 N-terminal sequence determination.

The purified enzyme was exhaustively dialysed against 0.5% ammonium bicarbonate before being freeze dried under vacuum. Material was then sent to Mr Brian Dunbar at Aberdeen University for sequencing by automated Edman degradation on an Applied Biosystems gas phase sequencer. Twenty three amino acids were determined although the first three were ambiguous. This was probably caused by interference by the ammonium bicarbonate. Protein was therefore purified again from a different experiment and this time dialysed against distilled water for sequencing at ICI Pharmaceuticals (Alderly Edge) by Ms J.Young.

This attempt produced 24 amino acids confirming the original sequence and giving the first three amino acids unambiguously. Figure 3.14 summarises the yields obtained for each cycle. Comparison of this sequence to that of the IDH-II from *Vibrio spp.* ABE-1 showed that the two sequences are 69% identical whereas there was no significant similarity to the *E.coli* sequence (see section 3.3.2 and figure 3.15).

a)



b)

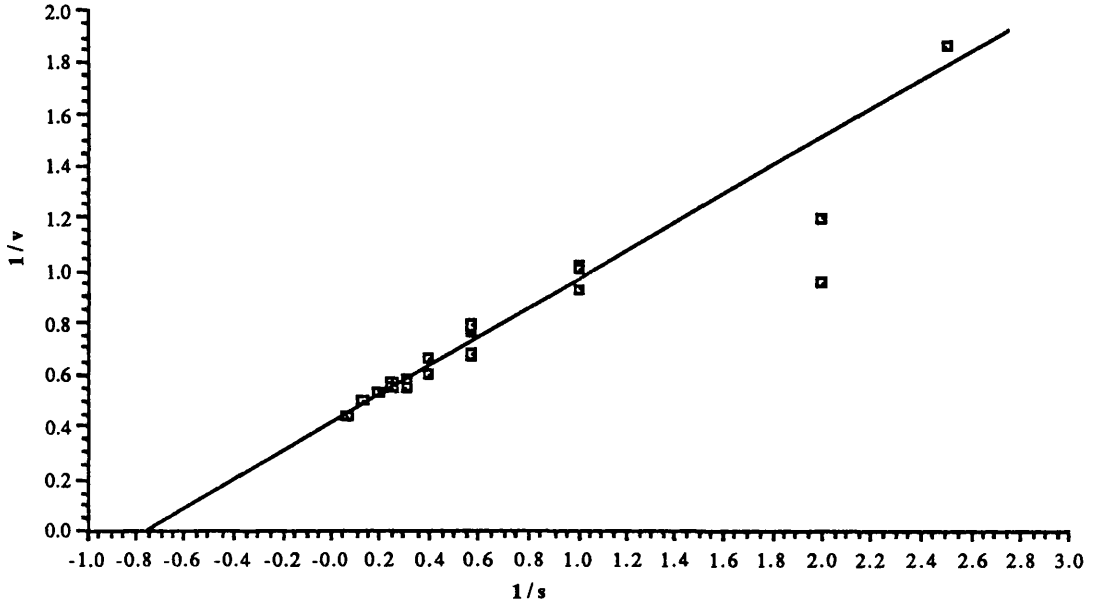
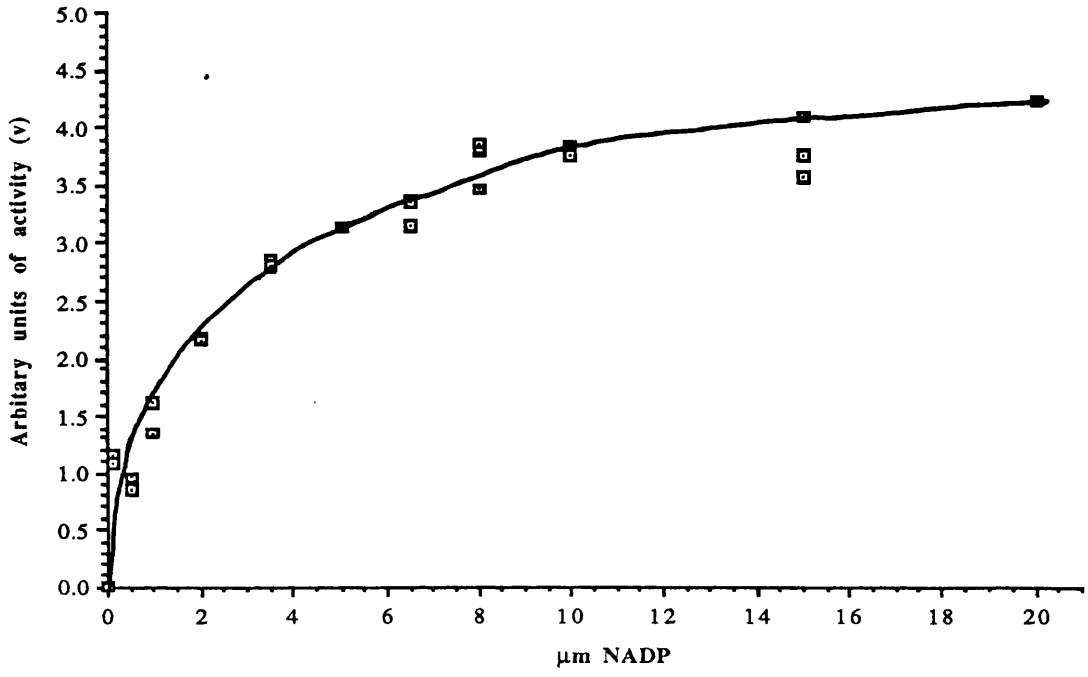


Figure 3.10. Determination of the K_m of IDH for isocitrate.

a) Shows a direct plot of IDH activity (where 'v' is an arbitrary activity unit measured on the spectrofluorometer; section 2.11.1b) against concentration of D-isocitrate. K_m was defined as being the concentration of D-isocitrate required to give an enzyme activity of half V_{max}.

b) The results shown in 3.10a were displayed as a double reciprocal plot (Lineweaver-Burk plot). The line intercepts the horizontal axis at $-1/K_m$. All curves and lines were drawn by eye. 's' is the substrate concentration in μM.

a)



b)

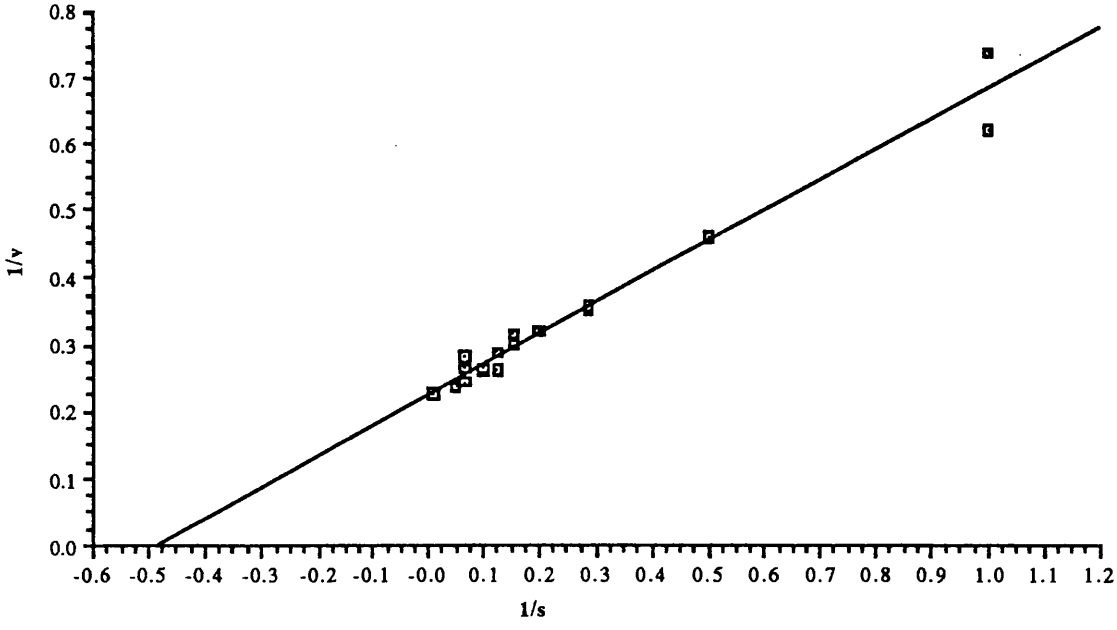
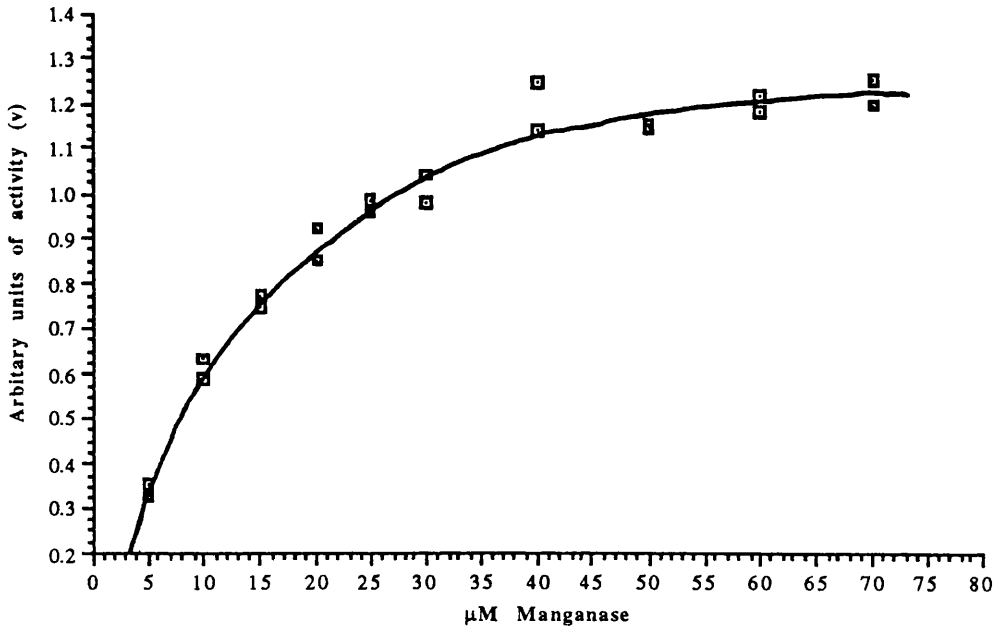


Figure 3.11. Determination of the K_m of IDH for $NADP^+$

a) Shows a direct plot of IDH activity (where 'v' is an arbitrary unit of activity measured on the spectrofluorometer; section 2.11.1b) against concentration of $NADP^+$. K_m was defined as being the concentration of $NADP^+$ required to give an enzyme activity of half V_{max}.

b) The results shown in 3.11a were displayed as a double reciprocal plot (Lineweaver-Burk plot). The line intercepts the horizontal axis at $-1/K_m$. All curves and lines were drawn by eye. 's' is the substrate concentration in µM.

a)



b)

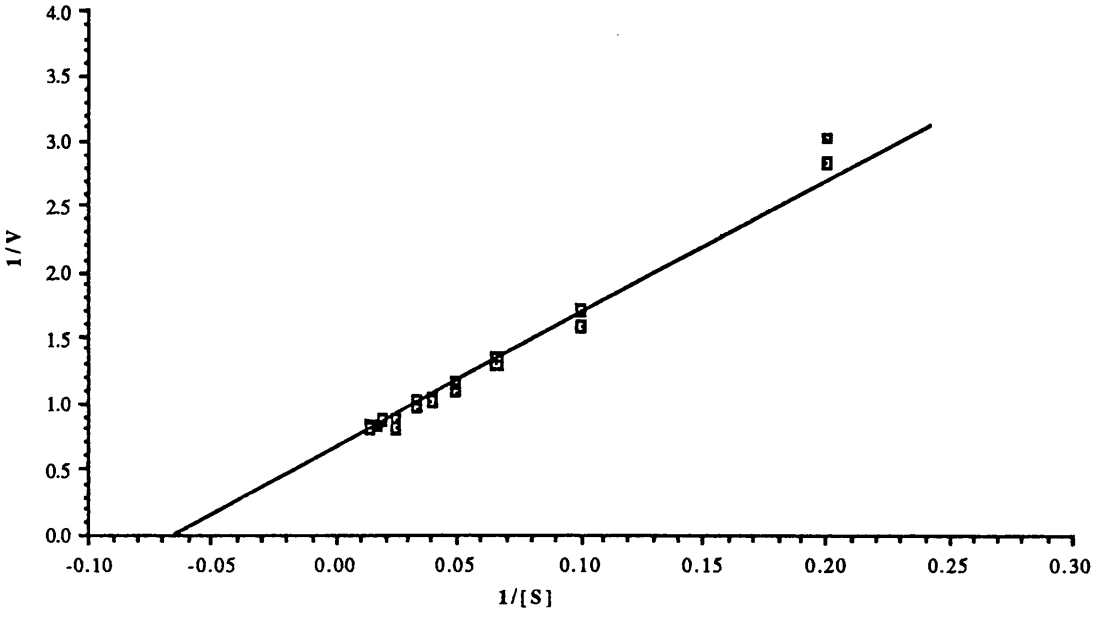


Figure 3.12. Determination of the K_m of IDH for Mn^{2+}

a) Shows a direct plot of IDH activity (where 'v' is an arbitrary unit of activity measured on the spectrophotometer ; section 2.11.1a) against concentration of Mn^{2+}. K_m was defined as being the concentration of Mn^{2+} required to give an enzyme activity of half V_{max}.

b) The results shown in 3.12a were displayed as a double reciprocal plot (Lineweaver-Burk plot). The line intercepts the horizontal axis at $-1/K_m$. All curves and lines were drawn by eye. 's' is the substrate concentration in µM.

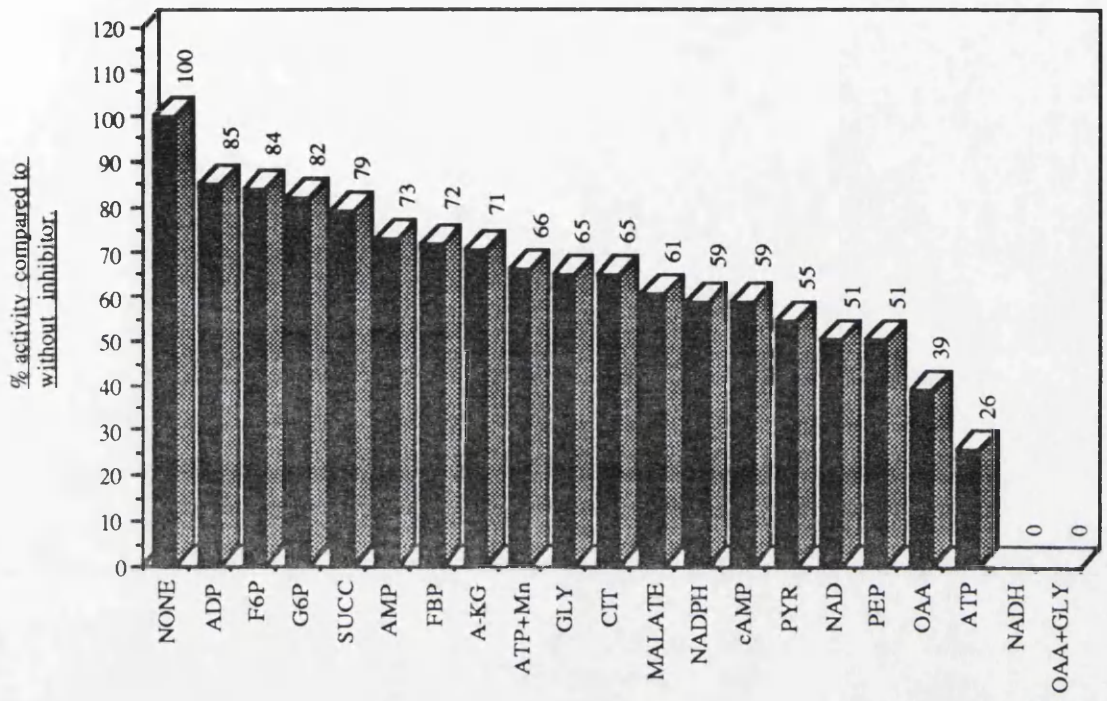


Figure 3.13 Effect of TCA cycle and glycolytic intermediates on IDH activity.

Metabolites were tested for their ability to increase or decrease *S.coelicolor* IDH activity compared to controls performed without metabolite. Assays were carried out in 150 mM TrisCl (pH 7.6) at 30°C, in the presence of 1 mM $MnCl_2$, 1.5 µM isocitrate and 2 µM $NADP^+$. Metabolites were used at a concentration of 1mM. All assays were performed on a spectrofluorometer (section 2.11.1b). The assay with ATP in the reaction mixture shows a reduction in IDH activity of 74%. However ATP is known to chelate $MnCl_2$, perhaps explaining this reduction in activity. Another assay was therefore performed in the presence of ATP with another 1mM $MnCl_2$, bringing the final concentration to 2 mM (labeled ATP+Mn).

Cycle	1	2	3	4	5	6	7	8	9	10	11	12
Amino acid	T	D	S	T	I	I	Y	T	H	T	D	E
pmols	134	180	188	88	170	154	124	95	50	34	94	23

Cycle	13	14	15	16	17	18	19	20	21	22	23	24
Amino acid	A	P	A	L	A	T	Y	S	F	L	P	V
pmols	83	47	45	40	22	18	27	10	19	20	7	15

Figure 3.14 N-terminal amino acid sequence of the *S.coelicolor* IDH.

Approximately 1 nmol of purified IDH was applied directly to an "Applied Biosystems" gas phase sequencer (section 3.2.8). The identity of the amino acid derivative produced by each cycle is shown using the single letter code for amino acids. The yield of each amino acid derivative is also shown.

Type I IDHs.

```
a) -TNLIIIPTTGDKITFIBGKLSVPNGPIIPYIZGDGIGVD
b) MZSLVVVPAQGLLITLQNGKLSVPGNPIIPYIZGDGIGVD
c) ---------MQGEKITVQNGVLNVPNNPIIPFIEGDGTGPD
d) ---------PLITTETGKKMHVLEDGRKLITVIPGDGIGPE
e) MYEKLQPPSVGSKITFVAGK
```

Type II IDHs.

```
f) -NEAPTIVWTRTDESPALASYSLLPIVQAFTETAGVSVVQ
g) -TD-STIIYTHTDEAPALATYSFLPVVRAYASQAGVAVET
h) STDNSLIIYTITD
```

Key.

Type I IDHs.

a) *Vibrio spp.* IDH-I
Fukanaga *et al.*, 1988.

b) *E.coli* IDH
Thorsness and Koshland, 1987.

c) *B.caldotenax* IDH
C.Wild, personal communication.

e) *T.thermophilus* IDH.
Miyazaki *et al.*, 1992.

d) *Synechocystis spp* .IDH
Muro-Pastor and Florencio, 1992.

Type II IDHs.

f) *Rhodomicrobium vannielii* IDH
Leyland and Kelly, 1991.

g) *S.coelicolor* IDH
This work.

h) *Vibrio spp* IDH-II
Fukanaga *et al.*, 1988.

Figure 3.15 Comparison of the N-terminal sequences from type I and II IDHs.

Unless otherwise stated amino acid sequences were obtained *via* direct sequencing of the proteins (see references). However residues 24-38 of the *S.coelicolor* sequence were predicted from the DNA sequence of the gene encoding IDH in*S.coelicolor* (see figure 5.2). Similarly residues 15-34 of the *B.caldotenax* IDH were predicted from the DNA sequence (C.Wild personal communication). The IDH from *T.thermophilus* was obtained entirely from the DNA sequencing of the gene.

IDH Class	Organism	Native Mr (kda)	Km isocitrate (μM)	Km $NADP^+$ (μM)	Vmax (U/mg)	pH	Temp (°C)	Buffer	Reference.
I	*Vibrio spp.* IDH-I	88 (D)	33	-	24.3	8.4	40	33mM Tris	Ochiai (1979)
I	*A.calcoaceticus* IDH-I	100 (D)	26	54	-	8	25	100mM Tris	Self and Weitzman (1972)
I	*T.thermophilus*	115 (D)	8.8	6.3	77.6	7.8	55	50mM Phophate	Eguchi *et al.* (1988)
I	*P.variotii*	99 (D)	12	4.4	45.2	7.5	30	80mM TES	Takao *et al.*, (1986)
I	*R.spheorodes*	105 (D)	8.7	7	45	7.0	30	50mM Phosphate	Chung and Braginski (1972)
I	*B.stearothermophilus*	92.5 (D)	1	3	34	7.5	36	20mM TEA	Howard and Becker (1970)
I	*E.coli*	80 (D)	10*1	5*1	216	7.5	37	150mM Tris	Borthwick *et al* (1984 c)
I	*Synechocysis spp.*	112 (D)	5.7	6	*c.*16	7.5	30	50mM Phosphate	Muro-Pastor and Florencio (1992)
II	*S.coelicolor*	90 (M)	1.3	2.05	36	7.5	30	50mM Tris	This Work
II	*Vibrio spp.* IDH-II	81 (M)	25	-	59.2	8.4	20	33mM Tris	Ochiai *et al.*, (1979)
II	*A.vinelandii*	78 (M)	18	18	722	7.5	37	50mM Tris	Barrera and Jurtshuck 1979
II	*Rhodomicrobium vannelii*	75-80 (M)	-	2.5	23.8	8.0	30	20mM Tris	Leyland and Kelly, 1992
ND	*A.niger*	60 (M)	77	20	14.4	7.0	28	50 mM Phosphate	Meixner-Monori *et al.*, 1986
ND	*A.calcoaceticus* IDH-II	300 (M)	13	100	-	8	25	100 mM Tris	Self and Weitzman (1972)

Figure 3.16 Comparison of some characteristics of other NADP-IDH's

Some of the characteristics of other bacterial NADP-IDH's are shown above as well as two examples of eukaryotic NADP IDH's, for comparison to the IDH from *S.coelicolor*.

The letters D or M in brackets indicates whether the enzyme is a dimer or a m onomer respectively. ND indicates that the class of enzyme that this enzyme belongs to has not been determined. TEA= Triethanolamine. TES= N-Tris(hydroxymethyl) methyl-2aminoethanesulphonic acid. HEPES=N-[Hydroxylethyl]piperazine-N'[2-ethanesulphonic acid].

***1** The K_ms for NADP and isocitrate were calculated from the data presented in figure 2a and b in Nimmo (1986).

3.3 Discussion.

3.3.1 Purification and physical characteristics.

IDH activity was purified from cell free extracts by a three step procedure that incorporated the use of ammonium sulphate fractionation, Mono Q ion exchange chromatography and gel filtration on a Superose 12 column. Use of the FPLC facilitated a rapid and convenient purification that could be performed within 48 hrs and produced increases in specific activity of up to 130 fold with up to a 30% yield.

Many of the IDHs that have been purified to date have employed reactive dye columns. Attempts were made to use one of these columns, Procion Red sepharose, with which activity could be eluted using 0.5 mM $NADP^+$ but not 0.5 M KCl. This may be a useful step, if optimised, for further purification of the enzyme should it be required. Another potentially useful step that has not been optimised was the use of hydrophobic column chromatography using either phenyl sepharose or the FPLC equivalent, Phenyl superose. Protein bound to these columns if the sample was made to 1.7 M with respect to ammonium sulphate, and applied to a column equilibrated in the same buffer. Protein with IDH activity was eluted using a gradient of decreasing concentration of ammonium sulphate.

SDS-PAGE analysis of the purified IDH protein showed that its subunit molecular weight is 80,000. The similarity to the value obtained for the native molecular weight (90,000) indicates that the *S.coelicolor* enzyme is a monomer similar to the IDH-II from *Vibrio spp.* (Ochiai *et al.*, 1984) and the IDH from *Rhodomicrobium vannielii* (Leyland and Kelly, 1991).

3.3.2 Classification of IDHs into distinct groups.

3.3.2.1 Type I IDHs.

It has been suggested by Fukanaga *et al.* (1988), that IDHs from *E.coli, Salmonella typhimurium* and *Vibrio spp.* can be grouped together by virtue of their subunit organisation and cross reactivity with antibodies. Antibodies raised against the *Vibrio spp.* IDH-I enzyme cross react with IDH from *E.coli* and the close relative of *E.coli, S.typhimurium* (Fukanaga *et al.*, 1988). It has now been shown that another dimeric IDH, this time from *Bacillus caldotenax* , can also cross react with the anti-*E.coli* IDH antibody (figure 3.8).

Comparison of the N-termini of the IDHs from *B.caldotenax, E.coli, Synechocystis spp.* and *T.thermophilus* and the IDH-I from Vibrio *spp.*, provides further evidence for the classification of these enzymes into a group of related enzymes (figure 3.15). I have therefore called this family the class I enzymes, after the IDH-I from *Vibrio spp.*

Phosphorylation of IDH has been demonstrated *in vivo* in *E.coli* (see section 1.2.1) and *S.typhimurium* (Wang and Koshland, 1982). The *B.caldotenax* IDH can be inactivated *in vitro* by phosphorylation using the *E.coli* IKP (Catherine Wild, Department of Biochemistry, Glasgow University, personal communication). This may suggest that many of the dimeric, type I, IDHs may be controlled by a phosphorylation mechanism.

3.3.2.2 Classification of the *S.coelicolor* IDH as a "type II IDH".

Polyclonal antiserum raised against the monomeric form of IDH from *Vibrio spp.* (IDH-II), showed no cross-reaction with the type I IDHs from *E.coli, S.typhimurium* or the IDH-I from *Vibrio spp.* (Fukanaga *et al.*, 1988). Based on this information and the monomeric structure of IDH-II, it seems that this enzyme belongs to a separate class of bacterial IDHs. Analysis of the N-terminal sequence of the *S.coelicolor* enzyme shows marked similarity to that of the *Vibrio spp.* IDH-II enzyme (see figure 3.15). It also has a similar quaternary structure to IDH-II and has been shown to not cross react with the anti-*E.coli* IDH antibody. It is clear that these two enzymes could be grouped into a second class of IDHs, that I will call the type II class, after the IDH-II from *Vibrio spp.*

Recent characterisation of the *Rhodomicrobium vannielii* (Leyland and Kelly, 1991) IDH supports this classification. This enzyme can also be classified as a type II IDH: it is a large monomeric protein (75-80 kDa) and possesses an N-terminal sequence very similar to both the *S.coelicolor* IDH and *Vibrio spp* IDH-II. Antiserum raised against

the *Rhodomicrobium vannielii* IDH was found not to cross-react with the *E.coli* IDH, further suggesting that the *E.coli* type I enzymes are dissimilar to the type II IDHs. Initial attempts to phosphorylate the *S.coelicolor* IDH using the *E.coli* IKP were unsuccesful indicating that the *S.coelicolor* IDH is not controlled by an *E.coli* -like phosphorylation mechanism. It will remain to be seen if this is another general property of type II enzymes.

In the absence of any western blotting data and N-terminal sequence information, it is difficult to say if the two *A.calcoaceticus* IDHs can also be classified in this way. It seems likely that the dimeric form may belong to the class I type of IDHs, having a similar subunit molecular weight. The second enzyme is reported to have an unusually large molecular weight of 300,000 and is unique in its stimulation by AMP and ADP. These properties make the enzyme more similar to the NAD^+ dependent enzyme found in higher organisms (Barnes *et al.* 1971) perhaps representing yet another class of IDH.

3.3.3 K_ms for $NADP^+$, isocitrate and Mn^{2+}

To see if enzymes could be classified further, based on their kinetic properties, we measured the K_ms of the *S.coelicolor* IDH for $NADP^+$, isocitrate and Mn^{2+} and compared them to the kinetic parameters of some other IDHs (compiled in figure 3.16). It would seem that there is little correlation between the class of IDH to which an enzyme belongs and the kinetic characteristics they possess. Direct comparisons are difficult though because of the wide range of assay conditions used. The K_m for $NADP^+$ for the type II IDH from *Rhodomicrobium vannielii* was however determined under very similar conditions and gave a similar value to that for the *S.coelicolor* IDH.

3.3.4 Inhibitors.

It was of interest to find out how the *S.coelicolor* IDH may be controlled. From preliminary experiments it seems unlikely that the *S.coelicolor* enzyme will be controlled by an *E.coli* like kinase/phosphatase. Not only is the enzyme different in its monomeric size and immunological properties but initial attempts to phosphorylate the enzyme using the *E.coli* kinase/phosphatase have proved unsuccessful. This does not however exclude the possibility of control by some sort of covalent modification catalysed by an enzyme with a different specificity.

It may be possible that the *S.coelicolor* IDH is allosterically controlled in a manner similar to that proposed for the *A.calcoaceticus* IDH-II (Self and Weitzman 1972 and Self *et al.*, 1973) and *P.variotti* (Takao *et al.* 1986). To investigate this possibility several metabolites were screened for any stimulatory or inhibitory effects. Most of the metabolites tested had some inhibitory affect on enzyme activity, although it is doubtful for most of them whether this is significant at the high concentrations used. However ATP, NADH, OAA and glyoxylate plus OAA did have a large effect on IDH activity.

Inhibition of enzyme activity by ATP has been found for IDHs from several organisms including *Azotobacter vinelandii* (Barrera and Jurtshuk, 1970), *A.calcoaceticus,* (Kleber and Aurich 1975) and *Aspergillus niger* (Meixner-Monori *et al.* 1986). Meixner-Monori *et al.* (1986) suggested that such effects are probably an artefact caused by the chelation of Mn^{2+} by ATP. For measurements with the *S.coelicolor* IDH, 1 mM ATP was found to inhibit enzyme activity by 61% when 1mM of Mn^{2+} was present. Addition of a further 1 mM of Mn^{2+} still produced an inhibition of 44%. One may expect that the use of 2 mM Mn^{2+} would negate any chelation effects of ATP (which is only at half the molarity), perhaps indicating that this effect is in fact real. Leyland and Kelly (1991) have reported similar effects. NAD and NADP-linked activities are inhibited by ATP with Kis of 1.3 mM and 4.4 mM respectively. Further studies are needed to be confirm the physiological significance of these effects.

OAA was found to inhibit the *S.coelicolor* enzyme activity by 60% at the 1 mM concentration used. This effect had only previously been reported for the IDH from *P.variotti.* (Takao *et al.*, 1986) whose activity was reduced by 70% using 1 mM OAA. Recently an inhibition of 44% has also been reported for the *Rhodomicrobium vannielii* IDH using 1 mM OAA (Leyland and Kelly, 1991).

S.coelicolor was also found to be completely inhibited by 1 mM glyoxylate plus 1 mM OAA. Such a potent concerted inhibition is not unusual and has been reported for $NADP^+$ dependent IDHs from mammalian sources (cited in Nimmo, 1986) and *A.niger* (Meixner-Monori *et al.*, 1986) as well as from all bacterial sources tested. The physiological significance of inhibition by either OAA alone or OAA plus glyoxylate is however doubtful when one considers the small inhibition likely to be caused by the low intracellular concentrations of these metabolites (Nimmo, 1986).

Inhibition by 1 mM NADH is a property of the *S.coelicolor* enzyme that has not been reported for any of the other $NADP^+$ dependant IDHs. NADH is required to provide reducing power for the generation of ATP *via* the electron transport chain. Transhydrogenases have been found in many of the organisms mentioned, including *S.coelicolor,* enabling them to reduce NAD^+ at the expense of NADPH (Ragland *et al*, 1966). It is thought that this activity enables organisms to use, indirectly, the NADPH produced by IDH for the production of ATP. Inhibition by NADH could therefore serve as a signal of excess reducing power which reduces the total flux through the TCA cycle. It is possible that the NADH acts as an analogue of $NADP^+$ which could bind to the co-factor binding site and thus prevent enzyme activity. This effect could once again be an artefact of the assay though. Further studies would need to be performed to determine if NAD^+ would still have an effect at a physiological concentration.

Some of these observations could have been investigated further although any data obtained would probably not provide any definitive information as to how IDH is controlled *in vivo*. It was therefore decided to progress onto the cloning of the gene as no gene has yet been studied that encodes a type II IDH.

Cloning of the gene will allow protein sequence information to be obtained from the genetic sequence. The cloned gene will also be a valuable tool for later analysis of gene function. This may provide valuable information as to how this type of enzyme and its gene is regulated and may thus help to prove the existence or absence of isoenzymes.

CHAPTER 4

Cloning of the *icd* B gene that encodes IDH in *S.coelicolor.*

Chapter 4. Cloning of the *icd* B gene that encodes IDH in *S.coelicolor.*

4.1 Introduction.

One of the aims of this project was to clone the gene encoding IDH in *S.coelicolor.* Cloning of this gene will enable us to use it as a tool for the study of expression levels and regulation at different times of the life cycle and will provide the means to overexpress the enzyme in *S.coelicolor.* Insertional inactivation experiments can also be performed using the cloned gene to create mutants with reduced or complete lack of IDH activity.

The most immediate benefit of cloning this gene would however be the ability to obtain the amino acid sequence of the protein for comparison to other proteins, *via* sequencing of the gene. When it was discovered that the major IDH activity in *S.coelicolor* was due to an enzyme that belongs to the type II class of IDHs, this aim became a high priority as no other gene of this type has been cloned. This gene will from now on be called *icd* B, in order to distinguish between this gene and those genes that code for type I enzymes (from now on referred to as *icd,* as named in *E.coli*).

As discussed in section 1.4 the best way to clone this gene in the absence of *S.coelicolor icd* B mutants, is by the so-called technique of reverse genetics which requires the construction of an oligonucleotide designed against the N-terminal amino acid sequence of the IDH protein. The oligonucleotide can then be used to screen for DNA containing the N-terminal region of the *icd* B gene. Two methods were used to select for such fragments of DNA.

One method involved the construction of a library in the lambda bacteriophage vector, λGEM-11, that fully represents the entire *S.coelicolor* genome. The oligonucleotide was then used to screen plaques to select for clones containing the N-terminal region of the gene.

The second method involved the screening of a partial library constructed with the plasmid vector pUC18. The library is called partial because it was constructed from only a small proportion of the fragments produced by restriction digestion of genomic DNA (gDNA). DNA was selected by screening digests of gDNA with the N-terminal oligonucleotide. DNA that hybrisises to the oligonucleotide was isolated by size selection and used to construct the partial plasmid library. This library was then screened using the same oligonucleotide to select for transformants that contained at least part of the *S.coelicolor icd* B gene.

Chapter 4. Cloning of the *icd* B gene that encodes IDH in *S.coelicolor.*

In this chapter the methods used to clone the gene are described, along with the construction of some of the subclones derived from this study. Attempts were also made complement a strain of *E.coli* that is devoid of any IDH activity.

4.2 Results.

4.2.1 Design of an oligonucleotide.

Amino acid sequence information obtained from the N-terminus of the purified IDH (see section 3.2.8) was used to design an oligonucleotide for use as a probe in subsequent screening procedures. The high G+C content of streptomycete DNA imposes a bias on the choice of codons used by the organism, with a strong preference being given to those that contain Gs or Cs in the third position of a codon. An oligonucleotide was therefore designed taking advantage of this bias using the streptomycete codon preference table compiled by Seno and Baltz (1989).

In designing an oligonucleotide regions that contain amino acids such as serines, leucines or arginines should be avoided, as these amino acids can be coded by any of six synonymous codons. Conversely, regions containing methionines and tryptophans were highly favoured as these amino acids are encoded by only one codon. Amino acids such as asparagine, lysine, isoleucine and aspartate are also favoured because they show greater than 90% preference for one codon. Where amino acids are encoded for by two codons with an almost equal frequency, mixtures of oligonucleotides can be synthesised that cover both possibilities. Taking these factors into account the oligonucleotide Oligo-N was synthesised with the sequence as shown in figure 4.1.

4.2.2 Screening of digested genomic DNA.

To determine which restriction enzymes would be useful for the construction of a partial plasmid library, genomic DNA was digested to completion using a variety of restriction enzymes and the fragments produced separated by agarose gel electrophoresis (see figure 4.2a).

DNA was transferred to Hybond-N nylon membrane by Southern blotting and the membrane then baked and prehybridised as described in section 2.26.1. After prehybridisation the radiolabeled oligonucleotide (oligo-N) was added to the bag containing the membrane and hybridisation allowed to proceed for 2 hours before the being washed three times (see section 2.26.2). Different hybridisation and washing conditions were tried, by gradually altering the salt concentration and temperature, until

```
a)  N -Thr Asp Ser Thr Ile Ile Tyr Thr His Thr Asp Glu Ala Pro Ala Leu Ala Thr Tyr Ser Phe Leu Pro Val-C
b)  5'-ACC GAC TCC ACC ATC ATC TAC ACC CAC ACC GAC GAG GCC CCC GCG CTG GCC ACC TAC TCC TTC CTG CCC GTC-3'
       73% 92% 37% 73% 92% 92% 93% 73% 89% 73% 92% 78% 60% 49% 60% 56% 60% 73% 93% 37% 97% 56% 49% 56%
c)                3'-TGG TAG TAG ATG TGG GTG TGG CTG CTC CGG GGG  CG- 5'
                                                                  C   C
```

Figure 4.1 Design of oligonucleotide used when screening for *icd* B.

a) N-terminal amino acid sequence obtained by automated Edman degradation of purified IDH from *S.coelicolor* (see section 3.2.9).

b) The codons most frequently used by streptomycetes to encode each of the amino acids in the sequence are shown. This constitutes a best guess for the sense strand encoding *icd* B. The frequencies with which each codon is used in streptomycetes is shown below the sequence. Codon frequencies were taken from the review of Seno and Baltz (1989).

c) The oligonucleotide, oligo-N, was designed to hybridise to the sense strand of *icd* B. The sequence was therefore the complement of the sense strand (*ie.* an antisense probe).

Proline is encoded by CCC with a frequency of 49% in genes from *Streptomyces spp.* and by CCG with a frequency at 43%. It is therefore difficult to predict with confidence which codon is likely to be used in the *icd* B gene. To allow for both possibilities, the oligonucleotide was constructed with a degeneracy at the third position of the codon (indicated above). A similar situation occurs for alanine which is encoded by GCG with a frequency of 60% and by GCC with a frequency of 35%. A degeneracy was therefore also included at the third position of this codon (indicated above). The addition of the degeneracies increases the chance of picking the correct codon to 92% for proline and 95% for alanine.

conditions were found in which only a single hybridising fragment was found for each restriction digest (see figure 4.2b and c).

The conditions which allowed visualisation of a unique hybridising DNA fragment were; hybridisation at 55°C, in 5xSSC and washing at 65°C in 2xSSC.

4.2.3 Cloning of *Sal* I gDNA fragments into pUC18.

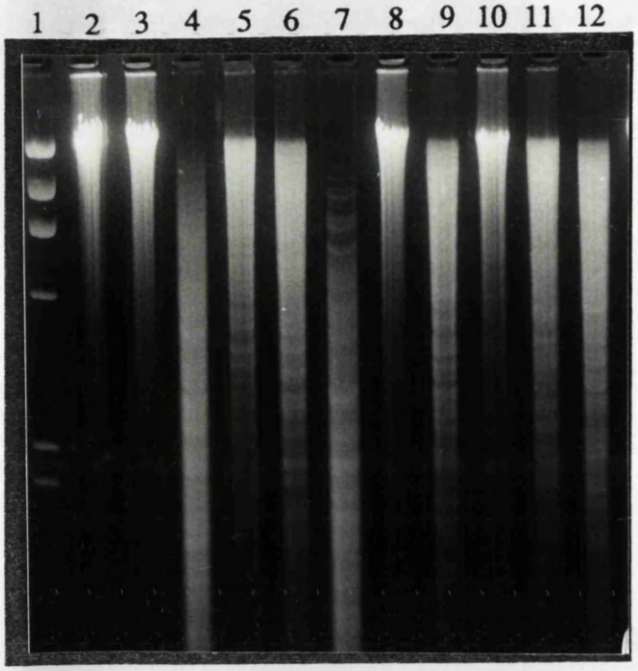
As the *S.coelicolor* IDH-II is 80-90 kDa, the maximum size of the *icd* B gene will be 2.7kb, assuming an average molecular weight for an amino acid of 100 Da. Most of the restriction digests produced a hybridising fragment that was either too small to contain the entire gene or were too large to conveniently clone into plasmids (see figure 4.2 b and c). However *Sal* I and *Pst* I produced 5 kb hybridising restriction fragments, which are large enough to potentially contain the whole *icd* B gene and yet is a convenient size for introduction into a plasmid vector.

In a scaled up reaction *Sal* I was used to digest 50 μg of gDNA to completion. Fragments were then separated by agarose gel electrophoresis using 0.8% (w/v) TAE gels. A slice of gel that contained DNA fragments in the size range of 4.5 kb to 5.5 kb was then excised and the DNA extracted by electroelution and purified as described in section 2.17.9.

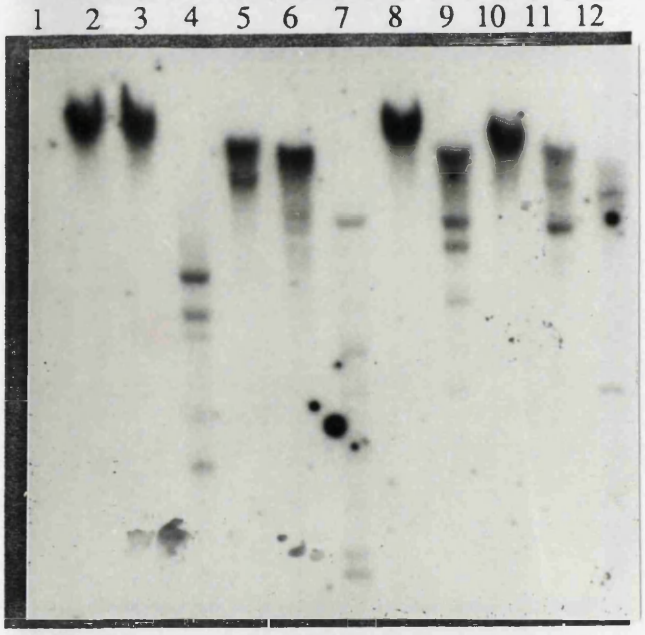
A ligation was then set up containing *c.*40 ng of the size selected DNA and 100ng of linearised pUC18 vector This vector was supplied by Pharmacia, already linearised by *Sal* I and treated with calf intestinal alkaline phosphatase. This treatment removes the phosphate group from the 5' ends of the linearised molecule, preventing both the intramolecular and intermolecular ligation of vector molecules. The efficiency of removal of the phosphate from the 5' ends of the vector meant that 99% of colonies obtained after transformation were expected to be recombinants.

Varying proportions of the ligation mix were used to transform a culture of *E.coli* DS941 that had been made competent by the method of Hanahan (see section 2.18.1a). Colonies that contained recombinant plasmids (detected by the white colour of colonies plated on media containing X-gal; section 2.18.3) were then picked for subsequent screening by hybridisation to oligo-N.

a)



b)



c)

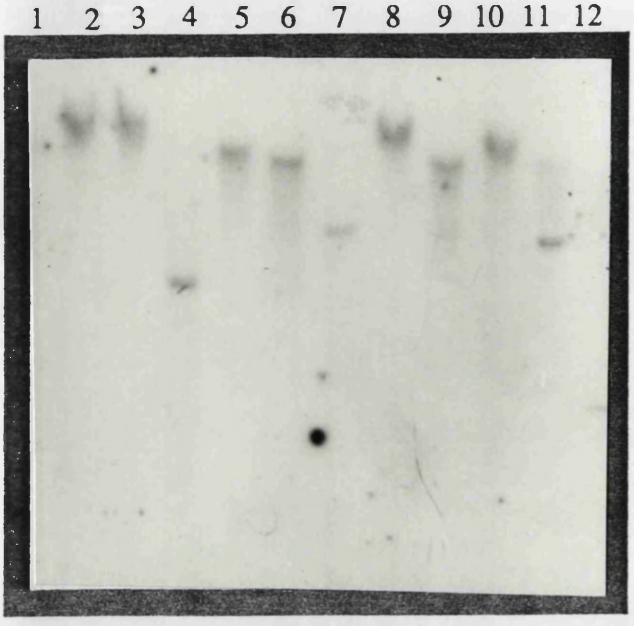


Figure 4.2 **Restriction digests and southern blotting of genomic DNA.**

a) Genomic DNA (6 µg) was digested for 2 hours at 37°C (except for *Sma* I which was digested at the recommended temperature of 25°C) using 20 units of each of the restriction enzymes shown below. Fragments were separated by electrophoresis through a 0.8% (w/v) agarose gel overnight (2 volts per cm length of gel) and photographed with UV illumination (section 2.17.7).

b) DNA was transferred to Hybond-N for hybridisation with oligo-N (section 4.2.2). Hybridisation was at 55°C (with 5xSSC buffer) for 2 hours. Washing was at 60°C (5xSSC). Several bands can be seen in each lane, showing that conditions are not stringent enough to allow specific hybridisation to the *icd* B gene alone.

c) The same filter (shown in b) was subjected to further washing at 65°C in 2xSSC. Only one DNA fragment is seen to hybridise to oligo-N in each lane showing that specific hybridisation conditions have been reached. Autoradiography was for 12 hours at -70°C (section 2.17.10). The size of each band is shown below. The size of each DNA fragment was calculated by comparison to the mobility of the λ *Hind* III DNA markers (section 2.17.1f) using a calibration graph (see section 2.17.7).

1. λ *Hind* III Markers	(see section 2.17.1f)	7. *Sal* I	(5 kb)
2. Undigested	(>23 kb)	8. *Hind* III	(>23 kb)
3. *Xba* I	(>23 kb)	9. *Kpn* I	(9.42-14 kb)
4. *Sma* I	(2.9 kb)	10. *Eco* RI	(>23 kb)
5. *Sst* I	(11-23 kb)	11. *Pst* I	(4.5 kb)
6. *Sph* I	(9.42-14 kb)	12. *Bam* HI	(1.6 kb)

4.2.4 Screening of partial plasmid library.

Three hundred of the transformants were picked and plated onto L-agar containing ampicillin as a 10x10 array on three plates (100 colonies per plate). Duplicate arrays were also made on Hybond-N and the colonies grown overnight before being treated ready for hybridisation (see section 2.27). Radiolabeled oligo-N was used as the probe and hybridisation and washing conditions were similar to those determined in section 4.2.2., although washing was initially performed at only 60°C

Autoradiography of the filters revealed two colonies that hybridised strongly to oligo-N, whilst two colonies were found to hybridise comparatively weakly (Figure 4.3 a). Each of these clones were then streaked to single colonies on L-amp plates and 10 purified colonies of each type plated as triplicate arrays for re-screening. Washing was this time performed at 65°C. This time only the recombinants DS941/pRT60 and DS941/pRT84 showed significant hybridisation to oligo-N (see figure 4.4b).

Plasmids isolated from these two recombinants were mapped with restriction enzymes and found to contain identical inserts in the same orientation relative to the vector. All further experiments were performed using the isolate DS941/pRT60.

4.2.5 Restriction mapping of pRT60.

pRT60 was digested with several restriction enzymes that were known to cut in the pUC18 polylinker. Fragments were separated by agarose gel electrophoresis and the sizes of the fragments noted. DNA fragments were then transferred to Hybond-N membrane by Southern blotting and oligo-N used in hybridisations to determine which fragments contained the oligo-N binding site. In this way a restriction map of the plasmid was deduced and the region containing the 5' end of the gene was determined (see figure 4.4a).

The oligonucleotide binding site was within the 0.8 kb *Bam* HI/*Sal* I fragment, on the extreme right of the insert (with respect to figure 4.4). As the oligonucleotide binds to the end of the gene that encodes the N-terminus of IDH there were two possibilities. If the rest of the *icd* B ORF lay to the left of this 0.8 kb fragment then the entire *icd* B gene would be contained within this 5 kb *Sal* I fragment. However, if the rest of the gene lay to the right of this oligonucleotide binding region only a part of the N-terminal region of *icd* B would have been cloned.

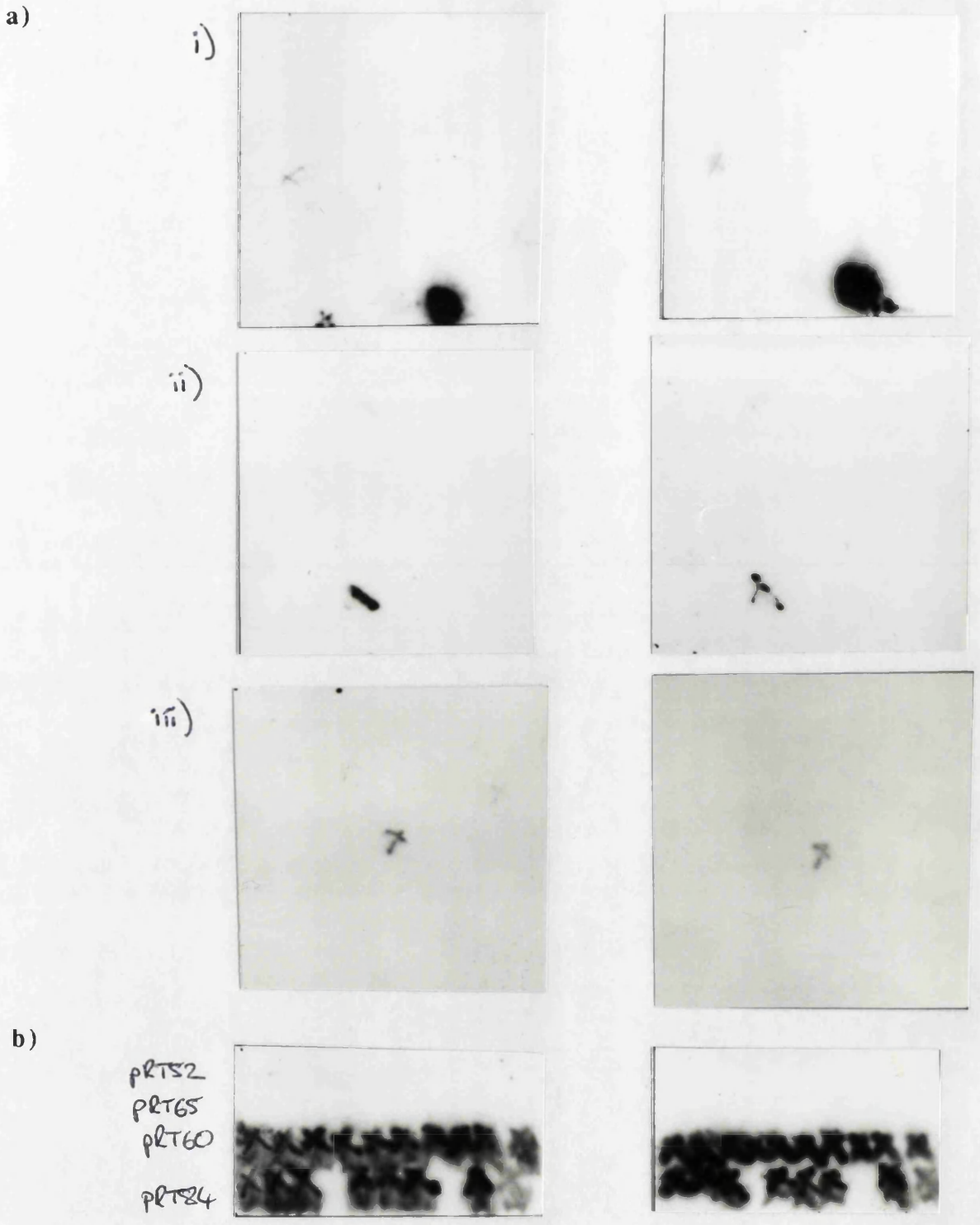


Figure 4.3. **Colony hybridisation.**

a) 300 *E.coli* colonies were grown on three seperate plates. DNA from the colonies was then transferred onto Hybond N in duplicate and the filter probed using oligo-N (section 4.2.4). Hybridisation was performed at 55°C in 5xSSC. Washing was at 60°C in 2xSSC. Autoradiography was for 2 hours at -70°C. The duplicate filters from each of the three plates (i to iii) is shown.

i) The colony showing the most hybridisation to oligo N was called pRT 60. The only other colony showing hybridisation in duplicate was called pRT 52.

ii) The colony hybridising to oligo N was called pRT 84.

iii) The colony hybridising to oligo N was called pRT 65.

b) The positive clones obtained in the primary screening were re-screened (section 4.2.4). Only pRTB 60 and pRT 84 still showed hybridisation when filters were washed at 65°C in 2xSSC. pRT 60 was used for further study.

4.2.6 Orientation of the *icd* B ORF.

To determine whether the whole gene had been cloned it was essential to find out in which direction the ORF was pointing. The 0.8 kb *Bam* HI/*Sal* I fragment was therefore introduced into both M13mp18 and 19 to produce the clones M13rt18 and 19 (see figure 4.4 b). Single stranded DNA prepared from one of these clones will contain the anti-sense strand of the gene, whilst the other will contain the sense strand.

Sequencing reactions were performed on single stranded DNA (ssDNA) prepared from each of these clones using oligo-N as the primer. As this anti-sense oligonucleotide would bind only to the template that contained the complementary sense strand, only one of the reactions would produce a sequencing ladder. A DNA sequence was obtained when M13rt19 was used as a template (showing that it contained the sense strand), whilst M13rt18 failed to produce sequence showing it contained the anti-sense strand.

Comparison of the orientation of those fragments in M13 to the restriction map of pRT60 allowed the deduction that the *icd* B ORF read from left to right (as shown in figure 4.4). This meant that only part of the 5' region of the gene had been cloned.

4.2.7 Construction of a genomic library in the vector λ-Gem 11.

As a contingency that the entire gene might not be cloned from the partial plasmid library, a lambda (λ) library was made that represents the entire *S.coelicolor* genome. This library will be a useful resource for the cloning of other genes.

To make a representative library, genomic DNA needs to be cleaved in a more or less random manner and introduced into a suitable vector ready for its replication in a host. Lambda replacement vectors are the vectors most commonly used for this purpose since relatively large (10-20kb) fragments of foreign DNA can be introduced and the libraries can be easily stored for long periods of time (Kaiser and Murray, 1985).

Limiting amounts of *Sau* 3A were used to partially digest genomic DNA (section 2.22.1). *Sau* 3A recognises a 4 bp sequence which occurs frequently in DNA sequences and produces a staggered end that is compatible with *Bam* HI staggered ends. Small scale digests were performed to determine the concentration of *Sau* 3A that produces the greatest amount of DNA in the size range suitable (9–20kb) for cloning into the lambda vector λ-GEM 11 (Promega "protocols and applications guide").

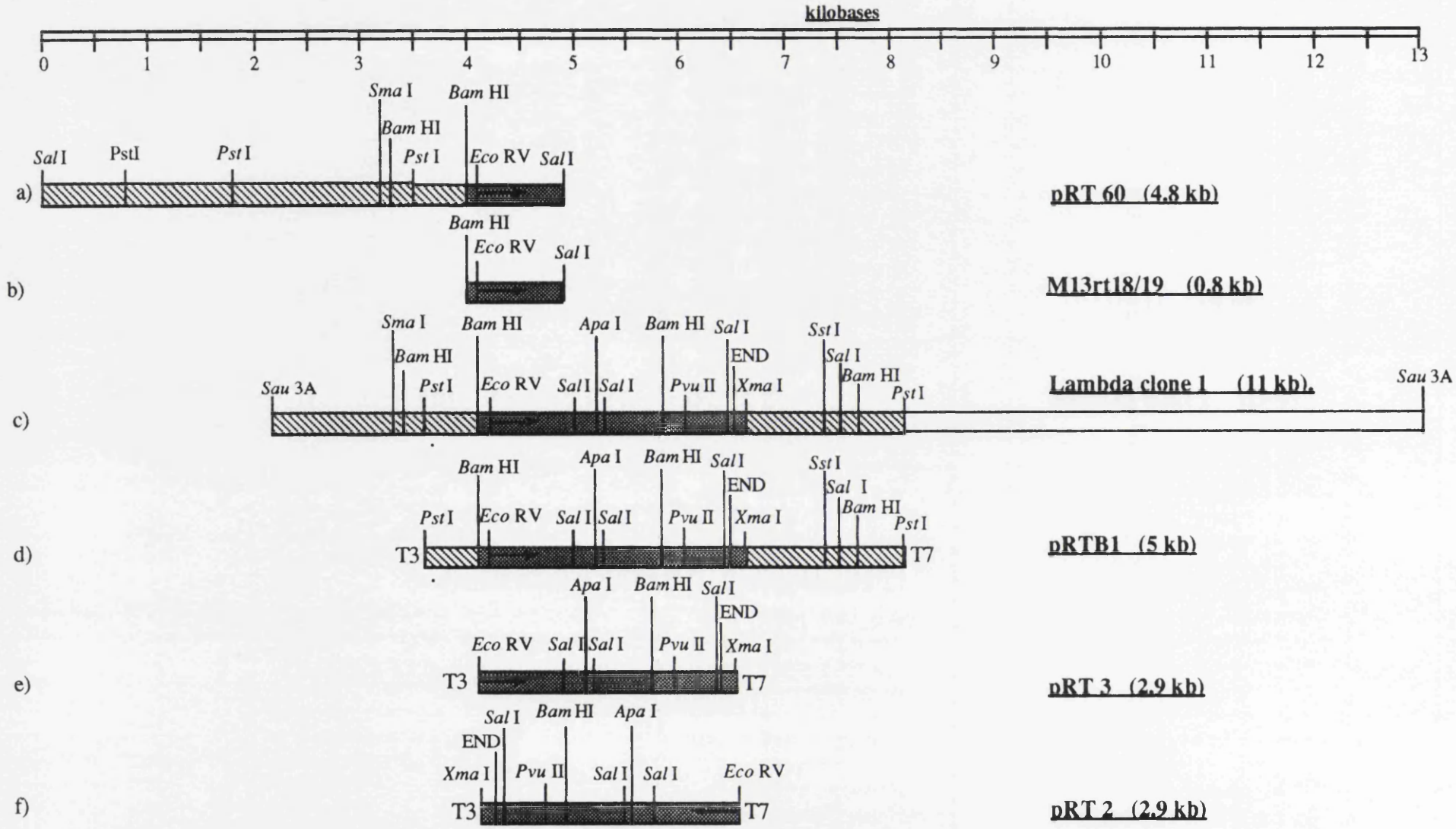

kilobases
0 1 2 3 4 5 6 7 8 9 10 11 12 13
a)
Sma I
Bam HI
Bam HI
Sal I
PstI
Pst I
Pst I
Eco RV
Sal I
pRT 60 (4.8 kb)
b)
Bam HI
Eco RV
Sal I
M13rt18/19 (0.8 kb)
c)
Sma I
Bam HI
Apa I
Bam HI
Sal I
Sst I
Bam HI
Sal I
END
Sau 3A
Pst I
Eco RV
Sal I
Sal I
Pvu II
Xma I
Pst I
Lambda clone 1 (11 kb).
Sau 3A
d)
Bam HI
Apa I
Bam HI
Sal I
Sst I
Sal I
END
Bam HI
Pst I
Pst I
Eco RV
Sal I
Sal I
Pvu II
Xma I
T3
T7
pRTB1 (5 kb)
e)
Apa I
Bam HI
Sal I
END
Eco RV
Sal I
Sal I
Pvu II
Xma I
T3
T7
pRT 3 (2.9 kb)
f)
Sal I
Bam HI
Apa I
END
Xma I
Pvu II
Sal I
Sal I
Eco RV
T3
T7
pRT 2 (2.9 kb)

Figure 4.4. Restriction maps of clones and subclones used in this study.

a) A clone containing the plasmid pRT60 was isolated (section 4.2.4). The 4.8 kb *Sal* I fragment had been introduced into the *Sal* I site of the plasmid pUC18. The *Hind* III site of the plasmid's polylinker lies to the right of this fragment (when orientated as shown) with the *Eco* RI site to the left.

b) The 0.8 kb *Bam* HI/*Sal* I fragment from pRT60 was cloned into M13mp18 and 19 to give M13rt18 and 19 respectively (section 4.2.6).

c) A λ clone was isolated that hybridised to oligo-N (4.2.9). The orientation of insert in λ–GEM-11 was not determined.

d) The 5 kb *Pst* I fragment from λ1 was subcloned into pBluescript II (KS^+) which had been digested with the same enzyme (section 4.2.10). Orientation of the insert in the vector is indicated by the position of the T7 and T3 promoters which flank the insert.

e) The 2.9 kb *Eco* RV/*Xma* I fragment from pRTB1 was cloned into pBluescript II (KS^+) which had been digested with the same enzymes (section 4.2.11) Orientation of the insert in the vector is indicated by the position of the T7 and T3 promoters which flank the insert.

f) The same 2.9 kb *Eco* RV/*Xma* I fragment from pRTB1 was also cloned into pBluescript II (SK^+) (section 4.2.11). Orientation of the insert in the vector is indicated by the position of the T7 and T3 promoters which flank the insert.

KEY.

▒▒ - Shows the region that was sequenced (chapter 5). Restriction site positions are therefore known accurately. Other restriction sites are present but are not shown because they cut too frequently.

▒▒. - The restriction sites shown in this region are known to exist from restriction mapping analysis. However it is also known that there are several other *Sma* I sites present whose positions were not determined and are therefore not shown.

▭ -. A restriction map for this region was not determined.

⟶ - Shows the direction in which the *icd* B open reading frame runs. The start codon is 56 base pairs from the *Eco* RV shown and the stop codon indicated by the label END.

All conditions were scaled up suitably for digestion of 200 μg of gDNA. Fragments were separated by electrophoresis on a 0.5% (w/v) agarose gel and fragments in the 11-18 kb size range purified from the gel by electroelution. 500 ng of λ-GEM 11 arms were then ligated to 500 ng of the prepared insert at a total concentration of 200 ng/μl. The ligation mixture was then packaged using Gigapack GOLD *in vitro* packaging extracts (Stratagene, see section 2.22.4) and the titre of recombinant phage determined (section 2.21.2). $2.3x10^5$ pfu were obtained with a background of wild-type phage (determined by ligation of arms without insert) of 2%.

4.2.8 Amplification of the library.

Libraries made by *in vitro* packaging of DNA start to lose viability significantly after four weeks in storage at 4°C. λ phage that have been packaged *in vivo* are however found to be far more stable. 12,000 pfu from the primary library were therefore used to infect a culture of *E.coli* NM621, which was then allowed to grow on plates. Phage harvested from these plates were used as the amplified library (see section 2.22.5) and could be stored for long periods. The equation below (Clarke and Carbon, 1976) can be used to calculate the probability of having every possible sequence from the *S.coelicolor* genome represented in a library made from a known number of plaque formnig units (pfu).

$$pfu = \frac{\ln(1-p)}{\ln[1-(\text{insert size}/\text{genome size})]},$$

Assuming a genome size of 8 x 10^6 bp for *S.coelicolor* (Hopwood and Keiser 1990) and a minimum insert size in the λ vector of 11kb, 12,000 pfu (of which 11,760 are recombinant) will give a probability (p) of 0.999999906 (ie/ virtually 1.0). This meant that a library made from 12,000 pfu almost certainly contained every possible sequence from the *S.coelicolor* genome and was therefore considered representative.

4.2.9 Screening of the amplified λ library.

If each DNA sequence was equally represented in this library then the number of plaques required to be screened would be the same as for screening a primary library. In other words 3,347 plaques would have to be screened in order to have a 99% chance of picking up any one sequence (using the equation above). Approximately 9,000 plaques

were therefore grown on a lawn of *E.coli* and the DNA transferred to Hybond-N nylon membrane as described in section 2.28. This is almost three the number required to have a 99% chance of picking up a particular sequence.

Plaque lifts were treated as described in section 2.27.and hybridisation experiments carried out using oligo-N as the probe. After hybridisation at 55°C in 6xSSC and eventual washing at 65°C in 2xSSC, 11 plaques were identified that bound to the radiolabeled probe (see figure 4.5 a-d).

Nine of these plaques were then isolated and phage streaked out on a lawn of *E.coli* to produce distinct single plaques. Twelve purified plaques from each of the isolates were then plated in a triplicate array. Two of the plates were used to take lifts of plaques onto Hybond-N and the third kept as a master plate. The duplicate membranes were then probed using oligo-N. Only two plaques were found to hybridise in duplicate to the probe. Both were derived from λ-1 (see figure 4.5 e).

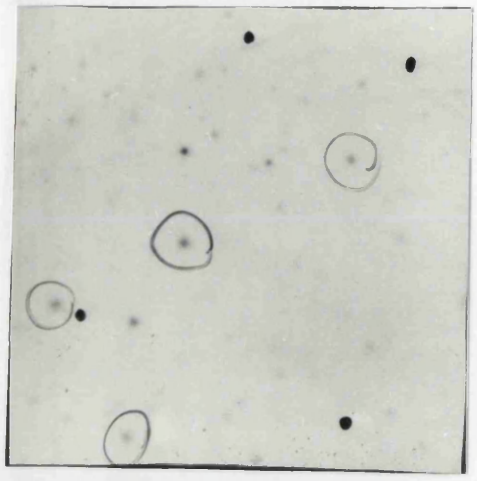
Restriction of DNA prepared from each of these plaques with several enzymes showed that both recombinant phage contained the same insert. This was to be expected as they were derived from the same original isolate. A Southern blot of the gel used in restriction analysis was screened using oligo-N to determine the fragment which contained the 5' region of the *icd* B gene. Digestion of λ-1 DNA with *Bam* HI, *Pst* I and *Sma* I produced fragments of 1.8 kb, 5 kb and 3.6 kb that hybridised to oligo-N. These are the same sizes as those obtained by hybridisation experiments with genomic digests (section 4.2.4). Digestion with *Sal* I however produced a hybridising fragment of only 2.8 kb, compared with the 5 kb fragment produced by *Sal* I digestion of gDNA. This means that one of the*Sau* 3A endpoints lies within the 5 kb *Sal* I fragment produced from genomic digests. The precise position of this endpoint relative to the insert found in pRT60 is shown in figure 4.4a and shows that the entire *icd* B gene must be contained within λ-1.

Southern blotting of genomic DNA digests had shown that oligo-N bound to a 5 kb *Pst* I fragment. Restriction analysis of pRT60 had shown that the start of the gene was within a *c.*1.5 kb *Pst* I/*Sal* I fragment (figure 4.4a). It was therefore deduced that a second *Pst* I site must lie 3.5 kb to the right of this *Sal* I site. Since the length of the gene can be no longer than 2.8 kb, then the end of the gene cannot lie further than 2.8 kb to the right of this same *Pst* I/*Sal* I fragment. Consequently we can conclude that the end of the gene must lie to the left of the second *Pst* I site and that the entire *icd* B gene must therefore lie within the 5 kb *Pst* I fragment.

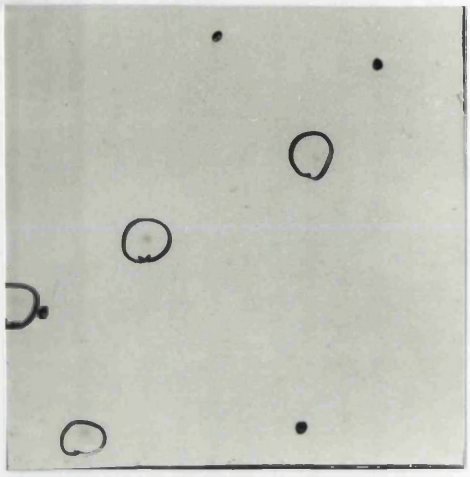
4.2.10 Cloning of the 5kb *Pst* I fragment into pBluescript II (KS+).

The recombinant, λ–1, contained an 11kb fragment which was too large to easily restriction map. λ–1 DNA was therefore digested to completion with *Pst* I and the 5 kb fragment (known to contain the entire *icd* B gene) was isolated by agarose gel electrophoresis followed by electroelution. This fragment was then ligated to pBluescript II (KS+) which had been restricted previously with *Pst* I. The resultant plasmids were introduced into *E.coli* DS941 which had been made competent by the $CaCl_2$ method. Three of the transformants containing plasmids with inserts were isolated and the plasmids mapped by restriction digestion. All three isolates contained plasmids with identical restriction maps indicating that all three contained identical inserts in the same orientation. The restriction map produced is shown in figure 4.4c.

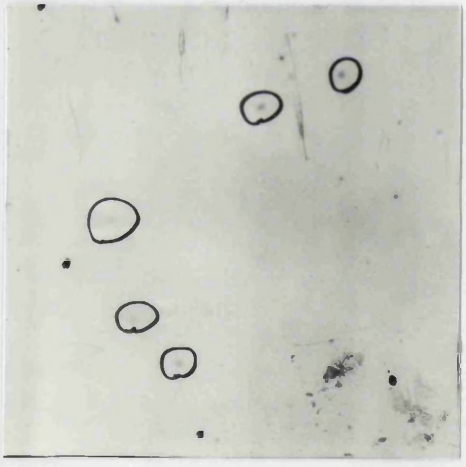
a)



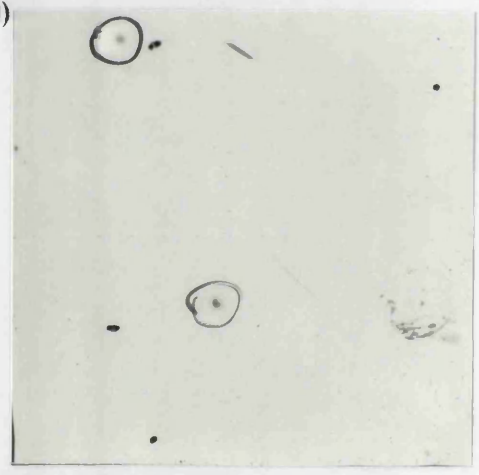
b)



c)



d)



e)

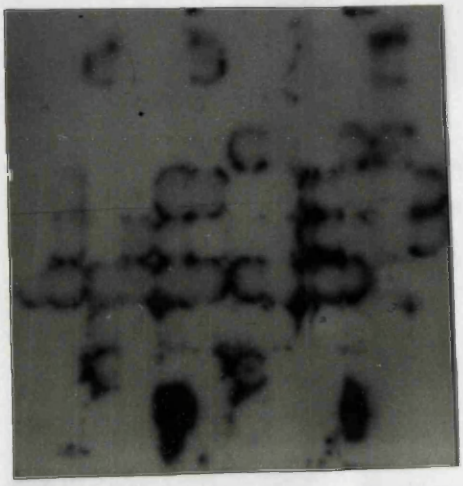


Figure 4.5 Plaque hybridisation with oligo-N.

Three filters (with approximately 3000 plaques on each) were screened by hybridisation to oligo-N. Duplicate filters were prepared by taking plaque lifts as described in section 2.28. All 6 filters were subjected to hybridisation at 55°C in 5xSSC and washed at the temperatures shown below. Only one of the duplicate filters is shown. Those plaques seen to hybridise on duplicate filters are circled.

a). All filters were originally washed at 60°C in 5xSSC. One of the filters screened at this temperature is shown here.

b). The same filter as in a) is shown after being washed at 65°C 2xSSC. It shows the decrease in the number of plaques hybridising to the oligonucleotide.

c) and d). Shows the other two filters that were treated in the same way as the filter shown in a) and b).

e). Nine of the plaques (from filters b), c) and d)) that were found to hybridise in duplicate were purified to single plaques. Twelve purified plaques from each isolate were plated (horizontally adjacent to each other in figure 4.5e) ready for re-screening (section 4.2.9). Only two plaques were found to still hybridise which were both derived from the same original isolate. These were called λ1 and λ2 and were later found to be identical (section 4.2.9).

4.2.11 Cloning of the 2.9kb *Eco* RV/*Sma* I fragment from pRTB1.

Sequencing of M13rt18 and 19 identified the translation start codon of the *icd* B gene, which is just 47 nt downstream of the *Eco* RV restriction site. This meant that the end of the gene could be calculated to be somewhere to the left of the unique *Sma* I site found in pRTB1, since a 90kDa protein would require a gene 2.7 kb long (see section 4.2.3).

The *Eco* RV/*Sma* I fragment was therefore used to clone the entire ORF of *icd* B, which may also contain some of the required upstream and downstream regulatory sequences. Both *Eco* RV and *Sma* I produce blunt ends which does not allow for the directional cloning of this fragment into a vector. However the isoschizomer of *Sma* I, *Xma* I, recognises the same site but produces staggered ends with a 5' overhang rather than blunt ends (see below).

5'CCC$^{\nabla}$GGG 3'	5'C$^{\nabla}$CGGG 3'
3'GGG$_{\Delta}$CCC 5'	3'GGGCC$_{\Delta}$C 5'
Sma I	*Xma* I

Xma I was therefore used instead of *Sma* I to enable directional cloning of the insert into both pBluescript II (KS+) and pBluescript II (SK+). Due to the expense of the enzyme it is provided at only low concentrations. However its thermostability allows digestion to be carried out overnight at 37°C.

Complete digestion of both of the vectors and pRTB1 by *Xma* I was followed by complete digestion with *Eco* RV. The appropriate fragment was then separated and purified by agarose gel electrophoresis followed by electroelution. The purified fragment was then ligated to each of the linearised pBluescript plasmids (SK+ and KS+) to create the recombinant plasmids pRTB2 and 3 respectively (see figure 4.4 e).

4.2.12 Attempts to complement the *E.coli icd*$^-$ mutant, EB106.

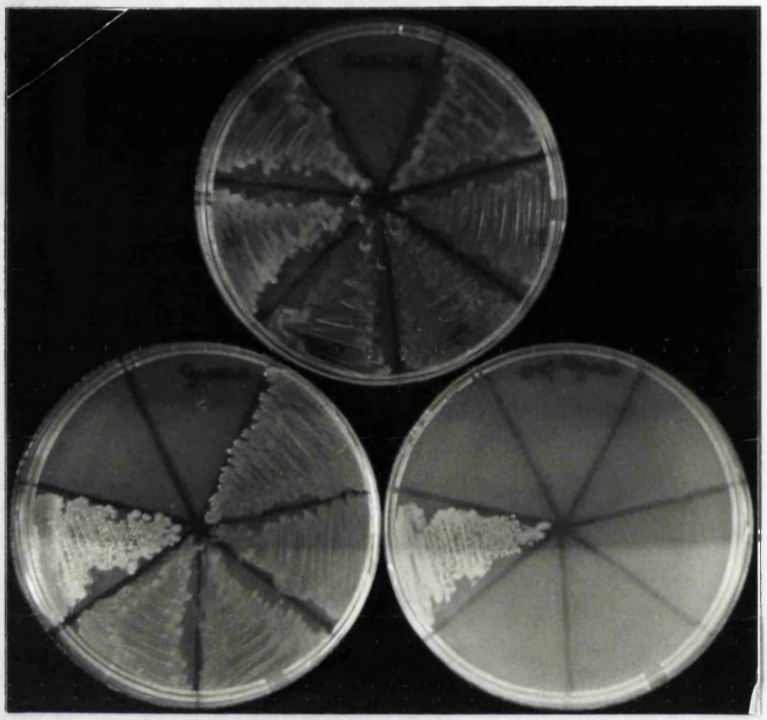
It was of interest to see if the *S.coelicolor icd* B gene could complement the *E.coli icd*$^-$ mutant, EB106, since this could provide an insight into the similarities between the IDHs from the different organisms and could provide final proof that a functional gene from *S.coelicolor,* that encodes an IDH, had been cloned.

It is worth pointing out at this stage that for reasons that are as yet unclear, *E.coli icd*$^-$ mutants, simultaneously confer resistance to naladixic acid (Helling and Kukora,

1971). This provides a useful way of maintaining selection for mutants which would otherwise rapidly revert back to wild type. EB106 was therefore made competent for transformation by the $CaCl_2$ method (naladixic acid was added to all media until the cells were harvested, see section 2.15c) and the plasmids pRTB1, 2, and 3, pBluescript II (KS+) and pCK505 introduced into the this strain. pCK505 is a pBR322 derived plasmid that was used as a positive control since it contains the whole of the wild type *E.coli icd* gene and is known to be able to complement this mutant (LaPorte *et al.*, 1985).

Several hundred transformants were obtained from each of these transformations. A single colony from each transformation was plated on to several selective media to test for complementation. It can be seen from figure 4.6 that none of the plasmids (other than the control plasmid pCK505) enabled EB106 to grow on minimal media without an amino acid supplement, showing that complementation had not occurred.

a)



b)

c)

Key.

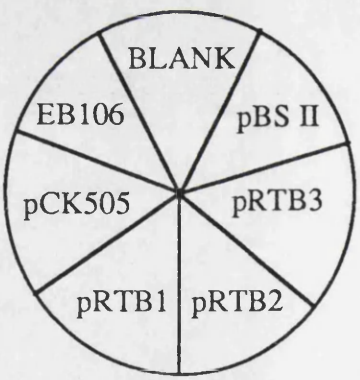


Figure 4.6 Complementation of the *icd* mutation in the *E.coli* strain EB106.

Plasmids pRTB1, pRTB2 and pRTB3 (see figure 4.4) were introduced into the *E.coli* mutant EB106 which is devoid of IDH activity (due to a point mutation in *icd*). Each of these plasmids contain the entire ORF for the *icd* B gene from *S.coelicolor* (section 4.2.11 and figure 4.4). As a positive control for complementation a plasmid (pCK505) that contains the wild type *icd* gene from *E.coli* was also introduced into the strain. As a negative control the vector pBluescript II (KS^+)was also introduced into EB106. Transformants were streaked onto the selective media below. None of the media contained naladixic acid (nal) in case any complementation caused reversion to nal sensitivity.

a) **L-agar:** All strains should grow on this complex media including EB106 (with no plasmid). This plate therefore confirms cell viability.

b) **MM+pro+amp:** Ampicillin was added to minimal media (section 2.15c) to select for cells that contained plasmids (pBluescript II confers ampicillin resistance to strains containing the plasmid). Proline was added as a supplement that will allow *icd* $^-$ strains to grow, provided that the strain possesses a plasmid.

c) **MM-pro+amp:** Ampicillin was also added to the minimal media. Proline was not added to the media. Only those strains that contain a plasmid capable of complementing the *icd* $^-$ mutation will be able to grow.

The results show that only plasmid pCK505 can complement the *icd* $^-$ mutation of *E.coli*. It was therefore concluded that the *icd* B gene from *S.coelicolor* cannot complement this mutation when introduced into EB106 on either of the plasmids used (section 4.2.12).

4.3 Discussion.

The cloning of the *icd* B gene that codes for the type II IDH from *S.coelicolor* would prove an invaluable tool for the study of the genes regulation and expression and the role of the enzyme in cellular function. It could be used to study the effect of over and under expression of the gene on cell growth and antibiotic biosynthesis on different growth media, especially by comparing the difference between cell growth on acetate and glucose. Complete inactivation of the wild type genomic copy could also be performed, using a disrupted copy of the cloned gene in insertional inactivation experiments.

The most immediate benefit derived from cloning of the gene would however be the ability to obtain the DNA and hence amino acid sequence, for comparison to other genes. Using the N-terminal amino acid sequence obtained from the type II IDH from *S.coelicolor*, an oligonucleotide was designed (using the codon bias found in streptomycete DNA) which was used as a probe for screening for the *icd* B gene. Initially digests of *S.coelicolor* genomic DNA were used to find empirically the hybridisation and washing conditions required to produce hybridisation to the *icd* B gene alone.

The most stringent conditions employed in these screening experiments were hybridisation at 55°C (with 6xSSC) and washing at 65°C (with 2xSSC). Raising of the temperature by another 5°C led to complete loss of signal. Binnie (1990) reported that final washing conditions are normally performed at 5°C below the Tm of a probe showing that the Tm for oligo-N is between 65 and 70°C. Knowing the Tm for the probe allows us to calculate the percentage similarity of oligo-N to the target sequence 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=0.33 (2xSSC), Tm=70°C, and G+C=65.7, this equation indicates that oligo-N is 91.4% identical to the target sequence, suggesting 2-3 mismatches. Later sequencing of this region shows that oligo-N does indeed have 3 mismatches when compared to the target sequence (see figure 5.3) indicating that the most stringent conditions possible were used in hybridisation experiments. If the final washing conditions were carried out at a temperature only 5°C lower, then other DNA

fragments were found to also hybridise to the oligonucleotide. This shows that maximum stringency is required in order to produce a unique single band.

It is worth pointing out that when determining hybridisation/washing conditions it is important to consider the combination of conditions used at each stage. In this case, hybridisation at lower stringencies produced signals that could not be washed off, even after exhaustive washing using the conditions of high stringency finally employed *ie.* 65°C (6xSSC). Conversely, hybridisation at 65°C (6xSSC) produced no signals at all, even though the subsequent washing conditions were no more stringent than those eventually used during washing. The combination of hybridisation at one stringency and washing at a higher stringency were the only set of conditions that would produce unique hybridisation signals using this probe. This highlights the difficulty often experienced when trying to find specific hybridisation conditions using oligonucleotide probes with an unknown number of mismatches.

Another major problem encountered during these screening procedures, was the occurrence of significant amounts of non-specific hybridisation to the membrane that could often obscure any specific signals. The use of analytical grade NaCl for making the SSC buffer was found to reduce this problem considerably, with general purpose NaCl giving variable results depending on the batch used (Dr H.Bramwell, Departments of Biochemistry and Genetics, University of Glasgow, personal communication). To increase the signal to noise ratio the amount of target DNA was increased by loading 5-10μg of digested genomic DNA onto gels, instead of the 1μg initially used. Around 6000 plaques from the *S.coelicolor* λ library were screened for hybridisation to oligo-N. This number of plaques contains a total of $6.6x10^7$ bp of insert DNA (assuming a minimum insert size of 11 kb) which is the equivalent of 8.25 genome lengths. Assuming equal representation of each sequence in the library we would therefore expect to obtain 8-9 positive signals.

When the λ library was screened 11 plaques were found to hybridise to oligo-N in the initial screening. However only one of these was found to hybridise during a re-screening procedure, less than the 8 plaques expected by calculation. Chance may have played a role in reducing the number of positives obtained but it must also be remembered that the library that was screened had undergone an amplification step. The presence of some inserts in a λ phage can lead to a reduced growth rate compared to other recombinant phage and can lead to their under-representation in the amplified library. Although this can be a problem for libraries amplified in liquid culture, it is generally considered that there is less of a problem when amplification is performed on solid media (Kaiser and Murray, 1987). Nevertheless it is possible that clones containing the *icd* B

gene (or a proximal sequence) could produce smaller plaques on plates and thereby be less represented than expected.

Another explanation is the possibility that the some of the wrong plaques had been picked from the original plates screened, since the line up of the autoradiograph to the original plate is not very accurate, especially on plates on which plaques are plated to high density.

Restriction digests of the clones that were isolated from this library was compared to the map of pRT60 and the information obtained from genomic digests. This showed that this clone contained the entire *icd* B ORF (c.2.7 kb) within a 5kb *Pst* I fragment which was cloned into pBlusescript to obtain the clone pRTB1. Further restriction mapping and subcloning allowed the introduction, in both orientations, of a 2.9kb fragment (containing the entire *icd* B gene) into pBluescript.

The availability of a strain of *E.coli* that is deficient in an active form of IDH (*icd*$^-$ mutant, EB106), prompted an attempt to complement the mutation using the *S.coelicolor icd* B gene. Introduction into EB106 of all of the plasmids available, that contained the whole of the *S.coelicolor icd* B ORF (pRTB1, 2 and 3), produced no complementation. It is possible that the constructs pRTB 2 and 3 do not contain the upstream (or downstream) regulatory sequences required for gene expression, since they have only short stretches of DNA up and down stream of the ORF. However it is almost certain that pRTB1 will contain those sequences required for expression in its native host at least.

It is well known though 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). The most likely explanation for lack of expression of this gene in *E.coli* is therefore a low level, or complete absence, of transcription from the gene. Indeed later sequence analysis would seem to confirm that the appropriate sequences normally required for expression of genes in *E.coli* are absent (see section 5.3.4). This gene will therefore require modification of its non-translated regions (*ie.* RBS and promoters) in order to allow its expression in *E.coli*.

CHAPTER 5.

DNA sequencing and analysis of *icd* B.

5.1 Introduction.

A fragment of DNA had been cloned which was expected to contain the whole gene encoding IDH, (section 4.2.9). As this is the first gene for a type II IDH to be cloned, it was of great interest to deduce its peptide sequence for comparison with other proteins, in the hope of gaining some insight into its function and mechanism of action.

It was also described in section 1.1.3 how RNA polymerase and promoter heterogeneity play a key role in the differential control of gene expression (reviewed by Strohl, 1992). The relationship between promoter structure and gene function and the level and time of expression is however still unclear. Such relationships will only be solved by the analysis of more promoters. It was therefore of interest to examine the upstream sequences of the *icd* B gene as a first step in the study of its promoter and to see if there were any similarities to other known promoters. A search for other regulatory sequences, such as ribosome binding sites and terminators, was also made.

At the time of writing the most up-to-date study of such regulatory sequences had been performed by Seno and Baltz (1989) which has also served as an excellent review of the knowlege accumulated so far about streptomycete sequences. The information about streptomycete regulatory sequences that is presented in this chapter has therefore been obtained from this publication unless otherwise stated.

In this chapter I describe the strategy used to obtain the DNA sequence of *icd* B and the surrounding region and the analysis of this sequence. Analysis of the amino acid sequence deduced from the DNA sequence of *icd* B is discussed in chapter 6.

5.2 Results.

5.2.1 Choice of sequencing strategy.

Probably the most reliable, easiest and most used method of sequencing DNA is by the di-deoxy method of Sanger *et al.*, 1977. Quality controlled kits can be purchased from various suppliers that contain all of the reagents required for performing these reactions, making this method both reliable and convenient. In this method the strand complementary to the one being sequenced is synthesised *in vitro* by extension from a DNA primer using a DNA polymerase (DNApol).

Four separate reactions are performed each containing all four dNTPs required for chain elongation. Each separate reaction also contains one of the four dideoxy nucleotide analogues (*ie.* ddATP, ddCTP, ddGTP or ddTTP). These analogues are used at a concentration that allows their incorporation at a random position in the growing chain.

For instance the presence of a cytosine in the template will allow the incorporation of a dGTP or a ddGTP molecule into the newly synthesised strand. If the nucleotide analogue is incorporated into the chain, further elongation is prevented. A family of DNA molecules is therefore generated each terminating at the position corresponding to where a cytosine was on the template. Similar reactions performed with the other analogues generate fragments that terminate at positions corresponding to the other three nucleotides.

The molecules in each reaction mixture can be separated on the basis of their length by denaturing polyacrylamide gel electrophoresis (see section 2.30). Running each reaction next to each other on a gel allows determination of the order in which each of the nucleotides occur in the growing chain. The use of a radioactive nucleotide during the reaction allows the newly synthesised strands to be visualised by autoradiography.

The most widely used DNApol has been derived from the T7 bacteriophage and has been modified to remove its 5' to 3' exonuclease activity. However streptomycete DNA has a propensity to form secondary structures, due to its high G+C content, which often causes the enzyme to pause for significant periods of time at these points. This manifests itself as bands in all four lanes of a sequencing autoradiograph with the same mobility (known as "four track stops") making determination of the true bases around this region very difficult. This problem can be circumvented to a large extent by using higher incubation temperatures for sequencing reactions, which helps to melt the intramolecular secondary structures. Most sequencing using the T7 polymerase therefore used a temperature of 42°C for the termination reactions, rather than the recommended temperature of 37°C (see section 2.29.1b).

Regions in which secondary structure was still a problem were sequenced using Taq polymerase. This thermostable DNApol, isolated from the extreme thermophile, *Thermus aquaticus* allowed reactions to be performed at temperatures of up to 70°C. However bands on autoradiographs were found to be thicker and less well defined when Taq pol was used in reactions compared to when T7 pol was used. As one reads DNA sequence from an autoradiograph bands get closer together near the top of the gel (where the larger DNA molecules are found), eventually getting so close that individual bands cannot be resolved. Consequently less sequence can be read from reactions produced by Taq pol as

the thicker bands produced merge sooner. Taq polymerase was therefore not routinely used.

Another problem incurred by the high G+C content of streptomycete DNA is the greater frequency with which compressions occur. Whereas "four track stops" are caused by the formation of secondary structures during the sequencing reaction, compressions are due to the formation of secondary structures as the DNA is being separated by electrophoresis. Compressions are identified as regions on an autoradiograph where bands become unevenly spaced over a localised region. A band from one of the reaction mixtures will have a slower mobility than normal whilst a band from one of the other nucleotide reaction mixtures will have a faster moblity than normal. If this aberrant behaviour is extreme, then two bands from different lanes can appear to have similar mobilities. This obviously makes determination of the sequence at this point difficult. Deaza nucleotide analogues were however routinely used in sequencing reactions which helped to prevent this phenomenon (Mizosawa *et al.*, 1986).

Sequencing reactions can be performed using either single or double stranded DNA using the methods described in manufacturers manuals (*e.g.* Promega T7pol sequencing kit). DNA sequencing using single stranded templates were found to be the most reliable method of producing clear, distinct bands and enabled 250-400 nt to be read from a single set of reactions.

Using the universal primer to sequence template will produce 200-400 nt of sequence from the region adjacent to the universal primer's binding site. A strategy is therefore required to sequence regions further away. One such method is to design oligonucleotides which will hybridise to regions of DNA that have already been sequenced. These can be used as primers in subsequent reactions to extend the sequence futher along the same strand. Oligonucleotides complimentary to sequence already obtained can be used to sequence the other strand in the reverse direction. In this way one can sequence a region of interest by "walking" along the template.

5.2.2 Construction of subclones used for DNA sequencing.

To determine in which direction the *icd* B ORF is reading the 0.8kb *Bam* HI / *Sal* I fragment was introduced into the sequencing vectors M13mp 18 and 19 (section 4.2.6). When oligo-N (the oligonucleotide designed from the N-terminal protein sequence) was used as a primer in sequencing reactions only M13rt18 produced sequence (section 4.2.6 and figure 5.1a). Both templates were also sequenced using the universal (-40) sequencing primer and the DNA sequence then extended sequentially with

oligonucleotide primers until the whole fragment had been sequenced on both strands (see Figure 5.1a).

The presence of a *Bam* HI site internal to the *icd* B gene enabled the insert to be cloned in two parts. Digestion of pRTB2 with *Bam* HI and *Xba* I produced two fragments derived from the 2.9 kb insert. One of the fragments (1.55 kb) had one *Xba* I staggered end and one *Bam* HI staggered end. Its ligation into M13mp19, which had been linearised with the same restriction enzymes, produced the sub clone M13rt21 (see figure 5.1b). The second fragment (0.8 kb) had *Bam* HI staggered ends at both termini and was introduced into the *Bam* HI site of M13mp18. Its introduction in both orientations produced the two subclones M13rt24 and M13rt24b.

The entire 2.7 kb *Eco* RV/*Sma* I fragment from pRTB2 was also subcloned into M13mp18 and 19 using the *Kpn* I and *Xba* I restriction sites which flank this region in pRTB2. These recombinants (M13rt 20 and 21) were also used for sequencing. Figure 5.1b summarises all of clones used for sequencing and figure 5.1a shows the oligonucleotides used on each template. The DNA sequence is shown in figure 5.3.

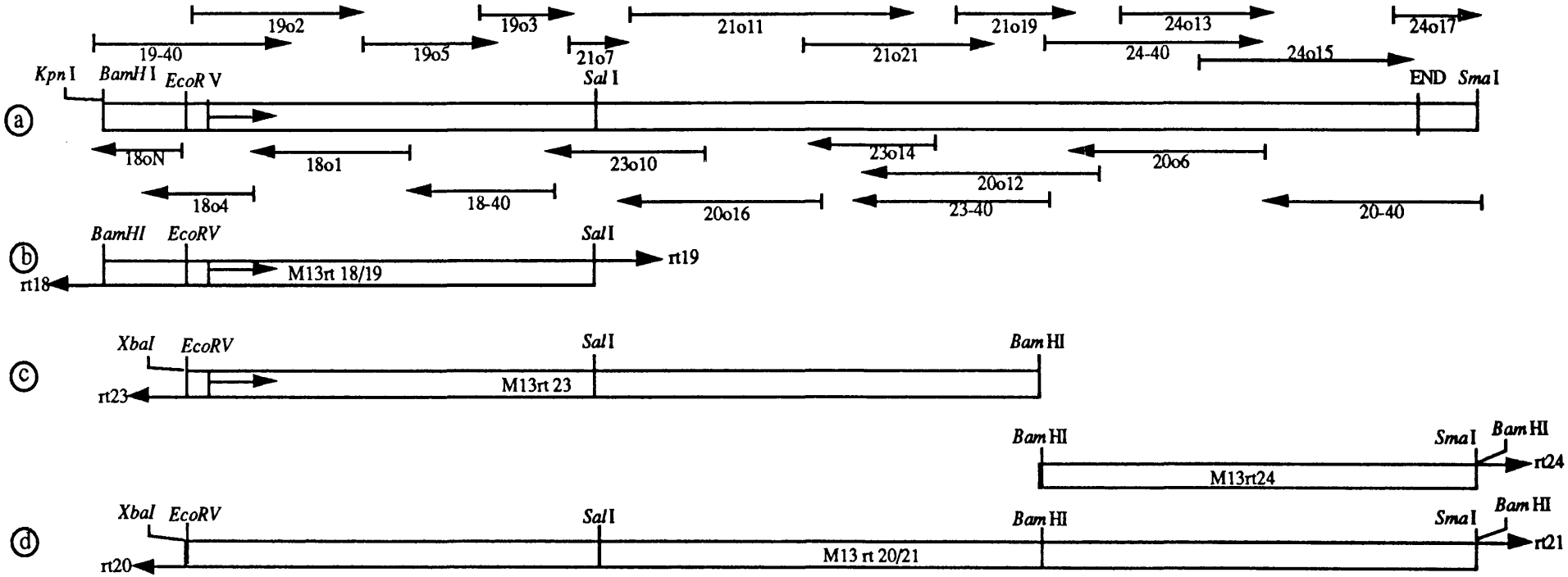

19o2
19o3
21o11
21o19
24o13
24o17
19-40
19o5
21o7
21o21
24-40
24o15
Kpn I
BamHI
EcoR V
Sal I
END
Sma I
a
18oN
18o1
23o10
23o14
20o6
20o12
18o4
18-40
20o16
23-40
20-40
BamHI
EcoRV
Sal I
b
rt18
M13rt 18/19
rt19
XbaI
EcoRV
Sal I
Bam HI
c
rt23
M13rt 23
Bam HI
Sma I
Bam HI
rt24
M13rt24
XbaI
EcoRV
Sal I
Bam HI
Sma I
Bam HI
d
rt20
M13 rt 20/21
rt21

Figure 5.1. Summary of M13 subclones and oligonucleotides used for sequencing the *S.coelicolor* gene, *icd* B.

(a)

The entire region of DNA sequenced is shown in here. Only the relevant restriction sites used for creating the subclones are shown (for a more comprehensive map see figure 4.4). The start codon for the *icd* B gene and direction of transcription is indicated by the arrow to the right of the *Eco* RV site. Single stranded DNA sequencing was performed using several oligonucleotides as primers. The length and direction of sequence obtained from each oligonucleotide is indicated. Labeling refers to the oligonucleotide and subclone used in the sequencing (*eg.* 21o11 indicates that oligonucleotide 11 was used in sequencing reactions with the subclone M13rt21).

(b), (c) and (d).

Fragments of DNA containing part (or all) of the *icd* B gene and surrounding regions were introduced into M13mp 18 and 19 for sequencing with oligonucleotides (see section 5.2.2. for description). The position of each of the clones relative to the 2,560 bp region sequenced is shown here. Arrows labeled with numbers indicate which M13 subclones sequence in which direction (*eg.* the subclone M13rt 18 was used to sequence the coding strand of the *icd* B gene and M13rt 19 used to sequence the non-coding strand).

5.3 Discussion.

5.3.1 Analysis of Open Reading Frames (ORFs).

The high guanosine plus cytosine (G+C) content of streptomycete DNA imposes an extreme bias on the distribution of nucleotides in each of the three positions of a codon within a coding sequence. This is most extreme in the third base of codons, which is the position that allows the greatest degeneracy. Approximately 90% of all codons within streptomycete coding regions containing a guanosine or a cytosine in this position whilst 70% contain G/C in the first position and 50% in the second (Bibb *et al.*, 1984). Non coding regions do not contain such a restrictive bias with the frequency of G+C being distributed in a random manner. It is therefore relatively easy to distinguish an ORF in a region of DNA by looking for this non-random distribution of nucleotides (Bibb *et al.*, 1984).

A computer program that performs this sort of analysis is provided in the "GCG" package as a program called "CODONPREFERENCE" (Devereaux, 1984). It calculates the mean G+C content in each of the three positions within a window of specified size. The window is moved sequentially along the DNA sequence and the results are usually plotted as three separate graphs representing the G+C content for each codon position, refered to as "phases". If the region contains a real ORF then the plot should show one of the phases to have a medium G+C content (*c.*70%), one that will have a low G+C content (*c.*50%) and a third that will have a high G+C content (*c.*90%). This well defined order (*ie.* -Medium-Low-High-Medium-Low-High-) not only indicates the presence of an ORF but also the direction in which it is transcribed and translated.

As a consequence of such a rigid distribution of the G+C content, a bias is also imposed on the codons used by streptomycetes. Synonomous codons containing Gs or Cs (especially at the third position), are used in extreme preference to those containing As and Ts. The frequency of codons within a potential ORF (PORF) should therefore follow the codon usage table that was compiled from 67 streptomycete genes (Figure 5.5a) and provides a further way of testing where a PORF is likely to encode a protein. The "CODONPREFERENCE" program also draws a plot (figure 5.3 for each phase of the sequence that provides a measure of how closely the codon usage of the predicted frames correlates with a codon usage table.

Non coding regions of DNA will produce predicted amino acid sequences with a more random codon usage than coding regions and would therefore be expected to use every codon with almost equal preference. This in turn means that non-coding regions

will use rare codons with a higher frequency than coding regions. The presence of a large number of rare codons in a putative coding region will therefore indicate that it is probably false. These are also highlighted by the program "CODONPREFERENCE" (figure 5.3)

Analysis of the "CODONPREFERENCE" plot produced for the region suurounding the *icd* B gene, shows that there is a region of DNA which shows extreme G/C bias and a codon usage similar to that used in streptomycetes in general. This indicates that this region seems to contain a true ORF. The nucleotide composition in phase 3 shows the high frequency of Gs and Cs that occur typically in the third position of a codon, whilst phases 1 and 2 show the medium and low frequency expected for the first and second bases of a codon respectively. This pattern shows that the direction of translation is from left to right (frame 1) as the sequence is written in figure 5.3. The paucity of rare codons observed in this frame would seem to confirm the presence of an ORF in this frame. The effect of such a high G+C content on the codon bias of the gene can be clearly seen, with the codon usage trace following the trace of the G+C bias almost exactly.

5.3.2 Translation start and stop codons.

Analysis of the DNA sequence (Figure 5.2), shows that there is a GTG start codon at position 198, which is consistent with the predicted start position indicated by the frame analysis program (Figure 5.3). The frequency with which GTG codons are used as a start codon is far higher in streptomycete genes than in *E.coli*, being used 18% of the time as compared to only 3% in *E.coli* (Seno and Baltz, 1989). The amino acid sequence deduced from the nucleotide sequence (Figure 5.2) had an N-terminus identical to the sequence determined by direct sequencing of the protein (figure 3.19) except that the first amino acid of the protein corresponded to the second codon of the ORF. The N-terminal formyl-methionine must therefore have been removed post-translationaly from the nascent polypeptide.

The first translation stop codon in the correct frame was found to be a TGA codon at position 2422 in Figure 5.2, which is at a position consistent with the loss of codon bias (indicated by figure 5.3). Translation of the 2,223 nucleotide region between these start and stop codons would result in a protein of 741 amino acids with a molecular weight of 79,916. This agrees quite well with the molecular weights determined by both gel filtration and denaturing SDS-PAGE (see section 3.2.3). This confirms that the whole gene is contained within this region (and therefore within plasmids pRTB 1, 2 and 3) and that the protein is indeed a large monomeric protein encoded by a single gene.

Figure 5.2 DNA and amino acid sequence for the *S.coelicolor icd* B gene.

a) Entire DNA sequence obtained from single stranded sequencing of the clones shown in figure 5.1 (section 5.2.1).

b) The translation of the putative ORF upstream of *icd* B is shown starting at base 3 and has been translated to the first stop codon at base 93. The possible terminator structure for this PORF is marked **A** (see figure 5.8).

c) The amino acid sequence for *icd* Bis shown starting from nucleotide 198 and finishing at the first stop codon at 2421. The possible terminator structure for this gene is marked **B** (figure 5.7) and the four possible promoters for *icd* Bare marked **I, II, III** and **IV** (figure 5.5).

d) The translation of the putative ORF downstream of *icd* Bis shown starting at the possible start codon at base 2423 and has been translated to the end of the sequence. The -35 region of the putative promoter is labeled as **V** and the -10 region as **VI.**

All amino acid sequence has been derived from translation of the DNA sequence using the GCG programme "TRANSLATE" with each amino acid written below the first base of the appropriate codon. Putative ribosome binding sites for *icd* B and the putative downstream ORF are marked **RBS**. Possible translation **start** and **stop** codons are shown in bold.

```
          10                  30                  50 -35 II
a) GGATCCCGTCGCCCGCCCGGACGGTCGCGCTCCAGCAGGGGGCCGTGGTCCTGCCGCAGC
b)   I  P  S  P  A  R  T  V  A  L  Q  Q  G  A  V  V  L  P  Q  Q

          70 -10 II          90 A                110
a) AGCGGGCGGGCGAGGACGTCGTCCCGCTCCCCTGAGCGCGGCGTGGCCGGAATCTGGGTG
b)   R  A  G  E  D  V  V  P  L  P          -35 IV              -10

          130          -35 III 150          170
a) CGACCCCTGTCGAGCCGGGGACGGTCACGAGATATCTTGATGTCGAGCAATGTTGCAGAC
    IV                                -35 I              -10 I

    RBS   190                  210                 230
a) GTGGAGCGGAGCACCCGGTGACTGACTCGACCATCATCTATACACACACTGACGAGGCCC
c)                    V  T  D  S  T  I  I  Y  T  H  T  D  E  A  P

          250                 270                 290
a) CGGCCCTGGCGACGTATTCGTTCCTGCCGGTGGTCCGGGCGTACGCCTCGCAGGCGGGTG
c)   A  L  A  T  Y  S  F  L  P  V  V  R  A  Y  A  S  Q  A  G  V

          310                 330                 350
a) TCGCGGTCGAGACCCGCGACATCTCGCTGGCCGGCCGCATCATCGCCCTGTTCCCGGAGT
c)   A  V  E  T  R  D  I  S  L  A  G  R  I  I  A  L  F  P  E  Y

          370                 390                 410
a) ACCTCACCGAGGACCAGCGCATCCCGGACGCCCTCGCGGAGCTGGGCGAGCTGGCCAAGA
c)   L  T  E  D  Q  R  I  P  D  A  L  A  E  L  G  E  L  A  K  T

          430                 450                 470
a) CGCCCGCCGCCAACATCATCAAGCTGCCGAACATCTCGGCGTCCATCCCGCAGCTCAAGG
c)   P  A  A  N  I  I  K  L  P  N  I  S  A  S  I  P  Q  L  K  A

          490                 510                 530
a) CCGCCGTCGCCGAGTTGCAGGGCCAGGGCTACGCGCTGCCGGCCTACCCGGACGACCCGA
c)   A  V  A  E  L  Q  G  Q  G  Y  A  L  P  A  Y  P  D  D  P  K

          550                 570                 590
a) AGACCGACGAGGAGCGCGACATCCGCGCCCGCTACGACAAGGTCAAGGGCTCCGCGGTCA
c)   T  D  E  E  R  D  I  R  A  R  Y  D  K  V  K  G  S  A  V  N

          610                 630                 650
a) ACCCGGTCCTGCGCGAGGGCAACTCCGACCGGCGCGCCCCGGCGTCGGTGAAGAACTACG
c)   P  V  L  R  E  G  N  S  D  R  R  A  P  A  S  V  K  N  Y  A

          670                 690                 710
a) CCAAGACCCACCCGCACCGCATGGGCGCCTGGACCGGCGAGTCCAAGACCAACGTGGCGA
c)   K  T  H  P  H  R  M  G  A  W  T  G  E  S  K  T  N  V  A  T

          730                 750                 770
a) CCATGGGCGAGAACGACTTCCGCTCCACCGAGAAGTCCGCGGTGATCGCCGAGGACGGCA
c)   M  G  E  N  D  F  R  S  T  E  K  S  A  V  I  A  E  D  G  T

          790                 810                 830
a) CCCTGCGGATCGAGCTGGTCGGCGACGACGGCACCACCACCGTGCTGCGCGAGTCGGTGC
c)   L  R  I  E  L  V  G  D  D  G  T  T  T  V  L  R  E  S  V  P

          850                 870                 890
a) CGGTGAAGAAGGACGAGGTCGTCGACGCCTCCGTACTGCGCGTCGCCGCCCTGCGCGAGT
c)   V  K  K  D  E  V  V  D  A  S  V  L  R  V  A  A  L  R  E  F

          910                 930                 950
a) TCCTCACCGCGCAGGTCGCCCGCGCCAAGGCCGAGGGCATCCTCTACTCGGTGCACCTGA
c)   L  T  A  Q  V  A  R  A  K  A  E  G  I  L  Y  S  V  H  L  K
```

```
      970                 990                1010
a) AGGCCACGATGATGAAGGTCTCCGACCCGATCATCTTCGGCCACGTGGTCCGCGCCTTCT
c)   A  T  M  M  K  V  S  D  P  I  I  F  G  H  V  V  R  A  F  F

         1030                1050                1070
a) TCCCGAAGACCTTCGCCCAGTACGGCGACGTGCTCGCCAAGGCCGGCCTCACCCCGAACG
c)   P  K  T  F  A  Q  Y  G  D  V  L  A  K  A  G  L  T  P  N  D

         1090                1110                1130
a) ACGGCCTGGGCGGCATCTACAAGGGCCTGGAGTCCCTGCCCGAGGGCGCCGAGATCAAGG
c)   G  L  G  G  I  Y  K  G  L  E  S  L  P  E  G  A  E  I  K  A

         1150                1170                1190
a) CGTCCTTCGACGCCGAGCTGGCCGAGGGCCCGGAGCTGGCCATGGTCGACTCCGACAAGG
c)   S  F  D  A  E  L  A  E  G  P  E  L  A  M  V  D  S  D  K  G

         1210                1230                1250
a) GCATCACCAACCTGCACGTGCCCTCGGACGTCATCATCGACGCGTCCATGCCGGCCATGA
c)   I  T  N  L  H  V  P  S  D  V  I  I  D  A  S  M  P  A  M  I

         1270                1290                1310
a) TCCGCACCTCCGGTCACATGTGGGGCGCGGACGGCGGCGAGCACGACGCCCTCGCCGTGA
c)   R  T  S  G  H  M  W  G  A  D  G  G  E  H  D  A  L  A  V  I

         1330                1350                1370
a) TCCCGGACTCCTCGTACGCCGGTGTGTACCAGGCCGTCATCGAGGACTGCCGCGCCAACC
c)   P  D  S  S  Y  A  G  V  Y  Q  A  V  I  E  D  C  R  A  N  R

         1390                1410                1430
a) GGCCGCTTCACGACCCCTCCACCATGGGTTCGGTCCCGAACGTCGGCCTCATGGCGCAGA
c)   P  L  H  D  P  S  T  M  G  S  V  P  N  V  G  L  M  A  Q  K

         1450                1470                1490
a) AGGCCGAGGAGTACGGCAGCCACGACAAGACCTTCGAGATCCCCACCACGGGCACCGTCC
c)   A  E  E  Y  G  S  H  D  K  T  F  E  I  P  T  T  G  T  V  R

         1510                1530                1550
a) GCCTGGTCGCACGACCGCGGGCAACGCGGGTGCTGGAGCAGACGGTCTCCGCCGGTGACA
c)   L  V  A  R  P  R  A  T  R  V  L  E  Q  T  V  S  A  G  D  I

         1570                1590                1610
a) TCTTCCGCGCCTGCCAGACCAAGGACGCGCCGATCCGCGACTGGGTGAAGCTGGCCGTCA
c)   F  R  A  C  Q  T  K  D  A  P  I  R  D  W  V  K  L  A  V  T

         1630                1650                1670
a) CCCGCGCGCGCGCCACCGGCGACCCGGCCGTCTTCTGGCTGGACGAGGGCCGCGCCCACG
c)   R  A  R  A  T  G  D  P  A  V  F  W  L  D  E  G  R  A  H  D

         1690                1710                1730
a) ACGCCAACCTGATCGCCAAGGTCAAGCAGTACCTGTCGGAGCACGACACCGAGGGCCTGG
c)   A  N  L  I  A  K  V  K  Q  Y  L  S  E  H  D  T  E  G  L  D

         1750                1770                1790
a) ACATCCGGATCCTCGACCCGGTCGAGGCGACCAAGCTGTCGGTGGAGCGCATCCGCCGCG
c)   I  R  I  L  D  P  V  E  A  T  K  L  S  V  E  R  I  R  R  G

         1810                1830                1850
a) GCGAGAACACCATCTCGGTCACCGGCAACGTGCTGCGCGACTACCTGACCGACCTGTTCC
c)   E  N  T  I  S  V  T  G  N  V  L  R  D  Y  L  T  D  L  F  P

         1870                1890                1910
a) CGATCCTGGAGCTGGGCACCAGCGCCAAGATGCTGTCGGTCGTCCCGCTGATGGCGGGCG
c)   I  L  E  L  G  T  S  A  K  M  L  S  V  V  P  L  M  A  G  G
```

```
     1930                 1950                 1970
a) GCGGCCTCTTCGAGACCGGTGCCGGCGGTTCGGCCCCGAAGCACGTGCAGCAGCTGGTCA
c)   G  L  F  E  T  G  A  G  G  S  A  P  K  H  V  Q  Q  L  V  K

          1990                 2010                 2030
a) AGGAGGACTACCTGCGCTGGGACTCGCTCGGTGAGTTCTTCGCCCTGGTGCCGTCCCTGG
c)   E  D  Y  L  R  W  D  S  L  G  E  F  F  A  L  V  P  S  L  E

          2050                 2070                 2090
a) AGAAGTACGCCGAGGCCACGGGCAACGGCAAGGCCAAGGTTCTCGCCGACACCCTCGACC
c)   K  Y  A  E  A  T  G  N  G  K  A  K  V  L  A  D  T  L  D  R

          2110                 2130                 2150
a) GCGCCACGGCGACGTTCCTCAACGAGGACAAGTCCCCGACCCGCCGCGTCGGCGGCATCG
c)   A  T  A  T  F  L  N  E  D  K  S  P  T  R  R  V  G  G  I  D

          2170                 2190                 2210
a) ACAACCGCGGCAGCCACTTCTACCTGTCCCTGTCTGGGCGCAGGGAGCTGGCGAAGCAGA
c)   N  R  G  S  H  F  Y  L  S  L  S  G  R  R  E  L  A  K  Q  T

          2230                 2250                 2270
a) CCGACGACGCGGACCTGGCCAAGGCCTTCGCCCCGCTCGCCGAGACGCTCACCGCGAGCG
c)   D  D  A  D  L  A  K  A  F  A  P  L  A  E  T  L  T  A  S  E

          2290                 2310                 2330
a) AGCAGAAGATCGTCGAGGAGCTGAACGCCGTCCAGGGCAAGCCGGCCGAGATCGGCGGCT
c)   Q  K  I  V  E  E  L  N  A  V  Q  G  K  P  A  E  I  G  G  Y

          2350                 2370                 2390
a) ACTACCAGCCCGACCCGGCCAAGGCCGCCAAGATCATGCGCCCGTCGACCACCTGGAACG
c)   Y  Q  P  D  P  A  K  A  A  K  I  M  R  P  S  T  T  W  N  E

          2410                 2430  ←  B  →  ←  2450
a) AGGCGCTGGCGTCCCTCGCCTGATCCTCGGGAGCCCCGCGCACCCCACGGGCACCCCGTT
c)   A  L  A  S  L  A                                  -35 V

          2470                 2490                 2510 RBS
a) TCCGTTCGGCGGAGGCGGGGTGTCCGTCTGTCTGCCGTCCCCCGACCTCCGGAGCACTCC
d)       -10 VI

          2530                 2550
a) TCATGACCGACGCCCCCGCCTCGTTCGCACTCGTCCCGGG
d)   M  T  D  A  P  A  S  F  A  L  V  P
```

5.3.3 Identification of Putative Ribosome Binding Sites (RBSs).

A conserved sequence of nucleotides is usually found just upstream of the start codon of many *E.coli* genes, that is normally required for efficient translation. It is complementary to at least part of the 3' end of the 16S rRNA, a constituent of the 30S ribosmal subunit complex required to initiate translation (Shine and Dalgarno, 1974). This sequence therefore provides a site which the 30S complex can recognise and bind to (helping it to distinguish between GUGs, UUGs and ATGs that are acting as start codons and those that encode amino acids internal to the coding region).

However exceptions have been cited where mRNA molecules have no RBS and yet are efficiently expressed. For instance the mRNA that is transcribed from the *prm* promoter of the *λC1* gene has no mRNA upstream of the start codon leaving no room for a normal SD ribosome binding site (Ptashne, 1976). This has also been found for the*aph* gene (neomycin resistance) from *Streptomyces fradiae*. When transcribed from the P1 promoter, it also lacks "leader" mRNA. Such transcripts seem to be translated efficiently anyway at the appropriate times of the life cycle (Janssen *et al.*, 1989). Janssen has also reported the removal of Shine-Delgarno sequences from *E.coli* genes (with reporter genes fused downstream) which have also been efficiently translated. It therefore seems that SD-sequences are not an absolute requirement for efficient translation raising questions as to the actual role that RBSs play (Janssen, ISBA conference 1991).

The sequence below shows a DNA sequence which is excactly complementary to the 16S ribosomal RNA from *S.lividans* and represents an ideal ribosome binding sequence for streptomycetes (Bibb and Cohen, 1982).

5'- AGAAAGGAGGTGATC -3'

Most sequenced streptomycete genes have been found to posses an upstream region of DNA that is similar to this sequence in a manner analogous to the Shine-Delgarno sequence found for most *E.coli* genes. Most of these SD sequence are found 1-18 bases away from the start codon (with an average of 10 from the G highlighted in bold above) and contain 4-12 matches to the above sequence (average of 6, Seno and Baltz 1989).

Analysis of the appropriate region upstream of the *S.coelicolor icd* B gene has revealed three possible RBSs (see figure 5.4). Sequence B could be expected to act as a RBS, being as similar to the ideal sequence (with a total of 5 matches) as many of the other streptomycete genes and having a distance from the start codon (9 nt) very close to the mean for streptomycetes. Sequence A, which is further away (13nt from the start codon), has 7 matches to the ideal sequence. The third sequence (sequence C) also has 7

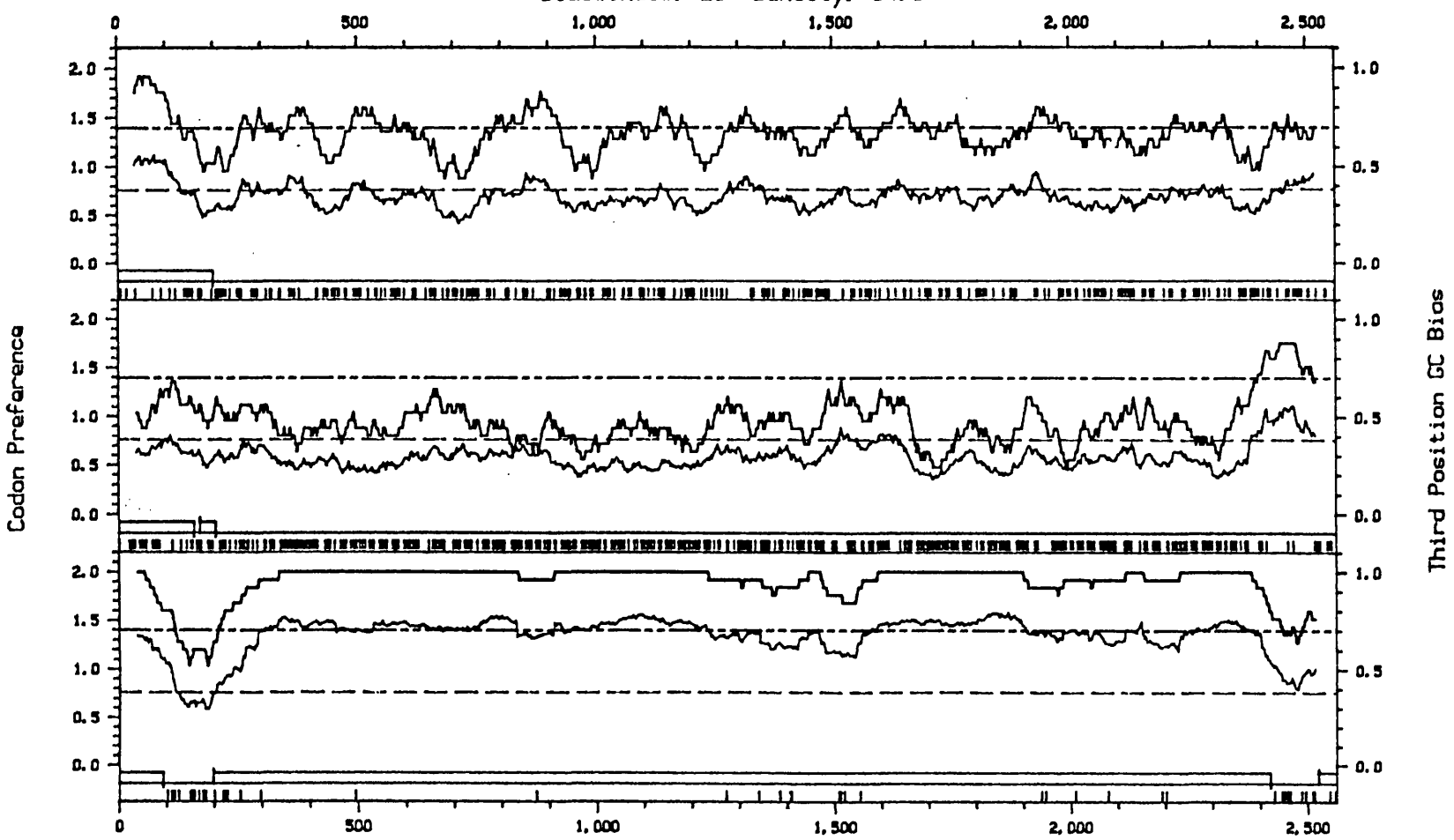

CODONPREFERENCE of: Nopol.Seq Ck: 1974, 1 to 2560 May 8, 1992 18:55
Codon Table: newone.cod PrefWindow: 25 Rare Codon Threshold: 0.10
BiasWindow: 25 Density: 84.0
Codon Preference
Third Position GC Bias
0
500
1,000
1,500
2,000
2,500
2.0
1.5
1.0
0.5
0.0

Figure 5.3 Analysis of the region of sequenced DNA for ORFs.

The computer program "CODONPREFERENCE" was used to analyse the region of DNA sequenced (see figure 5.2 and section 5.31). The plot is divided into three sections, each representing one of the three forward frames of the sequence. The first section shows the data calculated for frame 1, the second section for frame 2 and the third section for frame 3.

Each section contains two graphs. The upper of the two traces shows the G+C content for the third nucleotide of codons in each frame (*ie.* the fraction of codons within a window of 25 containing a G or a C in the third position).

The second of the two traces indicates how well the codon bias for the sequence follows the codon usage for the organism in general (comparisons were made to the codon usage table shown in figure 5.6a). The mean 'P' value for codons within the window size chosen (in this case a size of 25 codons) is plotted. Every synonomous codon is given a 'P' value that indicates how frequently the codon is used compared to one of the other synonomous codons. A 'P' value of greater than 1 indicates that the codon is used more frequently than would be expected in a random sequence (with the same nucleotide composition) and a value of less than 1 indicates that it is used less frequently than in a random sequence (see GCG reference manual for details).

Rare codons were also highlghted in these traces by placing a mark on the horizontal axis of each trace. Rare codons were defined in this case as those codons with a codons with a 'P' value of less than 0.1.

matches but overlaps with the start codon. As it is possible for ribosomes to bind to and translate mRNA with no leader sequences (see above), it is possible that this sequence could be functional.

It is impossible to tell from sequence analysis alone which of these sequences act as the functional RBS for the *icd* B gene, or if more than one is involved. It is possible that more than one is functional, although what function this would serve is unclear. Further studies would be required to solve this question maybe by using techniques such as footprinting (using the ribosome complex to protect mRNA) or by studying the effect of mutants created by site directed mutagenesis (*eg.* Janssen *et al.*, 1989).

5.3.4 Identification of putative promoter sequences.

It was described in section 1.1.3 how RNApol and promoter heterogeniety in streptomycetes play a key role in the differential control of gene expression. The relationship between promoter structure and gene function and the level and times of transcription is, however, still unclear. Such relationships will only be solved by the study of more promoters. It was therefore of interest to examine the upstream sequences of the *icd* B gene and compare it to known promoters as a first step in the study of the *icd* B promoter region .

Seno and Baltz (1989) compiled a consensus sequence from those streptomycete promoters that had been identified experimentally and found that the sequence is less representative than the *E.coli* counterpart. This is because promoter sequences from streptomycetes are variable, probably allowing for differential control of genes *via* control of transcription. However attempts to correlate gene funtion to the type of promoters used were unsuccessful (see section 1.1.3).

The DNA sequence upsteam of the *icd* B gene was examined by eye to look for sequences similar to known streptomycete promoters. Two sequences were identified that show some poor similarities to the -10 and -35 consensus sequences from both streptomycetes and *E.coli* (see Figure 5.5 a). Sequence I shows 50% identity to both consensus sequences with the gap between the -10 and -35 sequences being close to average. The close proximity to the RBS would require transcription to start straight after the last base of the -10 sequence (see Figure 5.2). Although this is unusual I have already described one example in streptomycetes (Janssen *et al.*, 1989) where mRNA has been transcribed with no leader sequence at all.

Sequence II would be more likely to act as a promoter. Although it only shows 42% identity to the *E.coli* sequence it shows 58% identity to the streptomycete sequence and lies further away from the putative RBS. A region was also found with 100% identity to 7 of the nucleotides in the -35 region of *gln* AP2 (Wray and Fisher, 1988) although no similarity was found to the -10 region of this promoter, (see figure 5.5 b). Yet another region was identified which matches the -10 region from the *aph* DP1 promoter in 8 out of 9 nucleotides (streptomysin resistance, Distler *et al.*, 1987 a and b) but shows no similarity to the -10 region of this promoter (Figure 5.5 c).

As described in section 1.1.3 more than one of these sequences may act as a promoter since there are now several examples of genes with multiple promoters. If this were the case then it may indicate that IDH is expressed at more than one stage of the lifecycle.

Unfortunately it is impossible to tell which, if any, of these putative promoters are functional as they have been identified solely by sequence similarities. It is possible that the real promoter lies further upstream from the sequence information available to us. Some promoters have been found to lie greater than 150 nucleotides upstream of the RBS. For example the P1 promoter of *gln* A lies 173 nt away from the RBS (Wray and Fisher, 1988) and *aph* D P2 is found 403 nt upstream (Distler *et al.*, 1987a). Analysis of the frame upstream of the *icd* B gene (section 5.3.7) may even suggest that the *icd* B is expressed as one of the distal genes within an operon, in which case there would be no promoter within the region of DNA sequenced.

These possibilities can be resolved only by further experiments. High resolution S1 mapping could be used to identify the point from which the mRNA is transcribed. The appropriate region upstream of this point (*ie.* at distances of approximately 10 and 35 nucleotides upstream of the mRNA start point) could then be examined to check for similarity to known promoter sequences. To check that any sequence(s) identified are functional as promoters, *in vitro* site directed mutagenesis could be performed to see if mutations will affect rates of transcription (*e.g.* Bibb *et al.*, 1985 and Janssen *et al.*, 1989). Footprinting experiments could also be performed, using the various streptomycete RNApol holoenzymes to protect the DNA (Buttner and Brown, 1985).

```
a)    177 AGACGTGGAGCGGAGCACCCGGTGACTGAC 206
b)    182 -----tGGAGcgGAgC-------------- 192
c)    188 ----------GGAGCacccg--------- 197
d)    193 ----------------AcccGGTGAcT--- 203
```

Figure 5.4 Three potential ribosome binding sites (RBSs) have been identified just upstream of the *S.coelicolor icd* B gene.

a) Shows the sequence of the region up and down stream of the *icd* B the translation start site (see figure 5.3). The GTG start codon is shown in **bold.**
b), c), d). The three potential RBSs are highlighted. The bases that show similarity to the streptomycete RBS consensus sequence (AGAAAGGAGGTGATC) are shown in CAPITAL letters (see section 5.3.3). The position of the first and last base in each sequence are shown, with the numbers that correlating to those used in figure 5.3.

```
a)                                   -35              -10
E.coli consensus                     TTGACA           TATAAT
Streptomycete consensus              TTGA CA          TAGG AT
I                              157   TTGAtg-(18)-     cAGacg-( 21)-GTG-198
II                              51   cTGcCg-(11)-     gAGGAc-(106)-GTG-198

b)
gln AP2 (-35)                        GGTCACG
III                            143-GGTCACG-(46)-GTG-198

c)                                    IV                        V
                               97-agCGCGGCGtggccggaatctGGGTGCgA-124
                                  TTCCCGGCG----------GGGTGCTAC
                                   tyl F                 aph DP1
```

Figure 5.5 Possible promoter sequences upstream of *icd* B.

The region upstream of the *icd* B translation start codon was examined for possible promoter sequences by comparing the DNA sequence to promoters from several other genes (reviewed by Seno and Baltz, 1989). Several sequences were identified that show some similarity to the promoters indicated below.

a) Shows the consensus sequences for the -10 and -35 regions of the *E.coli* and streptomycete promoters. The nucleotides most conserved between different promoters are shown in **bold.** Two different sequences were identified (labeled as I and II, also see figure 5.2) that showed some similarities to one or the other of the promoter consensus sequences.

b) A sequence was identified that possesses 100% identity to the -35 region of the second promoter (P2) for *gln* A. Nucleotides highlighted in **bold** indicate the similarity of this sequence to the -35 region of the *E.coli* promoter consensus sequence (figure 5.5a).

c) The region of DNA sequence shown here contains a 5nt sequence that is similar to the -35 region of the *tyl* F promoter (labeled as region IV). 13nt downstream from this possible -35 region there is a sequence (labeled as region V) that is identical to the -10 region of the *aph* DP1 promoter except for one mismatch. Sequence identities are shown in CAPITALS.

Numbers not in brackets indicate the position of bases with respect to figure 5.3.
Numbers in brackets indicate the number of nucleotides separating different regions.

5.3.5 A potential transcriptional terminator for *icd* B.

Terminator structures are required in order to signal the position at which the RNApol should stop transcribing DNA. Such signals are most commonly found in *E.coli* , in the form of a *rho* independent terminator, which is a sequence of DNA that has the ability to form a stable stem and loop structure (Rosenberg and Court, 1979). Very similar sequences are also found in many streptomycete genes, although in the case of the *hyg* gene the mRNA appears to terminate at the 5' side of the stem rather than on the 3' side which is the normal position (Zalacain, 1986).

A search for such a terminator structure around the 3' region of the streptomycete *icd* B gene was performed using the GCG program "TERMINATOR", which identifies structures using the method of Brandel and Trifonov, (1984 a and b). When this program is run with a stringency of 3.5 it is reported that 95% of all possible *rho* independent terminators are identified. The program identified only one potential terminator structure downstream of the *icd* B gene (see figure 5.7). The first base involved in binding in the stem of this structure, is found to be 6 nt away from the TGA stop codon and is therefore in the position expected for a true terminator structure.

5.3.6 Codon usage.

It has been suggested that if a particular tRNA species is present in limiting amounts, then the use of codons using this tRNA within a coding region would help to limit the expression of the gene. Such codons would be expected to occur only in genes for which this control of translation is required and would therefore occur less frequently than other synonomous codons used for incorporation of the same amino acids. In this way the codon usage of a gene would be able to affect the efficiency with which it is expressed (McCarthy and Gualerzi, 1990).

Such a situation has indeed been found in *S.coelicolor*. The TTA codon is very rare in streptomycetes and has to date been found only in genes that are involved in the regulation of development, antibiotic synthesis or resistance (Leskiw *et al.*, 1991b). In fact mutations in the *bld* A gene that encodes for $tRNA^{leu}_{TTA}$ prevent the organism from sporulating and producing antibiotics but have no visible effect on its vegetative growth (Lawlor *et al.*, 1987). The complete dependence of some genes on the presence of $tRNA^{leu}_{TTA}$ for expression was shown by the work of Leskiw *et al.* (1991a). They converted the TTA codon found in *car* B to a CTC codon (which also encodes for leucine). The *car* B gene was then found to express in *bld* A mutants in the absence of $tRNA^{leu}_{TTA}$.

a) Codon usage table compiled from 67 streptomycete genes.

	G		A		T		C		
G	Gly	19	Glu	81	Val	39	Ala	33	G
	Gly	10	Glu	19	Val	2	Ala	4	A
	Gly	8	Asp	4	Val	4	Ala	3	T
	Gly	64	Asp	96	Val	56	Ala	59	C
A	Arg	6	Lys	95	Met	0	Thr	30	G
	Arg	1	Lys	5	Ile	3	Thr	3	A
	Ser	3	Asn	4	Ile	5	Thr	2	T
	Ser	28	Asn	96	Ile	92	Thr	65	C
T	Trp	100	End	14	Leu	3	Ser	26	G
	End	83	End	3	Leu	0	Ser	3	A
	Cys	13	Tyr	5	Phe	1	Ser	1	T
	Cys	87	Tyr	95	Phe	99	Ser	39	C
C	Arg	36	Gln	93	Leu	55	Pro	52	G
	Arg	5	Gln	7	Leu	0	Pro	2	A
	Arg	7	His	6	Leu	2	Pro	3	T
	Arg	46	His	94	Leu	39	Pro	43	C

b) Codon usage table for the *icd* B gene from *S.coelicolor*.

	G		A		T		C		
G	Gly	2	Glu	100	Val	36	Ala	27	G
	Gly	0	Glu	0	Val	2	Ala	2	A
	Gly	14	Asp	0	Val	2	Ala	0	T
	Gly	84	Asp	100	Val	60	Ala	71	C
A	Arg	2	Lys	100	Met	100	Thr	20	G
	Arg	0	Lys	0	Ile	0	Thr	2	A
	Ser	0	Asn	0	Ile	0	Thr	4	T
	Ser	9	Asn	100	Ile	100	Thr	75	C
T	Trp	100	End	-	Leu	2	Ser	42	G
	End	100	End	-	Leu	0	Ser	0	A
	Cys	0	Tyr	10	Phe	0	Ser	2	T
	Cys	100	Tyr	90	Phe	100	Ser	47	C
C	Arg	16	Gln	100	Leu	70	Pro	85	G
	Arg	2	Gln	0	Leu	0	Pro	0	A
	Arg	0	His	0	Leu	2	Pro	0	T
	Arg	79	His	100	Leu	27	Pro	15	C

Figure 5.6. Comparison of codon usage in the *icd* B gene to that used by streptomycete genes in general.

a) The DNA sequences from 67 streptomycete coding regions were provided by Dr M Bibb from which a codon usage table was compiled using the program "xxxxxxxxxx" (provided as part of the GCG package). This table is therefore more representative than the table published by Seno and Baltz (1989), which was compiled from only 27 genes. However the new table was found to be very similar to the Seno and Baltz table, with the only significant difference being in the stop codons.

b) The frequency with which each codon is used in the *icd* B gene of *S.coelicolor* is shown in this table for comparison to the codon usage table shown in a).

The first nucleotide of a codon is shown in the first column of the tables above, the second nucleotide of a codon is shown in the top row and the third nucleotide shown in the last column. Cross referencing all three codon positions shows which amino acid is encoded and the frequency (expressed as a percentage) with which the codon is used to code for that amino acid.

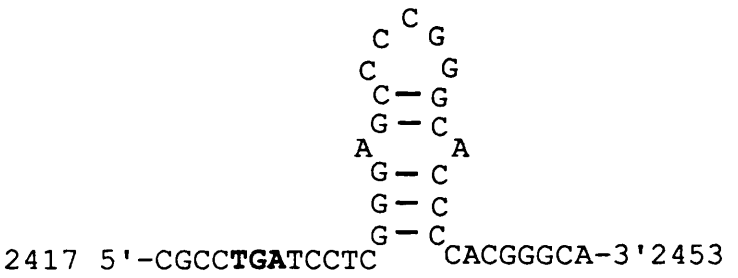


Figure 5.7 Possible terminator structure for the *icd* B gene.

Just downstream from the TGA stop codon (shown in **bold** type) for the *icd* B gene there is a sequence which has been identified as a potential *rho* independent terminator (see region **B** in figure 5.3).

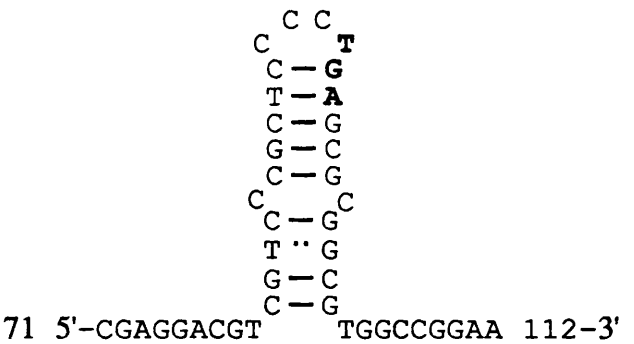


Figure 5.8 Possible terminator structure upstream of the *icd* B gene.

A sequence that is capable of forming a *rho* independent terminator has been identified upstream of the *icd* B start codon. It may form a terminator structure for a putative upstream ORF. The putative stop codon for this PORF (shown in **bold** type) would be included within the structure.

McCarthy and Gualerzi (1990) have suggested that the presence of rare codons would have the most effect if they were situated near to the 5' end of a gene, where a stalled ribosome could also affect the initiation of further transcription. In 7 out of the 11 genes containing TTA the codon occurs within the first 50 codons, allowing for such a steric interference at the ribosome binding site.

Other rare codons, such as CTA and TTT, could be expected to play a similar role. However these codons occur randomly both in genes required for vegatative growth and those used for other functions. It is therefore unlikely that they play a role in the control of either metabolic or morphological differentiation (Leskiw *et al.*, 1991b). It is possible however that they may play a role in limiting the amount of transcription rather than the timing of it.

The complete absence of TTA codons so far in DNA sequences of genes involved in primary metabolism made it of interest to check the codon usage of the *S.coelicolor icd* B gene. As IDH is almost certainly an absolute requirement for vegetative growth it would not be expected to contain any TTA codons, if the dogma postulated by Leskiw were to stand true. It would be expected that there will be few, if any, other rare codons in what is likely to be gene that is expressed to a reasonably high level. Comparison of the codon usage tables in Figure 5.6 shows that this is indeed the case with several codons, especially those with a high A+T content, not being used at all. This is consistent with the view of Leskiw that no vegatative genes in streptomycetes will contain TTA codons. It would be of interest to find out what the effect would be of introducing such a codon into a gene such as *icd* B that does not normally possess one.

Although the evidence of Leskiw *et al.* (1991b) would indicate that the level of $tRNA^{leu}_{TTA}$ could be used to differentially control genes at different times of the lifecycle, recent work by M.Bibb (John Innes Institute, Norwich, personal communication) has indicated that this tRNA species is in fact present throughout the lifecycle. This indicates that any such control by the *bld* A product is more complex than first thought, perhaps requiring other regulatory elements in order to exert its effect.

5.3.7 Possible ORFs up and downstream of the *icd* B gene.

Analysis of the codon usage and nucleotide composition in the region upstream from the GTG start codon of the *icd* B gene shows that another ORF may exist at this point (Figure 5.3). However due to the short length of this region it is difficult to tell from this data alone in which direction the possible open reading frame (PORF) would be transcribed. This region was therefore translated in all six possible reading frames and the amino acids sequences used to screen databases using the GCG computer program

"FASTA". No significant similarities to other proteins were identified. This may not be surprising because of the short length of sequence available and the fact that the extreme ends of proteins are often not well conserved.

The absence of rare codons in frame 1 (of the sequence shown in figure 5.2; phase 3 in figure 5.3) would however suggest that the upstream sequence could produce a PORF that is reading in the same direction as *icd* B and in the same frame. The only possible TGA stop codon is at position 93-95 and lies within a region (see Figure 5.2) that could form a terminator like structure (see figure 5.8).

Figure 5.3 shows that there may be another ORF downstream of the *icd* B gene (indicated by ORF indicator in phase 3). The absence of any strong G/C or codon usage bias in this region does not necessarily preclude this possibility, as the extreme N-terminal and C-terminal ends of proteins do not show the biases typical of the rest of an ORF. This is illustrated well by the lack of codon and G/C bias at the extreme ends of the *icd* B coding region, (shown by examination of figure 5.3). The only possible start codon found in this region was an ATG triplet at position 2523-2525. At the optimum distance away from this ATG (9 nt upstream of ATG) there is also a sequence that contains 4 of the residues that are often conserved within Shine-Dalgarno type RBSs (see Figure 5.9). The region between the putative start codon and the stop codon of the *icd* B gene was therefore examined for possible promoter sequence(s). A stretch of DNA sequence was revealed that resembles the *gln* A-P2 -35 sequence and a *vph* P1 -10 sequence (see Figure 5.10). This region begins just 3nt downstream of the putative *icd* B terminator structure.

Further sequence information could help decide if these PORFs are in fact real. Sequence similarities to another known protein(s) could indicate whether the PORFs are capable of encoding for a protein.

It was interesting to find that there is a possible ORF just upstream of the *icd* B gene and another just downstream. The proximity of these two PORFs to the *icd* B ORF suggests that one or both of them may form part of an operon with *icd* B. S1 mapping would therefore be required to determine the ends of the mRNA.

```
RBS        AAGGAGGT
2508- acctccGGAGcactcctcATGacc -2527
```

Figure 5.9 **Possible ribosome binding site for the PORF downstream of *icd*.**

A PORF was found just downstream of the *icd* B gene with a start codon at position 2522 (see figure 5.3). A search for a possible RBS upstream of this start codon revealed a sequence that showed some similarity to the consensus RBS (see section 5.3.3). The *S.coelicolor* sequence is shown below the consensus sequence with identical matches highlighted in CAPITALS.

```
          V                          VI
2449-GGgCACcccgtttccgttCGGCGG -2472
     GGTCAC------------CGGCGG
     gln A (-35)              vph (-10)
```

Figure 5.10 **Possible promoter for the putative ORF found downstream of *icd* B.**

A PORF was found just downstream of the *icd* B gene with a start codon at position 2522 (see figure 5.3). A search for a possible promoter sequence upstream of this start codon revealed a sequence that showed some similarity to the -35 region of the *gln* A promoter (labeled as region V).12nt downstream from this there is a sequence that is identical to the -10 region of the *vph* promoter (labeled as region VI). Numbering indicates the position of bases in figure 5.3.

The *S.coelicolor* sequence is shown below the consensus sequence with identical matches highlighted in CAPITALS.

5.3 Summary.

A region of DNA that covers 2,406 bp has been sequenced. This has been found to contain an ORF that could code for a protein of 741 amino acids with a molecular weight of approximately 80,000. This indicates that a gene has been cloned that encodes for a protein large anough to be the IDH.

Comparison of the N-terminal amino acid sequence, predicted from the DNA sequence, to the N-terminal sequence of the native protein, showed that they are identical. This confirms that the gene has been cloned, that encodes the protein subjected to N-terminal sequencing. Post-translational modification must have occurred to remove the N-terminal formyl methionine since this initiation amino acid is absent from the N-terminal sequence of the native protein.

Several possible RBSs and promoters have been identified in the appropriate positions upstream of the start codon although their functional significance has yet to be proved. A possible terminator has also been identified 6 nt downstream of the stop codon. The *icd* B gene possesses a codon usage very similar to the mean codon usage for streptomycetes.

Analysis of the DNA up and downstream of the *icd* B gene has revealed the presence of a possible ORF in each region. If real, both would be transcribed in the same direction as *icd* B.

CHAPTER 6.

Analysis of the amino acid sequence of IDH from *S. coelicolor*.

Chapter 6. Analysis of the amino acid sequence of IDH from *S.coelicolor.*

6.1 Introduction.

The IDH of *S. coelicolor* is the first large, type II IDH, for which the entire peptide sequence has been deduced. Analysis of the sequence may provide some insight into the relatedness and differences between the type I and II enzymes. Some of the possibilities that may be revealed by sequence analysis are given below.

a) The *S. coelicolor* IDH (and maybe type II IDHs in general), could have been derived from the fusion of two identical (or almost identical) gene sequences of the type I class. This could lead to the production of a type II monomeric protein that is, to all intents and purposes, virtually identical to a type I homodimer. In this case one would expect to find repetition in the amino acid sequence of the enzyme from *S. coelicolor* .

b) A similar fusion could have occurred between two gene sequences that were quite different to those that encode the type I monomer. If these sequences were similar then one would expect to find repetition in the amino acid sequence of the enzyme from *S. coelicolor* (*ie.* equivalent of a fused homodimer) If the sequences were different to each other then repetition in amino acid sequence would not be expected (*ie.* equivalent of a fused heterodimer). However in both cases one might expect the presence of two $NADP^+$ and two isocitrate binding sites, although it may not be possible to identify them at the primary sequence level alone.

c) The IDH from *S. coelicolor* could be very similar to the type I enzymes but with extra sequences distributed throughout the protein. These extra sequences might be present to substitute for regions which are responsible for subunit/subunit interactions in a homodimer and that might be required to allow the enzyme to assume the correct conformation for activity. This would assume that the the type II IDHs have evolved from the type I IDHs or from an ancestor that was common to both types. Alternatively the type I IDHs could have evolved from the type II enzymes; the ability to form a dimer could have obviated the necessity for some of the sequence found in the type II IDHs. In these cases one would expect to find only one $NADP^+$ and one isocitrate binding site in the IDH from *S. coelicolor.*

d) The IDH from *S. coelicolor* might be completely different to any of the IDH sequences known so far or may possess only some small regions of similarity that have a functional significance (*eg.* an $NADP^+$ or isocitrate binding site). This in itself would be interesting, maybe giving more insight into residues required for the function of IDH. With this arrangement it is possible that the enzyme could possess one or two binding sites for both $NADP^+$ and isocitrate.

Thorsness and Koshland (1987) had pointed out that the $NADP^+$ dependent IDH from (*E. coli* NADP-IDH) had no significant sequence conservation to any other dehydrogenase with the exception of the functionally similar isopropylmalate dehydrogenases (IMDHs), which use NAD^+ as a cofactor.

It is also clear from examination of the three dimensional structure of the *E. coli* enzyme that none of these IDHs or IMDHs have a classical Rossmann fold type nucleotide binding site (Rossmann *et. al.*, 1974). Examination of the protein sequence also shows that they do not possess either of the consensus sequences often quoted for $NAD^+/NADP^+$ dependent enzymes (Scrutton *et al.*, 1991 and Wierenga *et al.*, 1988).

It was therefore also of interest to find out if the *S. coelicolor* enzyme possesses one of the classical $NAD^+(P)$ binding sites. Alternatively it might possess a nucleotide binding site more similar to that of the IMDHs, or possibly no recognizable binding site at all. Attempts were also made to see if any conservation could be found for isocitrate binding sites.

Since this study was performed two other IDH sequences have become available which have confirmed some of the findings discussed in this chapter and are mentioned further in section 6.3.1., 6.3.2 and figure 6.7.

6.2 Results and Discussion.

6.2.1 Search for consensus sequences.

The Apple Macintosh program "MacPattern" was used to search for consensus sequences in the *S. coelicolor* enzyme sequence. The most relevant ones are described below.

6.2.1.a) Lactate dehydrogenase (LDH) consensus.

The consensus shown below is based around a histidine that is known to be required for the catalytic mechanism (Abad-Zapatero *et.al.*, 1987).

[LIVMA]-G-[EQ]-H-G-[DN]-[ST]

-where the residues labeled in brackets show the variation at individual sites.

6.2.1.b) IDH/IMDH consensus sequence.

This consensus was based on the region around the only histidine that was absolutely conserved amongst all of the IMDHs and the *E. coli* IDH. This residue was assumed to have a catalytic role based on the information available from LDH and shows a little resemblance to the LDH consensus sequence. This region has now been found to contain the bulk of the residues that contribute towards the binding of $NADP^+$ (Hurley *et.al.*, 1991).

E-[AP]-x-H-G-[ST]-A-P-x-x-[AP]

-where the 'x' can be any amino acid.

NOTE: This sequence is provided within the Apple Macintosh computer program "MACPATTERN" as shown above. However examination of the alignment of the IMDHs and the IDH of *E. coli* (shown by Hurley *et.al.*, 1989), shows that this sequence has been entered into the program incorrectly. The final residue highlighted in the sequence above (P) should be printed as a G.

No matches to either of these consensus sequences were found in the *S. coelicolor* enzyme using this program.

6.2.1.c) Nucleotide binding consensus sequences.

A modification of the NAD^+ binding site of Rossmann *et al.* (1974) was proposed by Wierenga *et al.* (1986) and Scrutton *et al.* (1990), for the nucleotide binding site of some $NADP^+$ dependent enzymes. The "FIND" program (supplied with the GCG package) was used to search for both of these consensus sequences in the *S. coelicolor* IDH sequence. A sequence identical to the $NADP^+$ binding site of Scrutton *et al.* was found as shown below. The numbers correspond to the first and last residues of the *S. coelicolor* sequence.

NAD^+ Consensus		G-X-G-X-X-G	
$NADP^+$ Consensus		G-X-G-X-X-A	
A region of the *S. coelicolor* IDH sequence.	580	-G-A-G-G-S-A-	585

None of these consensus sequences are found in any of the other sequences available (*ie.* the yeast and *E. coli* IDHs and the IMDHs). One cannot attach too much significance to this as there are no further obvious similarities to the rest of the proposed consensus sequences.

6.2.2 Comparison of the *S. coelicolor* enzyme to the other IDHs and IMDHs.

6.2.2.a) Comparisons of sequences using "DOTPLOT".

It is possible that there are only small regions of similarity between the *S. coelicolor* IDH and the other IDHs and IMDHs and that the proteins are otherwise dissimilar. Dotplots are a convenient way to compare such proteins as every region of one protein is compared to every region of another. The method used by Maizel and Lenk (1981) was used (supplied in the GCG package as a program called "COMPARE"). A comparison between the two proteins is made within a window of specified size. If the number of residues within this window equals or exceeds a specified threshold level (the stringency) then a match is scored. A window from one of the proteins is compared to every register of the second protein. The window is then moved successively by one amino acid at a time along the first sequence and the analysis repeated. In this way every register of one protein is compared to every register of the other and can reveal small regions of similarity, even if insertions or deletions are present in one of the proteins relative to the other.

Results are then displayed using the program "DOTPLOT" in which matches are represented by a dot placed in the appropriate position of a graph between two sequences. The dot is placed in a position corresponding to the middle of the matching window. Lines on the graphs representing regions of similarity therefore need to be extended by half a window length at each end, to deduce the actual coordinates of the matching residues.

When comparing the *S. coelicolor* sequence to the other IMDH and IDH sequences several window sizes were tried with different stringencies. A window size of 30 and a stringency of 15 to 17 were found to yield the best results. This reduced the amount of non-specific background matches, whilst leaving the most relevant matches more or less intact. The effect of such alterations are shown in figure 6.1. The validity of these conditions were tested by comparing the *E. coli* IDH to two of the IMDHs which are known to be similar. IMDH from *Bacillus subtilis* and *Thermus thermophilus* were used in all of the comparisons in this study because they are, respectively, the most and least similar in comparisons to the *E. coli* IDH (Hurley *et al.* 1989), providing a yard-stick for other comparisons (see figure 6.2 h and j).

It can be seen from these plots, that the *E. coli* IDH shows some regions of considerable similarity to each of the IMDHs and *E. coli* IDH. Similar plots were produced by comparing the yeast enzyme to the IMDHs. These plots suggest that all of the sequences are to some extent similar to each other. This is particularly true for the region marked 'B' which is common to all of the plots.

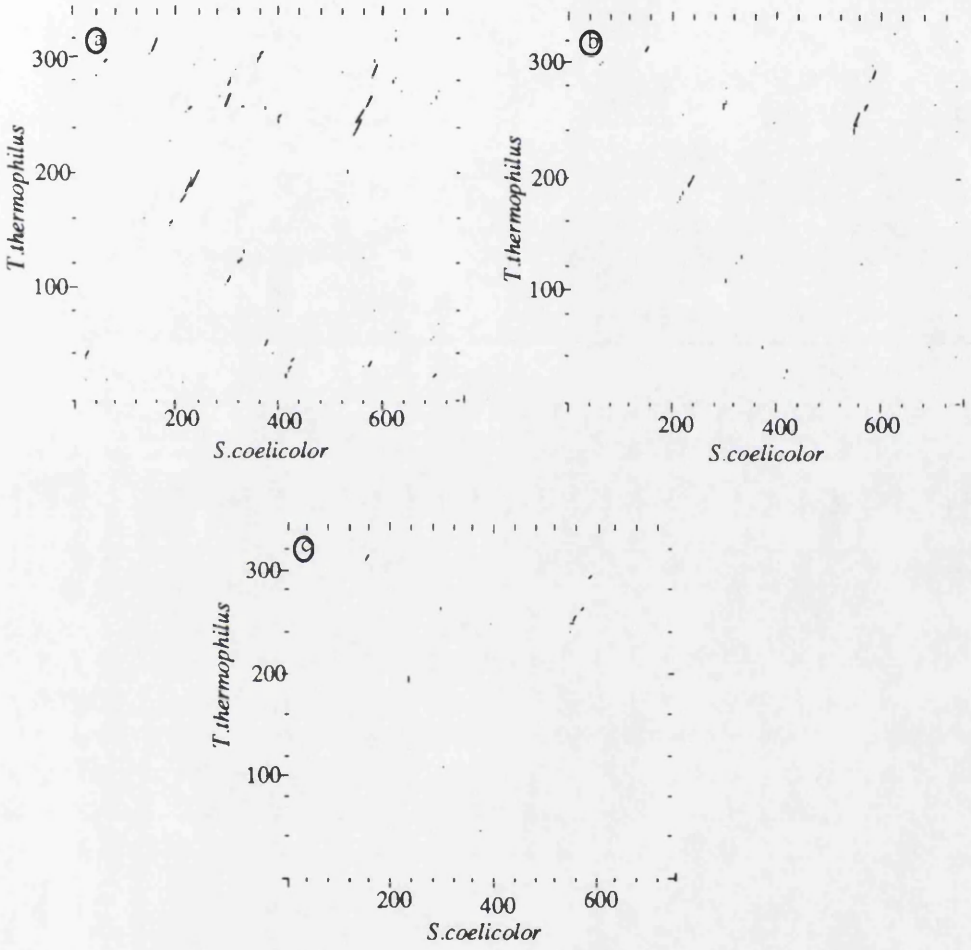


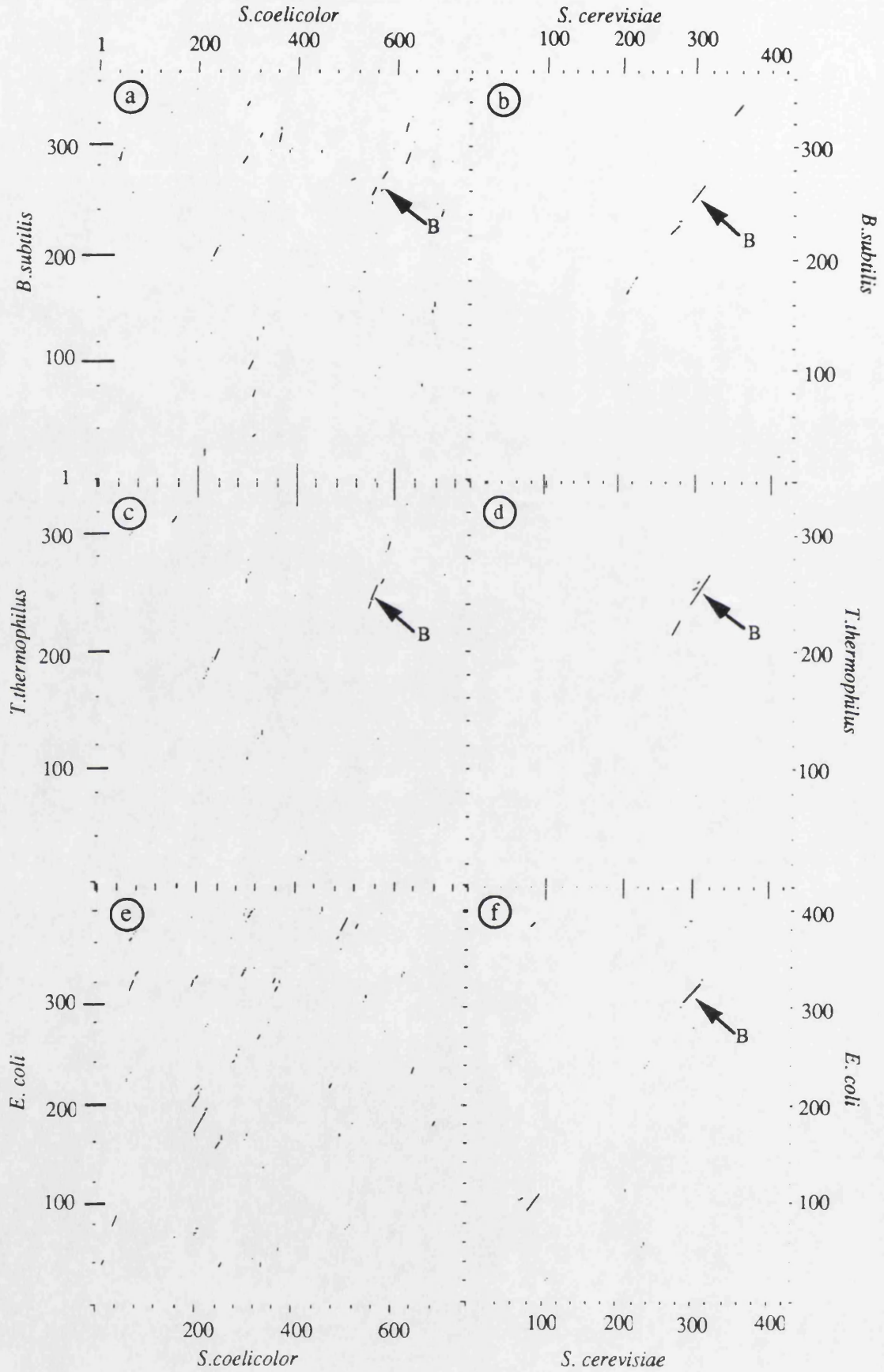
Figure 6.1 Comparison of the IDH sequence from *S. coelicolor* to the IMDH sequence from *T.thermophilus* using "DOTPLOTS".

The amino acid sequence of the IDH from *S. coelicolor* was compared to the amino acid sequence of the IMDH from *T.thermophilus*. The effect of altering the stringency parameter (using the "COMPARE" program) is shown using a window size of 30. Data was displayed using the "DOTPLOT" program.

Plot a) Was performed using a stringency of 15.

Plot b) Was performed using a stringency of 16.

Plot c) Was performed using a stringency of 17.


S.coelicolor
S. cerevisiae
B.subtilis
T.thermophilus
E. coli
a
b
c
d
e
f
B
1
200
400
600
100
300

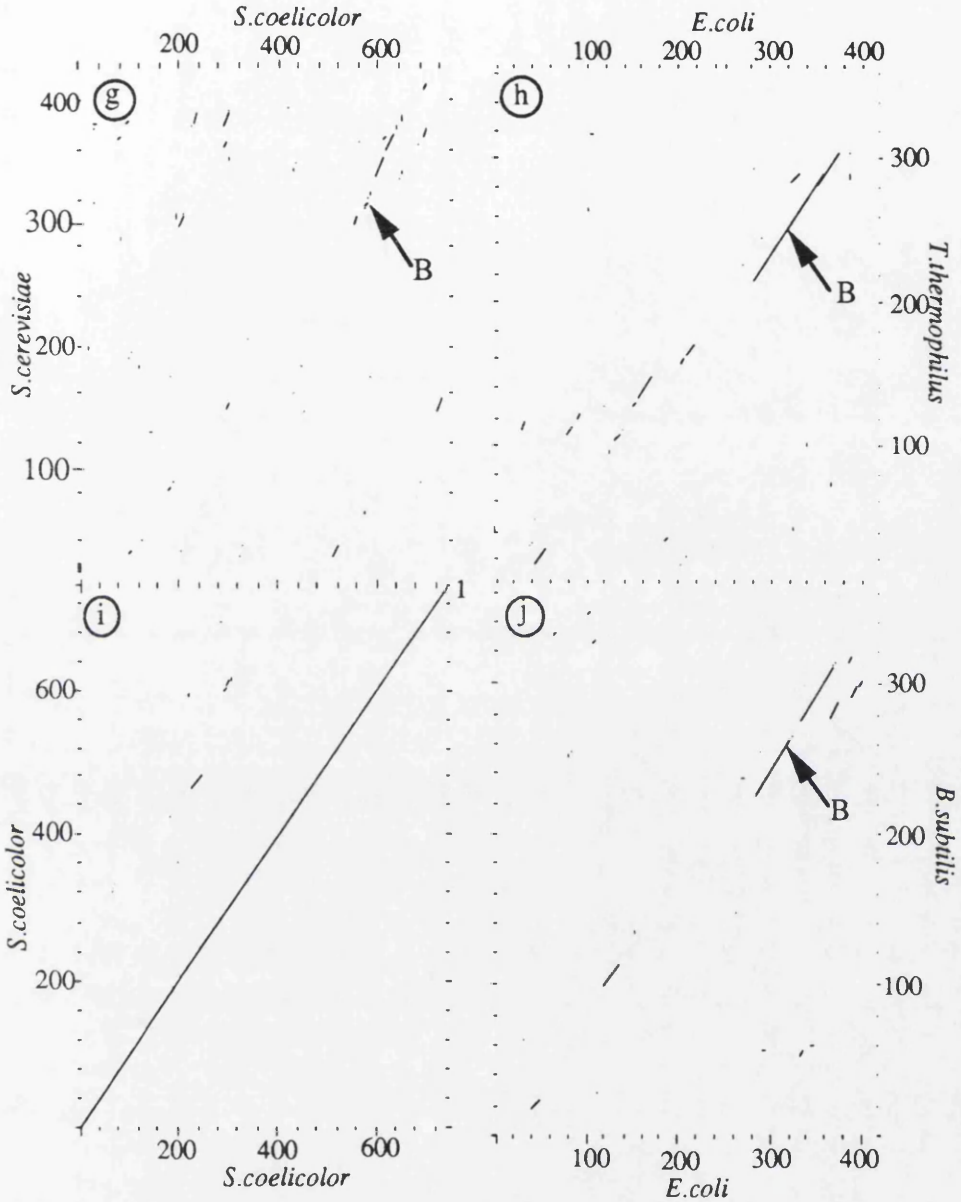


Figure 6.2 **Comparison of some IDH and IMDH sequences using "DOTPLOTS".**

The amino acid sequences of the IMDHs from *T.thermophilus* and *B.subtilus* were compared to the amino acid sequences of the IDHs from *S.cerevisiae, E. coli* and *S. coelicolor* , using the computer program "COMPARE" with a window size of 30. Results were displayed using the program "DOTPLOT". The stringency used for each plot is shown below in square brackets.

a)*S. coelicolor* IDH *vs. B. subtilis* IMDH [15]
b)*S. cerevisiae* IDH *vs. B. subtilis* IMDH [16]
c)*S. coelicolor* IDH *vs. T. thermophilus* IMDH [16]
d)*S. cerevisiae* IDH *vs. T. thermophilus* IMDH [16]
e)*S. coelicolor* IDH *vs. E. coli* IDH [16]
f)*S. cerevisiae* IDH *vs. E. coli* IDH [16]
g)*S. coelicolor* IDH *vs. S. cerevisiae* IDH [15]
h)*E. coli* IDH *vs. B. subtilis* IMDH [16]
i)*S. coelicolor* IDH *vs. S. coelicolor* IDH [17]
j)*E. coli* IDH *vs. T. thermophilus* IMDH [16]

Moreover, a similar region was also seen in a comparison of the yeast enzyme to the IDH from *S. coelicolor* (see plot g). This coincides with region 'B' of the yeast enzyme that is common to both IMDHs and the *E. coli* IDH. One would therefore expect to find this same region of similarity in comparisons between the *S. coelicolor* IDH and the IMDHs.

Examination of the plots between the IMDHs and the *S. coelicolor* enzyme confirms this, producing a staggered region of similarity, also labeled 'B' in plots a) and c). It is interesting to note that although the *S. coelicolor* IDH is similar to the yeast enzyme and both IMDHs, and the *E. coli* IDH shows similarity to the IMDHs and the *S.cerevisiae* enzyme, there is no obvious similarity between the *E. coli* and *S. coelicolor* enzymes.

Figure 6.2 shows that the region common to all of these proteins corresponds to the approximate region 310-330 of the *E. coli* sequence. This lies next to the region containing the bulk of the $NADP^+$ binding residues in the *E. coli* IDH (Hurley *et al.* 1991). It is therefore possible that this region is also an $NADP^+$ binding site in the enzymes from *S.cerevisiae* and *S. coelicolor* . The approximate positions of the equivalent common region of the other proteins are also shown below.

Enzyme		Approximate Position of Region B.
T. thermophilus	IMDH	260-280
B.subtilus	IMDH	260-280
E. coli	IDH	330-350
S. cerevisiae	IDH	310-330
S. coelicolor	IDH	570-590

6.2.2.b) Comparison of sequences using "BESTFIT".

To investigate these similarities further, the algorithm of Smith and Waterman (1981) was used to test if these protein sequences could be aligned. The algorithm is supplied as part of the GCG package as a program called "BESTFIT". It makes optimal alignments between two sequences, of as long a length as possible, with the best "quality" score possible, by inserting gaps within the sequences. Scoring is performed by comparing the similarity between amino acids within a window based on a similarity table (provided in the GCG package) with a penalty for any gaps that are added. It is possible to assess the relevance of such alignments by comparing the scores produced from the alignments of ten randomised sequences (using the program 'BESTFIT/RAN'). These sequences are produced from the shuffling of the query sequences. A "quality score" that is significantly larger than the "mean random score" would indicate an alignment of significance.

"BESTFIT" alignments were first performed between the yeast IDH sequence, the *E. coli* IDH and the IMDHs. In each case an alignment was made which was similar to the alignment produced between the *E. coli* IDH and the IMDHs (see section 6.2.3 for the alignments between the *E. coli* IDH and the IMDHs). The scores for these alignments (*eg.* 144.9 for the alignment between IDHs from yeast and *E. coli*) were less than the equivalent scores between the *E. coli* IDH sequence and the IMDHs (*eg.* 166.2 for the alignment between IMDH from *T.thermophilus* and the IDH from *E. coli*). However comparison to the randomised scores (*eg.* 129.0 for the alignment between IDHs from yeast and *E. coli*) suggest that the alignments were in fact significant. The data for every alignment are shown in figure 6.3.

Comparisons were therefore also carried out between the yeast enzyme and that from *S. coelicolor* , as dotplots had suggested a small amount of similarity. When many gaps were added the entire sequence of the yeast enzyme lined up to virtually the whole of the *S. coelicolor* enzyme (figure 6.4). Although a fair degree of significance is indicated by comparing the 'quality score' to the 'randomised score' (see figure 6.3a), it is possible that much of the line-up is spurious. If the alignment was really significant then the *S. coelicolor* enzyme has numerous extra sequences distributed throughout the protein but is otherwise similar. These extra sequences could enable the protein to fold in some way that enables it to function as a monomer, negating the requirement of a second subunit or the inserted sequences may have another function (see section 6.1c).

As the yeast IDH shows sequence similarity to the IMDHs and also to the *S. coelicolor* IDH, one would expect the *S. coelicolor* enzyme to show sequence similarity to the IMDHs, as already suggested by the dotplots. This was indeed found to be the case. Alignments of the two IMDH sequences to the *S. coelicolor* IDH were virtually identical, with the whole of the IMDH sequences being aligned to the last three quarters of the C-terminal end of the *S. coelicolor* IDH. These alignments are not shown because much of this alignment could be spurious as the "BESTFIT" program attempts to line up the whole of the *S. coelicolor* protein sequence to the much smaller IMDH sequences. This could have obscured any significant alignments and has probably led to the low score shown in figure 6.3.

The alignments at the C-terminal ends of the proteins were, however, very similar to those between the yeast enzyme sequence and all the other IDH/IMDH sequences. The alignments at the C-terminus was also similar to the alignments between the *E. coli* IDH and the IMDHs. This region also corresponds to the common region (region "B") suggested by the dotplots, which lies next to several residues involved in $NADP^+$ binding in the *E. coli* IDH. Alignment over this region could therefore have a functional significance .

With both dotplot and bestfit analysis, every combination of comparisons between the various proteins produced a line-up over this same area. The only exception to this is the direct comparison of the IDH from *E. coli* with that from *S. coelicolor*, even though the other line-ups suggested that a similarity of some degree would be expected.

An evolutionary relationship between the proteins could therefore be suggested from this, although one that is only very distant. The *S. coelicolor* and *E. coli* proteins could have diverged in different directions to such an extent that there is no obvious sequence similarity left between them. If the IMDHs and the *S.cerevisiae* IDH were to lie somewhere between these two, in evolutionary terms, then one might expect to see the sort of poor sequence similarity that has been found. This similarity, if true, is unexpected as it would imply an evolutionary relationship between type I and type II IDHs.

6.2.2c) Comparison of the C-terminal regions of proteins using "BESTFIT".

Because the programme "BESTFIT" attempts to line up the whole length of of two proteins, it is possible that much of the alignments between the *S. coelicolor* IDH and other proteins are spurious and have created the low scores shown in figure 6.3a (section 6.2.2c). Bestfit analyses were therefore repeated, this time using only the sequence flanking this C-terminal area of interest which includes Region 'B'. Figure 6.3b summarises the scores for the comparisons between all of the proteins and indicates that all these alignments are significant with the exception of the line up between the *E. coli* and *S. coelicolor* IDHs. Figure 6.5 however shows that the *E. coli/S. coelicolor* comparison shows some similarities to the other alignment.

From these data it could be concluded that there is significant sequence similarity between all of the proteins over a region that includes many of the residues required for $NADP^+$ binding in the *E. coli* IDH. This suggests that an $NAD(P)^+$ binding site common to all of the proteins has been found. Attempts were therefore made to line up all of these sequences together to try to deduce which residues were common to all of the enzymes. These residues may have a catalytic function.

a) **"BESTFIT" results between the whole of each protein.**

	B.subtilis				*T.thermophilus*				*E.coli*				*S.cerevisiae*			
	Q	M	I	S	Q	M	I	S	Q	M	I	S	Q	M	I	S
S.co	136.6	134.1 +/-3.9	23.6	45.0	128.0	126.8 +/-3.6	24.0	44.1	153.5	148.1 +/-4.1	19.0	39.4	166.3	144.8 +/-2.7	22.2	48.7
S.ce	142.7	118.7 +/-4.7	23.0	43.8	122.3	110.9 +/-2.8	21.7	45.2	144.9	129.0 +/-5.2	21.6	48.4				
E.co	175.8	122.4 +/-2.8	27.5	53.5	166.2	116.4 +/-3.3	29.1	54.1								

b) **"BESTFIT" results over the common region (region B).**

S.co	68.3	61.2 +/-1.8	23.8	42.7	60.7	53.8 +/-1.9	24.6	44.2	55.3	55.1 +/-2.8	21.8	45.8	65.8	29.2 +/-2.9	22.3	48.4
S.ce	57.7	45.3 +/-1.9	30.2	50.0	52.2	43.4 +/-2.0	27.0	48.3	57.8	49.5 +/-2.0	25.2	50.4				
E.co	81.4	48.3 +/-2.7	29.9	53.3	76.8	43.5 +/-2.7	31.0	59.5								

Figure 6.3 Scores produced by alignments of IDHs to IMDHs using "BESTFIT".

The amino acid sequences of the IMDHs and IDHs were compared to each other using the GCG program "BESTFIT". The scores produced are summarised in the table above along with scores produced by comparison of randomised sequences (see section 6.2.2b).

a) The entire amino acid sequences of the proteins were analysed producing the scores in this table. Individual alignments are not shown. However an alignment between the yeast and *E.coli* IDH sequences and the two IMDH sequences is shown in figure 6.6.

b) Analysis of the IDH and IMDH sequences to each other by both "DOTPLOTS" and "BESTFIT" had suggested that there was a conserved region of amino acid sequence at the C-terminal end of each protein. Analysis of this region by "BESTFIT" produced the alignments shown in figure 6.4 (see section 6.2.2c). The scores produced by these alignments are shown above.

Q= Quality score, M= Mean score produced from 10 randomised sequences, I= Percentage of residues that are identical in the alignment produced by "BESTFIT". S= Percentage of amino acids that are similar in the alignment produced (using the amino acid comparison table provided with the GCG package). *S.co =S.coelicolor, S.ce =S.cerevisiae E.co = E.coli.*

```
 49 IALFPEYLTEDQRIPDALAELGELAKTPAANIIKLPNISASIPQLKAAVA 98
    :.:::.  | .. |    ||.:::.:                    .:|.:|.
  1 MSMLSRRLFSTSR....LAAFSKI.................KVKQPVV 27
             .          .          .          .          .
 99 ELQGQGYALPAYPDDPKTDEERDIRARYDKVKGSAVNPVLREGNSDRRAP 148
    ||:|:: .                | :||:| . : | |          .
 28 ELDGDEMT.............RIIWDKIKKKLILPYL.........D 52
             .          .          .          .          .
149 ASVKNYAKTHPHRMGAWTGESKTNVATMGENDFRSTEKSAVIAEDGTLRI 198
    ..:| |. . . | :  |::. |. |. : ..:  . |:|.|..|:.
 53 VDLKYYDLSVESRDA..TSDKITQDAAEAIKKYGVGIKCATITPDEARVK 100
             .          .          .          .          .
199 ELVGDDGTTTVLRESVPVKKDEVVDASVLRVAALREFLTAQVARAKAEGI 248
    |:  ..     ::.|. ..  :::::.|:| : : . :.  |:| ....|
101 EFNLHK.....MWKSPNGTIRNILGGTVFREPIVIPRIPRLVPR.WEKPI 144
             .          .          .          .          .
249 LYSVHLKATMMKVSDPIIFGHVVRAFFPKTFAQYGDVLAKAGLTPNDGLG 298
    :.: | .:.  |..|.:| |.                        ::|:
145 IIGRHAHGDQYKATDTLIPGP.....................GSLE 169
             .          .          .          .          .
299 GIYKGLESLPEGAEIKASFDAELAEGPELAMVDSDKGITNL.HVPSDVII 347
     :||. :. ...::.   :| . :.|..:|| :.|.:|..: | . .: |
170 LVYKPSDPTTAQPQTLKVYDYK.GSGVAMAMYNTDESIEGFAHSSFKLAI 218
             .          .          .          .          .
348 DASMPAMIRTSGHMWGADGGEHDALAVIPDSSYAGVYQAVIEDCRANRPL 397
    | .:  ::.|.. ::   :|           .: :::|.|.|
219 DKKLNLFLSTKNTILKKYDG..........RFKDIFQEVYE........ 249
             .          .          .          .          .
398 HDPSTMGSVPNVGLMAQKAEEYGSHDKTFEIPTTGTVRLVARPRATRVLE 447
                  ||  ..::  :  :|                |:::
250 ..............AQYKSKFEQLGIHYE..............HRLID 269
             .          .          .          .          .
448 QTVSAGDIFRACQTKDAPIRDWVKLAVTRARATGDPAVFWLDEGRAHDAN 497
    : |  :::::.                                    .:.
270 DMV..AQMIKS.................................KGG 281
             .          .          .          .          .
498 LIAKVKQYLSEHDTEGLDIRILDPVEATKLSVERIRRGENTISVTGNVLR 547
    :|  :|.|                                   .|:| .
282 FIMALKNY.................................DGDVQS 295
             .          .          .          .          .
548 DYLTDLFPILELGTSAKMLSVVPLMAGGGLFETGAGGSAPKHVQQLVKED 597
    |.:.: |. |:|     | |::.  .| .: ..:| |...:| ... |::
296 DIVAQGFGSLGL.....MTSILVTPDGKTFESEAAHGTVTRHYRKYQKGE 340
             .          .          .          .          .
598 YLRWDSLGEFFALVPSLEKYAEATGNGKAKVLADTLDRATATFLNEDKSP 647
      . :|::.:||: .:| | :| .... :|:.|:.|| . :.:|
341 ETSTNSIASIFAWSRGLLKRGELDNTPALCKFANILESATLNTVQQDGIM 390
             .          .          .          .          .
648 TRRVGGIDNRGSHFYLSLSGRRELAKQTDDADLAKAFAPLAETLTASEQK 697
    |                   ::|| ..::.: . |:.. .| |.| |.:
391 T...................KDLALACGNNERS.AYVTTEEFLDAVEKR 419

698 IVEELNAVQ 706
    : .|:..::
420 LQKEIKSIE 428
```

Figure 6.4 Alignment between the IDHs from *S. coelicolor* and *S.cerevisiae* mitochondia using "BESTFIT".

S.cerevisiae sequence is shown on the bottom line and the *S. coelicolor* sequence on the top line. Numbers refer to the first and last amino acid at in each line. Identical matches are marked with a line (|). Matches with a score of greater than 0.5 are marked with a colon (:) and those that match with a score of greater than or equal to 0.1 are marked with a full stop (.), using the scoring table provided with the GCG package.

```
281 IADAFLQQILLRPAEYDVIACMNLNGDYISDALAAQVGGIGIAPGANIGD 330   Eco x Tth
    ..||:  ::: .|| :||:.. |: ||.:::  |. .|::|: |:|.:|
213 YVDAMAMHLVRSPARFDVVVTGNIFGDILGNLRADLPGSLGLLPSASLGR 262
          .         .         .         .         .
331 ECALFEATHGTAPKYAGQDKVNPGSIILSAEMMLRHMGWTEAADLIVKGM 380
    :.::||:.||.||.|||.:: ||.. ||||.||| ::     :: :.: :
263 GTPVFEPVHGSAPDYAGKGR.NPTAAILSAAMMLEQL....RPGGLARKV 307
          .         .         .
381 EGAINAKTVTYDFERLMDGAKLLKCSEFGDAIIENM 416
    |:|  ||.:  . .. ::|.   ....|...:: ::
308 EDA..AKALLETPPPDLGGS...EARAFTATVLRHL 338

511 TEGLDIRILDPVEATKLSVERIRRGEN.TISVTGNVLRDYLTDLFPILEL 559   Sco x Tth
    . |..   |:.  .. :.:. :|.... .: ||||:: |.|.:| :  :|
201 GRGYPDVALEHQYVDAMAMHLVRSPARFDVVVTGNIFGDILGNLRA..DL 248
          .         .         .         .         .
560 GTSAKMLSVVPLMAGGGLFETGAGGSAPKHVQQLVKEDYLRWDSLGEFFA 609
    ..|  :|. |..|  |..:||..  ||||.....        ::. :.::.
249 PGSLGLLPSASLGRGTPVFEPVH.GSAPDYAGK.......GRNPTAAILS 290
          .         .         .         .         .
610 LVPSLEKYAEATGNGKAKVLADTLDRATATFLNEDKSPTRRVGGIDNRGS 659
     . ||..       :: .|| .:: |. .:|:   .|.. :|| :.|:
291 AAMMLEQL.......RPGGLARKVEDAAKALLE...TPPPDLGGSEARA. 329
          .
660 HFYLSLSGRRELA 672
       :. .. |.||
330 ...FTATVLRHLA 339

501 KVKQYLSEHDTEGLDIRILDPVEATKLSVERI.RRGENTISVTGNVLRDY 549   Sco x Bsu
    ::.. :.|. .::::.   |:.: ... .:: |  ..: .: ||:|:: |.
200 RLWREVAEDVAQEFPDVKLEHMLVDNAAMQLIYAPNQFDVVVTENMFGDI 249
          .         .         .         .         .
550 LTDLFPILELGTSAKMLSVVPLMAGGGLFETGAGGSAPKHVQQLVKEDYL 599
    |.|  .:  |..|  ||. |..| .:|  : ... |||| ...  : :.:
250 LSDEASM..LTGSLGMLPSASLSSSGLHLFEPVHGSAPDIAGKGMANPFA 297
          .         .         .         .         .
600 RWDSLGEFFALVPSLEKYAEATGNGKAKVLADTLDRATATFLNEDKSPTR 649
       | : ::    :||. |.|.::: .||||.
298 AILSAAMLLRTSFGLEEEAKAVEDAVNKVLAS..................  329
          .         .         .         .         .
650 RVGGIDNRGSHFYLSLSGRRELAKQTDDADLAKAFAPLAETLTASEQKIV 699
                        :|.| | . ...|        .|.| |.
330 ......................GKRTRDLARSEEF........SSTQAIT 349
          .
700 EELNAVQGKPAEIGG 714
    ||:.|.  ....|:.
350 EEVKAAIMSENTISN 364
```

```
500 AKVKQYLSEHDTEGLDIRILDPVEATKLSVERIRRGENTISVTGNVLRDY 549   Sco x Sce
    ..| | :...:.  :.:: .|                      |:| .|.
270 DMVAQMIKSKGGFIMALKNYD.....................GDVQSDI 297
              .           .           .           .
550 LTDLFPILELGTSAKMLSVVPLMAGGGLFETGAG.GSAPKHVQQLVKEDY 598
    :.: |. |:| ||       : : ::|  ||.:|: |...:| .:. |::
298 VAQGFGSLGLMTS......ILVTPDGKTFESEAAHGTVTRHYRKYQKGEE 341
              .           .           .           .
599 LRWDSLGEFFALVPSLEKYAEATGNGKAKVLADTLDRATATFLNEDKSPT 648
     . :|::.:||: .:| | :| ....    :|:.|:.|| . :.:|   |
342 TSTNSIASIFAWSRGLLKRGELDNTPALCKFANILESATLNTVQQDGIMT 391
              .           .           .
649 RRVG.GIDNRGSHFYLSLSGRRELAKQTDDADLAKAFAPLA 688
    : :: :.:|.:.  |:.     |   :. :  |.|.: .:.
392 KDLALACGNNERSAYVT....TEEFLDAVEKRLQKEIKSIE 428

16 IRILDPVEATKLSVERIRRGENTISVTGNVLRDYLTDLFPILELGTSAKM 565   Sco x Eco
    | | |.:... |    :|.:| .: .. |:  ||:.| ::    |.:  :
275 IVIKDVIADAFLQQILLRPAEYDVIACMNLNGDYISDALAAQVGGIG..I 322
              .           .           .           .
566 LSVVPLMAGGGLFETGAGGSAPKHVQQLVKEDYLRWDSLGEFFALVPSLE 615
     ..: : .:.:||| :. |.|||...|                    :
323 APGANIGDECALFE.ATHGTAPKYAGQ.....................D 349
              .           .           .           .
616 KYAEATGNGKAKVLADTLDRATATFLNEDKSPTRRVGGIDNRGSHFYLSL 665
    | ..:.   .|.::   :::..|.    |  .. . |:|:.:.  :
350 KVNPGSIILSAEMMLRHMGWTEAA....DLIVKGMEGAINAKTVTY.... 391
              .           .
666 SGRRELAKQTDDADLAKAFAPLAETLTASEQKIVEEL 702
        ::.:  |:|.| |.        .. :: |:|::
392 ....DFERLMDGAKLLKC........SEFGDAIIENM 416
```

Figure 6.5 **Comparison of the C-terminal region of the IDH from *S.coelicolor* to other IDHs and IMDHs.**

The programme "BESTFIT" was used to line up the C-terminal region (amino acids 500-714) of the *S.coeicolor* IDH sequence to some other protein sequences. This region (shown boxed) is suspected to contain an NADP binding region over the putative NADP binding sites. Identical matches between amino acids in the alignments are marked with a line (|). Matches with a score of greater than 0.5 are marked with a colon (:) and those that match with a score of greater than or equal to 0.1 are marked with a full stop (.).

6.2.3 Sequence alignment of the IDHs / IMDHs.

Bestfit data between the yeast IDH, *E. coli* IDH and the IMDHs were very similar and produced alignments across the whole of each sequence (individual alignments not shown). This suggests that the *S.cerevisiae* enzyme is similar to each of the proteins across the whole of its length and not just over the putative $NADP^+$ binding sites. Attempts were therefore made to first line up all of these sequences to each other using the GCG program, "PILEUP".

As expected an alignment was produced that agrees well with those produced by the individual "BESTFIT" comparisons. Figure 6.6 shows these alignments. However one of the gaps introduced by the program was moved to a different position. The gap was moved so that it would lie over the region that encodes a clasp-like domain in the *E. coli* IDH. This clasp-like domain consists of two β-sheets that are known to be absent in the IMDH from *T.thermophilus* (Miyazaki *et al.*, 1992) and the IDH from the same organism (Imada *et al.*, 1991) and from the NAD-IDH from yeast (Cupp and McAlister-Henn, 1991) (see section 1.2.5). Residues identical in the yeast enzyme and at least two other sequences were highlighted along with the residues from *E. coli* that are known to contribute to the binding of substrate and co-factor or are involved in catalysis. There are also many conservative matches throughout the sequence, confirming the report by Hurley *et al.* (1991) that the *S.cerevisiae* IDH has sequence similarity to the *E. coli* enzyme. This would seem to suggest that this eukaryotic NADP-IDH can also be classified as a type-I IDH.

Such identities are even more apparent when one examines the residues that lineup with the *E. coli* sequence at positions required for $NADP^+$ or isocitrate binding. Figure 6.7c, shows the almost complete conservation of the residues required for isocitrate/isopropylmalate binding, which was first noticed by Hurley *et al.* (1991) between the IMDHs and the IDHs. Figure 6.7.a, shows that there is also a large amount of sequence conservation over those residues required for $NAD(P)^+$ binding which had not previously been noticed.

Attempts to include the *S. coelicolor* sequence in this alignment were not very successful as the "PILEUP" tried to line up the whole length of the *S. coelicolor* protein. Much of the protein is certain to be dissimilar to the rest of the proteins due to the extra sequence and could produce some spurious alignments. The various alignments produced by the bestfit analysis over the putative $NAD^+(P)$ binding site (figure 6.4) were therefore used to compile figure 6.7b). It can be seen that only a few residues show conservation when compared to those shown in figure 6.7a).

Complete conservation is seen though over the region 584-589, which coincides with the main cluster of $NAD^+(P)$ binding residues in the other sequences. An attempt was therefore made to derive a consensus sequence that could describe this conservation. No attempt was made to make a similar consensus for the isocitrate binding residues, since they are too distant from each other and at variable distances, making it difficult to identify residues that might be significant.

6.2.4 $NAD(P)^+$ consensus sequence.

The obvious conservation (shown in figure 6.7a and b) between the residues required for $NAD(P)^+$ binding, indicates that a consensus could be derived that is present in most of the sequences used in this analysis (consensus no. 1 below).

1) S-(26-27 X)-L-G-(17-19 X)-H-G-[ST]-A-(2 X)-[YH]-(5-12 X)-N-(37-47 X)-D-(2-5 X)-[RG]

x= any amino acid, *eg.* (26-27x) =26 to 27 amino acids
[Residues in square brackets indicate variability]

This consensus can be used to screen by eye putative $NAD^+(P)$ binding sites in other proteins. Its use for analysis by computer (using a program such as "BEST" or "FASTA") would be more difficult because of the variability in size of the gaps between the residues. Such variability is likely to be even more extreme in less conserved proteins. The consensus can be shortened to include only the core region in the middle of this consensus. Further modification to include other conserved residues in the proximity of this core region was also included to give the consensus below (consensus no. 2 below).

2) E - [AP] - (1-2X) - H - G - [ST] - A - P - X - [YH] -
3) E - [AP] - x - H - G - [ST] - A - P - X- X - [AG]

Note that the consensus sequence provided by the MacPattern program is incorrect (see section 6.2.1b). The corrected sequence is shown here in sequence 3.

It can be seen that this consensus is essentially the same as the sequence provided in the Apple-Macintosh computer program, "MACPATTERN" (see section 6.2.1b and consensus no. 3 above). Modifications to include the *S.cerevisiae* and *S. coelicolor* sequences have led to a more degenerate sequence which allows greater flexibility if a scoring system is used. For instance the *T.thermophilus* IMDH and the *E. coli* IDH sequence conform to the consensus perfectly, whilst the *B. subtilis* IMDH has one mismatch. *S.cerevisiae* and *S. coelicolor* IDHs conform less well, scoring six out of

```
     1                                                  *           50
Bsu  .......... .......... ....MKKRIA LLPGDGIGPE VLESATDVLK
Tth  .......... .......... ......MKVA VLPGDGIGPE VTEAALKVLR
Eco  MESKVVVPAQ GKKITLQNGK LNVPENPIIP YIEGDGIGVD VTPAMLKVVD
Sce  .......... .....MSMLS RRLFSTSRLA AFSKIKVKQP VVELDGDEMT

     51                                                    97
Bsu  .SVAERFNHE FEFEYGLI.. .....GGAAI DEHHNPLPEE TVAACKNADA
Tth  .ALDEA.EGG PTYEVFPF.. .....GGAAI DA.SAPFPEP TRKGVEEAEA
Eco  AAVEKAYKGE RKISWMEI.. .YTGEKSTQV YGQDVWLPAE TLDLIREYRV
Sce  RIIWDKIKKK LILPYLDVDL KYYDLSVESR DATSDKITQD AAEAIKKYGV

      98             +            +           +          139
Bsu  ILLGAVGGPK WDQNLSELRP EKGLLSIRKQ LDLFANLRPV KVFESLSDRS
Tth  VLLGSVGGPR LGRPSPQDPP GDGASFLKEK PDLFANLRPA KVFPGLERLS
Eco  AIKGPLTTPV GGGIRSLNVA ......LRQE LDLYICLRPV RYYQGTP..S
Sce  GIKCATITPD EARVKEFNLH K....MWKSP NGTIRNILGG TVFREPIVIP

     140              +       +                          182
Bsu  PLKKEYID.N VDFVIVRELT GGLYFGQPSKRYVNT
Tth  PLKEEIAR.G VDVLIVRELT GGIYFGEP.... RG
Eco  PVKH..PE.L TDMVIFRENS EDIYAGIEWK ADSADA.... .EKVIKFLRE
Sce  RIPRLVPRWE KPIIIGRHAH GDQYKATDTL IPGPGSLELV YKPSDPTTAQ

                                                           232
     182                                                    +
Bsu           EG EQEAVDTLFY KRTEIERVIR EGFKMA.ATR KGKVTSVDKA
Tth           MS EAEAWNTERY SKPEVERVAR VAFEAA.RKR RKHVVSVDKA
Eco  EMGVKKIRFP EHCGIGIKPC SEEGTKRLVR AAIEYAIAND RDSVTLVHKG
Sce  PQTLKVYDYK GSGVAMAMYN TDESIEGFAH SSFKLAI.DK KLNLFLSTKN

     232                                                   281
Bsu  NVLESSRLWR EVAEDVAQEFP... .......  ........ ..DVKLEHML
Tth  NVLEVGEFRR KTVEEVGRGYP... .......  ........ ..DVALEHQY
Eco  NIMKFTEGAF KDWGYQLARE EFGGELIDGG PWLKVKNPNT GKEIVIKDVI
Sce  TILKKYDGRF KDIFQEVYEA QYKSKFEQ.. .......... .LGIHYEHRL

  282 !          *               !!   !          **         331
Bsu  VDNAAMQLIY APNQFDVVVT ENMFGDILSD EASMLTGSLG MLPSASLSSS
Tth  VDAMAMHLVR SPARFDVVVT GNIFGDILGN LRADLPGSLG LLPSASL.GR
Eco  ADAFLQQILL RPAEYDVIAC MNLNGDYISD ALAAQVGGIG IAPGANI.GD
Sce  IDDMVAQMIK SKGGF.IMAL KNYDGDVQSD IVAQGFGSLG LMTSILVTPD

     332       * ***   *             **                    371
Bsu  GLHLFEP..V HGSAPDI.... ..AGKGMANP FAAIL..SAA MLLRTSFGLE
Tth  GTPVFEP..V HGSAPDY.... ..AGKG.RNP TAAIL..SAA MML.EQLRPG
Eco  ECALFEA..T HGTAPKY.... ..AGQDKVNP GSIIL..SAE MMLR.HMGWT
Sce  GKT.FESEAA HGTVTRHYRKY QKGEETSTNS IASIFAWSRG LLKRGELDNT

     372                          **       *                413
Bsu  EE....AKAV EDA.VNKVLA SGKRTRDLA. ..RSEEFS.. .....STQAI
Tth  GL....ARKV EDA.AKALL. .ETPPPDLG. ..GSEARA.. .....FTATV
Eco  EA....ADLI VKG.MEGAIN AKTVTYDFE. ..RLMDGAKL LKCSEFGDAI
Sce  PALCKFANIL ESATLNTVQQ DGIMTKDLAL ACGNNERSAY VTTEEFLDAV
```

Figure 6.6 Alignment of the yeast IDH sequence to the *E. coli* IDH and IMDH sequences.

The programme "PILEUP" from the GCG package was used to produce an alignment of the four sequences shown.

Sce = IDH from*S.cerevisiae* mitochondria, **Eco** = IDH from *E. coli* .
Tth = IMDH from *T. thermophilus* , **Bsu** = IMDH from *B. subtilus* .
(+) = amino acids contributing to binding of isocitrate in *E. coli* sequence. **(!)** Amino acids involved in coordinating Mg^{2+} **(*)** = amino acids contributing to binding of NADP in *E. coli* sequence.Numbering labels the residues of the *E. coli* IDH sequence (Hurley *et al.*, 1991).

a)

E.coli IDH	*B.subt* IMDH	*T.therm.* IMDH	*S.cerev* IDH	*T.therm.* IDH	*S.cerev* NAD-IDH
R-292	A-233	S-224	S-278	R-230	S-233
I-320	L-261	L-252	L-305	L-258	L-262
G-321	G-262	G-253	G-306	G-259	G-263
H-339	H-281	H-271	H-326	H-276	H-281
G-340	G-282	G-272	G-327	G-277	G-282
T-341	S-283	S-273	T-328	S-278	S-283
A-342	A-284	A-274	V-329	A-279	A-284
Y-345	I-287	Y-277	H-332	H-282	I-287
V-351	A-293	R-282	T-344	I-288	A-293
N-352	N-294	N-283	N-345	N-289	N-294
Y-391	R-334	P-320	K-392	T-329	R-333
D-392	D-335	D-321	D-393	D-330	-
R-385	R-338	G-324	G-399	V-333	-

b)

S.coeli *E.coli*	*S.coel-* *B.sub*	*S.coel-* *T.therm.*	*S.coel-* *S.cerev*
R-533	R-533	R-533	E-508
T-561	A-563	A-563	L-557
S-562	K-564	K-564	E-558
G-583	G-583	G-583	G-583
G-584	G-584	G-584	G-584
S-585	S-585	S-585	S-585
A-586	A-586	A-586	A-586
H-589	H-589	H-589	H-589
Y-617	K-595	W-601	W-601
A-618	E-596	D-602	D-602
F-661	D-676	R-649	R-649
E-670	D-677	R-650	R-650
K-673	L-680	G-653	D-655

C)

E.coli IDH	*B.subt* IMDH	*T.therm* IMDH	*S.cerev* IDH	*T.therm* IDH	*S.cerev* NAD-IDH
N-115	L-086	D-082	N-104	N-100	N-100
S-113	S-084	P-080	E-101	S-098	S-98
R-119	R-096	K-092	K-109	R-103	R-104
R-129	R-106	R-102	L-119	R-113	R-114
R-153	R-134	R-130	R-148	R-135	R-135
Y-160	Y-141	Y-137	Y-155	Y-143	Y-142
K-230	K-192	K-183	K-229	K-190	K-189
D-283	D-214	D-215	D-269	D-222	-
D-307	D-238	D-239	D-292	D-246	D-248
D-311	D-242	D-243	D-296	D-250	D-252

Key.

T.therm. = *T.thermophilus*
B.subt = *B. subtilis*
S.cerev = *S.cerevisiae*

All IDHs shown are NADP dependent except for *S.cerev.* NAD-IDH which is NAD dependent.

Figure 6.7 Summary of residues suspected to be involved in the binding of NADP and isocitrate.

a) Residues known to be important in the binding of NADP in *E. coli* IDH are shown in the first column . Those residues that line up with the *E. coli* sequence in figure 6.6 are shown for the IMDHs and *S.cerevisiae* IDH. Numbers represent the position of amino acids in each protein.

Sequences for *S.cerevisiae* NAD-IDH and the *T.thermophilus* NADP-IDH have recently been published and have been aligned with the *E. coli* IDH sequence. Using the alignments published by the respective authors (see section 6.3.2) residues lining up with the NADP binding residues were included in this table.

b) Comparison of the *S. coelicolor* IDH sequence to the other sequences produced different alignments (see figure 6.5). For each different alignment, the residues lining up with those proposed to be involved in NADP binding (inferred from figure.6.7.a) are shown.

c) Residues known to be important in the binding of isocitrate in *E. coli* are shown. Those residues that line up with the *E. coli* sequence are shown for the IMDHs and *S.cerevisiae* IDH. Numbers represent the position of amino acids in each protein and are derived from the alignment in figure 6.6.

Sequences for *S.cerevisiae* NAD-IDH and the *T.thermophilus* NADP-IDH have recently been published (see section 6.3.1). Those residues required for isocitrate binding in the *E. coli* sequence have been reported to also be conserved in these enzymes.

eight possible matches. Although consensus 2 is only a minor modification of the sequence in "MACPATTERN", it was obtained independently during this work. It utilised the data of Hurley *et al.* (1991) which has highlighted its importance in $NAD(P)^+$ binding. The consensus provided as part of "MACPATTERN" was assumed to have a some sort of catalytic role but was not considered for use as a consensus for nucleotide binding.

Such a consensus has indicated that all of these proteins do have an $NAD(P)^+$ binding site that can be identified at the primary sequence level. The conservation is not totally rigid however, especially outside the main cluster of binding residues. Recent publication of the sequence for the NAD dependent IDH from *S.cerevisiae* and for the NADP dependent IDH from *T. thermophilus* has shown that these sequences also possess a putative $NAD(P)^+$ binding consensus sequence.

6.2.5. Comparison of the *S. coelicolor* enzyme to itself.

As outlined in section 6.1. a) and b) , it is possible that the *S. coelicolor* enzyme could have evolved from the fusion of two identical genes to form a monomeric protein that has two identical halves. It was also discussed (section 6.1b) that the *S. coelicolor* enzyme may possess two NADP binding sites, even if the enzyme is not produced by the fusion of two genes.

In the first situation one might expect to find two regions of extensive similarity to other IDHs/IMDHS which would be revealed in the dotplots described in section 6.2.2a. In the second situation one may expect to find two small regions of similarity to the other proteins, representing two NADP binding sites (presuming that both NADP binding sites are similar). Neither situation was found indicating that the *S. coelicolor* IDH has not evolved from the fusion of two identical genes and that it only possesses one IDH/IMDH-like $NAD(P)^+$ binding site.

6.2.6 Database searching.

To test whether the *S. coelicolor* IDH was similar to any other proteins that have been sequenced several methods were used to screen various databases. The algorithm of Pearson and Lipman, "FASTA", was used as well as the more sensitive DAP algorithm of to screen the appropriate peptide sequence databases (*e.g.* PIR and SWISSPROT). TFASTA was also used to screen the more extensive DNA sequence databases. This algorithm is basically the same as the FASTA method but the DNA sequences in the database are first translated in all six reading frames before making the comparison. None of the searches identified any proteins with a significant sequence alignment.

6.3 Conclusions.

6.3.1 Relationships between the *S.cerevisiae* and *E. coli* IDHs and the IMDHs.

As discussed in section 1.2.5 the IMDHs and the *E. coli* IDH are related in both function and amino acid sequence. Comparison of the *S.cerevisiae* IDH to both classes produced a sequence alignment that resembles the alignment of the *E. coli* enzyme to the IMDHs.

NAD dependent IDHs from *E. coli,* pig heart and yeast, all have similar subunit compositions and molecular weights. Haselbeck and McAlister-Henn (1991) have also reported the similarity between the N-termini of the pig heart and yeast enzyme. Cross reaction of antibodies between the porcine and *E. coli* IDHs also indicates that they are related. One could infer from this that the *E. coli* and yeast enzymes must therefore be related in some way, a suggestion also made by Hurley *et al.* (1991).

From these data it would therefore appear that the yeast mitochondrial enzyme is in fact a member of the class I, 80 kDa, dimer type of IDH, of which the only previous examples have been in bacteria. Mitochondria are believed to have evolved from prokaryotes, probably providing an explanation for this, although it could be that the classification of $NADP^+$ dependent IDHs can be broadened to include those enzymes from eukaryotes.

Since the work described in thesis was completed other IDH sequences have become available. It was reported that the sequence for the NADP-IDH from *T.thermophilus* and the NAD-IDH2 subunit from yeast, show conservation with the residues required to bind isocitrate in *E. coli* (Miyazaki *et al.*, 1992 and Cupp and McAlister-Henn, 1992 respectively). The degree of conservation (33% identical) found for the yeast NAD-IDH2 was particularly surprising since this enzyme has a different quarternary structure and different coenzyme specificity to the type I IDH. These sequences have been included in figure 6.7c to illustrate the conservation over the substrate binding site for all of the IDH/IMDHs.

6.3.2 $NAD(P)^+$ binding sites.

A sequence relationship between all of the proteins studied was first noticed when a region of similarity was identified that was common to all of them. Close analysis of this region showed that significant alignments over this region could be made for almost every pair of proteins. This region contains within it 13 out of the 14 residues required for $NADP^+$ binding, although the main region of sequence conservation surrounds a cluster of 6 amino acids within a 7 amino acid region (residues 339-345 for the *E. coli* enzyme).

Figure 6.7 shows that there is some variability in the amino acids at each position, although the degree of conservation is very noticeable. For most positions there is a conservative or identical amino acid between at least two of the proteins and often there are also conserved residues in the proximity of these regions (figure 6.6). A consensus sequence was therefore made to represent each sequence as fully as possible.

Both the NAD-IDH2 *S.cerevisiae* (Cupp and McAlister-Henn, 1991) and the NADP-IDH from *Thermus thermophilus* (Miyazaki *et al.*, 1992) have sequences that conform to this consensus sequence, corroborating its validity as a consenus for this type of enzyme.

6.3.3 Relationship between the *S. coelicolor* IDH and the other proteins.

Other than the sequence similarity over the putative $NADP^+$ binding site, no other obvious regions of possible conservation could be found. Exactly why the *S. coelicolor* IDH (and type-II enzymes in general) is twice as big as the type-I IDHs (as discussed in section 6.1) remains a question to be answered.

Alignment of the whole of the *S. coelicolor* sequence to that of *S.cerevisiae* required the insertion of a great many gaps making it probable that at least some of the alignment was insignificant. However if it were true it would suggest that the *S. coelicolor* protein is very similar to the yeast IDH but with several extra lengths of sequence distributed throughout the sequence. This leads to the conclusion that type I IDHs (of which the *S.cerevisiae* IDH could be included as a member) and type II IDHs (of which the *S. coelicolor* IDH is as a member) are related. However other sequences will need to be found that produce similar alignments before the significance of the alignment in figure 6.4 can be decided.

Dotplot analysis ruled out the possibility of the *S. coelicolor* enzyme having evolved from the duplication and fusion of a single gene. There was little sequence similarity between the N-terminal and C-terminal halves of the protein.

In summary , it can be concluded that the *S.cerevisiae* IDH is more closely related to the IMDHs and the *E. coli* IDH than was previously suspected. From alignments produced between these enzymes a consensus could be found that is quite different to the classical -GXGXX(A/G)- consensus that normally indicates a nucleotide binding fold.

IDH from *S. coelicolor* shows no extensive sequence similarity to any of these proteins but does appear to be partially conserved over the putative $NADP^+$ binding site Most similarity is seen to the yeast mitochondrial IDH and alignments with this protein may indicate an evolutionary relationship.

CHAPTER 7

General discussion.

7.1 Introduction.

The TCA cycle plays an important role in the production of many precursors in many organisms. In streptomycetes some of these precursors are used for the biosynthesis of antibiotics making this cycle of commercial interest. In *E.coli* grown on acetate as a sole carbon source, one of the TCA cycle enzymes, IDH, has been shown to play a key role in controlling the flux of carbon between the glyoxylate bypass and the TCA cycle. It is assumed that IDHs may play a similar role in streptomycetes grown on acetate or oils, making this enzyme an obvious target for study in these organisms. Control of flux, possibly mediated by IDH, may also have an effect on secondary metabolism, as both the glyoxylate bypass and the TCA cycle are likely to compete with polyketide biosynthetic pathways for available acetyl-CoA.

Cloning the gene that encodes IDH would provide a means of altering the levels of IDH (*eg.* by over or under expression of the cloned gene) allowing the effects on cell growth and antibiotic biosynthesis to be observed. The aim of this project was therefore to clone the gene by "reverse genetics", a method which first requires purification of the enzyme. During the development of the purification procedure some of the physical and biochemical properties of the enzyme were studied.

When the gene had been cloned the nucleotide sequence was determined and analysed. Comparison of the predicted amino acid sequence of the enzyme to other proteins was also performed and the conclusions that were reached are discussed in this chapter.

7.2 Classification of the *S.coelicolor* IDH as a "Type II" IDH.

It was suggested by Fukanaga *et al.* (1988) that bacterial IDHs can be classified into two groups. One group consists of dimeric enzymes with subunit molecular weights of *c.*45 kDa and, where data are available, have similar N-terminal amino acid sequences. Antibodies raised against one of the members of this group (IDH-I from *Vibrio spp.* ABE-1) have been found to cross react with two other members from the same group (namely IDH from *E.coli* and *S.typhimurium*), but do not cross react with members from the second group (see below).

The second group of enzymes are monomeric with molecular weights of around 80,000 and, until recently, had only two members; *Azotobacter vinelandii* IDH and IDH-II from *Vibrio spp.* Studies of the IDH from *S.coelicolor* showed that this enzyme could also be classified as a member this group. Not only was the *S.coelicolor* IDH a

monomeric protein with a molecular weight of around 80 kDa but it also had an N-terminal sequence 69% identical to the IDH II from *Vibrio spp.* Like the IDH-II from *Vibrio spp.* it also showed no cross reaction with an anti-*E.coli* IDH antiserum. It was therefore suggested that this group of monomeric enzymes should be called the "type II" IDHs (after the IDH-II from *Vibrio spp.*) and that the group typified by the *Vibrio spp.* IDH-I be called the "type I" IDHs.

As this thesis was being prepared the *R.vannielii* IDH was reported also to have an N-terminal amino acid sequence and subunit size similar to the *Vibrio spp.* IDH-II. The similarity to the N-terminal amino acid sequence of the enzyme from *S.coelicolor* is also striking (63% identical) even though the *R.vannielii* enzyme has been found to possess dual coenzyme specificity. This gives credence to the classification of all monomeric enzymes into a group distinct from the type I IDHs. It is possible that the IDH from *S.coelicolor* (and maybe type II IDHs in general) also has dual coenzyme specificity, but experiments were not performed to look for the low levels of NAD dependent IDH activity found for the IDH from *R.vannielii* .

7.3 Comparison of IDH sequences from various sources.

Genes for several IDHs (*E.coli* NADP-IDH, *T.thermophilus* NADP-IDH, *S.cerevisiae* mitochondrial NADP-IDH and *S.cerevisiae* mitochondrial NAD-IDH2) have now been cloned and sequenced which has allowed the comparison of the enzymes at the level of amino acid sequence . It was perhaps of no surprise to learn that the *E.coli* and *T.thermophilus* IDHs were very similar as they can both be classified as type I IDHs. More surprising was the fact that the amino acid sequence for the IDH2 subunit (from the yeast mitochondrial NAD dependent IDH) was also very similar to the *E.coli* sequence (33% identical; Cupp and McAlister-Henn, 1991). Not only do the enzymes have a different cofactor specificity, but it has also been suggested that the yeast IDH is an octameric enzyme made from two non-identical subunits (see section 1.2.4.4 and Keys and McAlister-Henn, 1990). It is also controlled allosterically unlike the *E.coli* enzyme. Despite this, the similarity in amino acid sequence is obvious, with most of the residues required for isocitrate binding being identical between both the *E.coli* NADP-IDH and yeast NAD-IDH2 (Cupp and McAlister-Henn, 1991). These identities are summarised figure 6.7.

Cupp and McAlister-Henn (1991) also noticed similarity in the amino acid sequences between peptides from each of the three NAD-IDH subunits from pig and the NAD-IDH2 subunit from yeast, as well as with the *E.coli* NADP-IDH sequence. This would seem to indicate a common evolutionary origin for each of the subunits from the different

enzymes, even though the pig heart NAD-IDH has a quaternary structure which is different from either of the other enzymes (see section 1.2.4.4).

This similarity in sequence was also surprising as it was expected that the reaction mechanisms of the NAD and NADP dependent IDHs would be significantly different. This was suggested by the fact that no reports have been made of an NAD-dependent IDH being inhibited by glyoxylate plus OAA, although this phenomenon is common feature of all NADP dependent IDHs (Nimmo, 1986). The inhibition by glyoxylate plus OAA may therefore somehow interfere with the interactions between the 5'-carboxyl of NADP (that is not present in NAD) and the enzyme.

As each of the pig NAD-IDH subunits have amino acid sequences which are similar to each other, as well as to the NAD-IDH2 subunit from yeast, it would be interesting to find out if the two yeast subunits (IDH1 and IDH2 from NAD-IDH) are also similar to each other. Keys and McAlister-Henn (1990) reported that the two subunits are immunologically distinct and have different N-terminal amino acid sequences. However it is possible that similarities can be found between the IDH1 and IDH2 subunit sequences (even though this is not suggested by immunological experiments) as antibodies raised against one protein do not usually cross react with another unless the proteins are at least 50% identical.

Haselbeck and McAlister-Henn (1991) had reported that there was no significant homology between the NADP-IDHs from yeast and *E.coli* despite the fact that they are both 80 kDa homodimers and that neither is subject to allosteric control. As part of the work described in this thesis the yeast and *E.coli* NADP dependent IDHs were compared to each other, as well as to two NAD dependent IMDH sequences (section 6.2.3 and 6.3.1). The alignments produced suggested that the *E.coli* and yeast NADP-IDHs are in fact similar at the amino acid sequence level. It was therefore suggested from the amino acid sequence information, coupled with the similarity in subunit size, that the yeast mitochondrial NADP-IDH can be classified as a type I IDH, despite its eukaryotic origins. This is perhaps not too surprising as mitochondria are believed to have evolved from prokaryotic origins. The similarity of the pig NADP-IDH amino acid sequence (from peptide sequence) to the NADP-IDH from yeast, may suggest that this IDH is also related.

From sequence comparisons it therefore appears that subunits for NADP dependent IDHs from both eukaryotes and prokaryotes are related to each other and also to each of the subunits from NAD dependent IDHs. The obvious similarity of the IMDHs to the *E.coli* and yeast NADP-IDHs (sec 1.2.5) suggests that even these enzymes are related. It therefore seems that each type of enzyme has evolved using the same type of subunit, but

that as the enzymes diverged each type of subunit acquired specialist functions. The degree of similarity between the NAD^+ and $NADP^+$ dependent enzymes was unexpected

In *E.coli* IDH both subunits are involved in catalysis and both are subject to control by phosphorylation. Other type-I enzymes may have similar mechanisms for control and catalysis, since virtually all of the residues required for isocitrate and Mn^{2+} binding are conserved in the *T.thermophilus* IDH and yeast NADP-IDH (figure 6.7).

Equilibrium dialysis experiments on yeast and pig heart NAD-IDHs have shown that the number of substrate and effector binding sites per active enzyme is less than the number of subunits (Keys and McAlister-Henn, 1990 and Huang and Colman, 1990). Subunits could have become specialised so that some could be involved in catalysis whilst others could be involved in binding effectors and exerting allosteric control. For instance it is conceivable that the ancestral NADP binding site could have evolved into an AMP binding site for some of the subunits. Although the NAD-IDHs from these two organisms have some similar properties, the IDH from yeast has been reported to be a octamer, based on initial studies of other eukaryotic NAD-IDHs such as that from pig heart (Keys and McAlister-Henn, 1990). However the pig heart NAD-IDH has recently been re-axamined and reported to be a tetramer (Huang and Colman, 1990). It is therefore possible that the yeast NAD-IDH is also a tetramer.

Attempts to align the NADP-IDH from *S.coelicolor* to the NAD dependent IMDHs and the NADP dependent IDHs from yeast and *E.coli* are also discussed in this thesis. The similarity of many of the proteins is low indicating that the type II IDHs are quite different to the other enzymes, although a distant relationship could be suggested by the alignments between the *S.cerevisiae* and *S.coelicolor* NADP-IDHs. If this relationship is significant then the *S.coelicolor* IDH (and maybe type II IDHs in general) has evolved with extra sequences distributed throughout the protein. These extra sequences could substitute for regions which would be responsible, in type I IDHs, for subunit-subunit interactions and in forming the correct enzyme configuration. If such a relationship does exist between the yeast and streptomycete IDHs then one could conclude that type II IDHs are distantly related to type I IDHs in general, since it was earlier suggested that the yeast NADP-IDH is probably also related to the *E.coli* IDH.

7.4 Identification of a putative novel nucleotide binding site.

When comparing the amino acid sequences of various IDHs and IMDHs a sequence was found that is very similar to and contained within all of the other proteins compared. This region coincides with the region in the *E.coli* IDH that contains many of the residues required for binding NADP (Hurley *et al.*, 1991). It was therefore suggested that all of the proteins analysed during this work have similar NAD(P) binding sites. Comparisons of the recently published sequences for the NAD-IDH from yeast and the NADP-IDH from *T.thermophilus* have been included in figure 6.7a and would seem to confirm the conservation of NAD(P) binding residues in the IDH and IMDH type enzymes.

Miyazaki *et al.* (1992) noticed that a region of the *T.thermophilus* IDH showed sequence similarity to isopropylmalate isomerase, as well as to several IMDHs. They concluded that this region must be therefore be involved in isopropylmalate or isocitrate binding in the appropriate enzymes, as the two similar substrates seemed to be the only common link between the enzymes. However the region identified by Miyazaki *et al.* (1992) aligns with the region that, in the *E.coli* IDH, is involved in the binding of NADP (Hurley *et al.*, 1991). The alternative explanation for a region of conservation is that the isomerase and IMDHs have a common evolutionary origin. However this possibility also seems unlikely, as there is no other similarity reported between the isomerase and dehydrogenase sequences (Andreadis *et al.*, 1984).

A consensus sequence for this putative NAD(P) binding site was therefore proposed in section 6.2.4, describing the common sequence found in all of the proteins analysed. It is clear that the proposed consensus is different to the any of the other NAD(P) binding sites (discussed in section 1.3). The diagram below illustrates that the proposed IDH/IMDH-NADP binding sites are quite different to the other NAD(P) binding sites. However there are some features of these binding sites that are conserved.

For instance the glycine at position 3 in the alignment above is conserved in all of the sequences, suggesting that this amino acid may have a functional role. Indeed the residue at this position in the Wierenga, LDH-type nucleotide binding site is required to allow the close approach of the 5' pyrophosphate group of NAD(P) to the α-helix, αB. A similar role was also proposed for this residue in FNR type binding folds. Another common feature is also found between the NAD(P) binding sites of the LDH type enzymes, the FNR enzyme and the *E.coli* IDH; all make use of a positively charged dipole, found at the N-terminus of an α-helix, to stabilise interactions between the protein and the negatively charged 5' pyrophosphate of NAD(P).

	1 2 3 4 5 6
Wierenga NAD consensus	G-X-G-X-X-G
Wierenga NADP consensus	G-X-G-X-X-A
Spinach FNR	G-T-G-I-A-P
E.coli IDH	T-H-G-T-A-P
B.subtilis IMDH	V-H-G-S-A-P
T.thermophilus IMDH	V-H-G-S-A-P
S.coelicolor IDH	A-G-G-S-A-P

In conclusion it can be said that there is no consensus sequence that can be used to describe all NAD(P) binding sites. However some sequence similarities and inferred structural similarities have been noticed between nucleotide binding sites from different enzyme families. This is likely to reflect functional constraints on the types of residues required rather than an evolutionary relationship.

7.5 Future work.

Now that the IDH from*S.coelicolor* has been successfully purified and the gene encoding this enzyme cloned and sequenced, there are numerous experiments that can be performed at both the biochemical and genetic levels. It has been shown that the enzyme of *S.coelicolor* is a type II IDH based on quarternary structure, subunit size, N-terminal amino acid sequence and immunological properties. Compared to the type I IDHs, little is known about how type II IDHs function or how they are controlled.

To gain an insight into the role (if any) that IDH plays in controlling metabolism in *S. coelicolor,* it will be neccesary to measure the level of IDH activity at different times of the life cycle and under different growth conditions. There would be a special interest in any differences found in the levels of IDH activity found between mycelia harvested during the growth and antibiotic production phases and any differences found for mycelia grown on media with glucose compared to acetate as the sole carbon source. Any change found in the total cellular IDH activity, could be due to alterations in the activity of the enzyme (by covalent modification or allosteric effectors), or could be due to changes in the quantity of IDH protein.

Changes in the level of protein synthesis could be detected using an antiserum raised against the *S.coelicolor* IDH. Alternatively the antiserum raised against the type II IDHs from *R.vinelandii* (Leyland and Kelly, 1991) or *Vibrio spp.* ABE-1 (Ishii *et al.,* 1987) could probably be used. The cloned *icd* B gene could be used as a specific probe for measuring the levels of IDH mRNA thereby detecting any changes in the rate of

transcription. Any changes in protein synthesis not accounted for by fluctuating transcription levels would probably be accounted for by changes in the rate of translation.

From the data presented in this work (and that of Leyland and Kelly, 1991), it was found that certain metabolites (such as OAA) have an effect on the activity of the IDHs from both *S.coelicolor* and *R.vinelandii*. The relevance of such effects needs to be addressed further by measuring the intracellular concentrations of these metabolites and observing the inhibitory effects of the metabolites on IDH activity at these concentrations.

The putative multiple promoters identified during this work may play a role in controlling the levels of transcription at different stages in the life cycle and/or under different growth conditions. However it is possible that other functional promoters lie elsewhere and that some or all of the sequences identified are non-functional. The functional promoter(s) could be identified by searching for promoter sequences upstream of the transcription start point(s). The mRNA start point(s) could be identified by experiments such as S1 mapping or primer extension mapping. Determination of both ends of the transcript that contains the coding information for IDH could also help to resolve whether *icd* B is part of an operon (*ie.* whether the transcript is long anough to encode other proteins up and/or downstream of *icd* B).

The validity of putative ORFs up and downstream of the *icd* B gene could also be investigated by sequencing further up and downsream of *icd* B. This is of particular interest since it was found that the gene encoding citrate synthase lies immediately upstream of the *icd* gene in *B.caldotenax* (Catherine Wild, personnal communication) perhaps suggesting some sort of concerted regulation.

A 30 fold over-expression has recently been achieved in the heterologous host, *S. lividans* (P.Chapman, Institute of Genetics, Glasgow University). The 5 kb *Pst* I insert from pRTB1 was ligated into the multicopy number streptomycete vector pIJ486 (Hopwood *et al.*, 1985) and introduced into *S.lividans*. The strain grew more slowly than the wild type strain, probably due to the pressure of overproduction rather than an effect caused by the activity of the *S. coelicolor* IDH.
It would now be interesting to observe the effect on cell growth and antibiotic production, of over and under-expressing the *S.coelicolor* IDH in the native host. It is likely that overexpression will have little effect on this host either. However complete loss of all IDH activity (using a disrupted copy of the cloned *icd* B gene in genomic replacement experiments) is likely to make the strain auxotrophic for glutamate or proline, analogous to the situation for *E.coli icd* mutants. It is possible that if such mutants are provided with supplements, then the strain will grow but may have altered levels of antibiotic biosynthesis.

More subtle changes in IDH activity could be made by introducing the cloned gene into an *S.coelicolor icd* B mutant, on a plasmid that allows controllable expression of the gene. Alternatively it may be possible to express the antisense strand of the *icd* B gene in a controllable manner. Antisense mRNA will bind to sense mRNA preventing further translation.

It would also be interesting to observe the effect of expression of the *S.coelicolor* gene in the *E.coli icd* mutant (especially during growth on acetate) as *E.coli* may not be able to control the activity of the heterologous enzyme by the phosphorylation mechanism used to control its own IDH. This would probably require modification of the *icd* B promoter sequence(s) to enable expression of the gene by the *E.coli* RNApol. Expression of the *E.coli* gene in a streptomycete *icd* B mutant (both with and without the *E.coli* IKP) could also prove interesting, perhaps providing further insights into how the enzymes from both organisms are controlled.

The number of NADP and isocitrate binding sites per monomer is unknown for the *S.coelicolor* IDH or indeed any of the type II IDHs. It would be interesting to determine this property, as it is conceivable that the type II enzymes may possess more than one NADP binding site. Equilibrium dialysis experiments could be used to determine the number of binding sites for isocitrate and indeed the number of binding sites for any effectors that may be identified.

Analysis of the amino acid sequence for the IDH from *S.coelicolor* has revealed a region that may be involved in the binding of NADP. Several experiments could be performed to try and confirm or disprove this suggestion. Modifiction of the enzyme using chemicals that specifically modify different amino acids could help to identify residues required for enzyme activity. Protection against such modification by the presence of NADP bound to the enzyme could indicate which residues are adjacent to or within the NADP binding site. Similar experiments could be performed to also identify residues involved in isocitrate binding. Another method would be to substitute specific amino acids thought to be involved in NADP binding for amino acids with different properties. This could be achieved by introducing specific mutations into the cloned *icd* B gene by site directed mutagenesis. Any change in the ability of the enzyme to bind NADP would provide evidence that the amino acid changed site is involved in coenzyme binding.

The ultimate method for identifying the NADP binding site would be by determining the position of NADP within the three-dimensional structure of the enzyme using techniques such as X-ray crystallography. Production of sufficient quantities of pure

protein should be relatively easy now that the enzyme has been overexpressed in *S. lividans*. Determination of the crystal structure would also be of interest for comparison to the structure of type I IDHs (based on the *E.coli* IDH). It is possible that the core structure of the type II IDH is similar to that of the type I IDHs with intervening sequences being required for the correct folding and activity of the enzyme as a monomer.

References.

Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. and Rossmann, M.G. (1987). Refined crystal structure of dogfish M4 apo-lactate dehydrogenase. J. Mol. Biol. **198**, 445-467.

Albretsen, 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.

Amersham International plc. Blotting and hybridization protocols for Hybond membranes. In: Hybond-N protocol manual.

Apostolakos, D., Menter, P.A., Rampsch, B.J., Reeves, H.C. and Birge, E.A. (1982). Genetic map position of the cistron coding for isocitrate dehydrogenase in *Escherichia coli* K-12. Cur. Microbiol. **7**, 45-47.

Barford, D. (1991). Molecular mechanisms for the control of enzymic activity by protein phosphorylation. Biochim. et Biophys. Acta **1133**, 55-62.

Barrera, C.R. and Jurtshuk, P. (1970). Characterization of the highly active isocitrate (NADP) dehydrogenase of *Azotobacter vinelandii*. Biochim. et Biophys. Acta **220**, 416-429.

Bibb, and Cohen, (1982). Gene expression in *Streptomyces* construction and application of promoter probe plasmid vectors in *Streptomyces lividans*. Mol. Gen. Genet. **187**, 265-277.

Binnie, C. (1990). The use of oligonucleotides as hybridisation probes. In: Advances in Gene Technology. Volume I. JAI Press Ltd. pp 135-154.

Bennett, P.M. and Holms, W.H. (1974). Reversible inactivation of the isocitrate dehydrogenase of *Escherichia coli* ML308 during growth on acetate. J. Gen. Micro. **87**, 37-51.

Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nuc. Acid Res. **7**, 1513-1523.

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.

Borthwick, A.C., Holms, W.H. and Nimmo, H.G. (1984). The phosphorylation of *Escherichia coli* isocitrate dehydrogenase in intact cells. J. Biochem. **222**, 797-804.

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. & Trifonov, E.N. (1984). A computer programme for testing potential prokaryotic terminators. Nuc. Acids Res. **12**, 4411-4427.

Buttner, M.J. & Brown, N.L. (1985). RNA polymerase interactions in *Streptomyces*. *In vitro* studies of a *S. lividans* plasmid promoter with *S. coelicolor* RNA polymerase. J. Mol. Biol. **185**, 177.

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

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

Clarke, L. and Carbon, J. (1976). A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**, 91.

Chater, K.F. (1984). Morphological and Physiological differentiation in *Streptomyces*. In: "Microbial Development." Losick, R. and Shapiro, L. (eds). Cold Spring Harbor Laboratory Press, New York. pp. 89-115.

Chung, A.E. and Braginski, J.E. (1972). Isocitrate dehydrogenase from *Rhodepseudomonas spheroides*: purification and characterization. Arch. Biochem. and Biophys. **153**, 357-367.

Cupp, J.R. and McAlister-Henn, L. (1991). NAP^+-dependent isocitrate dehydrogenase: cloning, nucleotide sequence and disruption of the *idh2* gene from *Saccharomyces cerevisiae*. J. Biol. Chem. **266** (33), 22199-22205.

Dean, A.M., Lee. M.H.I. and Hoshland, D.E. Jr. (1989). Phosphorylation inactivates *Escherichia coli* isocitrate dehydrogenase by preventing isocitrate binding. J. Biol. Chem. **264** (34), 20482-20486.

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.

Devereux, J., Haeberli, P. and Smithies O. (1984). A comprehensive set of sequence analysis pprograms for the VAX. Nuc. Acid res. **12**, 387-395.

Distler, J., Mansouri, K., Possiwotzki, K., Heinzel, P., Mayer, G. and Pipersberg, W. (1988). Organisation and expression of streptomycin production genes in *S. griseus*. In: "Abstracts of the 108th Meeting of the Genetical Society", Norwich, UK.

Distler, J., Ebert, A., Mansoori, K., Possowatzki, K., Stock, M. and Piepersberg, W. (1987). Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: nucleotide sequence of three genes and analysis of transcriptional activity. Nuc. Acid Res. **15**, 8041-8056

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.

Eguchi, H., Wakagi, T. and Oshima, T. (1989). A highly stable NADP-dependent isocitrate dehydrogenase from *Thermus thermophilus* HB8: purification and general properties. Biochim. et Biophys. Acta **990**, 133-137.

Fukunaga, N., Yoshida, S., Ishii, A., Imagawa, S., Takada, Y. and Sasaki, S. (1988). Isolation and characterization of a mutant defective in one of two isozymes of the isocitrate dehydrogenase from a psychrophilic bacterium *Vibrio* sp. strain ABE-1. J.Biochem

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.

Gibson, T.J. (1984). "Studies on the Epstein-Barr virus genome". PhD Thesis, Cambridge University, England.

Grafe, U., Bormann, E.J. and Truckenbrodt. G. (1980). Control by phospho-adenosinediphospho-ribose of NADP-dependent isocitrate dehydrogenase and 6-phosphogluconate dehydrogenase in *Streptomyces griseus*. Zeitschrift fur Allgemeine Mikrobiologie **20** (10), 607-611.

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

Hawley, D.K. and McClure, W.R. (1983). Compilation and analysis of *E. coli* promoter DNA sequences. Nuc. Acid Res. **11**, 2237-2255.

Haselbeck, R.J. and McAlister-Henn, L. (1991). Isolation, nucleotide sequence and disruption of the *Saccharomyces cerevesiae* gene encoding mitochondrial NADP(H)-specific isocitrate dehydrogenase. J. Biol. Chem. **266** (4), 2339-2345.

Helling, R.B. and Kukora, J.S. (1971). Nalidixic acid-resistant mutants of *Escherichia coli* deficient in isocitrate dehydrogenase. J. Bacteriol. **105** (3), 1224-1226.

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.

Hong, S-K., Kito, M., Beppu, T. and Horinouchi, S. (1991). Phosphorylation of the AfsR Product, a global regulatory protein for secondary-metabolite formation in *Streptomyces coelicolor* A3(2). J. Bacteriol. **173** (7), 2311-2318.

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.

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. (1985a). In: "Genetic manipulation of *Streptomyces*: a laboratory manual." The John Innes Foundation.

Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A.M., Floss, H.G. and Omura, S. (1985b). Production of hybrid antibiotics by genetic engineering. Nature **314**, 642-646.

Horinouchi, S. and Beppu, T. (1990). Autoregulatory factors of secondary metabolism and morphogenesis in actinomycetes. Biotechnology **10** (3), 191-204.

Howard, R.L. and Becker, R.R. (1970). Isolation and some properties of the triphosphopyridine nucleotide isocitrate dehydrogenase from *Bacillus stearothermophilus*. J. Biol. Chem. **245** (12), 3186-3194.

Huang, Y-C. and Colman, R.F. (1990). Subunit location and sequences of cysteinyl peptides of pig heart NAD-dependent isocitate dehydrogenase. Biochem. **29** (36), 8266-8273.

Hurley, J.H., Thorsness, P.E., Ramalingam, V., Helmers, N.H., Koshland, D.E.Jr. and Stroud, R.M. (1989). Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. Proc. Natl. Acad. Sci. USA **86**, 8635-8639.

Hurley, J.H., Dean, A.M., Sohl, J.L., Koshland, D.E.Jr. and Stroud, R.M. (1990a). Regulation of an enzyme by phosphorylation at the active site. Science **249**, 1012-1016.

Hurley, J.H., Dean, A.M., Thorsness, P.E., Hoshland, D.E.Jr. and Stroud, R.M. (1990b). Regulation of isocitrate dehydrogenase by phosphorylation involves no long-range conformational change in the free enzyme. J. Biol. Chem. **265** (7), 3599-3602.

Hurley, J.H., Dean, A.M., Koshland, D.E. Jr. and Stroud, R.M. (1991). Catalytic mechanism of $NADP^+$-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and $NADP^+$ complexes. Biochem. **30**, 8671-8678.

Hyde, G.E., Crawford, N.M. and Campbell, W.H. (1991). The sequence of squash NADH: Nitrate reductase and its relationship to the sequences of other flavoprotein oxidoreductases. J. Biol. Chem. **266** (35), 23542-23547.

Ikeda, T. and La Porte, D.C. (1991). Isocitrate dehydrogenase kinase/phosphatase: *aceK* alleles that express kinase but not phosphatase activity. J. Bacteriol. **173** (5), 1801-1806.

Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. and Oshima, T. (1991). Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2 A resolution. J. Mol. Biol. **222**, 725-738.

ISBA (International Symposium on Biology of Actinomycetes). University of Wisconsin, Madison, 1991. Abstract book.

Ishii, A., Ochiai, T., Imagawa, S., Fukunaga, N., Sasaki, S., Minowa, O., Mizuno, Y. and Shiokawa, H. (1987). Isozymes of isocitrate dehydrogenase from an obligately psychrophilic bacterium, *Vibrio* sp. strain ABE-1: purification and modulation of activities by growth conditions. J. Biochem. **102**, 1489-1498.

Janssen, G., Ward, J.M. and Bibb, M. (1989). Unusual transcriptional and translational features of the aminoglycoside phosphotransferase gene (*aph*) from *Streptomyces fradiae*. Genes and Development **3**, 415-429.

Jaurin, B. & Cohen, S.N. (1985). *Streptomyces* contain *Escherichia coli*-type A+T rich promoters having novel structural features. Gene **39**, 191.

Kaiser, K. & Murray, N.E. (1985). The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. In: DNA cloning Vol. I. (Glover, D.M., ed). IRL Press.

Karplus, P.A., Daniels, M.J. and Herriott, J.R. (1991). Atomic structure of ferredoxin-$NADP^+$ reductase: prototype for a structurally novel flavoenzyme family. Science **251**, 60-66.

Kelly, M.T. and Brock, T.D. (1969). Molecular heterogeneity of isolates of the marine bacterium *Leucothrix mucor*. J. Bacteriol. **100**, 14-21.

Khokhlov, A.S., Anisova, L.N., Tovarova, I.I., Kleiner, E.M., Kovalenko, I.V., Krasilnikova, O.I., Kornitshaya, E. Ya. and Pfiner, S.A. (1973). Effect of A-factor on the growth of asporogenous mutants of *Streptomyces griseus*, not producing this factor. Mikrobiol. **13**, 647.

Khokhlov, A.S. (1982). Low molecular weight microbial bioregulators of secondary metabolism. In: Overproduction of Microbial Product (Krumphanzl, V., Sikyta, B. and Vanek, Z., eds.). Academic Press, London. pp 96.

Kim, J-J.P. and Wu, J. (1988). Structure of the medium-chaim acyl-CoA dehydrogenase from pig liver mitochondria at 3-A resolution. Proc. Natl. Acad. Sci. USA **85**, 6677-6681.

Kleber, H.P. and Aurich, H. (1976). Control of $NADP^+$-specific isocitrate dehydrogenase from *Acinetobacter* by nucleotides. FEBS **61** (2), 282-285.

Kornberg, H.L. (1966). Anaplerotic sequences and their role in metabolism. Essays in Biochem. **2**, 1-31.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**, 680-685.

Lakshmi, T.M. and Helling, R.B. (1976). Selection for citrate synthase deficiency in *icd* mutants of *Escherichia coli*. J. Bacteriol **127**, 76-83.

La Porte, D.C., Thorsness, P.E. and Koshland, D.E. Jr. (1985). Compensatory phosphorylation of isocitrate dehydrogenase: a mechanism for adaptation to the intracellular environment. J. Biol. Chem. **260** (19), 10563-10568.

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

Lineweaver, H. and Burke, D. (1934). J. Amer. Chem. Soc. **56**, 658.

Leyland, M.L. and Kelly, D.J. (1991). Purification and characterization of a monomeric isocitrate dehydrogenase with dual coenzyme specificity from the photosynthetic bacterium *Rhodomicrobium vannielii*. . **202**, 85-93. Eur. J. Biochem.

Lilley, K.S., Baker, P.J., Britton, K.L. Stillman, T.J., Brown,P.E., Moir, A.J.G., Engel, P.C., Rice, D.W., Bell, J.E. and Bell, E. (1991). The partial amino acid sequence of the NAD^+-dependent glutamate dehydrogenase of *Clostridium symbiosum*: implications for the evolution and structural basis of coenzyme specificity. Biochim. et Biophys. Acta **1080**, 191-197.

Losick, R. and Pero, J. (1981). Cascades of sigma factors. Cell **25**, 582-584.

Losick, R., Youngman, P. and Piggot, P. (1986). Genetics of endospore formation in *Bacillus subtilis*. Ann. Rev. Genet. **20**, 625-667.

McKee, J.S. and Nimmo, H.G. (1989). Evidence for an arginine residue at the coenzyme-binding site of *Escherichia coli* isocitrate dehydrogenase. Biochem. J. **261**, 301-304.

Maizel and Lenk (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. (USA) **78** (12), 7665-7669.

Malpartida, F. and Hopwood, D.A. (1984). Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. Nature **309**, 462-464.

Meixner-Monori, B., Kubicek, C.P., Harrer, W., Schreferl, G. and Rohr, M. (1986). NADP-specific isocitrate dehydrogenase from the citric acid-accumulating. J. Biochem. **236**, 549-557.

Miyake, K., Horinouchi, S., Yoshida, M., Chiba, N., Mori, K., Nogawa, N., Morikawa, N. and Beppu, T. (1989). Detection and properties of A-factor-binding protein from *Streptomyces griseus*. J. Bacteriol. **171**, 4298.

Miyake, K., Kuzuyama, T., Horinouchi, S. and Beppu, T. (1990). The A-factor-binding protein of *Streptomyces griseus* negatively controls streptomycin production and sporulation. J. Bacteriol. **172** (6), 3003-3008.

Miyazaki, K., Eguchi, H., Yamagishi, A., Wakagi, T. and Oshima, T. (1992). Molecular cloning of the isocitrate dehydrogenase gene of an extreme thermophile, *Thermus thermophilus* HB8. App. Environ. Micro. **58** (1), 93-98.

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. Nuc. Acid Res. **14**, 1319-1324.

Muro-Pastor, M.I. and Florencio, F.J. (1992). Purification and properties of NADP-isocitrate dehydrogenase from the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. Eur. J. Biochem. **205**, 99-105.

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. (1984). Control of *Escherichia coli*: isocitrate dehydrogenase: an example of protein phosphorylation in a prokaryote. TIBS(9), 475-478.

Nimmo, H.G. (1986). Kinetic mechanism of *Escherichia coli* isocitrate dehydrogenase and its inhibition by glyoxylate and oxalacetate. Biochem J. **234**, 317-323.

Nimmo, H.G., Borthwick, A.C., El-Mansi, E.M.T., Holms, W.H., MacKintosh, C. and Nimmo, G.A. (1987). Regulation of the enzymes at the branchpoint between the citric acid cycle and the glyoxylate bypass in *Escherichia coli*. Biochem. Soc. Symp. **54**, 93-101.

Ochiai, T., Fukunaga, N. and Sasaki, S. (1979). Purification and some properties of two $NADP^+$-specific isocitrate dehydrogenases from an obliquely psychrophilic marine bacterium, *Vibrio* sp., strain ABE-1. J. Biochem. **86**, 377-384.

Ochiai, T., Fukunaga, N. and Sasaki, S. (1984). Two structurally different NADP-specific isocitrate dehydrogenases in an obligately psychrophilic bacterium *Vibrio* sp, strain ABE-1[1]. J. Gen. Appl. Microbiol. **30**, 479-487.

Promega Corporation, USA (1991). Promega Protocols and Applications Guide, 2nd edition.

Ptashne, M., Backman, K., Humazaun, M.Z., Jeffrey, A., Maurer, R. and Sauer, R.T. (1976). Autoregulation and function of a repressor in bacteriophage lambda. Interactions of a regulatory protein with sequences in DNA mediate intricate patterns of gene regulation. Science **194**, 156-161.

Ramachandran, N. and Colman, R.F. (1980). Chemical characterisation of distinct subunits of pig heart DPN-specific dehydrogenase. J. Biol. Chem. **255**, 8859-8864.

Ragland, T.E., Kawasaki, T. and Lowenstein, J.M. (1966). Comparative aspects of some bacterial dehydrogenases and transhydrogenases. J. Bacteriol. **91**, 236-244.

Reeves, H.C., O'Neil, S. and Weitzman, P.D.J. (1986). Changes in NADP-isocitrate dehydrogenase isoenzyme levels in *Acinetobacter calcoaceticus* in response to acetate. FEMS **35**, 229-232.

Rondeau, J-M., Tete-Favier, F., Podjarny, A., Reymann, J-M., Barth, P., Biellmann, J-F and Moras, D. (1992). Novel NADPH-binding domain revealed by the crystal structure of aldose reductase. Nature **355**, 469-472.

Rossmann, M.G., Moras, D. & Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. Nature **250**, 194-199.

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. **74**, 5463-5467.

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.

Scrutton, N.S., Berry, A. and Perham, R.N. (1990). Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature **343**, 38-43.

Self, C.H. and Weitzman, P.D.J. (1972). The isocitrate dehydrogenases of *Acinetobacter lwoffi*. Separation and properties of two nicotinamide-adenine dinucleotide phosphate-linked isoenzymes. J. Biochem. **130**, 211-219.

Self, C.H., Parker, M.G. and Weitzman, P.D.J. (1973). The isocitrate dehydrogenases of *Acinetobacter lwoffi*. Studies on the regulation of nicotinamide-adenine dinucleotide phosphate-linked isoenzyme. J. Biochem. **132**, 215-221

Sherman, D.H., Malpartida, F., Bibb, M.J., Bibb, M.J., Kieser, H.M., Hallam, S.E., Robinson, J.A., Bergh, S., Uhlen, M., Simpson, T.J. and Hopwood, D.A. (1988). Cloning and analysis of genes for the biosynthesis of polyketide antibiotics in *Streptomyces* species. Proceedings of the 8th International Biotechnology Symposium, Paris. (Durand, G., Bobichon, L., Florent, J., eds). pp 123-137.

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. **71**, 1342-1346.

Smith, T.F. & Waterman, M.S. (1981). Adv. Appl. Math **2**, 482-489.

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

Takao, S., Iida, T. and Tanida, M. (1986). Purification and characterization of $NADP^+$-linked isocitrate dehydrogenase from *Paecilomyces varioti*. Agric. Biol. Chem. **50** (6), 1573-1579.

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

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.

Tomich, P.K. (1988a). *Streptomyces* cloning: useful recombinant DNA systems and a summation of cloned genes. Antimicro. Agents and Chemother. **32** (10), 1465-1471.

Tomich, P.K. (1988b). *Streptomyces* cloning: possible construction of novel compounds and regulation of antibiotic biosynthetic genes. Antimicro. Agents and Chemother. **32** (10), 1472-1476.

Voronina, O.I., Tovarova, I.I. and Khokhlov, A.S. (1978). Studies on the A-factor induced inhibition of glucose-6-phosphate dehydrogenase in *Actinomyces streptomycini*. Bioorg. Khim. **4**, 1538.

Wang, J.Y.J. and Koshland Jr., D.E. (1982). The reversible phosphorylation of isocitrate dehydrogenase of *Salmonella typhimurium*. Arch. Biochem. and Biophys. **218** (1), 59-67.

Westpheling, J.E., Raines, M. and Losick, R. (1985). RNA polymerase heterogeneity in *Streptomyces coelicolor*. Nature **313**, 22-27.

Whittaker, P.A., Campbell, A.J.B., Southern, E.M. and Murray, N.E. (1988). Enhanced recovery and restoration mapping of DNA fragments cloned into a new lambda vector. Nuc. Acids Res. **16**, 6725-6734.

Wierenga, R.K. and Hol, W.G.J. (1983). Predicted nucleotide-binding properties of p21 protein and its cancer-associated varient. Nature **302**, 842-844.

Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986). Prediction of the occurrence of the ADP-binding BaB-fold in proteins, using an amino acid sequence fingerprint. J. Mol. Biol. **187**, 101-107.

Williams, S.T., Goodfellow, M., Alderson, G., Wellington, E.M.H., Sneath, P.H.A. and Sackin, M.J. (1983). Numerical classification of streptomyces and related genera. J. Gen. Microbiol. **129**, 1743-1813.

Wray, L.U. Jr. and Fisher, S.H. (1988). Cloning and nucleotide sequence of the *S. coelicolor* gene encoding glutamine synthetase. Gene **71**, 247-256.

Zalacain, M., Gonzales, A., Guerrero, M.C., Mattaliano, R.J., Malpartida, F. and Jimenez, A. (1986). Nucleotide sequence of the hygromycin B phosphotransferase gene from *S. hygroscopicus*. Nuc. Acids Res. **14**, 1565-1581.