THE STRUCTURAL GENES FOR METHYLMALONYL-CoA METABOLISM IN PROPIONIBACTERIUM SHERMANII

by

Nigel Keith Davis

A dissertation submitted to the University of Cambridge in candidature for the degree of Doctor of Philosophy

March, 1986

Corpus Christi College and
Department of Biochemistry,
Cambridge.

PREFACE

I would like to express my gratitude to Dr. P.F. Leadlay, my supervisor, for his guidance and encouragement. I would also like to thank D. Murfitt and J.Q. Fuller for expert technical assistance. In addition, I would like to thank Dr. J.E. Walker of the MRC Laboratory of Molecular Biology and Professor J. Fothergill of Aberdeen University for performing N-terminal protein sequence analyses. Finally I would like to thank my wife, Elaine, for her encouragement and support.

I declare that this thesis is the result of my own work, and includes nothing which is the outcome of work done in collaboration, except where indicated otherwise.

Financial support was provided by the award of an SERC Studentship, and, in the latter stages, by a grant from the British Technology Group.

SUMMARY

1. The methylmalonyl-CoA epimerase gene of *P. shermanii* was cloned into *E. coli* on a 10.4kb restriction fragment using oligonucleotide probes directed against part of the N-terminal peptide sequence. The gene was not expressed in *E. coli* but the identity of the gene was confirmed by *in vitro* transcription-translation of the cloned DNA by an *S. lividans* cell free system followed by immuno-precipitation of the products with anti-epimerase antiserum.

2. A 2.3kb fragment of the original clone was shown to carry the entire epimerase gene by coupled transcription-translation of the DNA. A larger, 5.4kb fragment was subcloned into *S. lividans* in the high copy number plasmid pIJ702. Epimerase was expressed from its own promoter to over 10% of the soluble protein. The protein was purified and had enzymic activity.

3. The DNA sequence of the 2.3kb fragment was determined. Within this sequence an open reading frame was found with the correct N-terminal peptide sequence, amino acid composition and subunit molecular mass for epimerase. Other reading frames on this fragment showed no significant homology with any published sequences.

4. Methylmalonyl-CoA mutase was purified from *P. shermanii* by an improved method yielding a larger protein than previously reported (M_r 165 000). The protein was an $\alpha\beta$-dimer with subunits M_r 79 000 and 67 000. Several peptides from a proteolytic digestion of mutase were purified and then sequenced by the manual DABITC method.

5. Oligonucleotide probes complementary to the DNA coding for some of the mutase peptides were used to probe *P. shermanii* DNA. Small, hybridising restriction fragments were cloned into *E. coli*. Clones obtained with one probe caused synthesis of polypeptide precipitable with anti-mutase antiserum in the *in vitro* transcription-translation system. A peptide of approx. M_r 65 000 was encoded by the largest cloned fragment corresponding to nearly one complete subunit of mutase.

CONTENTS

ABBREVIATIONS

ATP	adenosine triphosphate
DABITC	dimethylaminoazobenzene isothiocyanate
DABTH	dimethylaminoazobenzene thiohydantoin
DMSO	dimethylsulphoxide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FPLC	fast protein liquid chromatography
HPLC	high-pressure liquid chromatography
IPTG	isopropyl-β-D-thiogalactoside
LGT	low gelling-temperature
NADH	reduced nicotine adenine dinucleotide
NEM	N-ethylmaleimide
PITC	phenylisothiocyanate
PMSF	phenylmethanesulphonyl fluoride
SDS	sodium dodecyl sulphate
TE	10mM Tris-HCl/0.1mM EDTA
TLE	thin-layer electrophoresis
T_m	melting temperature of DNA duplex
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
MSNT	1-(mesitylene sulphonyl)-3-nitro-1,2,3-triazole

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1 The propionate pathway of Propionibacterium shermanii

Propionibacterium shermanii (syn. freudenreichii) is a microaerotolerant Gram positive bacterium which is found naturally in the rumen of dairy animals and in dairy products. Although it can grow on carbohydrates it appears to have found an ecological niche where it uses lactate, which is produced by many other rumen microorganisms as a waste product, as its main carbon source, fermenting it to propionate and acetate. The metabolism of P. shermanii is shown in Figure 1.1.

P. shermanii oxidises lactate to pyruvate from which it derives its energy by oxidative decarboxylation using pyruvate dehydrogenase yielding acetyl-CoA. The cleavage of the thioester bond is coupled to ATP synthesis and the acetate released is excreted. This pathway yields energy but requires two oxidations using NAD so some pyruvate is metabolised via a different pathway to maintain the oxidation-reduction balance.

The propionate pathway involves two reduction steps and is not energy requiring (or yielding) because two steps that normally require energy input are mediated by transferases coupling them to other steps in the pathway. Transcarboxylase carboxylates pyruvate by transferring the carboxyl group from (2S)-Methylmalonyl-CoA avoiding the kinase step that would normally be used. The oxaloacetate produced by this reaction then follows the reverse of the citric acid cycle through two reductions until succinate is formed. This is converted to succinyl-CoA by transfer of coenzyme A from propionyl-CoA, which was produced by the transcarboxylase reaction. Succinyl-CoA is isomerised to (2R)-methylmalonyl-CoA by methylmalonyl-CoA mutase and the (2R)-methylmalonyl-CoA is epimerised by methylmalonyl-CoA epimerase to the (2S)- form. This is then decarboxylated by transcarboxylase to yield propionyl-CoA. The carboxyl group is transferred to another pyruvate molecule. The coenzyme A group is transferred to another succinate molecule, and the resulting propionate is excreted.

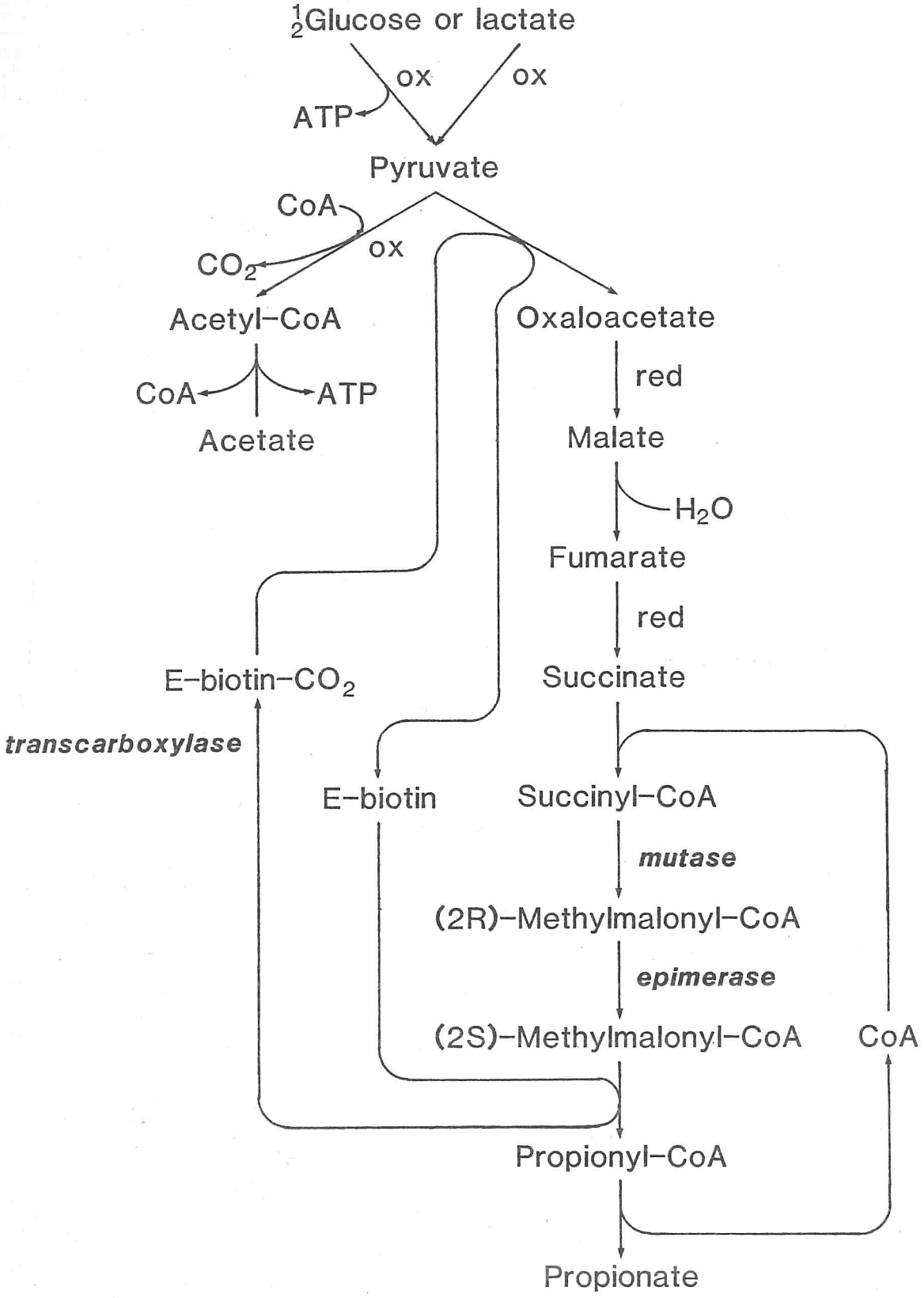


Figure 1.1 The propionate pathway of P. shermanii

The overall pathway from lactate to propionate, therefore, involves one oxidation and two reduction steps, but the lactate to acetate pathway involves two oxidation steps, so, to maintain an overall balance, two lactate molecules are converted to propionate for every one converted to acetate and every ATP formed. The overall reaction to produce one ATP molecule is:

$$3 \text{ lactate} \rightarrow 2 \text{ propionate} + 1 \text{ acetate} + 1CO_2 + 1 H_2O$$

The low energy yield means that P. shermanii grows fairly slowly. Holes in Swiss cheeses are caused by the carbon dioxide released by the late fermentation of P. shermanii. The propionate also adds to the flavour.

In the rumen the propionate excreted is absorbed by the host animal and converted back to succinate. This pathway requires ATP for a kinase reaction to produce propionyl-CoA, although GTP is regained from hydrolysis of the succinyl-CoA bond. ATP is also required for the carboxylation of propionyl-CoA to (2S)-methylmalonyl-CoA. The epimerisation and isomerisation are direct reversals of the bacterial reactions. This mammalian pathway is also used for the metabolism of propionyl-CoA produced from the breakdown of odd-chain fatty acids and branched chain amino acids.

Three of the enzymes of the P. shermanii propionate pathway have been intensively studied from the point of view of their catalytic mechanisms. Methylmalonyl-CoA transcarboxylase is a biotin-containing enzyme that is unique because a carboxy group is transferred between two acceptor molecules rather than to or from bicarbonate. Methylmalonyl-CoA mutase is a coenzyme B_{12} dependent enzyme catalysing the unusual carbon skeleton rearrangement of succinyl-CoA to give (2R)-methylmalonyl-CoA. The reason why these two enzymes use different epimers of methylmalonyl-CoA is not clear, but the epimerase is a simple enzyme requiring no cofactor and catalysing the interconversion of the two forms.

1.1.1 Transcarboxylase

Methylmalonyl-CoA-pyruvate transcarboxylase (E.C. 2.1.3.1) catalyses the reaction shown in Figure 1.2. It is a large complex (M_r 1.2×10^6) constructed from three types of subunit. A central core of six copies of one subunit has

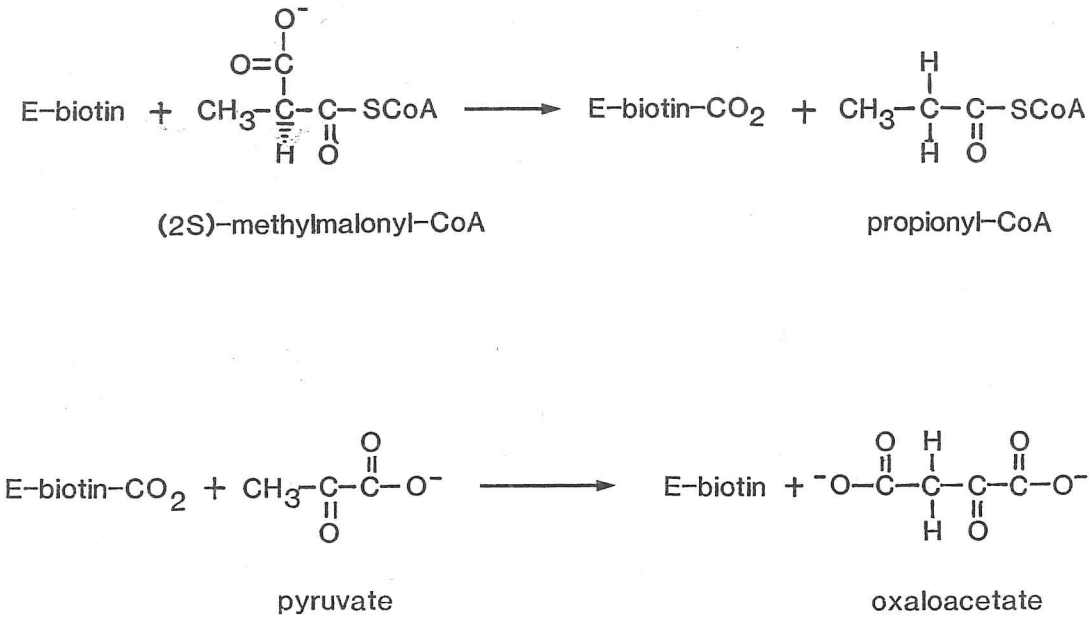


Figure 1.2 The reaction catalysed by transcarboxylase

The two half-reactions shown are catalysed by different subunits of transcarboxylase. The carboxyl group is carried between these sites by biotin attached to a third subunit.

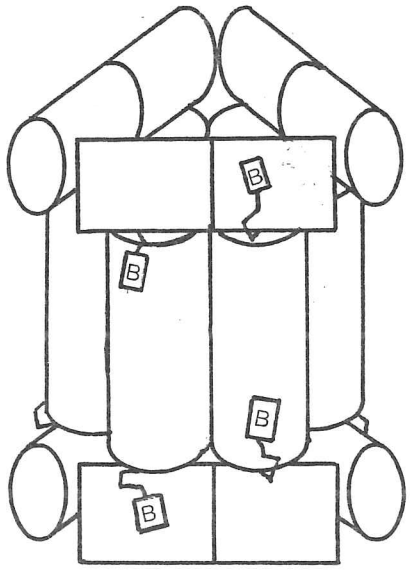


Figure 1.3 The structure of transcarboxylase

The central core of six subunits has twelve active sites catalysing the decarboxylation of (2S)-methylmalonyl-CoA. The twelve outer subunits catalyse the carboxylation of pyruvate. The carboxyl group is carried between the active sites by biotin (B) attached to twelve small subunits, which also bind the outer subunits to the core.

bound to it six dimers of a second subunit type through twelve small subunits as shown in Figure 1.3 (Wood and Zwolinski, 1976). The small subunit (M_r 12 000) carries the biotin group and acts as carboxyl group carrier between the two sorts of active site on the other subunits. At one of these, in the central core, (2S)-methylmalonyl-CoA is decarboxylated yielding propionyl-CoA. The carboxyl group is then carried to another active site, on an outer subunit, where pyruvate is carboxylated to give oxaloacetate. In other biotin-dependent enzymes the carboxybiotin is produced from ATP and bicarbonate, except for the irreversible reaction catalysed by malonyl-CoA decarboxylase from Micrococcus lysodeikticus (Wood and Barden, 1977).

The central core of six polypeptides has twelve active sites, one at each end of each polypeptide. These two active sites on the same polypeptide chain catalysing the same reaction may have arisen by gene duplication although proteolytic digestions of this subunit have not yielded any evidence to support this (Zwolinski et al., 1977). The two faces of the core, which each bind three dimers of the outer subunit through six biotinyl subunits, are non-equivalent: dissociation of external subunits takes place preferentially from one face.

The small, biotin-carrying subunit has been sequenced and, around the lysine to which the biotin is bound, contains an amino acid sequence that is conserved in most biotin-dependent enzymes (Maloy et al., 1979). The amino acid compositions of the other two subunits are quite similar to one another, and there is some evidence for immunological cross-reaction between them, suggesting that they may contain related sequences. This might be expected since both must bind to the biotin containing part of the small subunit during the reaction. The amino terminal amino acid of the polypeptide comprising the central core is reported to be alanine but that of the outer subunit is blocked (Zwolinski et al., 1977).

The outer subunits are bound to the central core through the small, biotinyl subunit. Knowledge of the amino acid sequence of this biotin-carrying subunit has allowed use of peptide fragments and a synthetic peptide to show that the first fourteen residues bind to the central core and the next twelve bind to the outer subunit (Kumar et al., 1982). After most of the

work reported in this thesis was completed the gene for this subunit was cloned and sequenced (Murtif et al., 1985). This confirmed the peptide sequence but only a small fragment was cloned and the genes for the other subunits were not present.

As well as this intensive study of the structure of the transcarboxylase multisubunit complex there has been much work directed towards elucidating the stereochemical course and catalytic mechanism of the enzyme. The question of the concertedness of proton removal from substrate and transfer of the carboxy group from carboxybiotin has now been settled in favour of a non-concerted mechanism (Stubbe et al., 1980; Kuo and Rose, 1982; O'Keefe and Knowles, 1986).

1.1.2 Methylmalonyl-CoA epimerase

Methylmalonyl-CoA epimerase (E.C. 5.1.99.1) catalyses the deprotonation and reprotonation at C-2 of methylmalonyl-CoA. Unlike transcarboxylase, this enzyme is found in mammalian tissue as well as bacteria because it is involved in the normal metabolism of propionate produced by breakdown of odd-chain fatty acids and branched chain amino acids. It has been purified from sheep liver (Mazumder et al., 1962) and rat liver (Stabler et al., 1985) as well as from P. shermanii (Allen et al., 1963; Leadlay, 1981). The bacterial enzyme has a native M_r of 35 000 by gel filtration, 33 000 ± 2 000 from sedimentation equilibrium and has an $S^{\circ}_{20,w}$ of 3.05S. However, on SDS-polyacrylamide gels it migrates as a single band of M_r 17 000 ± 1 000. Cross-linking with dimethylsuberimidate causes a band of M_r 34 000 to appear on SDS-polyacrylamide gels as the M_r 17 000 band disappears but no higher oligomers are seen. This suggests that the epimerase is a globular dimer of two identical subunits. The amino acid composition contains a total of 15-16 arginine and lysine residues per M_r 16 500. After tryptic digestion a maximum of 14-15 fragments could be resolved supporting the argument that the two subunits are identical (Leadlay, 1981).

Each subunit contains two cysteine thiol residues which both react cleanly with iodoacetic acid or performic acid under denaturing conditions. The native enzyme reacts very slowly with iodoacetate, with concomitant loss of

work reported in this thesis was completed the gene for this subunit was cloned and sequenced (Murtif *et al.*, 1985). This confirmed the peptide sequence but only a small fragment was cloned and the genes for the other subunits were not present.

As well as this intensive study of the structure of the transcarboxylase multisubunit complex there has been much work directed towards elucidating the stereochemical course and catalytic mechanism of the enzyme. The question of the concertedness of proton removal from substrate and transfer of the carboxy group from carboxybiotin has now been settled in favour of a non-concerted mechanism (Stubbe *et al.*, 1980; Kuo and Rose, 1982; O'Keefe and Knowles, 1986).

1.1.2 Methylmalonyl-CoA epimerase

Methylmalonyl-CoA epimerase (E.C. 5.1.99.1) catalyses the deprotonation and reprotonation at C-2 of methylmalonyl-CoA. Unlike transcarboxylase, this enzyme is found in mammalian tissue as well as bacteria because it is involved in the normal metabolism of propionate produced by breakdown of odd-chain fatty acids and branched chain amino acids. It has been purified from sheep liver (Mazumder *et al.*, 1962) and rat liver (Stabler *et al.*, 1985) as well as from *P. shermanii* (Allen *et al.*, 1963; Leadlay, 1981). The bacterial enzyme has a native M_r of 35 000 by gel filtration, 33 000 ± 2 000 from sedimentation equilibrium and has an $S^{o}_{20,w}$ of 3.05S. However, on SDS-polyacrylamide gels it migrates as a single band of M_r 17 000 ± 1 000. Cross-linking with dimethylsuberimidate causes a band of M_r 34 000 to appear on SDS-polyacrylamide gels as the M_r 17 000 band disappears but no higher oligomers are seen. This suggests that the epimerase is a globular dimer of two identical subunits. The amino acid composition contains a total of 15-16 arginine and lysine residues per M_r 16 500. After tryptic digestion a maximum of 14-15 fragments could be resolved supporting the argument that the two subunits are identical (Leadlay, 1981).

Each subunit contains two cysteine thiol residues which both react cleanly with iodoacetic acid or performic acid under denaturing conditions. The native enzyme reacts very slowly with iodoacetate, with concomitant loss of

enzyme activity, but much faster with thiol-directed reagents such as Ellman's reagent (Nbs_2) and 4,4'-dithiobis(pyridine). One thiol is modified rapidly with loss of 90% of the enzyme activity (P.F. Leadlay, personal communication). Substrate protection against modification of this thiol is poor, making it unlikely that this cysteine thiol group is the enzyme base actually involved in proton transfer at the active site, although a water molecule or another enzyme base may be able to take over the catalytic role to give the residual activity. It is not yet known whether there are one or two active sites per dimer, so only one thiol in each subunit may be involved in catalysis if thiols are used. Strangely, if the enzyme is previously denatured in 6M guanidinium chloride, the second thiol group still does not react with Nbs_2 or 4,4'-dithiobis(pyridine). However, if the native enzyme is treated with 4,4'-dithiobis(pyridine) and excess reagent is removed, addition of urea to 2M final concentration induces rapid release of a second equivalent of the chromophore 4-thiopyridine ($t_{\frac{1}{2}}$=2min at 30°C, pH7.5) and loss of the residual 10% of the enzyme activity. The subunits are not cross-linked by this process so it appears that the two cysteine residues in each polypeptide are close enough together in space to form an intrachain disulphide bond upon mild denaturation with 2M urea.

Experiments with (2R)-methylmalonyl-CoA specifically tritiated at C-2 showed that none of the label was transferred to (2S)-methylmalonyl-CoA (Leadlay and Fuller, 1983). Reactions of unlabelled (2R)-methylmalonyl-CoA in tritiated water produced tritiated product but no label was found in the residual substrate (Fuller and Leadlay, 1983). These results and the accompanying isotope effects indicated that two enzyme bases were involved as shown in Figure 1.4.

One base removes the proton from C-2 of the substrate and the second, protonated base, protonates C-2 from the opposite side to give the inversion. The bases do not pass the proton from one to the other during or between reactions otherwise label would be found in the product when using tritiated substrate. Also the bases cannot equilibrate with the medium while substrate is bound or else label would be found in the residual substrate when the reaction is performed in tritiated water. The isotope effects also suggest that the bases do not carry any other protons so they could be thiol, carboxy,

imidazole or, possibly, phenolic hydroxy groups.

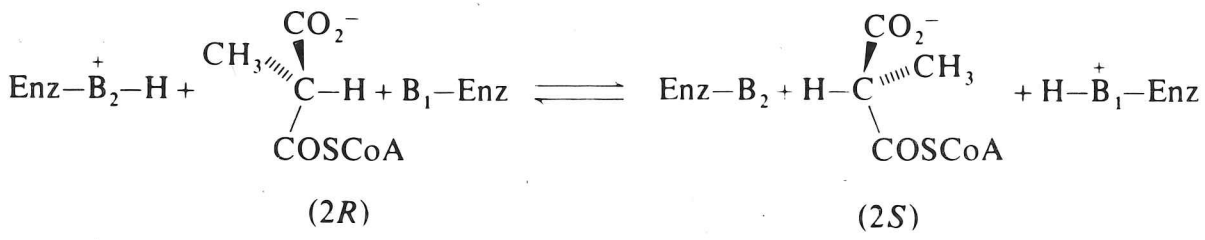


Figure 1.4 The mechanism of the epimerase reaction

Two enzyme bases are involved in the catalysis: one removes the proton from C-2 of the substrate and the other reprotonates to give the inversion. The figure is taken from Fuller and Leadlay (1983).

These results suggest that the mechanism of epimerase resembles the two-base mechanism established for another pyridoxal phosphate-independent racemase, proline racemase (Cardinale and Abeles, 1968; Rudnick and Abeles, 1975).

Evidence has been obtained for the involvement of divalent metal ions in the enzyme activity (Leadlay, 1981; Stabler et al., 1985). The ion may help to bind the substrate and polarise the thioester or carboxyl group carbonyl bonds to promote the removal of the C-2 proton, but the precise role of the metal is unclear.

1.1.3 Methylmalonyl-CoA Mutase

Methylmalonyl-CoA mutase contains 5'-deoxyadenosylcobalamin (coenzyme B_{12}) and catalyses the rearrangement of methylmalonyl-CoA to succinyl-CoA. The enzyme has been purified from several sources. The sheep liver enzyme (Cannata et al., 1965) has M_r 165 000 and contains one adenosylcobalamin per M_r 75 000. The holoenzyme is not sensitive to thiol-directed reagents, but the apoenzyme is. The enzyme from human liver has M_r 150 000 (Fenton et al., 1981), consisting of two, apparently identical, subunits. These enzymes were isolated largely as the holoenzyme form and were stable on storage.

Mutase has also recently been purified from the intestinal worm Ascaris lumbricoides (Han et al., 1984) as a potential target for chemotherapeutic agents. The enzyme from this source was also a dimer of apparently identical

subunits (M_r 75 000) with a total M_r of 143 000 and contained one tightly-bound adenosylcobalamin per subunit.

The enzyme purified from P. shermanii (Zagalak et al., 1974) was smaller than the eukaryotic enzymes (M_r 124 000) and was apparently a dimer of two different subunits (M_r 66 000 and 61 000). This enzyme had little activity without added cofactor, but with adenosylcobalamin present had a specific activity of 14U mg^{-1}, similar to the human enzyme. This enzyme was not very stable in solution.

Methylmalonyl-CoA mutase has also been partially purified from another bacterium, Streptomyces erythreus, where it is implicated in providing methylmalonyl-CoA precursors for the synthesis of the macrolide antibiotic erythromycin. This enzyme had just one subunit size (M_r 63 000) but the size of the native enzyme was not reported (Hunaiti and Kolattukudy, 1984a).

Thus, the bacterial enzymes were apparently smaller than those purified from eukaryotic sources and the enzyme from P. shermanii was unique in having two different subunits. The coenzyme is not so tightly bound in the P. shermanii enzyme, which is also less stable than the eukaryotic enzymes.

The reaction catalysed by mutase involves the movement of a hydrogen atom between two adjacent carbon atoms and the movement in the opposite direction of the entire -COSCoA group, as shown in Figure 1.5 (see Halpern, 1985 for a recent review of coenzyme B_{12}-dependent rearrangements).

Other coenzyme B_{12} dependent rearrangements also involve the movement of a hydrogen atom and the migration in the opposite direction of $-CCH_2COOH$ by α-methyleneglutarate mutase (Kung and Tsai, 1971), $-CHNH_2COOH$ by glutamate mutase (Barker et al., 1964), -OH by diol dehydrase (Abeles et al., 1960), or $-NH_2$ by ethanolamine ammonia lyase (Kaplan and Stadtman, 1968). Methylmalonyl-CoA mutase is the only one so far found in mammalian systems.

The coenzyme B_{12} (5'-deoxyadenosylcobalamin, see Figure 1.6) required as cofactor for these rearrangements generates free radicals by homolytic cleavage of the cobalt-carbon bond. The free radical abstracts a hydrogen atom from the substrate which then rearranges before the hydrogen is returned

and the cobalt-carbon bond reforms. Adenosylcobalamin is also involved in ribonucleotide reductase in some organisms but in this reaction the hydrogen atom abstracted from a thiol group in reduced thioredoxin is transferred to a different substrate, the nucleotide. Methylcobalamin is used in the transfer of 1-carbon units, and may also have a mechanism involving radicals.

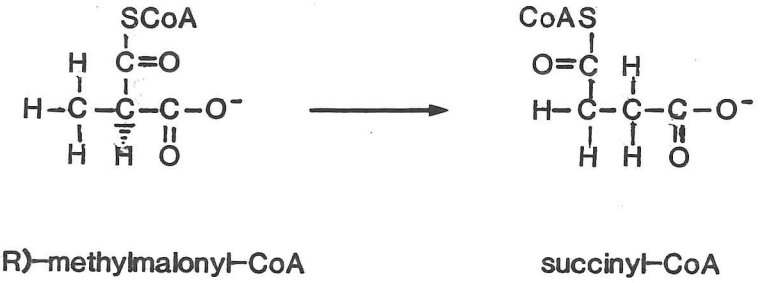


Figure 1.5 The reaction catalysed by mutase

A radical produced by coenzyme B_{12} in mutase abstracts a proton from the methyl group of (2R)-methylmalonyl-CoA. The substrate radical then rearranges: the -COSCoA group migrates onto the old methyl group. The proton is then returned. The groups which move are shown above the main carbon skeletons.

The cobalt-carbon bond in adenosylcobalamin is unusually long due to steric crowding between the adenosyl moiety and the corrin ring. Normally the dissociation energy of a cobalt-carbon bond would be too high to allow the rates of reaction observed with the mutases, but it is reduced in the coenzyme by the steric hindrance. The bond energy would still be slightly too high but a slight distortion of the corrin ring when the cofactor is bound to the protein could lower the dissociation energy sufficiently. Homologous series of model compounds with increasing bulk of the group bound to the cobalt show the amount of distortion necessary to give various dissociation energies, and can approach the dissociation energy calculated for the enzyme-bound cofactor (Halpern, 1985).

Direct evidence for the existence of adenosyl ($AdCH_2\cdot$) and cobalamin (B_{12r}) radicals derived from adenosylcobalamin has been obtained for some of the rearrangements, although not methylmalonyl-CoA mutase, using electron paramagnetic resonance spectroscopy. The radicals in the methylmalonyl-CoA mutase reaction may not be sufficiently long-lived to be detected by this technique. The $AdCH_2\cdot$ radical can be shown to abstract a proton from the

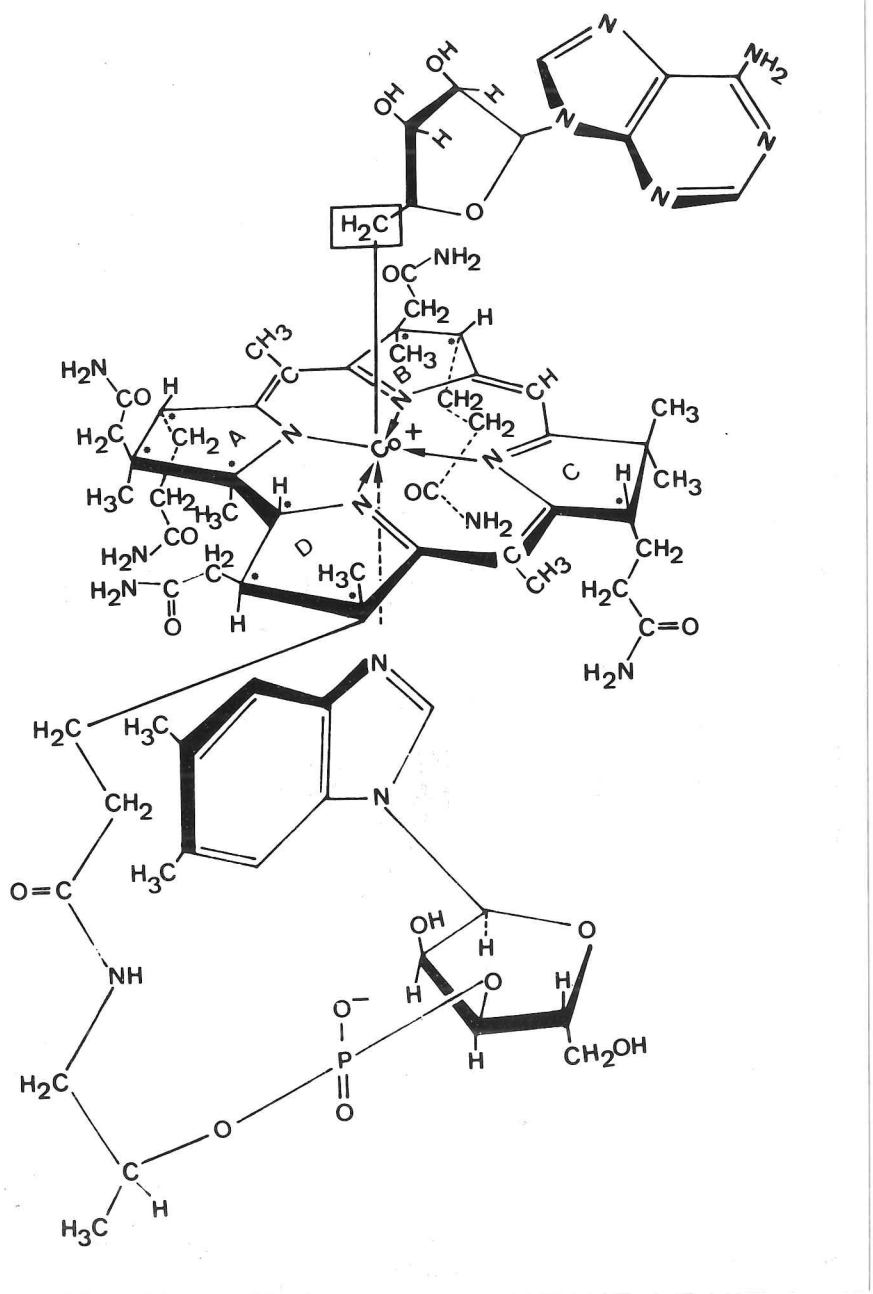


Figure 1.6 Coenzyme B_{12} (from Zagalak and Friedrich, 1979)

The methylene group of adenosylcobalamin, which becomes a $-CH_2\cdot$ radical after homolytic cleavage of the cobalt-carbon bond, is shown enclosed in a box.

substrate by using deuterium in the substrate. This becomes scrambled with the protons on $AdCH_2\cdot$ consistent with the formation of $AdCH_3$ in which all hydrogen atoms on the methyl group become equivalent. The substrate radical then undergoes a rearrangement before the $AdCH_3$ returns a hydrogen atom and the cobalt-carbon bond reforms.

The actual rearrangement of the substrate is the least understood part of the mechanism of these enzymes. Thioester analogues of methylmalonyl-CoA have been shown to undergo the 1,2 migration as in the methylmalonyl-CoA reaction when a substrate radical is generated (Halpern, 1985). This was done in the absence of cobalt compounds so organocobalt complexes are not necessary. Other model systems in which cobalt is present can also give the rearrangement but these may also be due to radical production. The 1,2 migration of the thioester group has been suggested to proceed via a cyclopropyloxy radical intermediate (the analogous 1,2 migration of vinyl groups via cyclopropylmethyl radicals is well established).

The migration of saturated groups (-OH, $-NH_2$ or $-CHNH_2COOH$) is not expected theoretically, and has not been shown in model compounds using radicals, but is facile via a carbonium ion intermediate. The cobalt-carbon cleavage to form $AdCH_2^+$ is very unfavourable, but the stability of the carbonium ion may counteract this. However, there is no evidence yet for this mechanism in these rearrangements.

As well as providing substrate specificity the role of the protein in methylmalonyl-CoA mutase appears to be to provide sufficient strain on the cobalt-carbon bond to promote radical formation and fast reaction rates. The enzyme probably also guides the correct rearrangement. It would be harmful for the cell if the free coenzyme could produce radicals at an appreciable rate as they could cause many undesirable reactions and rearrangements, so the enzyme contains the potentially dangerous radicals, preventing their release during the reaction. There is no evidence yet for direct involvement of enzyme side-chains in the catalysis but this cannot be ruled out. Cleland (1982) explains extremely high kinetic isotope effects on tritium transfer from coenzyme to product (300), by the suggestion that the radical, initiated by the coenzyme, is propagated through 8-10 reaction turnovers by a protein-

based radical before the cobalt-carbon bond is reformed. O'Brien et al. (1985) have also obtained evidence for a second, protein-based carrier of the transferred hydrogen in ethanolamine deaminase.

1.1.4 Reasons for cloning these enzymes

The three methylmalonyl-CoA utilising enzymes of the propionate pathway of P. shermanii have all been the subjects of a large amount of research. From a structural point of view most is known about the transcarboxylase, which is the largest (10 times the size of mutase) and most complex (a total of 30 polypeptides of three types). The sequence of the biotin containing subunit has allowed further progress, but to understand more completely the subunit interactions the sequences of the other two subunits are required. Some work has also been done on the structure of the smallest of the enzymes, epimerase. Here, again, the peptide sequence would help in the interpretation of the protein chemistry and suggest further experiments to elucidate the structure. Very little work has been done on the protein component of mutase, although there has been much interest in the coenzyme. The primary structure of the protein would provide a good basis for beginning to understand the role of the peptide in this enzyme.

The peptide sequences can help in the understanding of the protein's mode of action by allowing identification of particular amino acids or regions that can be labelled with specific reagents, or that can bind other subunits or specific antibodies. In particular, active site residues may be located using tools such as suicide inhibitors, photoactive substrate analogues or substrate protection against modification.

Peptide sequences can, in theory at least, be obtained directly from the protein. However, it is often more convenient to deduce the peptide sequence from the DNA sequence of the cloned gene. As well as giving the peptide sequence, the cloned genes may be deliberately overexpressed in a new host to obtain, not only large amounts of the enzyme, but also enzyme free of contamination by other enzymes of the pathway. This would facilitate structural and enzymological studies. Also, it may be possible to crystalise the enzyme, and the primary structure would help the interpretation of

crystallographic data.

A further benefit of having clones of these enzymes is that they could be delibarately mutated to observe the effects on catalysis of changes in the amino acid sequence. Base changes can be made in the sequence by priming DNA synthesis on the clone with an oligonucleotide containing a mismatch. Tyrosyl-tRNA synthetase is one enzyme that has been studied by this approach (Winter et al., 1982). One application of this technique would be to modify each of the cysteine residues in epimerase. Converting one or other of these to serine or glycine should allow identification of the one that is accessible to thiol-directed reagents. Levels of catalytic activity in various mutant enzymes would help to establish whether or not the thiols are involved in catalysis.

1.2 Polyketide antibiotics

Primary metabolism occurs in all microbes. In some, like Escherichia coli, this is tightly regulated and efficient. The production of energy and intermediates, which are then used in biosynthesis, are well controlled, so that no intermediates accumulate, even when the food supply is exhausted. In other microbes that possess genes for the production of secondary metabolites some of the steps are less well regulated and some intermediates accumulate. When the growth rate decreases due to lack of food these accumulating pools may be channelled into secondary metabolism. This may be partly to remove the accumulating metabolites as an alternative to tighter regulation, but many secondary metabolites may serve a useful function so the pools of intermediates are important rather than unwanted.

Antibiotics are a major class of secondary metabolites which may serve to protect a dwindling energy supply from other microbes or, indeed, to kill other microbes to obtain more food. Streptomycetes produce a huge number and variety of antibiotics, in fact the majority of those known. The reason for this can be understood from their growth habit. The bacteria grow as a mycelial mat on solid media. Their natural habitat is the soil. When the food supply runs out they produce aerial hyphae which form spore chains so that the strain can spread to a new site. The aerial hyphae are at some distance

from the food supply, which is nearly exhausted anyway, so the substrate mycelia are lysed releasing a rich pool of metabolites and macromolecules on which the aerial mycelia may feed. This rich food supply is likely to attract other soil microbes so the Streptomycetes produce antibiotics during this phase of differentiation to protect their energy supply for the production of spores.

1.2.1 Polyketide biosynthesis

The polyketide antibiotics are a large, diverse group consisting of classes such as the macrolides and polyenes. These very different types of chemical all have a major similarity. The basic structure in each case is composed of several acyl groups joined head to tail. These groups, which may be acetate, propionate, malonate, methylmalonate or butyrate, are polymerised in a manner analogous to the synthesis of fatty acids from acetate and malonate. Generally, however, each added monomer is not fully reduced, as they would be in fatty acid synthesis, so keto-, hydroxy and carbon-carbon double bond functional groups are left. In later steps extra methyl, hydroxy, glycosyl or epoxy groups may be introduced to produce the antibiotic.

The macrolide antibiotic erythromycin, for example, consists of an aglycone core (erythronolide) and two unusual sugars (desosamine and either mycarose or cladinose). The first-formed aglycone, 6-deoxyerythronolide b, is formed by the successive addition of six methylmalonyl groups onto a propionyl 'starter' unit, followed by cyclisation to give a 14-membered ring. Corcoran (1981) has reviewed the biosynthesis of erythromycin. The structure of 6-deoxyerythronilide b is shown in Figure 1.7.

The biosynthetic origin of all the carbon and oxygen atoms of 6-deoxyerythronolide b has been determined using isotopically labelled precusors. Friedman et al. (1964) used ^{14}C-labelled precursors to show that all the carbon atoms in erythronolide could be derived directly from propionate, but that methylmalonate was a preferred precursor for all but the terminal 3-carbon unit. Cane et al. (1981) also demonstrated that all the oxygen atoms in 6-deoxyerythronolide b were derived directly from propionate and methylmalonate, rather than by oxygenation of a fully reduced precursor,

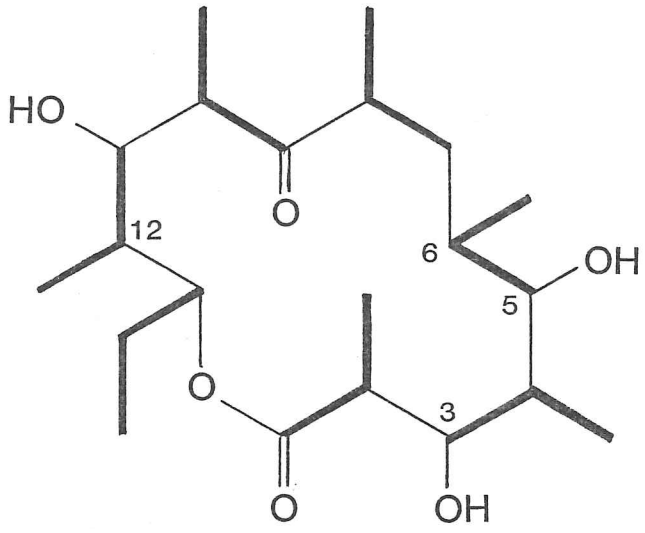


Figure 1.7 The structure of 6-deoxyerythronolide b

The seven three-carbon units which make up the lactone ring of 6-deoxyerythronolide b are shown by bold lines. The chain consists of a propionate unit onto which are condensed six methylmalonate units. After the lactone has been formed carbon 6 is hydroxylated. To form the family of erythromycins carbon twelve may also be hydroxylated and two glycosylations occur. Desosamine is added at position 5 and either mycarose or cladinose at position 3.

using substrates enriched with ^{13}C and ^{18}O and analysing the products by NMR. No intermediate between propionate and 6-deoxyerythronolide b has been isolated, and it has not been possible to subdivide the class of mutants blocked at this stage by cross-feeding experiments which would require an enzyme-free intermediate.

These results suggest an analogy between the putative erythronolide synthase and fatty acid synthase. The analogy is strengthened by the fact that cerulenin (1,2-epoxy-3-oxo-7,10-dodecadienamide), a potent site-specific inhibitor of the condensing enzyme of all types of fatty acid synthase (Omura, 1981), also inhibits erythronolide synthesis in resting cells of S. erythreus (Roberts, personal communication). Whether or not the aglycone synthase is the same enzyme as fatty acid synthase or a related, but different enzyme, is not clear. The macrolide synthase may be a simpler enzyme, because reductase and dehydrase enzyme activities may not be required. Neither of the enzymes has been purified to homogeneity from antibiotic producing organisms.

It is possible that one enzyme catalyses both reactions. The fatty acid synthase from the uropygial gland of the goose can either use acetate and malonate to form the normal, saturated fatty acids, or it can use acetate or propionate to initiate the chain with methylmalonate as chain-extending substrate to form multibranched fatty acids (Buckner and Kolattukudy, 1976). No mixed products have been observed so it was postulated that the first chain-extending unit locks the enzyme into making branched or unbranched fatty acids and that the control over which was made depended on the substrate supply. The fatty acids produced are fully saturated so this situation is not completely analogous to that in polyketide biosynthesis. Penicillium patulum contains two separable multi-enzyme complexes, one of which synthesises fatty acids and the other synthesises the polyketide-based 6-methylsalicylate (Scott et al., 1974). These two enzymes were of similar size but not identical.

The possibility exists that antibiotic synthesising enzymes are not synthesised completely de novo at the onset of antibiotic production. They could contain many of the same components as the normal fatty acid synthase, but with antibiotic specific subunits or modifications, such as

phosphorylation, which alter the substrate specificity and inhibit the dehydration and reduction activities. Ohlrogge and Kuo (1985) have found two forms of acyl carrier protein in plants. Both forms occurred in the leaves but only one in the seeds and roots.

The monomers for the synthesis of polyketides are assumed to be the corresponding coenzyme A derivatives as they are for fatty acid synthesis. The sources of these are not clear, but an acyl-CoA carboxylase has been purified from S. erythreus suggesting that the methylmalonate can be derived from propionate (Hunaiti and Kolattukudy, 1982). Propionate or propanol added to the medium stimulates the production of erythromycin. Radiolabelled propionate was incorporated intact showing that it was not metabolised via succinate. No epimerase activity has been detected but it is likely to be present since both isomers of methylmalonate are incorporated into erythronolide and Cane et al. (1981) showed that all the oxygen atoms of 6-deoxyerythronolide b could be derived directly from methylmalonate, indicating that isomerisation at some positions during, or after, polymerisation is unlikely. In fatty acid synthesis it is not yet clear whether all the malonate addition reactions proceed with the same stereochemistry. Only the first has been established but there is evidence to suggest that they are not all the same (Anderson and Hammes, 1984). Methylmalonyl-CoA mutase has also been purified from S. erythreus (Hunaiti and Kolattukudy, 1984a), and could provide an alternative source of methylmalonate from succinate. The activity of this enzyme was low but peaked at the onset of antibiotic production.

Other evidence that succinate can be a direct precursor of erythronolide comes from the incorporation of succinate, labelled with ^{14}C in the two carboxyl groups and with ^{3}H on the methylene groups. The $^{14}C/^{3}H$ ratio was halved in the erythromycin produced, suggesting that one carboxyl was lost in the condensation of methylmalonyl-CoA formed by direct isomerisation, rather than via extensive degradation when much of the tritium would be lost. Lancini and Grandi (1981) and Omura et al. (1976) have shown that succinate can be the source of methylmalonate for the production of two other macrolides (rifamycin and tylosin respectively) by ^{13}C-NMR.

Probably the source of methylmalonyl-CoA will depend on the growth conditions and supplied nutrients and may be one, or more, of succinate, propionate and branched-chain amino acids.

1.2.2 Regulation of antibiotic biosynthesis

Antibiotics are generally only produced by the microorganism when the food supply is exhausted and the growth rate slows down. Rapidly metabolised carbon, nitrogen and phosphate sources all reduce the amount of antibiotic produced. As an example, the production of tylosin (a macrolide antibiotic) by S. fradiae is affected by glucose levels. If glucose is added late in growth, when antibiotic is being produced, the production ceases until the glucose is consumed, when it resumes. The glucose level is reflected by the total level of adenylate nucleotides. At the onset of antibiotic production this level drops and when glucose is added it rises. The phosphate charge ratio on the adenylates, however, remains constant showing that this ratio does not act as a signal of the cell's energy status (Vu-Trong et al., 1980).

Extra phosphate added to the medium also increases the adenylate pool and stimulates growth and glucose uptake. However, if added late in the growth phase, after antibiotic production has started, it has little effect on antibiotic biosynthesis. It is not clear whether adenylate levels are related directly to antibiotic producton or are just a separate reflection of the growth rate. Methylmalonyl-CoA carboxyltransferase activity and some propionyl-CoA carboxylase activity were present in this strain. The enzyme activities peaked at the start of antibiotic production and the synthesis of both enzymes was repressed by added phosphate (Vu-Trong et al., 1981).

Rapidly metabolised carbon sources and high phosphate levels have similar effects on many other antibiotic producing strains (Omura and Tanaka, 1983). Rapidly metabolised nitrogen also has an effect similar to phosphate, repressing synthesis of enzymes rather than inhibiting those that are already formed (Flores and Sanchez, 1985). For the commercial production of antibiotics the growth media are optimised with sufficient levels of nitrogen and phophate for good growth, while also allowing high levels of production. Slowly metabolised nutrients may be used and in some cases the bacteria may

be transferred from a good growth medium to a good production medium when they stop growing.

As well as the control exerted by the medium there is important regulation over the amount of antibiotic production by the synthetic pathway itself. The levels of the precursor pools can have a large effect on the amount of flux through the pathway. Many strains developed for higher antibiotic yield have increased amounts of the precursors present, and feeding of precursors, such as propionate, can also increase yields (Martin, 1977). The relative levels of production of the related paulomycins A and B can be selectively altered by feeding precursors specific to one or other antibiotic (Marshall et al., 1984). Regulation of these pools can be by feedback inhibition or enzymes that remove the precursors. Either of these may be affected in higher yielding strains. Increased channelling of precursors into secondary rather than primary metabolism may also occur. For instance, a malonyl-CoA decarboxylase has been purified from S. erythreus. This enzyme will remove one of the substrates of fatty acid synthase causing a reduction in fatty acid synthesis while antibiotic production can proceed (Hunaiti and Kolattukudy, 1984b).

It is often forgotten while studying the effects of various liquid growth media that the natural habitat of Streptomycetes is the soil, and so they grow more naturally on solid media. In liquid media Streptomycetes do not differentiate properly, whereas on solid media spores are formed when the nutrients are exhausted, and antibiotic production is closely linked with differentiation. Many mutants defective in differentiation and the formation of spores also lose the ability to synthesize antibiotics, whereas the reverse is not true (Hopwood and Merrick, 1977). This finding suggests that there is genetic control of antibiotic production during differentiation and not just catabolite effects.

An autoregulating chemical, A-factor, has been found in several Streptomycetes. S. coelicolor strains defective in the synthesis of this compound do not sporulate or synthesise antibiotics unless A-factor is added to the medium (Hara and Beppu, 1982). However, an S. lividans mutant defective in A-factor synthesis was not defective in sporulation. The gene cloned by

complementing the loss of antibiotic production in this strain was shown to regulate the synthesis of both antibiotic and A-factor (Horinouchi et al., 1983).

Westpheling et al. (1985) have shown that there are at least two forms of RNA polymerase in S. coelicolor, and that these recognise different promoter sequences. RNA polymerases may, therefore, be involved in the regulation of gene expression as they are in Bacillus. The complex interrelationships of gene regulation during the Streptomycete life-cycle are not yet understood and may vary between strains. The situation is further confused by the incomplete differentiation in liquid cultures.

1.2.3 Modification of antibiotic producing strains

Many antibiotic producing strains have been improved to give higher yields by mutagenesis with UV light or chemicals followed by laborious screening. Any improved strain may have been modified in one of several ways: for example, increased levels of precursors or reduced feed-back inhibition. Each improved strain is then used as the basis for further mutagenesis. Other benefits that could be obtained by this route include the reduction or elimination of other, unwanted products which might be closely related compounds or just other chemicals that may interfere with the recovery processes.

Although levels of production of an antibiotic and relative levels of related compounds produced by the same organism can be altered by this mutagenic approach, the synthesis of novel compounds is not often found. For instance, only one mutant is known in which the chromophore of alternating double bonds in polyene macolide antibiotics has a different length (Martin, 1977).

One approach to obtaining novel antibiotics, which has been successful with penicillin, is the feeding of modified precursors. Addition of phenylacetic acid to cultures of Penicillium chrysogenum causes synthesis of penicillin G, an analogue of the natural penicillin N. This technique has not been used with the macrolides, but novel antibiotics can be produced by hybrid biosynthesis. The aglycone produced by one strain, blocked in the late

steps of synthesis, can be fed to another strain, blocked in aglycone formation by inhibition with cerulenin. It may then be glycosylated and otherwise modified by the late enzymes of the pathway in the second organism (Omura and Tanaka, 1983).

The development of gene cloning vectors in Streptomycetes now permits similar experiments to be performed without the need for the intermediates to be purified from one strain and then fed to another. Notable success was achieved by Malpartida and Hopwood (1984) who cloned the entire biosynthetic pathway for the polyketide antibiotic actinorhodin from S. coelicolor. This large gene cluster caused production of actinorhodin by S. parvulus, which is not a closely related species and did not produce a similar antibiotic before introduction of the cloned DNA. One of the genes on this cloned fragment encoded a hydroxylase. When this gene was introduced into Streptomyces sp. AM-7161, which produces the related antibiotic medermycin, a novel antibiotic was formed. This compound, mederrhodin A, was hydroxylated in the analagous position to the hydroxylation of actinorhodin by the cloned enzyme (Hopwood et al., 1985). Interestingly, when the entire actinorhodin pathway was cloned into Streptomyces sp. AM-7161 both actinorhodin and medermycin were produced, but there was no synthesis of hybrid antibiotics. This may be due to the hydroxylase being part of an actinorhodin-synthesising complex which prevents medermycin being modified by blocking access to the hydroxylase.

This result is encouraging for the production of novel secondary metabolites, but shows that the synthetic pathways are complex and poorly understood. Multienzyme complexes involved in antibiotic biosynthesis may channel intermediates between active sites without releasing them, and so may prevent many modifications by cloned enzymes. However, cloning of the entire actinorhodin pathway into the granaticin and dihydrogranaticin producer S. violaceoruber Tu22 caused production of a new antibiotic, dihydrogranatirhodin, as well as actinorhodin, with very little of the normal, host strain antibiotics. This new antibiotic is generated by a different part of the actinorhodin pathway to mederrhodin, and the production of a granaticin hybrid antibiotic with the complete actinorhodin pathway present may reflect a different substrate specificity or accessibility of the particular enzyme responsible.

Specific clones, such as the actinorhodin hydroxylase, may now be transferred into many other antibiotic producing strains in the search for novel products. However, it is also likely that, with access to facilities for rapidly screening thousands of strains, pharmaceutical companies may use less directed methods and shot-gun clone fragments of DNA from one organism into another in the hope of obtaining new compounds.

Clones of the three methylmalonyl-CoA utilising enzymes of P. shermanii could also be used to improve antibiotic producing strains. For some macrolides the major source of methylmalonate appears to be propionate. However, insertion of methylmalonyl-CoA mutase and epimerase, coupled to feeding with succinate should increase the methylmalonate precursor pool level. Also, transcarboxylase and epimerase together should increase the flux from propionate to methylmalonate. It should be possible to construct gene cartridges of these enzymes which can then be inserted into suitable vectors for any strain. It might prove necessary to replace the P. shermanii promoter with another to allow coordinate regulation of the genes with those for antibiotic production.

Another use of such clones would be the provision of high levels of methylmalonate in strains producing antibiotics based on acetate rather than propionate units and which normally contain little or no methylmalonate. In some cases this might cause methyl branches to be inserted into a metabolite. It should also be possible to use the cloned genes as probes for the genes in Streptomycetes to help study the synthesis of secondary metabolites. They could be used to follow the levels of mRNA for these enzymes under various conditions, and to discover the distribution of these enzymes amongst Streptomyces species. If there were sufficient homology in some cases, the chromosomal genes might be inactivated by inserting a modified, cloned gene into the chromosome by recombination. Removal of the epimerase and mutase from S. erythreus, for example, would allow only one isomer of methylmalonate to be formed. If erythronolide could still be made in this situation then it would show that the synthase itself can isomerise methylmalonate to give the product in which both isomers are present, even though both substrate isomers are incorporated if available.

1.3 Gene cloning with oligonucleotides

In order to clone any gene it is necessary to be able to select the desired clone from thousands of others. In some cases it is possible to obtain mutants defective in the gene and then complement this lesion by the insertion of cloned DNA into the organism. This approach requires that there be suitable vectors that can carry the gene and allow its expression in the mutant strain. No gene cloning has been done in *P. shermanii* and it is not known whether vectors from any other source can be used in this species. It might be possible to find a plasmid or phage in a *Propionibacterium* and develop a vector from this. Alternatively, vectors from other species could be tested in *P. shermanii*. This type of approach would also necessitate the development of a transformation protocol to get the vector into the cells. Mutants in each gene would be required, which should be deletions to avoid recombination with the cloned DNA, along with the ability to screen for the regaining of each gene. All this would be very time consuming, and carry no guarantee of success.

Therefore, rather than cloning the genes in the organism from which they derive it is better to clone into a heterologous host. Again, a selection or screening procedure is needed. This could be complementation of a deficiency in the host or an assay for the new enzyme activity. Of more general application would be screening with antibodies to the protein of interest to detect its presence in the new host. This can be done by immobilising the cell extracts on a solid support, such as nitrocellulose, and detecting the presence of a particular antigen with radioactively labelled antibody, or with a second antibody or Protein-A that has been labelled with radioactivity (Clarke *et al.*, 1979; Ehrlich *et al.*, 1979) or coupled to an enzyme catalysing a chromogenic reaction (Kaplan *et al.*, 1983). These methods require that the cloned gene be expressed in the new host, but, because no genes had been cloned from *P. shermanii*, it was not known which host strains could recognise *P. shermanii* promoters. *Streptomyces* vectors have been developed over several years and these bacteria are now being found to be good hosts for cloning and expressing genes from many organisms (Bibb and Cohen, 1982; Bibb *et al.*, 1983). *P. shermanii* is also Gram positive and has DNA with a high G+C content, similar to the Streptomycetes, suggesting that an antibody screening method

might be successful using Streptomycete hosts.

Although an antibody screen of a library of P. shermanii DNA cloned into a Streptomycete might be successful, there might be no expression of the cloned DNA causing the approach to fail. A screening method that detects the DNA required, rather than the product of expression of this DNA, avoids any possible problems with expression. If expression is not obtained the DNA can be transferred to other hosts, or various promoters from other genes can be spliced in front of the gene.

Oligonucleotides provide such a means of detecting the DNA of interest. If some of the peptide sequence of the protein is known, then the DNA sequence coding for it can be deduced. The oligonucleotide is a short piece of DNA and will form a duplex with a complementary piece of single-stranded DNA, so, if labelled, the oligonucleotide can be used as a probe for the complementary DNA sequence amongst the denatured DNA of recombinant clones.

1.3.1 Oligonucleotides as probes

Oligonucleotides, strictly oligodeoxyribonucleotides, have been used as probes in many ways (see Itakura et al., 1984 for a recent review). These various approaches have been different ways of overcoming the degeneracy of the genetic code. This degeneracy means that there are several possible coding sequences for every peptide sequence, so there are many possible oligonucleotides, only one of which will be the correct probe for that gene.

Kinnaird et al. (1982) were able to clone the glutamate dehydrogenase of Neurospora crassa using a single probe, because comparison of the amino acid sequences of the wild-type protein and one with two compensating frame-shift mutations allowed the deduction of an unambiguous DNA coding sequence. This sort of analysis requires good genetic knowledge of the organism from which a gene is to be cloned.

A more general method involves the synthesis of a mixture of all possible sequences. Wallace et al., (1979) showed that a single mismatch in an oligonucleotide had a significant effect on the stability of the hybrid formed with the complementary DNA, so that by using suitable hybridisation

conditions only the correctly matched oligonucleotide should bind. However, the strength of G-C base pairs is stronger than that of A-T base pairs, adding approx. 4°C to the T_m, rather than 2°C, for each pair (McFarland et al., 1982). This means that if there are several ambiguities in a sequence there will be a wide range of T_m, and a hybridisation temperature that allows the most A+T rich oligonucleotide to bind will not be stringent for the rest, and so will allow more mismatched hybridisation. Wallace et al. (1981) obtained specific binding with a mixture of eight 14-mers, but the correctly binding sequence was the most G+C rich. One of the sequences with one mismatched base was made independently and hybridised at 8% of the level of the independently synthesised correct oligonucleotide. The mixture of eight oligonucleotides only hybridised at 90% of the level of the correct sequence, perhaps due to competition for the binding sites by the mismatched sequences which are then lost during the washing step before autoradiography.

Carroll and Porter (1983) used a mixture of 16 14-mers to clone a human complement C4 gene. To reduce the background they washed the filters at a temperature above the T_m of the most A+T rich probe but still obtained two different clones. These both contained the same sequence as the most G+C rich probe, but only one encoded the remainder of the known peptide sequence. A longer probe should be used to decrease the probability of the sequence occurring by chance in the DNA being probed, but this increases the number of degenerate positions. Bell et al. (1984) successfully used a mixture of 256 23-mers. The genes for two proteins with an identical peptide sequence were the targets, but, again, only half of the clones obtained were those sought and at least one of the correct sequences had one of the highest hybridisation temperatures. Cohn et al. (1983) needed only eight 17-mers to cover all possibilities for part of the luciferase gene from Vibrio harveyi. In this case only the correct gene was cloned and the hybridising oligonucleotide did not have the highest T_m.

In specific cases the hybridisation conditions can be adjusted to allow absolute selection between the correct oligonucleotide and one with one mismatch. This normally requires the mismatch to be in the centre of the probe. Two oligonucleotides differing at only one position can be used to show whether a patient is homozygous or heterozygous for the sickle cell, β^S-

globin, allele (Conner et al., 1983). In this case it was known that each oligonucleotide would hybridise stringently to one allele so the conditions could be adjusted for maximum discrimination. When using a mixture of oligonucleotides to clone a gene whose sequence is unknown the conditions cannot be so optimised. Raising the temperature may remove non-specific binding and reduce background, but it may also prevent the correct probe from binding. It is not possible to accurately define a T_m for each oligonucleotide without measuring the amount of binding at various temperatures because it is dependent on the sequence as well as the composition of the oligonucleotide. G+C rich oligonucleotides, in particular, tend to deviate from the usual rule of thumb based on length and composition.

Recently a modified hybridisation potocol has made the T_m of each oligonucleotide in a mixture the same, dependent only on the length. This is done by using tetramethylammonium chloride in place of sodium chloride in the buffer. The cation binds to double-stranded DNA differently depending on whether there is a G-C or A-T base pair present. At a suitable concentration this has the effect of making both types of base pair equally strong (Wood et al., 1985).

This procedure should increase the chance of finding the correct clone rather than mismatched sequences. However, with a mixture of probes it is not possible to know whether all the sequences are actually present. Ike et al. (1983) showed that in the phophotriester method of synthesis purines react at only 70% of the rate of pyrimidines, so that pyrimidine rich sequences are favoured. This can be avoided to some extent by using mixtures of dimers or trimers as building blocks. If the first base is the same, but the second or third different, the rates of reaction are similar. Differential reaction rates are not observed using phosphite-triester chemistry (Elmblad et al., 1982). The number of different sequences in a mixture can be reduced by introducing inosine at some degenerate positions. This forms similar strength base pairs with A, C and T bases and so reduces the degeneracy at that position (Ohtsuka et al., 1985).

The use of inosine or tetramethylammonium chloride may now make the use of mixtures more reliable, but, at the start of this project, although several

mixtures had been successfully used to clone genes (for example Singer-Sam et al. (1983) used 32 16-mers), they were not ideal.

An alternative to making all possible mixtures is to make the single most likely sequence. This can be done by taking account of codon usage data, if any is available, for the organism from which the gene is to be cloned. Obviously, however, there is only a small chance that the guessed sequence is actually correct. To increase the specificity of the probe sufficiently for it to be useful a long probe has to be made. Probes of length 50-100 bases can be used if the overall homology is greater than approx 70% (Lathe, 1985). Anderson and Kingston (1983) used an 86 base probe to clone bovine pancreatic trypsin inhibitor. They found that mixtures of probes were too non-specific, so they made a long probe instead. This was done by ligating 10 oligonucleotides together to form an 86bp double-stranded piece of DNA. This was ligated into M13mp8 and the radioactive probe was then prepared from this. The probe was later found to be 74% homologous to the cloned gene. Jaye et al. (1983) succesfully used a 52 base probe that was synthesised as two 26-mers and then ligated together using a 13-mer to hold the oligonucleotides correctly end-to-end. Long oligonucleotides are now even more practical because automatic oligonucleotide synthesisers can now make sequences of 100 bases in good yield. However, this type of approach demands that a sufficiently long peptide sequence be known (20-30 amino acids) and, therefore, requires that more work may have to be done on the protein before it can be cloned than if shorter oligonucleotides are used.

1.3.2 Synthesis of oligonucleotides

Oligonucleotides have only become widely used as accurate probes for specific DNA sequences over the past six years. Organic chemists had developed much of the chemistry prior to this, but when recombinant DNA techniques became available biochemists found a practical use for them. To make sufficiently long probes the nucleotide synthetic chemistry was modified to enable solid-phase synthesis using the techniques previously developed for peptide synthesis.

Solid-phase synthesis has several advantages over solution chemistry. The product does not have to be purified at each stage so losses are minimised. Reagents can be washed away after each step and so may be used in large excess, decreasing the reaction times required. The synthesis only involves addition of reagent solutions and washing, requiring only simple apparatus and allowing automation. However, reagents must be purer to maintain high yields because of the excesses used. These are now available commercially and over the past four years have become purer and more reliable, although the purity must still be checked and some reagents further purified if necessary.

The solid support must be non-compressible to avoid restricting the solvent flow and should not affect the reactions. The first good support was a composite of polydimethylacrylamide and Kieselguhr (Gait et al., 1982a and b). Ester groups are incorporated into the support by adding acryloylsarcosine to the polymerisation reaction which forms the polydimethylacrylamide in situ in pores in the Kieselguhr. To these are added a spacer arm consisting of ethylene diamine and two glycine residues. The first nucleoside is then coupled to this through a succinate linkage. The long spacer is required to hold the growing chain away from the resin which tends to retard the coupling reaction. This resin has the highest loading capacity, and so it is used for synthesising large amounts of oligonucleotide for crystallographic and NMR studies, but for smaller quantities, such as are required for cloning, other supports are now favoured.

Long chain alkylamine controlled pore glass has now superseded the Kieselguhr resin because, as the name suggests, the pore size is more constant, giving more consistent reaction and flow characteristics. The support also has less effect on the chemistry (Sproat and Bannwarth, 1983). Long oligonucleotides are generally synthesised on this support. The succinate derivative of the first nucleoside is again coupled to a long spacer.

Both of these supports give quite large amounts of oligonucleotide (100µg or more) which is far more than is necessary for cloning a gene (only approx. 10ng are required for each hybridisation). Also, only a few oligonucleotides can be made at once (as many as there are reaction columns). For these reasons

the use of paper discs as supports is attractive. Using four columns, one for each nucleotide, and sorting the discs into the relevent column before each round of the synthesis, over 100 oligonucleotides may be made simultaneously using only as much of each reagent as would normally be required to make four oligonucleotides (Matthes et al., 1984). The nucleoside succinate is coupled directly to the paper rather than through a linker arm, so there is more inhibition due to the support, and the bond is an ester which forms less efficiently than the amide on the other supports. Despite this 1-2μg of a 17-mer can be formed on a 5mm disc of Whatman 1 filter paper. This is enough for at least 100 hybridisations. The apparatus used in the synthesis is shown schematically in Figure 1.8.

Three different types of chemistry are used in oligonucleotide synthesis: phosphotriester, phosphite-triester and phosphoramidite. These latter methods have faster reactions but involve more steps. Shorter cycle times and higher coupling yields make them more suitable than the phosphotriester method for automated synthesis. The simpler phosphotriester method is the most practical for manual synthesis. In each chemistry improved catalysts and solvents and purer reagents allow synthesis of long oligonucleotides from monomers, whereas until recently dimers, trimers and longer units had to be used to obtain long oligonucleotides in good yield (Gait et al., 1982b).

An improved catalyst, N-methylimidazole, reduces the coupling time for the phosphotriester method to 15min instead of the original 45min (Efimov et al., 1983). When synthesisng many oligonucleotides by the paper disc method the longest part of the cycle is sorting the discs, but over 100 oligonucleotides can be made in two days, which was the time required to make one 17-mer previously.

The principle of the elongation reaction is that all reactive groups on the incoming nucleotide and the growing chain are protected except for the 5' hydroxyl of the growing chain and the phosphate group on the nucleotide. A dehydrating agent, MSNT, promotes the condensation reaction. After each coupling, excess reagents are washed away and then the new 5'-hydroxyl group is exposed by removing the acid-labile protecting group, dimethoxytrityl. After removal of the acid another monomer can be added. One coupling reaction

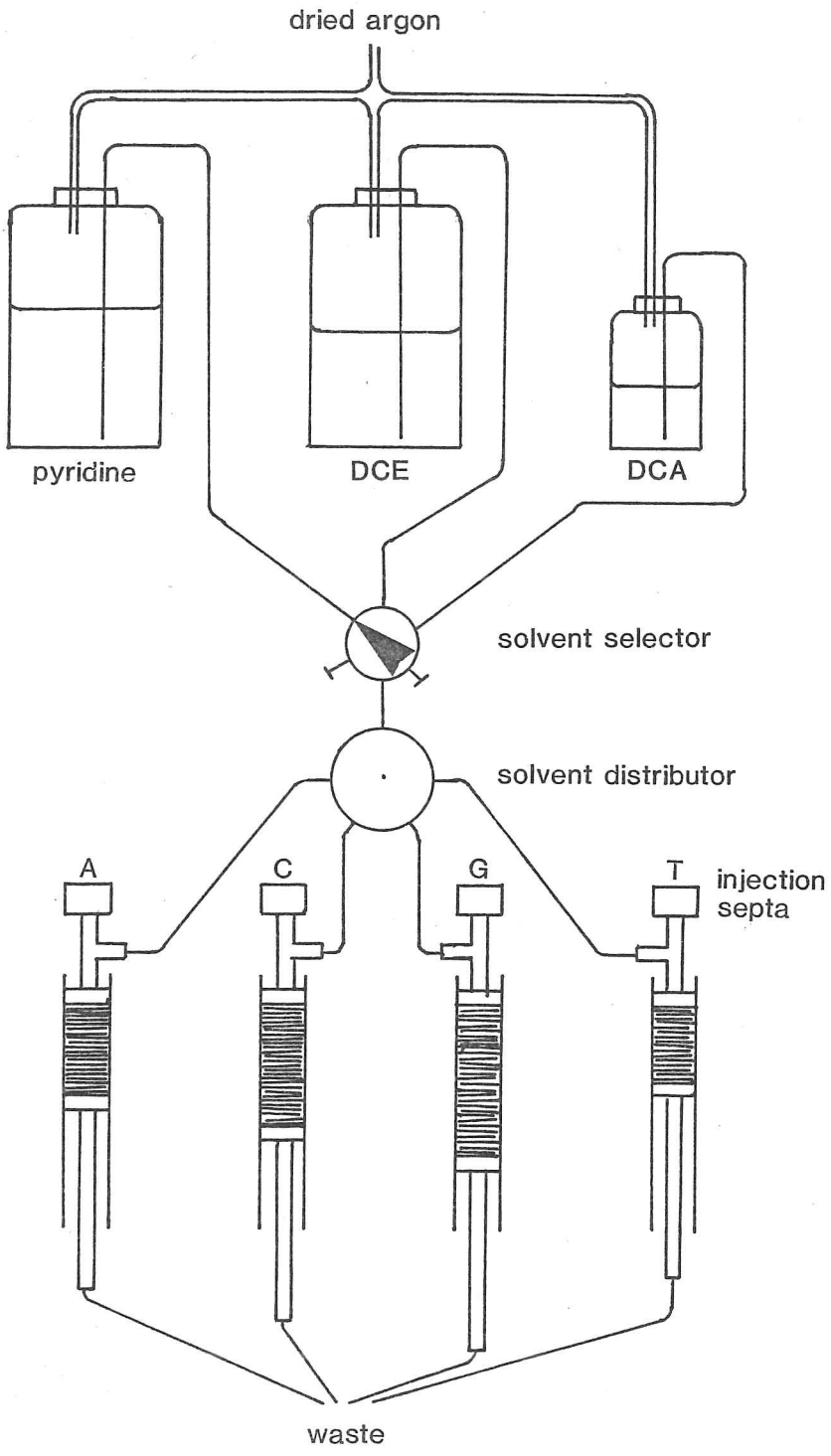


Figure 1.8 The apparatus used for synthesising oligonucleotides

The apparatus shown is that used for synthesising many oligonucleotides simultaneously by the paper disc method. Between each coupling the discs are sorted into the relevent column for the next round. After deprotection the solvent flow is stopped and a nucleotide solution injected into each column. The reagents are then washed away and the discs resorted.

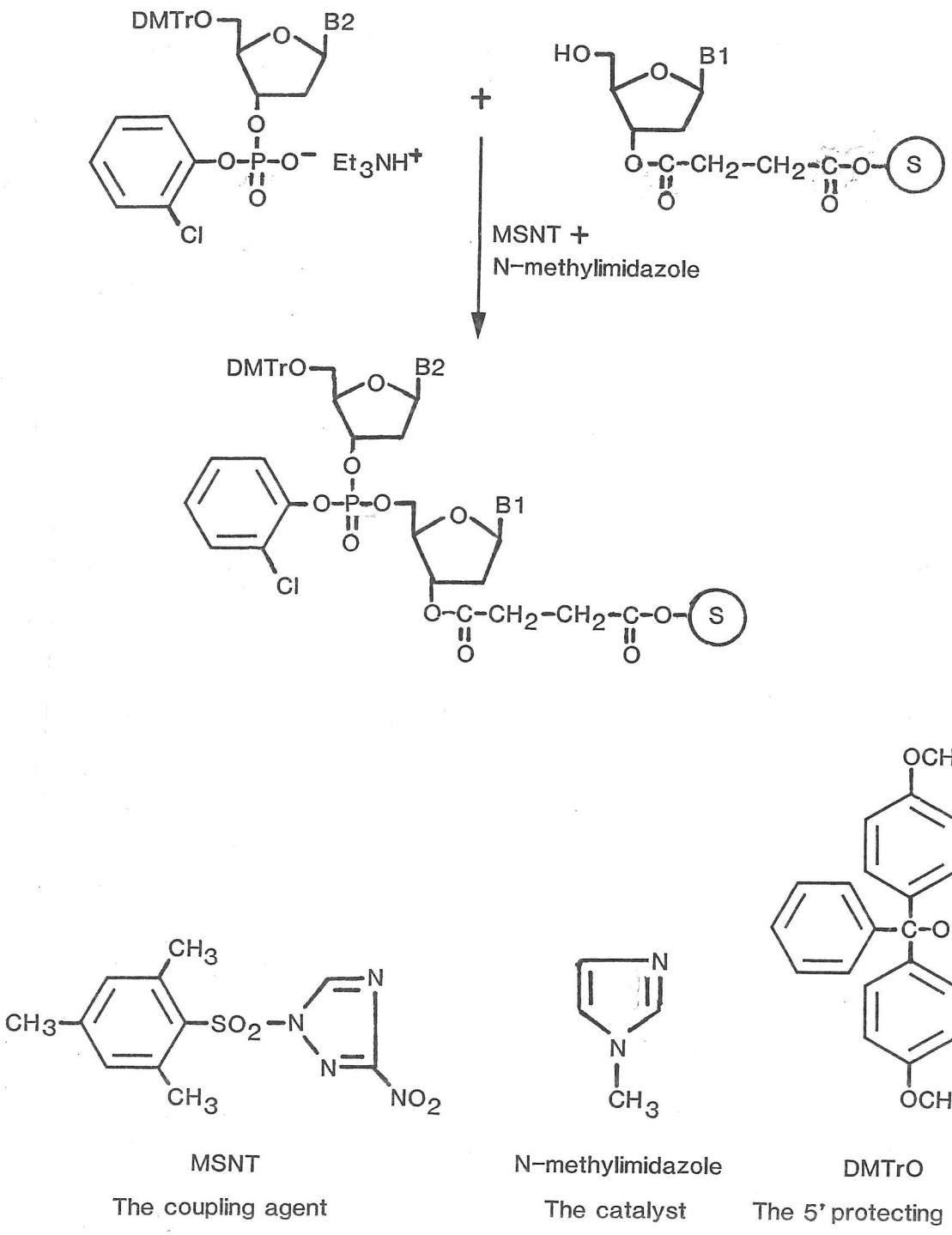


Figure 1.9 One nucleotide coupling reaction

The first nucleotide coupling reaction is shown. The fully protected nucleotide is coupled to the 5'-deprotected chain-initiating nucleoside, which is attached to the support. The ester linkage shown is for coupling to a paper support, amide linkages are used on other supports.

is shown in Figure 1.9.

In the early methods a capping step was included after each coupling reaction. This blocked any unreacted 5'-hydroxyl groups preventing chains with deletions from being formed. After the synthesis all the protecting groups except the 5'-dimethoxytrityl group were removed and the oligonucleotides cleaved from the support. Only the correct sequence should have the 5'-dimethoxytrityl which is a highly hydrophobic group and retards the oligonucleotide on an HPLC reverse-phase column. The final protecting group could then be removed. However, capping is no longer necessary, except after coupling the nucleotide succinate to the support, because the internucleotide coupling reactions are very efficient. The products of the synthesis are removed from the resin and fully deprotected. The desired sequence is the longest of those present in the mixture and may be purified by HPLC ion-exchange chromatography using a salt gradient. A quicker method for purifying many oligonucleotides is by gel-electrophoresis in the presence of urea (Matthes et al., 1984).

The finished oligonucleotide has no 5'-phosphate group and may be radioactively labelled using γ-[^{32}P]-ATP and T4 polynucleotide kinase (Maniatis et al., 1982). Recently biotin has been incorporated into oligonucleotides chemically and enzymatically (see e.g. Chollet and Kawashima, 1985; Kempe et al., 1985). The binding of the oligonucleotide is not impaired and the hybrid can be detected using streptavidin coupled to horseradish peroxidase or alkaline phosphatase and a chromogenic substrate. Biotinylated oligonucleotides do not yet have the sensitivity of radioactively labelled probes but, if this can be achieved they will be much more convenient to use because the labelled oligonucleotides could be stored indefinitely. The two-week half-life of ^{32}P and a high level of autoradiolysis make it impossible to store the labelled oligonucleotide, and the products of the breakdown of [^{32}P]-ATP inhibit the kinase reaction so experiments requiring high-sensitivity must be done with fresh label. There can be variations between kinase reactions so a stable labelled oligonucleotide would provide greater consistency as well as convenience.

1.4 Aims of this work

At the beginning of this work the three enzymes of the propionate pathway of *P. shermanii* that use methylmalonyl-CoA had been the study of much mechanistic work. The protein structure of transcarboxylase had also been studied in detail, but the only primary structure known was that of the smallest subunit. None of the peptide sequences of the other proteins were known. The first aim of this project was to clone and sequence one, or more, of these enzymes to aid the interpretation of, and provide the basis for, further mechanistic experiments.

Some early experiments were aimed at cloning transcarboxylase using the known sequence of one subunit to predict oligonucleotides that would hybridise to the gene. None of the oligonucleotides synthesised hybridised to *P. shermanii* DNA. The reason for this is not clear because the gene has now been cloned and the sequence is as expected (Murtif *et al.*, 1985). These workers used oligonucleotides and also had difficulties. They have only cloned the small subunit so far.

This project was concentrated on the epimerase and mutase genes. The first objective was to obtain some peptide sequences from which oligonucleotide probes could be deduced. These would be used to clone the genes which could then be sequenced. The clones would also enable overexpression of the enzymes, and alteration of specific amino acids by *in vitro* mutagenesis.

During the project it became apparent that, because methylmalonyl-CoA is a precursor for many antibiotics, the cloned genes might also have a use in altering antibiotic producing organisms. For this reason the second major aim of this project was to clone at least one of the genes into a Streptomycete to show whether active enzyme could be expressed. Later, the genes could be subcloned into various Streptomycetes in different combinations to observe the effect on secondary metabolite production.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

Water was deionised and then purified by a Milli-Q apparatus (Millipore S.A., Molsheim, France).

Most chemicals were Analar grade, otherwise the best grade commercially available. Methanol and acetic acid for gel fixing were General Purpose grade. SPS grade chemicals and triple distilled pyridine, used for oligonucleotide synthesis, were from Cruachem Chemical Co. Ltd., Livingstone, Scotland. Pierce and Warriner (UK) Ltd., Chester, Cheshire, supplied Sequanal grade trifluoroacetic acid, pyridine and PITC, used for DABITC sequencing. Butylacetate was Analar grade, distilled from DABITC. DMSO for transformations was glass-distilled and stored in small aliquots at -70°C. ATP (100mM and 10mM) and DTT (100mM) were also stored in aliquots at -70°C. Once thawed each aliquot was discarded after use. Urea and agarose were Ultrapure grade from BDH, Poole, Dorset. Sea Plaque, low melting agarose, was from FMC Corporation, Marine Colloids Division, Rockland, Maine, USA. Culture media were obtained from Difco Laboratories, Detroit, Michigan, USA and agar was Oxoid No.1 from Oxoid Ltd., Basingstoke, Hants. Thiostrepton was from E.R. Squibb & Sons Inc., Princeton, New Jersey, USA, and ampicillin was from Beecham Research Laboratories, Brentford, Middlesex.

2.1.2 Radiochemicals

The following radiolabelled chemicals were obtained from Amersham International (Amersham, Bucks):

Adenosine 5'-(γ-[^{32}P]-)triphosphate, >5 000Ci mmol^{-1},

Adenosine 5'-(γ-[^{32}P]-)triphosphate, ~3 000Ci mmol^{-1},

Deoxyadenosine 5'-(α-[^{35}S]-thio)triphosphate, 650Ci mmol^{-1},

[^{14}C]-iodoacetic acid, 50mCi mmol^{-1}, and

[^{35}S]-methionine, >800Ci mmol^{-1}.

2.1.3 Enzymes

Proteases were obtained from Sigma Cemical Co. (Poole, Dorset) except Endoproteinase Arg-C which was from BCL-The Boehringer Corporation (London) Ltd., Lewes, East Sussex. Restriction enzymes and T4 DNA ligase were from Amersham, NBL (Cramlington, Northumberland), and Pharmacia Ltd. (Milton Keynes). Staphylococcal nuclease and Bal31 nuclease were from BCL, T4 DNA ligase and the Klenow enzyme, the large fragment of DNA polymerase I, were from Dr. T. Hunt, Biochemistry Department, University of Cambridge.

2.1.4 DNA

Phage lambda DNA and HindIII digested λDNA came from Bethesda Research Laboratories, Gibco Ltd., Paisley, Scotland. Plasmid pUC12 DNA was the gift of S.M. Ward. Plasmid pUC13 DNA, EcoRI cut and dephosphorylated, and BamHI cut and dephosphorylated were from Pharmacia. M13mp10 DNA, SmaI cut and dephosphorylated, was from Amersham.

2.1.5 Strains and media

P. shermanii strain N.C.I.B. 9885 was obtained from the National Collection of Industrial Bacteria, Aberdeen, Scotland. It was grown in defined medium (Brown *et al.*, 1973) at 30°C. The pH was maintained at pH6.5 by addition of ammonia solution (sp.gr. 0.880).

E. coli strain TG1 (K12, Δ(*lac-pro*), *supE*, *thi*, *hsdD5* / F' *traD36*, *proA*$^{+}$*B*$^{+}$, *lacI*^{q}, *lacZ* ΔM15) was maintained on minimal medium lacking proline to ensure retention of the F-factor. It was grown at 37°C. Liquid cultures were grown in 2xTY (2xTY: 10g tryptone, 10g yeast extract and 5g NaCl in 1l). Plates contained 1.5% agar in the same medium. Ampicillin was added to 100μg ml^{-1} when selecting for cells transformed with recombinant pUC12 or pUC13 and to maintain these plasmids both on plates and in liquid culture. For transformation SOB and SOC media were as described by Hanahan (1983).

S. lividans strains TK24 (str-6, harbouring no known plasmids) and TK24/pIJ702 were obtained from Professor D.A. Hopwood, John Innes Institute, Norwich. The liquid medium, YEME, and plate medium, R5, were as described by

Hopwood et al. (1985). All cultures were grown at 30°C and liquid cultures were given as much aeration as possible by using only 200ml in a 2l flask or using a fermentor with antifoam (polypropylene glycol 1025) added and air blown through at 10l min^{-1}. Thiostrepton was added at 5µg ml^{-1} for liquid cultures or 50µg ml^{-1} for plate cultures of pIJ702 and its derivatives.

2.1.6 FPLC and HPLC.

A Pharmacia FPLC apparatus (with GP-250 gradient programme and P-500 pumps) was used with HR5/5 Mono-Q (anion exchange) and Mono-S (cation exchange) columns for protein purification. A BRL NACS-20 column (3.4ml) was used for nucleic acid purifications.

The HPLC used was a Varian 5000 instrument (Walton on Thames, Surrey) fitted with a Valco inlet and a Kratos SF769 UV detector (Schoeffel Instruments). A 5µ C_{18} Micropak MCH-5 (N cap) reverse-phase column (36cm x 4mm) was from Varian and a Chrompak Lichrosorb 10 RP8 reverse-phase column (25cm x 4.6mm) was from Waters.

2.1.7 Miscellaneous

DE81 filters, Whatman 1 and 3MM paper, 541 filters, GF/C filters, DE52 and Sephadex G75-superfine were from Whatman Biochemicals (Maidstone, Kent). Nitrocellulose BA85 sheets were from Schleicher and Schull, distributed by Andermann, Kingston upon Thames, Surrey. Protein-A sepharose was from Sigma.

Fuji RX film was used for autoradiography.

2.2 Oligonucleotides

2.2.1 Synthesis

Oligonucleotides were synthesised by the phosphotriester method (Sproat and Gait, 1984) on Kieselguhr-dimethylmethacrylate resin, long-chain alkyl controlled-pore-glass, or Whatman 1 paper discs as solid supports.

The sequencing primer and the mixture of 32 probes for epimerase were synthesised on a resin support (Gait et al., 1982a) and the single epimerase probe was made on a controlled pore glass support (Sproat and Bannwarth, 1983). These were purified by gel filtration and gel electrophoresis. Many oligonucleotides were synthesised simultaneously using the method of Matthes et al. (1984). These were purified by ethanol precipitation and gel electrophoresis. All oligonucleotides were stored in H_2O at -20°C.

Before synthesis the nucleotide succinates were checked for purity by TLC, the melting point of the MSNT checked and the pyridine tested for the presence of amines as described by Sproat and Gait (1984).

The primer was further purified on a NACS-20 FPLC column using a gradient of 0.4-0.6M NaCl in 12mM NaOH pH11.8. The fractions were neutralised with 1M Tris-HCl pH7.5 and three volumes of ethanol (96%) added to precipitate the oligonucleotide. Primer was also made by the paper disc method. This was suitable for sequencing without extra purification.

2.2.2 Labelling

Oligonucleotides were radiolabelled using T4 DNA kinase and $\gamma[^{32}P]$ATP. For probing chromosomal digests and screening colonies the highest specific activity $\gamma[^{32}P]$ATP available was used (>5 000Ci $mmol^{-1}$). For some plasmid mapping gels a lower specific activity was used (3 000Ci $mmol^{-1}$). Normally 10ng of oligonucleotide (20µg giving an absorbance of 1 at 260nm), which is approx. 2pmol of 17-mer, was labelled with 2pmol of $\gamma[^{32}P]$ATP, generally 1µl of the high specific activity solution. The reaction was performed in 50µl kinase buffer/4mM DTT with 8-12U T4 DNA kinase (10xkinase buffer: 0.5M Tris-HCl, 1mM EDTA, 0.1mM spermidine, 0.1mM $MgCl_2$, pH7.6).

To label more oligonucleotide the total volume was increased to 100µl and more label and enzyme were used. When labelling many oligonucleotides for multiple probing of small slices of nitrocellulose or dried gel a mix was made of all the components of the reaction except the probe. This contained 0.2µl ATP and 0.5µl kinase for each reaction and 20µl of this was added to 2ng of each oligonucleotide.

All reactions were incubated at 37°C for 1hr. Oligonucleotide was separated from unincorporated label using a DE52 column. A 0.5ml bed volume column in a pasteur pipette plugged with siliconised glass wool was equilibrated in TE. The unincorporated label was removed by 4-5 1ml washes with TE/0.2M NaCl until very few counts were eluting from the column. The oligonucleotide was then eluted with 4-5 0.5ml washes with TE/0.5M NaCl. The incorporation of label into the oligonucleotide was generally 40-70%. The reactions were stopped by boiling for 5min when the oligonucleotide was not to be purified. The column procedure gave a vast reduction in the background and non-specific hotspots obtained in hybridisations.

2.3 DNA techniques

2.3.1 General

DNA was stored in TE (10mM Tris-HCl/0.1mM EDTA pH8.0) at -20°C. Ethanol precipitations were performed by adding 1/10 volume 3M sodium acetate pH6.5 and 2.5-3 volumes of 96% ethanol and incubating at -20°C overnight. The precipitate was collected by 10min centrifugation at 13 000g in an Eppendorf microfuge.

Phenol (Analar) was equilibrated with TE containing 0.1% (w/v) 8-hydroxyquinoline and was stored at 4°C. Chloroform contained 4% isoamyl-alcohol. For extraction with either of these, or a 50:50 mix, an equal volume was added to the DNA solution and vortexed. The phases were separated by 2min centrifugation in an Eppendorf microfuge.

Restriction digests were performed as described in Maniatis *et al.* (1982) using a 3-10 fold excess of enzyme and/or time. DTT was added to 5mM for overnight digestions only.

Ligations were also as described in Maniatis *et al.* (1982). Reaction volumes were 10µl. ATP (10mM) and DTT (100mM) were stored in small (100µl) aliquots at -70°C which were discarded after use and not refrozen. The ligation of the 10.4kb EcoRI fragment isolated from LGT agarose into pUC12 was done with 30ng plasmid and 0.5, 1 or 2µl of insert DNA. The ligations of

various DNA size fractions into pUC13 were done using 1µl of each insert and 100ng of vector.

2.3.2 Purification

P. shermanii is resistant to lysozyme treatment, because the cell wall contains directly cross-linked peptidoglycan with meso-(RS)-diaminopimelate at position 3 (Schleifer et al., 1968), so 2g of cells were ground to a powder under liquid nitrogen with a pestle and mortar. DNA was then prepared as described by Maniatis et al. (1982).

Plasmid DNA was prepared on both a small and a large scale by the method of Kieser (1984). The large scale preparation of pND1 was given the extra phenol extraction to remove single stranded DNA. More RNA was also removed by precipitation with a final concentration of 2.5M ammonium acetate at 0°C for 20min. The precipitate was discarded and 100U RNAse-T_1 added to the supernatant. After 15min at 37°C the DNA was precipitated with ethanol. From a 500ml culture of pND1/TG1 grown overnight at 37°C 4.4mg of plasmid DNA was obtained. One fifth of this was further purified on a NACS-20 column. The sample was loaded in 0.5M NaCl/10mM Tris-HCl/1mM EDTA, pH7.2, and eluted with a gradient up to 1M NaCl in the same buffer. This removed remaining RNA and any chromosomal DNA. The fractions were ethanol precipitated and the DNA resuspended in TE.

2.3.3 Gel isolation

DNA fragments were isolated from gels by using low gelling temperature (LGT) agarose or by binding to DE81 paper.

For the isolation of the 10.4kb EcoRI fragment carrying the epimerase gene a 0.7% gel was cast using Sea Plaque LGT-agarose. P. shermanii DNA (80µg) digested with EcoRI was electrophoresed and then visualised with minimum exposure on a 300nm transilluminator. A slice of gel containing DNA fragments of 8-15kb was cut out (approx. 1ml volume) and 200µl 3M sodium acetate, pH6.5 added. After 30min the agarose was melted at 70°C and then added to 5ml TE + 0.5ml 3M sodium acetate. When this had cooled to room temperature it was vortexed with 2ml phenol and then centrifuged at 5 000rpm for 5min in a

Sorvall SS34 rotor. The phenol was back extracted with 1ml TE. The aqueous phases were pooled and extracted with 5ml phenol/chloroform and then 5ml chloroform. The DNA was precipitated by addition of 20ml 96% ethanol and incubation overnight at -20°C. The pellet from centrifugation at 10 000rpm for 20min in the Sorvall SS34 rotor was rinsed with 96% ethanol then dissolved in 1ml TE. The DNA was then precipitated by addition of 100µl 100mM spermine and incubation on ice for 30min, followed by 10min centrifugation in an Eppendorf microfuge. The pellet was washed with 75% ethanol/25% 0.3M sodium acetate for 2min then redissolved in 50µl TE.

Most DNA fragments were purified by binding to DE81 paper as described by Dretzen et al. (1981). Chromosomal digests contained approx. 100µg DNA and were loaded in two 13mm wells. For preparation of the fragment to subclone into S. lividans 40µg pND1 was digested with BglII and run in two 13mm tracks and for preparation of the fragment to sequence approx. 50µg was digested with XhoI and also run in two 13mm tracks.

The gels were run and stained and the DNA visualised on a 300nm transilluminator. Slits were cut above and below the band of interest and pieces of DE81 paper that had been previously soaked for several hours in 2.5M NaCl, followed by 20min in the electrophoresis buffer, were inserted in these slits. The electrophoresis was then continued for 30-60min at 100-150V for the 15cm gels or 250V for the 20cm gels. It could be confirmed using the transilluminator that the DNA had all been run onto the paper. The paper inserted above the band of interest stopped larger fragments being collected and could now be discarded. Alternatively, for several size fractions from one digest of chromosomal DNA a slot was cut parallel to the digest and the gel rotated by 90° before continuing the electrophoresis. The positions of the standards were marked with Indian ink so that after running the DNA into this long slot containing pieces of DE81 paper the paper could be divided into known size fractions.

Fragments of less than approx. 8kb were eluted by incubating the paper (no more than 3cm²) with 400µl 1.5M NaCl/20mM Tris-HCl/1mM EDTA, pH7.5 at 37°C for 5hr. Breaking up the paper at the beginning of the incubation aided recovery. For larger fragments 1M NaCl/1.6mM arginine/1.6mM EDTA, pH11.5 was used. The

2.3kb XhoI fragment of pND1 and its sonicated pieces for sequencing were eluted with 2µl 5mg ml^{-1} tRNA added. This did not give any better recovery and may have been the source of some of the small clones that were not part of the main sequence so was not used in other experiments.

After elution the paper was removed by centrifugation or by filtering through siliconised glass wool. Ethidium bromide was removed by two extractions with 1ml water saturated n-butanol. The aqueous phase was extracted with phenol/chloroform followed by chloroform and then precipitated overnight at -20°C with 1ml 96% ethanol. The DNA was washed with ethanol, dried and dissolved in 20µl TE (30µl for the sonicated fragments for sequencing).

2.3.4 Bal31 nuclease digestion

P. shermanii DNA (50µg) was digested with EcoRI and then extracted with phenol, phenol/chloroform and chloroform. It was precipitated with ethanol and redissolved in 40µl TE. Then 10µl 1mg ml^{-1} bovine serum albumen and 50µl 2xBal31 buffer were added. (2xBal31 buffer: 40mM Tris-HCl, 0.4M NaCl, 24mM $CaCl_2$, 24mM $MgCl_2$, 2mM EDTA, pH8). Bal31 (5U) was added and the reaction incubated at 37°C. Samples (5µl) were taken at intervals over 1hr and 5µl 40mM EGTA added to stop the reaction The samples were then run on an agarose gel.

2.3.5 Transformation

E. coli was transformed with plasmid DNA and transfected with M13 DNA by the high-efficiency method of Hanahan (1983). Up to $2x10^7$ transformants were obtained per µg of uncut pUC12 DNA.

S. lividans protoplasts were prepared and transformed with plasmid DNA by a rapid, small scale method (Hopwood et al., 1985).

2.3.6 Sequencing

DNA sequencing was done by the Sanger dideoxy chain termination method using single-stranded phage templates (Sanger et al., 1977, 1980). Shot-gun clones were prepared from sonicated DNA, single stranded template prepared

from these, and sequencing reactions performed as described by Bankier and Barrell (1983).

Sequence data was read from the autoradiograms using the program GELIN written by Dr. M. Bishop (Cambridge) on a BBC microcomputer equipped with a Torch discdrive and GP7 sonic digitiser (Science Accessories Corporation). The data was processed on an IBM3081 computer using the DBUTIL and DBAUTO programs of Staden (1982).

2.3.7 Transcription-translation and immunoprecipitation

A cell-free transcription-translation extract was prepared from *S. lividans* as described by Thompson *et al.* (1984). The extract was programmed with approx. 1µg DNA (3µg chromosomal DNA) as described by these authors, except that a final concentration of 15mM $MgCl_2$ gave better incorporation of label than 12mM, and [^{35}S]-methionine (>800Ci $mmol^{-1}$) was used undiluted.

Specific translation products were immunoprecipitated with the relevant antiserum and Protein-A sepharose beads. To the 32µl final volume of the transcription-translation reaction was added 8µl 10% SDS. The sample was boiled for 5min and then 200µl PTE was added (PTE: 50mM sodium phosphate, pH7.0, 150mM NaCl, 1% Triton X-100, 5mM EDTA). Preimmune antiserum (4µl) was added followed by 15µl of a 12.5% (w/v) suspension of Protein A-sepharose beads in 25mM Tris-HCl, pH7.5. This was mixed by turning end-over-end for 1hr at room temperature. The solution was briefly centrifuged in a microfuge and the supernatant transferred to a new tube. Immune serum (4µl) and more Protein A-sepharose (15µl) were added and the incubation continued for a further 1hr with mixing. After brief centrifugation the supernatant was removed. The pellet was washed with 0.4ml PTE four times and then twice with 0.4ml SPNE (SPNE: 50mM sodium phosphate, pH7.0, 150mM NaCl, 0.1% SDS, 2mM EDTA). The pellet was boiled with 20µl of gel sample buffer for 5min and this solution was analysed by SDS-polyacrylamide gel electrophoresis. The gel was soaked in Amplify (Amersham) and dried before autoradiography using preflashed film.

2.4 Gel electrophoresis

2.4.1 Agarose gels

Agarose gel electrophoresis was performed as described by Maniatis et al. (1982). The gels were normally 0.7%, but 1.0 and 1.5% were also used. They were approx. 6mm thick with wells 1x5x5mm or 1x13x5mm. Two sizes were used; approx. 11x15cm (100ml) and 20x20cm (250ml). The gels were run, submerged in buffer, for 3hr at 100V for the small gels and 200V for the large gels or 20V and 50V respectively overnight. When two rows of wells were cast in the 20x20cm gels they were run at 150V for approx. 3hr. All gels were run in TAE (50xTAE stock: 242g Tris base, 18.6g disodium EDTA brought to pH7.7 with glacial acetic acid in a final volume of 1l). Some were run with 0.5µg ml^{-1} ethidium bromide in the gel and running buffer for quick results, but most were stained for 20min in 1µg ml^{-1} ethidium bromide (approx. 200ml) after running. Sample buffer (40mM Na_2EDTA, 0.1% SDS, 20% Ficoll 400, 2.5mg ml^{-1} bromophenol blue) was added to 1/5 or 1/4 of the volume of the DNA sample.

2.4.2 SDS-polyacrylamide gels

Protein samples were separated on 15cm SDS-polyacrylamide slab gels using a buffer system based on that of Laemmli (1970). The gel compositions used are shown in Table 2.1. The solutions were degassed before the addition of 100µl 10% ammonium persulphate and 10µl TEMED (50µl and 5µl respectively for the stacking gel) immediately prior to pouring. The separating gel was overlaid with water-saturated n-butanol to exclude air and ensure a level surface while it was setting. This was washed away with water before the stacking gel was poured. Samples were boiled for 5min with an equal volume of 375mM Tris-HCl pH8.8/4% SDS/35% glycerol/2.86M 2-mercaptoethanol/0.01% bromophenol blue prior to loading. The gels were run in buffer containing 6g Tris-HCl, 28.8g glycine and 10ml 10% SDS in 1l, at 25mA while stacking and then 35mA for approx. 5hr.

Table 2.1 Composition of polyacrylamide gels

Gel concentration	12%	15%	20%	24%	5% Stacker
40% Acrylamide (ml)	-	-	-	18.0	-
30% Acrylamide (ml)	12.0	15.0	20.0	-	1.67
1% Bisacrylamide (ml)	3.25	2.60	1.95	2.16	1.30
1.5M Tris-HCl, pH8.7 (ml)	7.5	7.5	7.5	7.5	-
1.5M Tris-HCl, pH6.8 (ml)	-	-	-	-	1.25
H_2O (ml)	7.14	4.79	0.44	2.23	5.7

After running the gels were treated in one of three ways:

a) Fixing and simultaneously staining with Coomassie blue. The gel was soaked in 250ml 5:1:5 methanol/acetic acid/water containing 0.25% Coomassie brilliant blue (filtered) with agitation for 45min at 55°C or overnight at room temperature. It was then destained in 2-3x250ml of the same 5:1:5 mix without Coomassie blue at 55°C or room temperature.

b) Silver staining. The gel was fixed in 4% sulphosalicylic acid/10% acetic acid at 55°C for 30min, soaked overnight in 50% methanol and then silver stained by the method of Wray et al. (1981).

c) Preparing for autoradiography. The gel was fixed for 30min at room temperature in 500ml 5:1:5 methanol/acetic acid/water and then soaked in 100ml Amplify (Amersham) for 30min before being dried onto Whatman 3MM paper with a gel drier at 80°C. The dried gel was autoradiographed with preflashed film at -70°C.

2.4.3 Urea-polyacrylamide gels

Oligonucleotides were dissolved in formamide-dye mix (100ml deionised formamide, 0.1g xylene cyanol FF, 0.1g bromophenol blue, 2ml 0.5M disodium EDTA). They were then electrophoresed on 20% polyacrylamide gels (19:1 acrylamide:bisacrylamide) containing 8M urea, 40cm long and 1mm thick. They were visualised by shining a short wavelength UV lamp through the gel onto a silica TLC plate (HF_{254}) containing a fluorescent indicator. The oligonucleotides absorb the radiation so are seen as shadows. Analytical gels for running radiolabelled oligonucleotides were of the same composition but 0.4mm thick. All gels were run at 37W in TBE (10xTBE: 108g Tris base, 55g boric acid, 9.3g disodium EDTA in 1l). The analytical gels were fixed for 10min in

10% acetic acid/10% methanol before drying at 80°C onto Whatman 3MM paper.

Sequencing gels were of the same size and composition as the analytical gels but with a buffer gradient of 0.5-5xTBE prepared as described by Bankier and Barrell (1983). The samples were heated at 80°C for 20min with 2µl of the formamide-dye mix and the gels were fixed and dried as above.

2.5 Hybridisation

2.5.1 Preparation of gels

In early experiments DNA from agarose gels was transferred to nitrocellulose sheets by the method of Southern (1975). Later, the gels were dried onto 3MM paper using a gel-drier at 60°C. The DNA was then denatured by soaking the gel in 250ml 0.5M NaOH/0.15M NaCl for 20min during which time the paper could be removed and discarded. The gel was neutralised for 20min in 250ml 0.5M Tris-HCl, pH7.5/0.15M NaCl and was then ready for hybridisation (Tsao *et al.*, 1983).

This method is much quicker than Southern blotting, avoiding an overnight step. The dried gel is stronger than nitrocellulose and the positions of the standards can be accurately marked by stabbing the gel with a syringe needle containing Indian ink before drying the gel. Also less background hybridisation is observed. Stronger signals are obtained because most of the DNA remains in the gel, whereas not all the DNA is transferred by Southern blotting and the efficiency of transfer depends on the size of the DNA fragments. Also DNA becomes attached to nitrocellulose at many points so that often the oligonucleotide binding site may be obscured. In the dried gel the DNA is only entrapped so most sites should be available for hybridisation (Meinkoth and Wahl, 1984).

2.5.2 Preparation of colony filters

Filters were labelled with pencil and autoclaved before use. Colonies were either patched onto the filter on a fresh plate and incubated overnight for *E. coli* or for two days for *S. lividans*, or the filters were pressed directly

on the colonies to be screened and then treated immediately for E. coli colonies or laid spore side up on fresh plates and incubated for two days for S. lividans colonies.

At first, nitrocellulose filters were used. Those carrying S. lividans colonies were pretreated by laying them colony side uppermost on 3MM paper soaked in TE/4mg ml^{-1} lysozyme for 30-45min. All filters were then laid colony side up on 3MM paper soaked in 10% SDS for 5min and then transferred to 3MM paper soaked in 0.5M NaOH/1.5M NaCl for 5min. They were then immersed in the same solution for 1-2min followed by 0.5M Tris-HCl pH7.5/1.5M NaCl for 2-3min. Colony debris was rubbed off and the filters immersed in fresh neutralising solution for 1-2min. They were then dried and baked under vacuum at 80°C for 2hr.

Later Whatman 541 paper was used in place of the nitrocellulose. These filters were treated in the same way but were steamed for 3min while they lay on the 0.5M NaOH/1.5M NaCl soaked paper (Maas, 1983). This was done by placing the filters on 3MM paper in a crystallising dish, which was in turn placed in a pressure cooker. The lid was loosely fitted and no weight applied. They were ready to use immediately after drying, which could be speeded up by adding an ethanol rinse to the procedure. The 541 filters were much stronger than nitrocellulose and gave much stronger signals because of the better lysing and denaturation. The transfer of colonies and spores to the 541 paper was also more even than to the nitrocellulose and the preparation of the filters was quicker.

2.5.3 Hybridisation

Nitrocellulose was prehybridised in 20ml 6xSSC/0.05% sodium pyrophosphate/5xDenhardt's/50µg ml^{-1} tRNA/10% SDS for 2-4hr at the hybridisation temparature (20xSSC: 3M NaCl, 0.3M Na_3citrate; 100xDenhardt's: 2% bovine serum albumen fraction V, 2% Polyvinylpyrollidone 760 000, 2% Ficoll 400 000). This solution was then removed and probe added in 6xSSC/0.05% sodium pyrophosphate/5xDenhardt's. For 541 paper and dried gels no prehybridisation was done and the probe was added in the same solution as for nitrocellulose. After hybridisation overnight all filters were washed twice in 500ml

6xSSC/0.05% sodium pyrophosphate for 30min at the hybridisation temperature. Filters were blotted on absorbent paper and air dried and the gels briefly dried in the gel drier taking care not to allow the temperature to rise above the hybridisation temperature.

Most hybridisations were done in 5ml in bags for gels and in petri dishes for filters. For large numbers of filters 20ml of hybridisation solution was used. The NaCl in the probes prepared by DE52 column chromatography was allowed for by reducing the amount of 20xSSC used to keep the sodium concentration the same.

A more even distribution of label throughout a stack of many filters was observed when no prehybridisation was used as all parts of the filters were wetted with the solution containing the label rather than the oligonucleotide having to diffuse through the prehybridisation solution that remained on the filters. However, greater background was observed with 541 paper than with nitrocellulose. This was of no consequence since the signals from positive colonies were extremely strong, but might be reduced by adding tRNA or carrier DNA either to the hybridisation solution or in a prehybridisation step to block the non-specific sites on the filters to which the oligonucleotide can bind (Gergen *et al.*, 1979).

2.6 Protein methods

2.6.1 Protein purification

Methylmalonyl-CoA epimerase was purified from *P. shermanii* as described by Leadlay (1981).

The cloned epimerase was purified from *S. lividans* by a modification of this procedure. All operations were carried out at 0-4°C. The cells (160g) were ground in 50mM Tris-HCl pH7.5/1mM EDTA/0.1mM PMSF/1mM benzamidine hydrochloride using a KDL Dyno-Mill bead mill (W. Bachofen, Zurich). Streptomycin treatment and fractionation with poly(ethylene glycol) were omitted. Protein precipitated by addition of ammonium sulphate (55-90% saturation at 0°C) was resuspended in 50ml of the same buffer and gel-

6xSSC/0.05% sodium pyrophosphate for 30min at the hybridisation temperature. Filters were blotted on absorbent paper and air dried and the gels briefly dried in the gel drier taking care not to allow the temperature to rise above the hybridisation temperature.

Most hybridisations were done in 5ml in bags for gels and in petri dishes for filters. For large numbers of filters 20ml of hybridisation solution was used. The NaCl in the probes prepared by DE52 column chromatography was allowed for by reducing the amount of 20xSSC used to keep the sodium concentration the same.

A more even distribution of label throughout a stack of many filters was observed when no prehybridisation was used as all parts of the filters were wetted with the solution containing the label rather than the oligonucleotide having to diffuse through the prehybridisation solution that remained on the filters. However, greater background was observed with 541 paper than with nitrocellulose. This was of no consequence since the signals from positive colonies were extremely strong, but might be reduced by adding tRNA or carrier DNA either to the hybridisation solution or in a prehybridisation step to block the non-specific sites on the filters to which the oligonucleotide can bind (Gergen et al., 1979).

2.6 Protein methods

2.6.1 Protein purification

Methylmalonyl-CoA epimerase was purified from P. shermanii as described by Leadlay (1981).

The cloned epimerase was purified from S. lividans by a modification of this procedure. All operations were carried out at 0-4°C. The cells (160g) were ground in 50mM Tris-HCl pH7.5/1mM EDTA/0.1mM PMSF/1mM benzamidine hydrochloride using a KDL Dyno-Mill bead mill (W. Bachofen, Zurich). Streptomycin treatment and fractionation with poly(ethylene glycol) were omitted. Protein precipitated by addition of ammonium sulphate (55-90% saturation at 0°C) was resuspended in 50ml of the same buffer and gel-

filtered on a 1650ml Sephadex G75-superfine column. An additional 10ml ($2cm^2$x5cm) DE-Sepharose Fast-Flow column (Pharmacia) was run with a linear gradient (1l) of 0-1M NaCl in the same buffer. This concentrated the sample but gave little purification. The purest fractions were dialysed against 10mM sodium phosphate, pH6.5 and then fractionated on a hydroxyapatite column (2.3cm x 4.5cm) with a linear gradient (1l) of 10-200mM sodium phosphate, pH6.5. The pure protein was concentrated by ultrafiltration in an Amicon apparatus fitted with a PM-10 membrane.

Epimerase activity was assayed as described by Leadlay (1981).

The purification of methylmalonyl-CoA mutase from P. shermanii described here is also presented by Francalanci et al. (1986). Cell paste (100g) was suspended in 200ml 0.2M potassium phosphate, pH7.0/1mM EDTA/0.1mM PMSF/1mM benzamidine hydrochloride/1mM DTT/5g activated charcoal and the cells were broken by two passes through a KDL Dyno-Mill bead mill at 2°C. Debris was removed by centrifugation at 12 000g for 30min. The pellet was washed with 100ml buffer and recentrifuged. The pooled supernatants were then added to 0.5kg of moist DEAE-cellulose and the proteins eluted by a batch procedure as described by Kellermeyer and Wood (1969). Solid ammonium sulphate was added to 55% saturation at 0°C (32.6g/100ml) and the precipitate removed by centrifugation at 12 000g for 20min. The supernatant was taken to 85% saturation by further addition of ammonium sulphate (19.7g/100ml). The precipitate, collected by centrifugation, was resuspended in 15ml 50mM Tris-HCl, pH7.5 (at 20°C)/2mM EDTA/1mM benzamidine/1mM DTT and dialysed against this buffer for 2hr. This solution was clarified by centrifugation at 25 000g for 10min and then applied to a 5cmx90cm Biogel A1.5m gel-filtration column. Active fractions were loaded onto a 2.5cmx20cm DEAE-cellulose column equilibrated in the same buffer and the enzyme was eluted by a linear gradient (1l) of 0-0.25M KCl in the buffer. Active fractions were pooled and concentrated by ultrafiltration using an Amicon apparatus fitted with a PM-10 membrane.

Mutase was assayed as described by Zagalak et al. (1974) except that assays were conducted at 30°C. The formation of (2R)-methylmalonyl-CoA from succinyl-CoA was coupled to the oxidation of NADH through epimerase,

transcarboxylase and malate dehydrogenase. All operations with 5'-deoxyadenosylcobalamin were performed in dim red light.

2.6.2 Preparation of antisera

Antibodies were raised in rabbits as described by Hudson and Hay (1976).

2.6.3 Western blotting

Proteins were separated by SDS-polyacrylamide gel electrophoresis and then blotted to nitrocellulose by the method of Towbin et al. (1979), as modified by Burnette (1981). Specific protein products were detected using antisera, a horseradish peroxidase-antibody conjugate and diaminobenzidine as described by Hawkes et al. (1982).

2.6.4 S-carboxymethylation

Proteins were reduced with dithiothreitol and treated with iodoacetic acid in the presence of 5M guanidinium chloride prior to amino acid analysis and sequencing as described by Perham (1978). One sample of epimerase (100µg) was reacted with [^{14}C]-iodoacetic acid (8Ci mol^{-1}) so that the cysteine residues could be more readily detected by automated sequencing.

2.6.5 Sedimentation equilibrium

Sedimentation equilibrium was performed at 20°C in a Beckman model E analytical centrifuge equipped with Rayleigh interference optics following the procedures of Nureddin and Johnson (1977). The apparent weight average molecular weight was calculated from plots of ln(c) against (x^2) (where c is the protein concentration and x the radial distance) obtained from a full length photograph of the interference fringes taken after equilibrium was established. Mutase was at 0.9mg ml^{-1} in buffer of ionic strength I=0.1 and pH7.8 containing, in 1l, 5.79g $Na_2HPO_4.12H_2O$, 0.197g KH_2PO_4 and 2.923g NaCl.

2.6.6 Preparation of mutase peptides by proteolysis

A freshly prepared 1mg ml^{-1} solution of Endoproteinase Arg-C in 0.5% ammonium bicarbonate was used at 1.5% w/w (enzyme/substrate) to digest approx. 12mg (80nmol) carboxymethylated mutase taken up in 1.25ml of the same buffer. During the digestion most of the protein that was originally insoluble was dissolved. A second, equal amount of protease was added after 7hr and the digestion continued overnight. Samples were run on a TSK 3000SW HPLC column in 0.1M potassium phosphate, pH7.0 to follow the progress of the digestion. When no further change occurred the reaction was stopped by lowering the pH to pH2 with trichloroacetic acid and then freeze-dried.

The peptides soluble in 1ml of 25mM ammonium bicarbonate were gel filtered after removal of insoluble peptides by centrifugation. A 1.8cm^2x32cm Sephadex G75-superfine column was run in 25mM ammonium bicarbonate. Fractions containing material absorbing at 220nm were pooled in five groups and freeze dried.

2.6.7 HPLC and FPLC purification of mutase peptides

Each of the four pools containing larger molecular weight peptides from the gel filtration column was fractionated on an HPLC RP8 reverse phase column. They were loaded in 0.1% trifluoroacetic acid and eluted by a gradient of acetonitrile/0.1% trifluoroacetic acid collecting peaks by hand. Many of these were rechromatographed on the same column with a different buffer system (10mM ammonium acetate, pH5.8 and a gradient of acetonitrile). Some were run on a C18 column with the first solvent system. A few that contained mostly positively charged peptides at neutral pH were fractionated on an FPLC Mono-S cation exchange column. A linear gradient of 10mM to 1M NH_4HCO_3 gave poor resolution and had a high background absorbance at 230nm. Substituting triethylammonium acetate at pH6.5 gave good resolution, although still having a high absorbance. The triethylamine had to be redistilled otherwise some residue was left when the fractions were dried.

2.6.8 Thin layer electrophoresis of mutase peptides

To assess the purity and concentration of the peptides, samples, dissolved in 0.1% trifluoroacetic acid, were analysed by thin-layer electrophoresis at pH2 or pH6.5 and the peptides visualised by staining with fluorescamine as described by Perham (1978). Some of the plates were then treated with p-dimethylaminobenzaldehyde to detect tryptophan containing peptides, also as described by Perham (1978).

2.6.9 DABITC sequencing of peptides

Peptides (approx. 5-10nmol) were sequenced by a manual microscale method using dimethylaminoazobenzene isothiocyanate (DABITC) as described by Chang (1983).

CHAPTER 3

MOLECULAR CLONING OF METHYLMALONYL-CoA EPIMERASE

3. MOLECULAR CLONING OF METHYLMALONYL-CoA EPIMERASE

Methylmalonyl-CoA epimerase has been purified from *P. shermanii* and shown to be an α_2-dimer of subunit M_r 17 000 (Leadlay, 1981). In order to clone the gene coding for epimerase it was necessary to obtain some peptide sequence, enabling oligonucleotides to be made and used as probes.

3.1 N-terminal amino acid sequence of epimerase

A sample of reduced, carboxymethylated epimerase was kindly sequenced by Dr. J. Walker (MRC Laboratory of Molecular Biology, Cambridge) on an automated solid-phase sequenator. The first 31 amino acids (about one fifth of the protein) were obtained and are shown in Figure 3.1. Later, a second protein sample, carboxymethylated with [^{14}C]iodoacetic acid, was also sequenced. This revealed that two residues previously designated serine were in fact cysteine and contained the radioactivity in the protein. This then agreed with the nucleic acid sequence (see Chapter 5). Also, the amino acid composition of epimerase shows two cysteines and none were found elsewhere in the nucleic acid sequence.

Ser Asn Glu Asp Leu Phe Ile Cys Ile Asp His Val Ala Tyr Ala Cys
(Met) (Ser) (Ser)

Pro Asp Ala Asp Glu Ala Ser Lys Tyr Tyr Gln Glu Thr Phe Gly ...

Figure 3.1 N-terminal amino acid sequence of epimerase
The results of two automated sequence determinations agree except for those residues in brackets. The differences are explained in the text.

The second N-terminal sequence determination obtained Ser-Asn-Glu for the first three amino acids with no trace of methionine, whereas the first determination had a trace of methionine in the first position which masked the normally weak serine signal to give the apparent sequence Met-Asn-Glu. The two samples sequenced were from different preparations of the enzyme. The sample for the first determination may, therefore, have contained some epimerase that had not been fully processed to leave serine at the N-terminus or, possibly, a minor contaminant with methionine at the N-terminus. The

nucleic acid sequence confirmed that the full sequence was indeed Met-Ser-Asn-Glu as expected (Chapter 5).

3.2 Oligonucleotide probes synthesized for epimerase

The best portion of the N-terminal amino acid sequence of epimerase for making oligonucleotide probes is Lys-Tyr-Tyr-Gln-Glu-Thr with only two codons for each amino acid except the terminal threonine which has four. By making 17-mers the wobble base of this threonine can be omitted so there are only 32 possible DNA sequences that can code for this peptide sequence.

Initially, a mixture of all 32 possible 17-mers was made on the Kieselguhr resin support, but this did not bind to *P. shermanii* chromosomal DNA. This may have been due to a fault in the synthesis and it is difficult to check the presence and purity of all the correct sequences in a complex mixture. Thus, although some workers have successfully used mixtures to clone genes (e.g. Carroll and Porter, 1983; Cohn *et al.*, 1983; and Singer-Sam *et al.*, 1983) they are not always reliable. Two other mixtures made for the biotin binding subunit of transcarboxylase also failed to show any hybridisation to the chromosomal DNA, whereas a single oligonucleotide, the M13 sequencing primer, made with the same reagents and resin support, bound well to M13 DNA suggesting that the synthesis gave the correct product. This oligonucleotide was also good for use in DNA sequencing, but the stringency of binding required for hybridisation experiments is greater than that required for sequencing, because a transiently binding primer will become more tightly bound as the extension proceeds.

To choose a single sequence to make from the 32 possible sequences codon usage data would normally be used. Johnson and Cummins (1972) showed that the G+C content of *P. shermanii* DNA was 64-67%, so, although no genes had yet been sequenced from *P. shermanii*, the high G+C content of the DNA made choices possible. The G+C content of the second position in codons is determined by the amino acid composition of the protein and is usually in the range 35-45%. To maintain an overall G+C content in the gene of near 70% the wobble base will usually be G or C in preference to A or T, as is the case for Streptomycete genes (Bibb *et al.*, 1984). The G+C content in the first position

is also slightly higher than in average G+C content genomes because leucine and arginine can use the codons with a C in the first position in preference to T or A. In the best peptide sequence from epimerase (Lys-Tyr-Tyr-Gln-Glu-Thr) the first five amino acids have only two codons each and in each case G or C can be used in place of A or T. If the third base of the threonine codon is also omitted, then only one 17-mer need be made (see Table 3.1).

Table 3.1 Oligonucleotides made as probes for epimerase

Peptide sequence	NH_2	Lys	Tyr	Tyr	Gln	Glu Thr
Possible DNA coding sequences	5'	$AA\frac{G}{A}$	$TA\frac{C}{T}$	$TA\frac{C}{T}$	$CA\frac{G}{A}$	$GA\frac{G}{A}$ AC
32x17-mer oligonucleotides	3'	$TT\frac{C}{T}$	$AT\frac{G}{A}$	$AT\frac{G}{A}$	$GT\frac{C}{T}$	$CT\frac{C}{T}$ TG
Single 'high G+C' 17-mer	3'	TTC	ATG	ATG	GTC	CTC TG

If the last codon position is 90% G+C then this single oligonucleotide has a 60% chance of being correct and a 33% chance of having one mismatch to the correct sequence.

This single sequence was made on the controlled-pore glass support that was then available, because this gave better yields and was easier to use than the Kieselguhr resin support. P. shermanii chromosomal DNA digested with various restriction enzymes was run on an agarose gel, blotted to a nitrocellulose membrane, and then hybridised with the probe, which had been end-labelled with radioactive phosphate from γ-[^{32}P]-ATP. The oligonucleotide hybridised to one band in each of the digests (Figure 3.2), showing that it bound specifically to one sequence on the P. shermanii chromosome.

Later, when a procedure for making many different oligonucleotides at one time became available it was possible to make all 32 different sequences on separate paper discs (Matthes et al., 1984). Hybridisation at 12-20°C below the T_m of each oligonucleotide showed that the sequence which had been made previously bound most strongly, and those sequences with one base change from this bound weakly, confirming that the correct sequence had been chosen. Figure 3.3 shows the hybridisation of these oligonucleotides to slices of a Southern blot of one restriction digest run in a wide track on a gel.

The correct oligonucleotide made in the paper disc synthesis (P100) was used in all subsequent experiments.

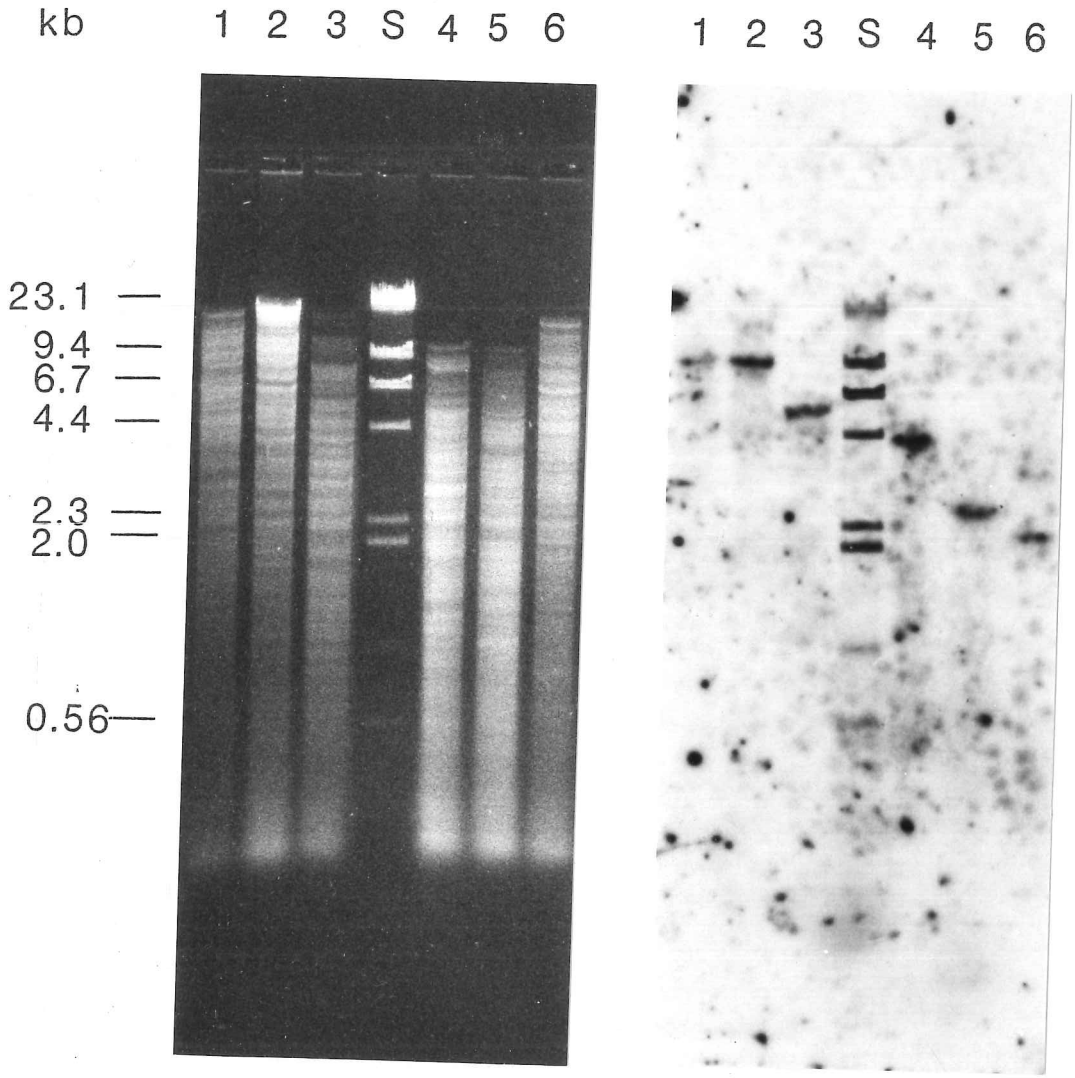


Figure 3.2 Hybridisation of the epimerase probe to chromosomal DNA

P. shermanii chromosomal DNA, digested with six different restriction enzymes and separated on an agarose gel was blotted to nitrocellulose. The filter was probed with the single, 'high G+C' oligonucleotide for epimerase. The digests are: 1, BamHI (GGATCC); 2, EcoRI (GAATTC); 3, PstI (CTGCAG); 4, PvuII (CAGCTG); 5, SmaI (GGGCCC); 6, XhoI (CTCGAG). The standards (S) are λDNA digested with HindIII and were labelled with α-[^{35}S]-dATP using Klenow enzyme and the other dNTPs. The gel is shown on the left and the autoradiogram on the right.

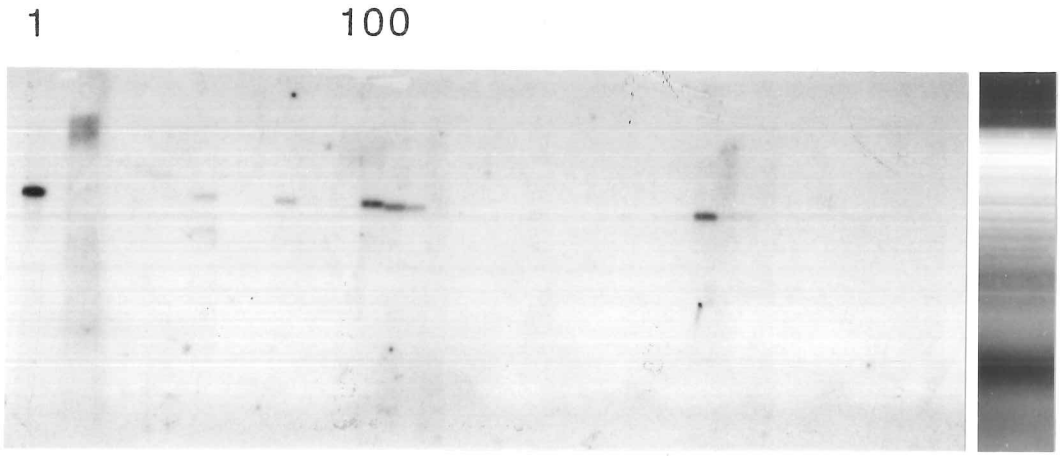


Figure 3.3 Hybridisation with the 32 oligonucleotides for epimerase

P. shermanii chromosomal DNA was digested with XhoI and electrophoresed in a wide track on an agarose gel. The DNA was blotted to nitrocellulose and the filter then cut into strips. Each strip was hybridised with an oligonucleotide. The single, 'high G+C', oligonucleotide made previously was hybridised to the strip labelled 1. The same sequence made by the paper disc synthesis was 100. The hybridisation temperature also allowed the five probes with one mismatch to bind. A strip of a photograph of the gel is shown alongside the autoradiogram.

3.3 Choice of restriction enzyme for cloning

As can be seen in Figure 3.2 the signal to noise ratio of the early hybridisations was not good. This would be likely to make colony or plaque screening of a total genomic library difficult. For this reason a different approach was taken. Restriction fragments of the size which hybridised with the oligonucleotide probe were isolated from the complete digest and inserted into the appropriate site in the polylinker of the multicopy plasmid pUC12 (Viera and Messing, 1982). This plasmid was used because the high copy number should give strong hybridisation, the polylinker allows use of several enzymes, recombinant plasmids can be recognised by inactivation of the β-galactosidase gene into which inserts are cloned, and inserts tend to be more stably maintained than in many other pBR322 derived vectors. The recombinant clones were patched out before screening, to give lines on the autoradiograms instead of dots.

Figure 3.2 shows some of the possible restriction digests and the sizes of the fragments which hybridise with the probe. Only about 500bp of DNA would be needed to encode the epimerase so one of the smaller fragments could be cloned. However, it was possible that by cloning a large piece one or other of the mutase and transcarboxylase genes might be found to be adjacent to the epimerase gene on the chromosome. Two or more genes of a bacterial metabolic pathway are often arranged in an operon or clustered, although frequently the complete pathway is spread between more than one group on the chromosome.

A second reason for cloning a large fragment is the distribution by size of the fragments in a restriction enzyme digest. An enzyme with a 6bp recognition site would normally cut the DNA on average every 4^6 or 4 096bp. If the *P. shermanii* chromosome is the same size as that of *E. coli*, approx. 4×10^6bp, then 6bp cutting enzymes should give about 1 000 pieces. These fragments would be distributed by size according to the Poisson distribution with mean length 4 096bp. There are always some differences in the distributions produced by different enzymes. These are partially due to statistical variation, but codon usage and methylation also have effects. Digestions of the high G+C content *P. shermanii* DNA give widely different distributions from one enzyme to another, as can be seen in Figure 3.2 in

which all the enzymes recognise 6bp sites. This variation is largely due to the number of G and C bases in the recognition site. EcoRI (GAATTC) cuts less frequently than XhoI (CTCGAG). Figure 3.4 shows the theoretical distributions of fragments produced by the standard 6bp cut (1 in 4 096), and those for DNA with a G+C content of 70% with either two or four G and C bases in the recognition site. A logarithmic scale is used for the fragment sizes because this is how the fragments are distributed when separated by gel electrophoresis. The apparent distribution of the stained DNA in a gel depends on the amount of DNA at each point (dependent on the sizes as well as the number of the fragments). This is also plotted in Figure 3.4 and agrees well with the distributions observed in a gel (Figure 3.2).

A restriction enzyme recognising a site with four G or C and two A or T bases will cut on average every 2 962bp in 70% G+C DNA $\{(\frac{100}{35})^4(\frac{100}{15})^2\}$ to give 1 350 pieces, whereas an enzyme recognising a sequence containing two G or C, and four A or T, bases would be expected to cut on average every 16 125bp $\{(\frac{100}{35})^2(\frac{100}{15})^4\}$ giving only 248 fragments. An enzyme recognising a sequence containing six G or C bases, e.g. SmaI (GGGCCC), would cut very frequently giving over 7 000 fragments of average length 544bp, whereas an enzyme recognising a 6bp sequence containg only A or T bases would hardly cut at all (approx. 46 fragments).

As can be seen in Figure 3.2 there is some variation between enzymes with the same number of G and C bases in the recognition sequence. This is probably due to different frequencies of use as codons of the triplets overlapped by the sequence, for instance, HindIII (AAGCTT) cuts even less often than EcoRI (GAATTC).

The approach of cloning a size fraction of DNA fragments from a restriction digest requires that the DNA be fractionated. Gels are the most accurate way of doing this and in practise the slice of gel corresponded to 1-2 divisions on the graph of Figure 3.4 (0.1-0.2 units of $\log_{10}$ fragment length). Cloning the 2.3kb XhoI or 2.5kb SmaI fragments is, therefore, undesirable, as the XhoI fraction would contain close to the maximum number of fragments in any fraction of that digest (65-140) and the SmaI fraction would contain even more. Using the formula of Clarke and Carbon (1976), over

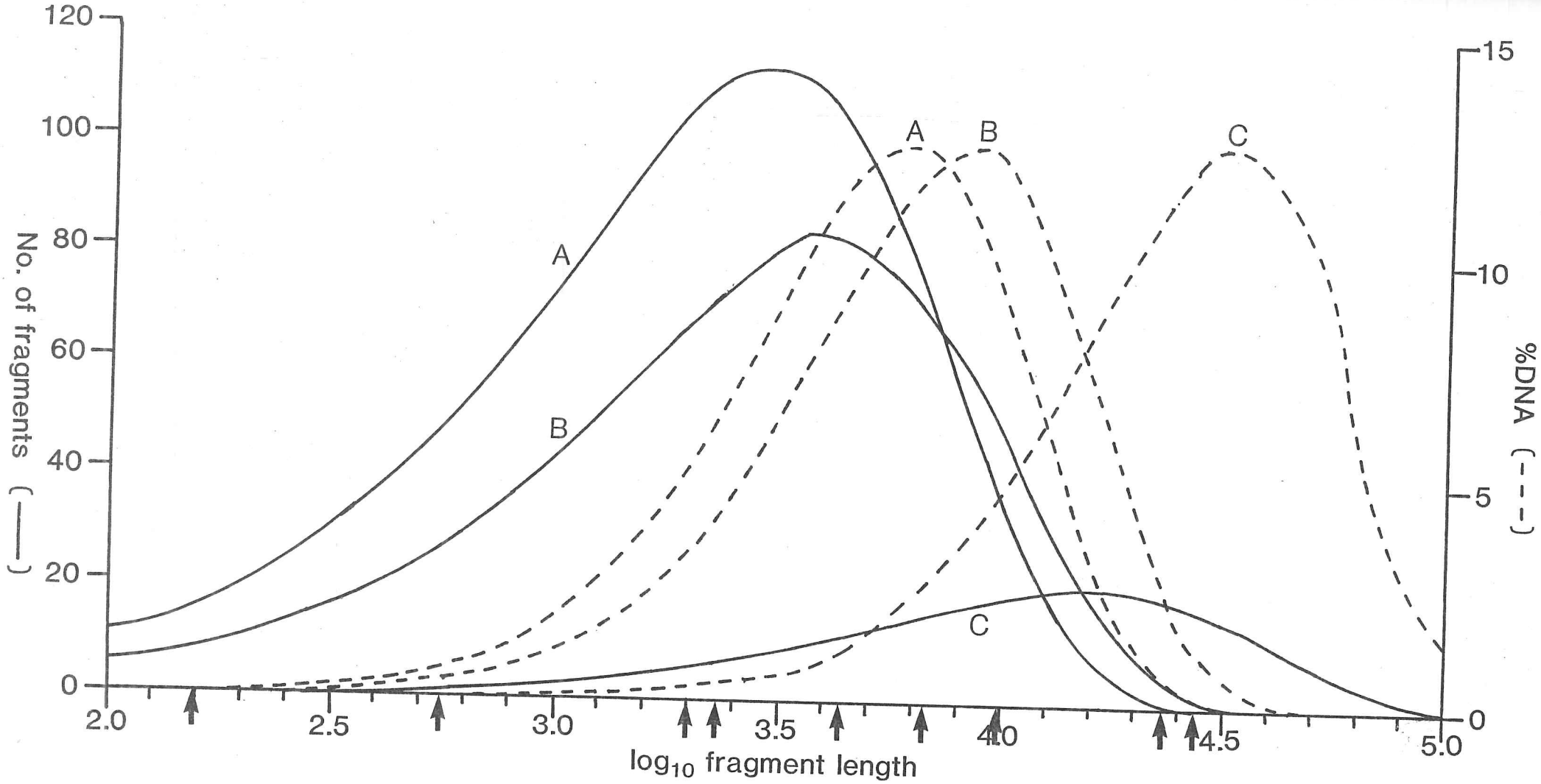


Figure 3.4 Distribution of restriction enzyme fragments in gels

The distributions, after electrophoresis, by size of fragment and amount of DNA, of restriction fragments from a 4×10^6bp chromosome are shown for various digestions. B shows the distributions expected from digestion with an enzyme recognising a 6bp sequence in DNA with a 50% G+C content. A and C show the situation in digests of 70% G+C DNA. A shows the case when the recognition site contains 4 A or T and 2 G or C. C shows the distribution when the site contains 2 A or T and 4 G or C. The arrows indicate the positions of HindIII fragments of λDNA (0.12, 0.56, 2.0, 2.3, 4.4, 6.7, 9.4, 23.1 and 27.5kb). DNA fragments larger than approx. 20kb are usually not resolved on gels.

640 clones would have to be screened to ensure 99% chance of finding the correct fragment (more if the chromosome is larger than $4x10^6$bp).

It would be preferable to clone a fragment of over 10kb produced by an enzyme recognising a site containing four G or C, and two A or T, bases, of over 3kb from an enzyme cutting a sequence of six G or C bases, or of any size from an enzyme cutting at a sequence containg two G or C, and four A or T, bases. The fragment must also be shorter than about 20kb to reduce the risk of deletions occurring within the recombinant plasmid.

The restriction enzyme EcoRI (GAATTC) gave a hybridising fragment of 10.4kb and cuts rarely because of the number of G and C bases in the recognition sequence. A fraction of fragments around this size, equivalent to two divisions in Figure 3.4, would contain approx. 37 different pieces, so 170 colonies should be screened for 99% certainty of finding the desired fragment. This number is quite practical for screening by patching out recombinants before probing them.

3.4 Localising the probe sequence within the 10.4kb EcoRI fragment

Although the EcoRI fragment was very much larger than the epimerase gene, it was still possible that the hybridising sequence was very close to one end of the fragment and that the gene was not all present. To check this, *P. shermanii* DNA was restricted with EcoRI and then digested with the exonuclease Bal31. This enzyme digests double stranded DNA, chewing in from free ends. The fragments, therefore, become progressively shorter until they are finally reduced entirely to mononucleotides. Samples taken after various times of digestion were run on an agarose gel. This gel was dried, rather than blotted to nitrocellulose, and the DNA denatured *in situ* (Tsao *et al.*, 1983). The dried gel was then probed with the radiolabelled oligonucleotide (Figure 3.5). This showed that the 10.4kb fragment could be reduced to less than 4kb without losing the sequence that hybridises to the oligonucleotide, indicating that this sequence was at least 3kb from each end and, therefore, that the whole epimerase gene would be present. There would also be room on each side for at least part of the genes for mutase or transcarboxylase if they were located adjacent to the epimerase gene. The dried gel procedure for

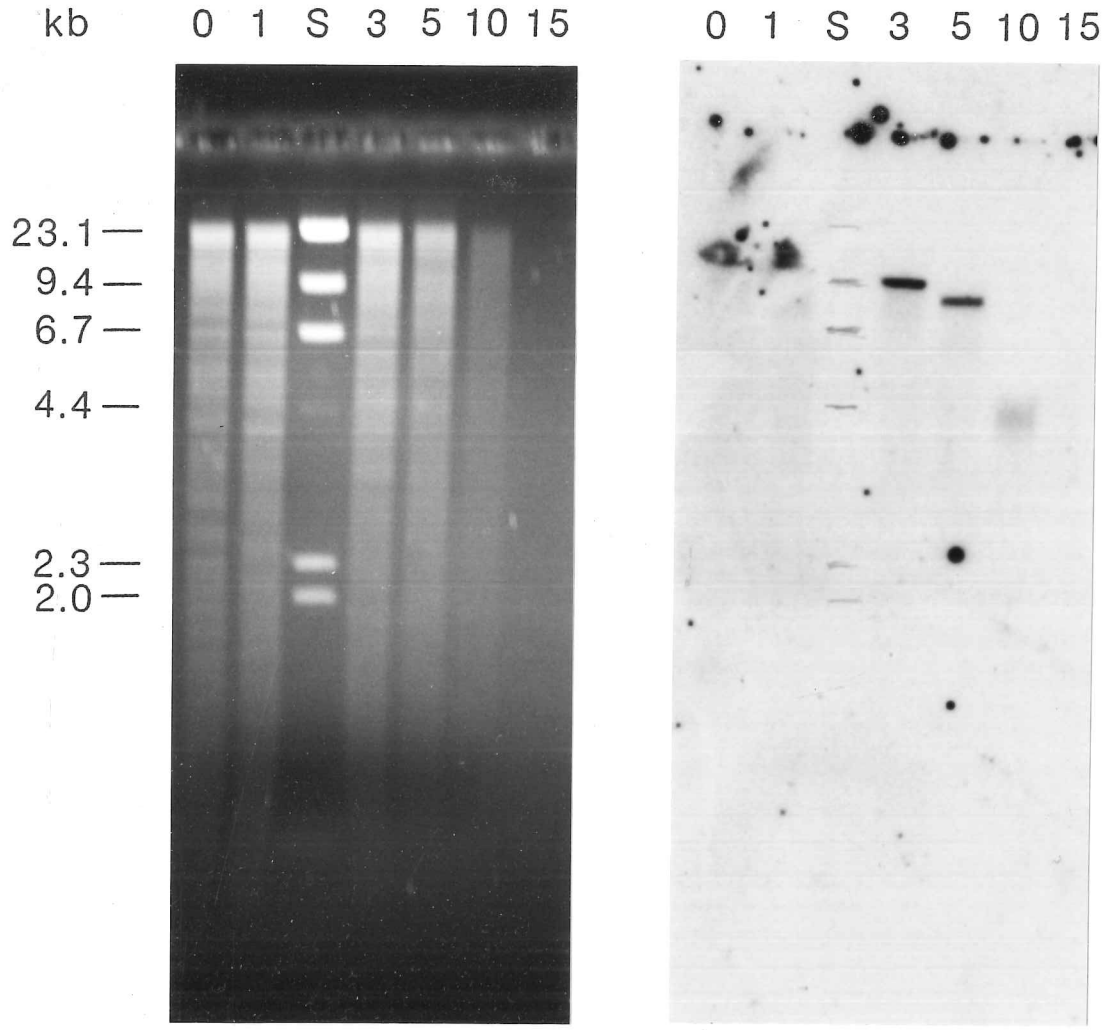


Figure 3.5 Use of Bal31 to show the region on the 10.4kb EcoRI site to which the oligonucleotide bound

P. shermanii DNA was digested with EcoRI. A series of time points were then taken during digestion of this DNA with Bal31. The samples were run on a gel which was dried and probed with oligonucleotide 100. The samples were taken after 0, 1, 3, 5, 10 and 15min of digestion. The standards (S) were λDNA digested with HindIII. Their positions were marked on the gel with Indian ink and then pencilled onto the autoradiogram. Unincorporated label was not removed from the probe, and the gel was damaged before drying.

probing DNA was more convenient than blotting to nitrocellulose as well as giving less background hybridisation. The signals obtained were also stronger (Meinkoth and Wahl, 1984). All subsequent gels were treated in this manner.

3.5 Cloning the 10.4kb EcoRI fragment

P. shermanii DNA was digested with EcoRI and fractionated on an agarose gel. Sea Plaque low gelling temperature agarose was used so that the DNA fraction could be extracted by melting a slice of the gel. The presence of the hybridising fragment in this fraction was confirmed by hybridisation to a sample of the eluted DNA run on another gel.

The vector, pUC12, DNA was also digested with EcoRI and half was then treated with phosphatase to prevent self-ligation. The insert DNA was ligated into both phosphatased and unphosphatased vector, because phosphatase treatment sometimes damages the DNA preventing ligation (caused by contaminating endonucleases in the phosphatase). The products were used to transform *E. coli* strain TG1 and ampicillin resistant colonies were selected.

Recombinant plasmids should be recognisable by the loss of β-galactosidase activity due to insertion of DNA into the *lacZ* gene at the polylinker. On plates containing the chromogenic substrate X-Gal and an inducer of the gene, IPTG, recombinants remain white whereas the parental vector gives blue colonies. Approximately 140 white colonies were patched onto a master plate for storage and onto a nitrocellulose filter on a second plate. The majority of these colonies came from the ligations with phosphatased vector.

The filters were probed with the oligonucleotide and one hybridising patch was found. After purifying this strain by streaking out and rescreening, a rapid plasmid preparation showed that the insert was the same size as the hybridising band in the chromosomal digest. This plasmid was designated pND1 and was used in all subsequent experiments.

3.5.1 Repeat screening for epimerase clones using Whatman 541 paper

Methods for colony hybridisations using Whatman 541 filter paper in place of the normal nitrocellulose have been published (Gergen et al., 1979; Carroll and Porter, 1983), and just at the time of the above experiments a modification was published in which the filters were steamed at the sodium hydroxide denaturation step (Maas, 1983). This procedure caused the cells to be lysed much more efficiently, and approx. 100 times more DNA was stuck to the filters, giving a much stronger signal in hybridisations. Approx. 600 new transformants were screened directly, without patching out, using Whatman 541 paper and the steaming procedure. Five extremely hot spots were observed on the autoradiogram. These were true positives, because, when streaked out to ensure pure colonies and then rescreened, the same strong signal was again observed. The plasmids were purified from two of these and the inserts were found to be the same size as that in pND1. This showed that the insert was stably maintained in the E. coli plasmid.

The colonies were rescreened by patching to both nitrocellulose and 541 paper filters, which were then hybridised at the same time. The 541 paper gave much stronger and more even signals. The strong signals obtained were partly due to the high copy number of the plasmid, and those on 541 paper were stronger due to the better denaturation. The background in these hybridisations was much lower than in the first hybridisation to chromosomal DNA, because unincorporated label was efficiently removed by a DEAE-Sepharose column before the probe was added to the hybridisation mix, whereas previously a short Sephadex G25 column was used which did not work so well.

3.6 Restriction map of pND1

A large scale preparation of plasmid pND1 DNA was made. Additional treatments with ammonium acetate and RNAse-T_1 were used to remove most contaminating RNA, and some of the preparation was further purified to remove remaining RNA and chromosomal DNA by NACS-20 column chromatography.

A restriction map was made by performing single, double and some triple digests with several six base-pair cutting restriction enzymes. The digests

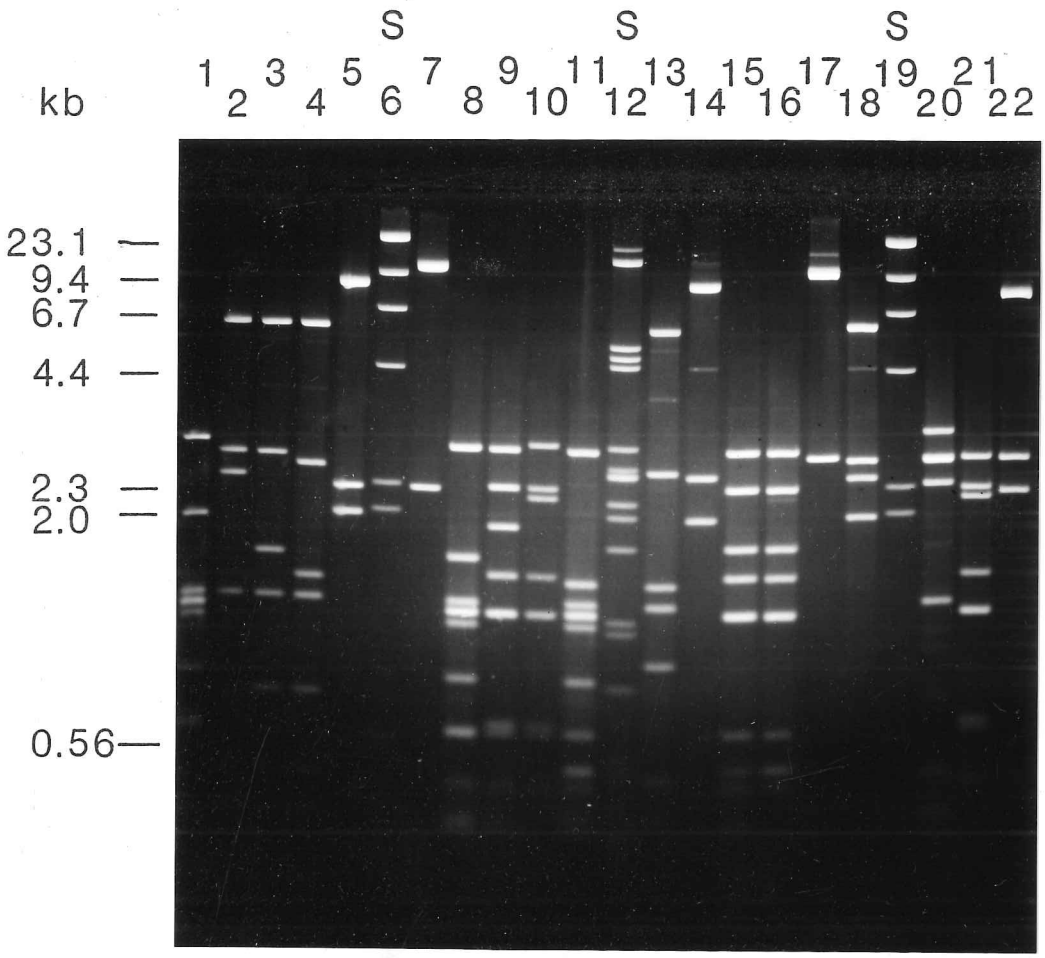


Figure 3.6 One of the gels for determining the restriction map of pND1

The gel was 20x20cm containing 250ml 1% agarose and was run for 18hr at 50V. Tracks 6 and 19 are λDNA digested with HindIII and track 12 is λDNA digested with PstI. The digests of pND1 are:1, SacI + XhoI; 2, SacI; 3, SacI + BstI; 4, SacI + BstI + PstI; 5, BstI + PstI; 7, BstI; 8, SacI + BstI + XhoI; 9, BstI + XhoI; 10, XhoI; 11, SacI + PstI + XhoI; 13, SacI + PstI; 14, PstI; 15, BstI + PstI + XhoI; 16, PstI + XhoI; 17, EcoRI; 18, PstI + EcoRI; 20, SacI + EcoRI; 21, XhoI + EcoRI; 22, BstI + EcoRI. BstI is identical to BamHI and SacI is identical to SstI.

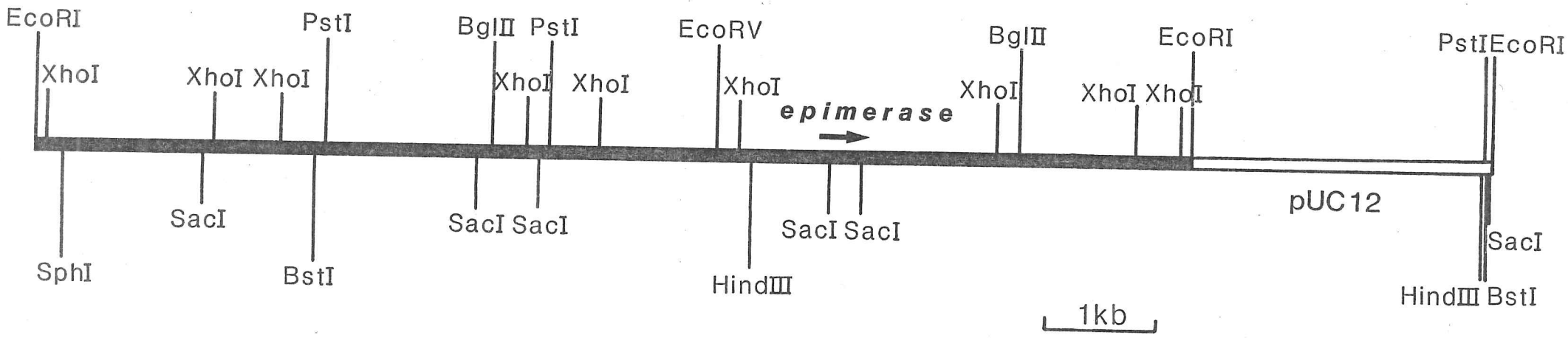


Figure 3.7 Restriction map of pND1

Plasmid pND1 is shown linearised through one of the two EcoRI sites. The insert was an EcoRI fragment inserted in the EcoRI site of pUC12. This site is at one end of the polylinker in that plasmid. The restriction map was constructed from analysis of single, double and triple digestions. Sites for the following enzymes were mapped: BglII, BstI (identical to BamHI), EcoRI, EcoRV, HindIII, PstI, SacI (identical to SstI), SphI, XhoI.

were fractionated on 0.7 and 1.0% agarose gels and the sizes of the fragments estimated from the mobilities relative to HindIII and PstI cut λDNA. One of these mapping gels is shown in Figure 3.6 and the map obtained from all the digestions is presented in Figure 3.7.

As an aid to creating the map the gels were probed with the oligonucleotide so that the hybridising fragment in each digest could be identified. In all the single digests the hybridising fragments were the same size as the hybridising band in the corresponding chromosomal digestion if the fragment was entirely contained within the insert, or of a size consistent with the corresponding band if the fragment extended beyond the length of the insert, showing that there had been no deletion or rearrangement in the cloned DNA.

3.7 Discussion

Several oligonucleotides were synthesised. These were complementary to the possible DNA coding sequences deduced from part of the epimerase N-terminal amino acid sequence determined on an automated solid-phase sequenator. The oligonucleotide in which each A/G and T/C ambiguity was represented by G or C respectively was found to be the correct sequence binding stringently to one size fragment in each restriction digest of *P. shermanii* chromosomal DNA. It was shown that the hybridising sequence lay in the middle of a 10.4kb EcoRI restriction fragment. This fragment was cloned into *E. coli* using the vector pUC12, and a restriction map was made.

A DEAE-sephadex column to remove unincorporated label from the probe greatly reduced background hybridisation. Dried gels were probed directly with radiolabelled oligonucleotide, rather than blotting the DNA to nitrocellulose. Replacing nitrocellulose with Whatman 541 paper for colony screens allowed the filters to be steamed, giving better lysis and thus stronger signals.

CHAPTER 4

EXPRESSION OF THE CLONED EPIMERASE GENE

4. EXPRESSION OF THE CLONED EPIMERASE GENE

An oligonucleotide had been made for the DNA sequence deduced from part of the N-terminal amino acid sequence of the epimerase protein. This had been used as a probe to clone a restriction fragment carrying that sequence into the *E. coli* plasmid pUC12 as descibed in Chapter 3.

To check that the cloned fragment did indeed carry the epimerase gene the DNA could be sequenced. This would reveal whether or not the translation of the sequence agreed with the N-terminal protein sequence of epimerase, with an open reading frame of the correct length and amino acid composition beyond the known amino acid sequence. This was done (Chapter 5), but to show that the whole gene was present intact, and could direct the synthesis of active enzyme, it was necessary to express the gene.

Expression of the proteins coded by the cloned DNA might also reveal the presence of the mutase or transcarboxylase genes by probing with the relevant antiserum, whereas the nucleotide sequence would not do so without, at least, partial knowledge of their amino acid sequences. If other genes of interest were not present on the large fragment it would only be necessary to sequence that part encoding the epimerase.

E. coli was not expected to recognise the promoters of *P. shermanii*, because *E. coli* often does not recognise foreign promoters, particularly those from Gram positive and high G+C content bacteria (see e.g. Thompson *et al.*, 1984), although the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* is well expressed from its own promoter when cloned in *E. coli* (Barker, 1982). A western blot of total protein from the *E. coli* strain carrying pND1 showed that no protein was synthesised that cross reacted with an anti-epimerase antiserum raised in a rabbit (Figure 4.5).

4.1 In vitro expression of epimerase

In order to express the proteins encoded by pND1 an *S. lividans* transcription-translation system was used (Thompson *et al.*, 1984). This Streptomycete cell-free system was employed because it is a Gram positive bacterium with similar G+C content to *P. shermanii*, and also because it was hoped to clone the gene into Streptomycetes later.

The DNA to be tested was incubated with the extract and [^{35}S]-methionine to radiolabel the products. The resultant mixture was then incubated with antiserum and Protein-A Sepharose. Any protein bound to the antibodies could then be isolated from the mixture by centrifugation, because it would be bound to the Sepharose beads through the Protein-A. Bound proteins were solubilised from the pellet and electrophoresed on an SDS-polyacrylamide gel, which was autoradiographed. Any radioactive protein seen must have been bound by the antibody.

The proteins synthesised by the extract when programmed with pND1 DNA were incubated in separate experiments with antisera, raised in rabbits, to each of the proteins mutase, epimerase and transcarboxylase. The immunoprecipitated products showed that epimerase was made, but not mutase or transcarboxylase (Figure 4.1).

4.1.1 Transcription-translation of digested pND1

The restriction map of pND1 (Chapter 3) suggested that the epimerase gene might be contained in a 2.3kb Xho1 fragment. DNA of pND1 was digested with Xho1 and then used to direct protein synthesis in the transcription-translation system (Figure 4.2). Full size epimerase was obtained, indicating that all the information necessary for the expression of the epimerase gene was present on the fragment. Therefore, the extract was recognising the *P. shermanii* promoter and had not been expressing the epimerase by readthrough from an *E. coli* promoter in the vector. The epimerase was the same size as an authentic sample of enzyme that had been purified from *P. shermanii* and labelled with [^{14}C]-NEM.

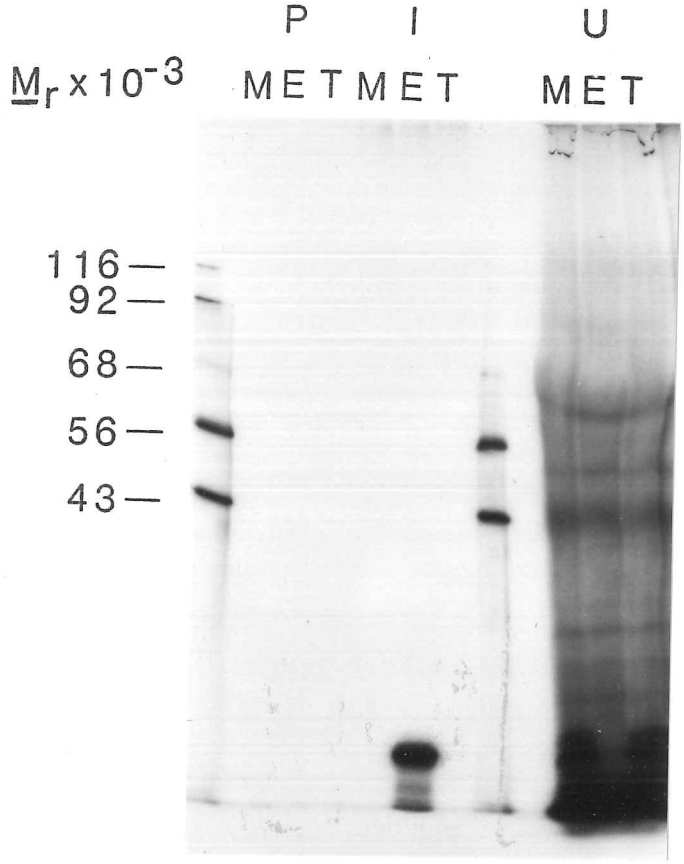


Figure 4.1 Transcription-translation of pND1

The products of transcription-translation of pND1 by an S. lividans cell-free extract were immunoprecipitated with preimmune (P) and immune (I) antisera raised against mutase (M), epimerase (E) and transcarboxylase (T). The unprecipitated proteins (U) were then precipitated with 7.5% trichloroacetic acid for 15min at 70°C to hydrolyse nucleic acid. These precipitates were washed with 7.5% trichloroacetic acid followed by acetone. All samples were then electrophoresed on a 12% SDS-polyacrylamide gel.

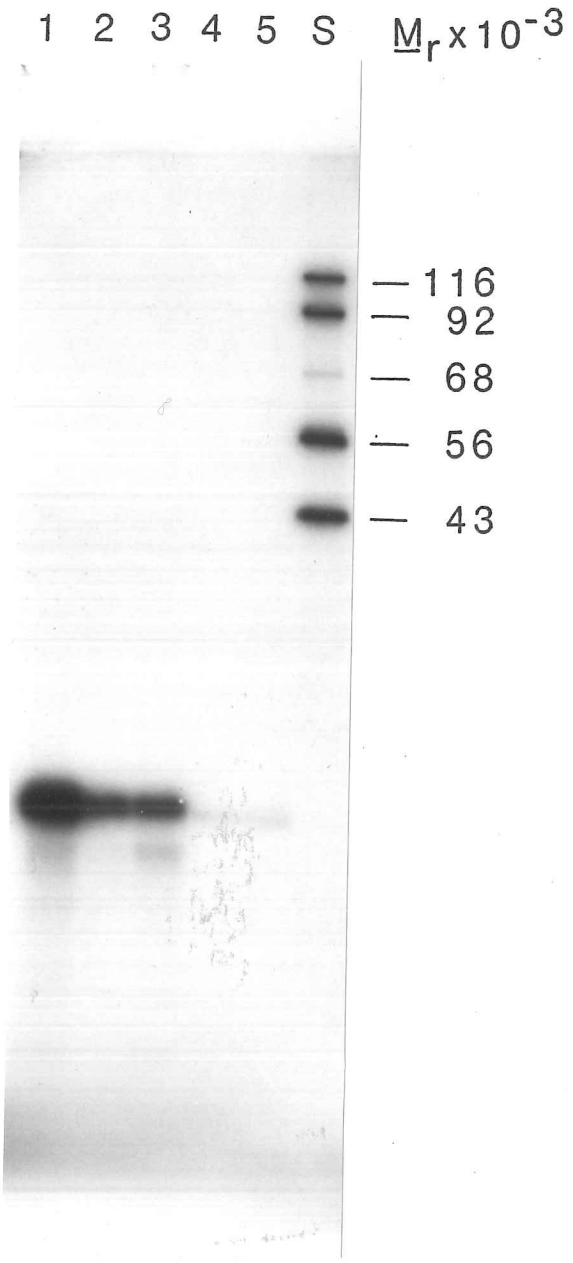


Figure 4.2 Transcription-translation of digested pND1

Plasmid pND1 was digested with XhoI prior to transcription-translation. The products were immunoprecipitated with anti-epimerase antiserum, and electrophoresed on a 20% SDS-polyacrylamide gel. Track 1, the products of undigested pND1; 2, the products of XhoI digested pND1; 3, the products of pND1 partially digested with SmaI, the digest did not go to completion; 4, authentic epimerase labelled with [^{14}C]-NEM; 5, as 4 but immunoprecipitated; S, standards.

4.1.2 Transcription-translation of P. shermanii chromosomal DNA

The total products made by the cell-free system with pND1 showed that epimerase was by far the major one. The next strongest band is due to the vector, and is probably the ampicillin resistance protein, which is degraded by this system (Thompson et al., 1984). The large amount of epimerase made indicates that this promoter is very strong in S. lividans. This may be fortuitous, or it may that the promoter is one of the strongest in P. shermanii. Possibly the promoter could be down-regulated in vivo by the product of another gene that has not been cloned on this fragment. Alternatively, all P. shermanii promoters might be well recognised by S. lividans but other genes on this fragment are read from promoters outside the region cloned.

Total P. shermanii chromosomal DNA was used to direct protein synthesis in the cell-free system and the products immunoprecipitated with the three antisera. Epimerase was just detectable after a very long exposure, also confirming that the extract could recognise at least one P. shermanii promoter. However, mutase and transcarboxylase were not detected, suggesting that their promoters are not so strong or not so well recognised (a promoter of less than one fifth the strength of that for epimerase would not have been detected by this exposure).

Possibly if some promoters are very strong they may reduce expression at weaker promoters by sequestering the RNA polymerase. Immunoprecipitated products from transcription-translation of pND1 were also given the long exposure and still no mutase or transcarboxylase were seen. It was concluded that the promoters for these genes were absent, because, unless they were well over 1 000 times weaker than the epimerase promoter, some expression should be seen at that sensitivity, although it is possible that S. lividans does not recognise these promoters. There was no expression of mutase or transcarboxylase by read-through from vector promoters either, so the genes were probably completely absent.

4.2 In vivo expression of epimerase

The *in vitro* transcription-translation showed that an *S. lividans* cell extract could make epimerase from the cloned gene. Epimerase should, therefore, be expressed *in vivo* when the gene was cloned into *S. lividans*.

The 4.7kb BglII fragment of pND1 that includes all of the 2.3kb XhoI fragment, and thus the epimerase gene, was isolated from an agarose gel using DE81 paper, and ligated into BglII-digested and phosphatase-treated pIJ702. The products were transformed into *S. lividans* strain TK24 and plasmid containing cells were selected with thiostrepton.

This high copy number plasmid contains a tyrosinase gene which produces melanin from tyrosine, turning the colonies black. The gene is disrupted by insertions into the BglII site giving white recombinants. However, this screen was not found to be necessary, as using phosphatase treated vector few black colonies were obtained. The transformants were allowed to sporulate (approx. 5 days) and then a piece of nitrocellulose was laid on top. This was removed and laid, spore side up, on a fresh plate and incubated for two days. Pieces of Whatman 541 paper were laid on the same colonies and the spores allowed to grow on these also. The cells on each filter were then lysed and the DNA that bound to the filter was probed with the oligonucleotide. Positive colonies were picked and streaked out, then patched out and rescreened to ensure purity. The autoradiogram of the initial screen (Figure 4.3) shows that the 541 paper gives stronger, more even signals than the nitrocellulose.

Plasmid DNA was prepared from the positive strains. Digestions with BglII and XhoI showed that the expected insert had been obtained in both orientations in different clones (pND2 and pND3) as shown in Figure 4.4.

The positive strains were grown in YEME liquid culture for 2 days and expressed a major protein of the same molecular weight as epimerase. A Western blot of the total soluble proteins from these strains showed that this protein species could bind anti-epimerase antibodies (Figure 4.5). TK24/pND2 and TK24/pND3 both gave the same level of expression, indicating that epimerase is expressed *in vivo* from its own promoter, and that the

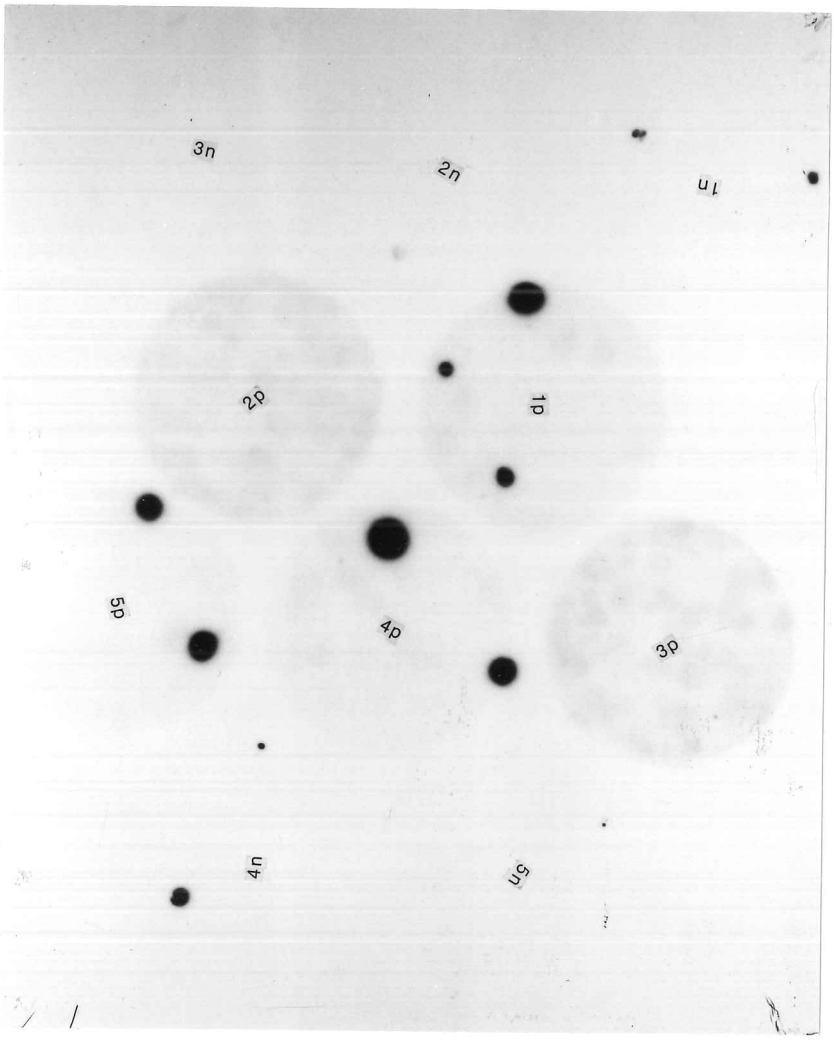


Figure 4.3 Nitrocellulose and 541 paper screens of S. lividans clones

Five plates of transformants from separate ligations of the epimerase-carrying BglII fragment of pND1 into pIJ702 were screened. The colonies were allowed to sporulate and then nitrocellulose (n) and 541 paper (p) filters were laid in turn on the plates. The colony replicas were incubated for two days before screening. The numbering on each pair of filters is in the same orientation with respect to the original plate.

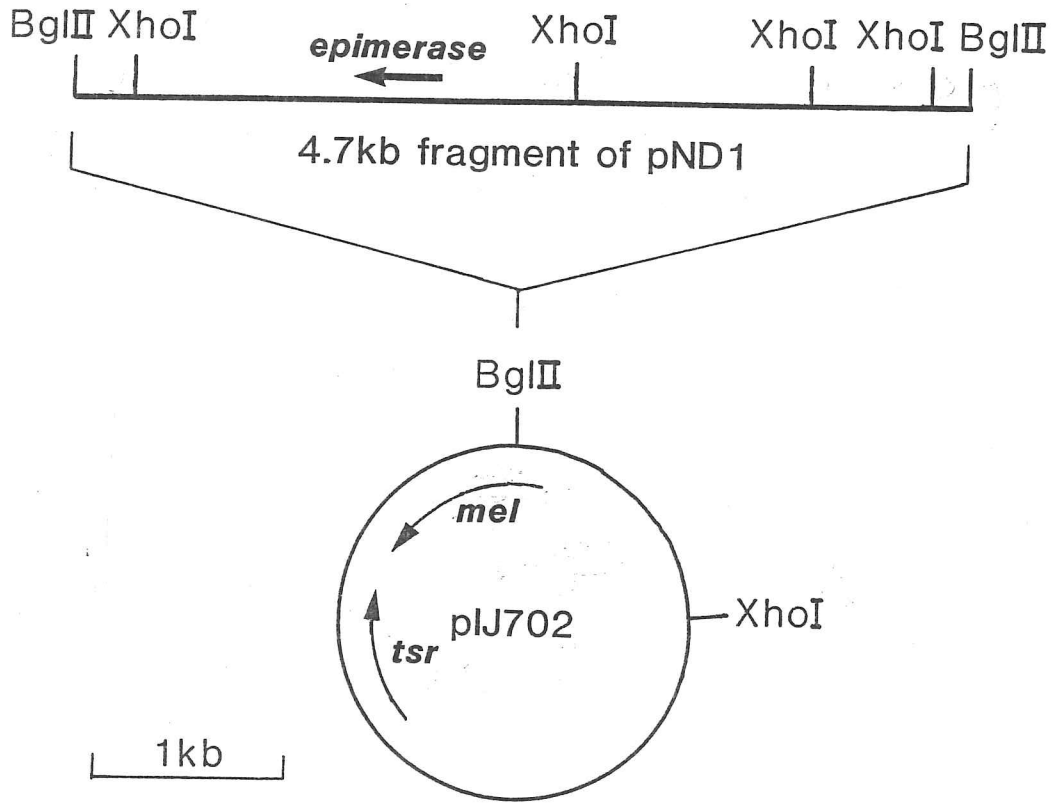


Figure 4.4 Construction of the S. lividans subclones

The BglII fragment carrying the epimerase gene was taken from pND1 and inserted at the BglII site of pIJ702. The insert in pND2 was in the orientation shown and the insert in pND3 was in the opposite orientation. Insertion at the BglII site of pIJ702 inactivates the melanin gene (mel). Plasmid-containing cells could be selected by thiostrepton resistance (tsr).

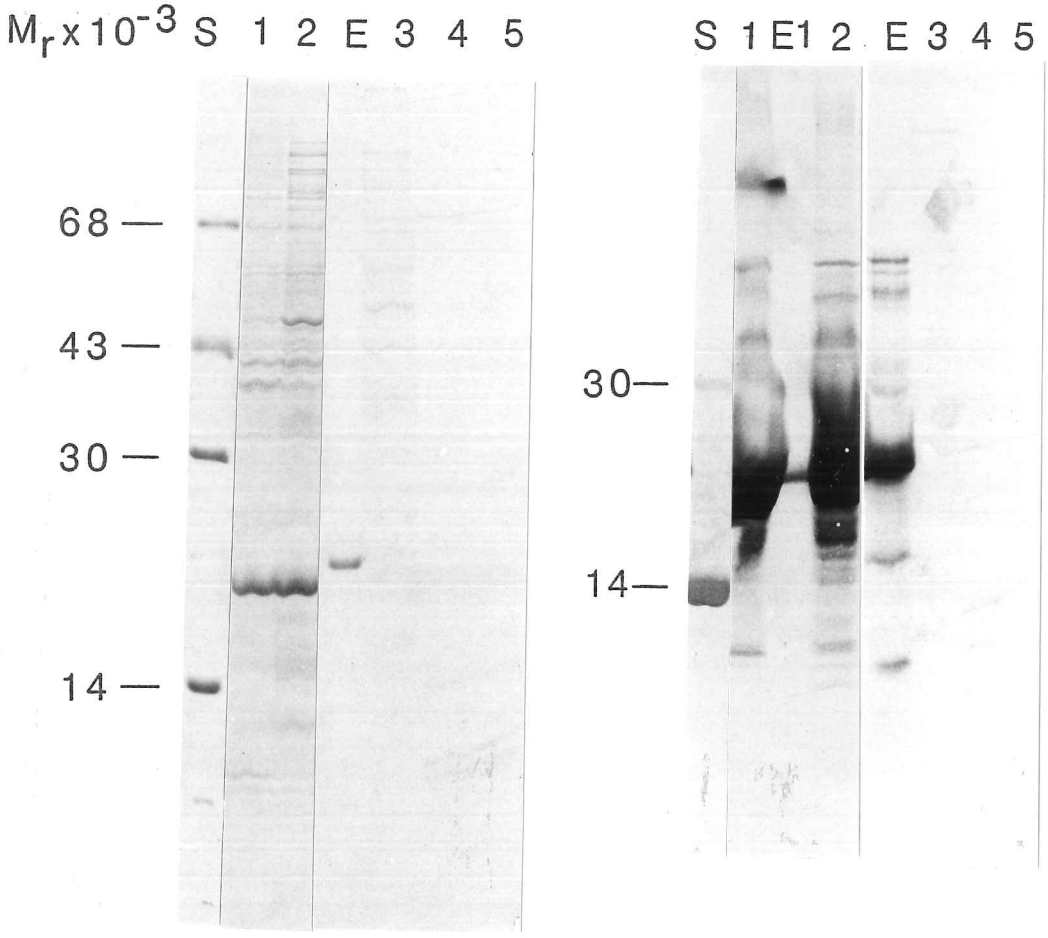


Figure 4.5 Western blot of the proteins from S. lividans clones

Total soluble protein from sonicated cells was electrophoresed on a 20% SDS-polyacrylamide gel and electroblotted to nitrocellulose. Anti-epimerase antiserum and a second-antibody-horseradish peroxidase complex were used to detect epimerase. Track 3 contains extract of S. lividans strain TK24, and tracks 1 and 2 contain TK24/pND2 and TK24/pND3 extracts respectively. Track 4 contains E. coli strain TG1, and track 5 TG1/pND1. Epimerase purified from P. shermanii was in track E (3µg) and track E1 (30ng).

strength of this promoter does not adversely affect plasmid functions.

4.2.1 Purification of the expressed epimerase

The epimerase assay is an NADH-linked assay but crude S. lividans extract is rich in NADH-oxidase activity. This makes the assay unusable on the crude extract, so the enzyme was purified. Pure enzyme could also be used to determine the amino acid composition and N-terminal sequence to confirm the identity of the product. Epimerase purified from a background containing no transcarboxylase or mutase would aid enzymological studies on all three enzymes, because there would be no contamination with the other enzymes. Also, because the enzyme is so abundant in the clone, large amounts could be purified, enabling work on crystallisation of the protein to begin. If crystals are obtained they should yield a great deal of information, because the enzyme is small and the full sequence is known (Chapter 5), which will greatly aid the interpretation of the crystallogaphic data. The enzyme catalyses a single substrate reaction in which the product is virtually identical to the substrate with an equilibrium constant close to one. Comparison of crystals with and without the substrate present should help identify the active site, catalytic residues, and regions that move when the substrate is bound.

The two-day culture of TK24/pND2 was used to innoculate a 15l culture in a fermentor. This was harvested after 64 hours and contained a high proportion of the epimerase-sized protein (as expected from Figure 4.5) Epimerase was purified essentially as described by Leadlay (1981). The supernatant from broken cells was fractionated by precipitation with ammonium sulphate and then on a Sephadex G75-superfine column. At this stage the epimerase containing fractions contained no NADH-oxidase activity and catalysed the epimerase reaction, showing that the cloned enzyme was active (Figure 4.6).

The active fractions, containing 90-95% pure epimerase were loaded on a DE-sepharose column and eluted with a salt gradient. This concentrated the protein but afforded little purification, possibly due to overloading the column. The purest fractions were further chromatographed on a hydroxyapatite column using a salt gradient. This procedure yielded pure epimerase with a

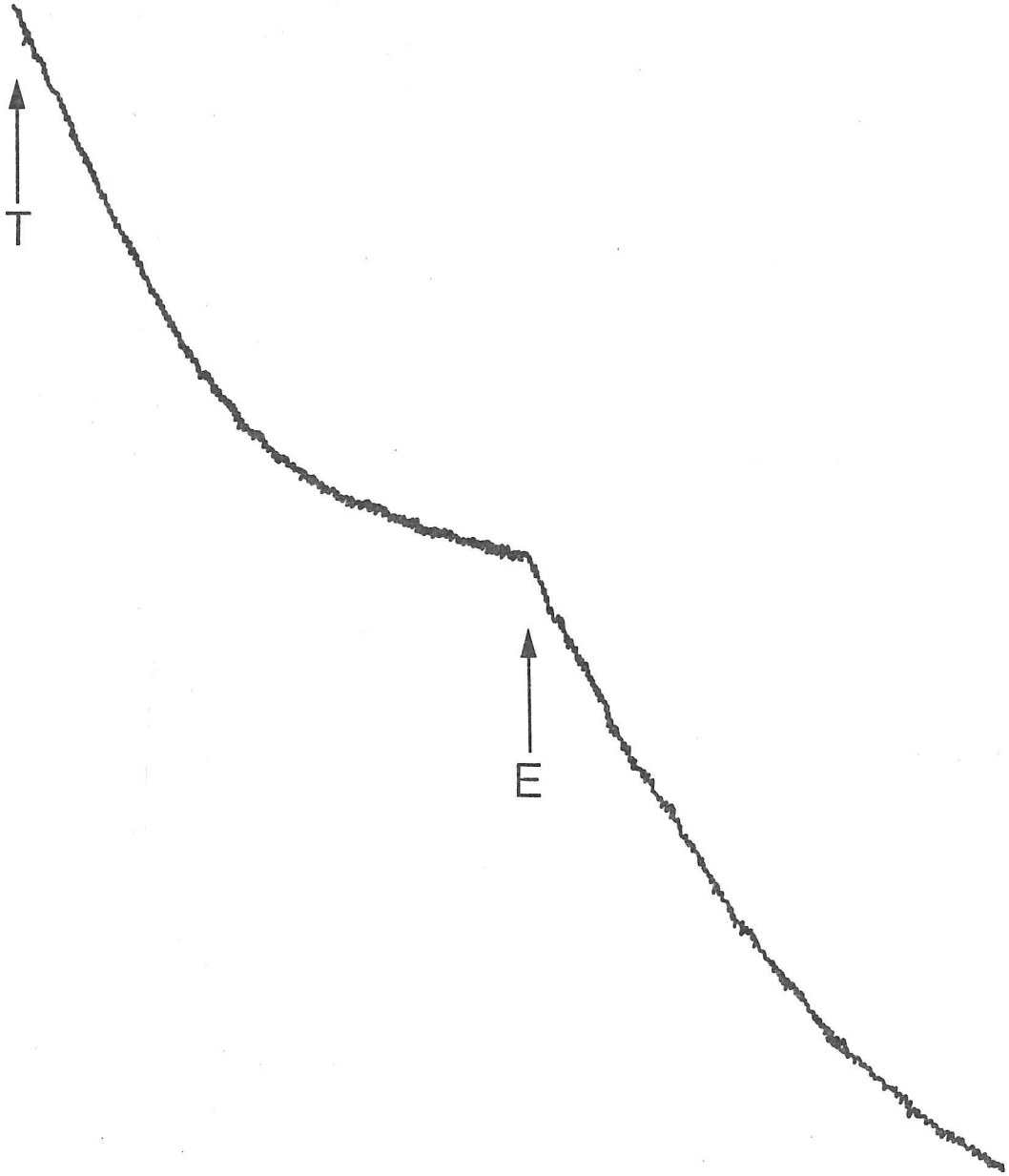


Figure 4.6 Assay of the cloned epimerase

The reduction in NADH absorbance measures the amount of succinate reduced to malate by malate dehydrogenase. Succinate is produced from pyruvate by transcarboxylase as (2S)-methylmalonyl-CoA is decarboxylated. The (2S)-form is removed from (RS)-methylmalonyl-CoA when transcarboxylase is added (T). Once this isomer has been removed epimerase is added (E), converting the (2R)- form to the (2S)- form, and so causing more succinate to be formed. Transcarboxylase can take the (2R)- form at a low rate but the enzyme used here was contaminated with some epimerase.

small amount of one smaller band that co-chromatographed with epimerase on each column, and is probably the product of slight proteolysis of one subunit in some of the epimerase dimers.

A sample of the purified epimerase was sequenced on an automated sequenator by Prof. Fothergill at Aberdeen. The first eleven residues were obtained, and agreed with the sequences previously obtained from the DNA and the protein purified from P. shermanii. The protein was in two forms: 60% had methionine at the N-terminus, whereas 40% was one residue shorter and started with serine.

It was not possible to accurately quantitate the enzymatic activity of the epimerase as the only available sample of the coupling enzyme, transcarboxylase, was not very active and contaminated with epimerase. However, the epimerase gene cloned in S. lividans clearly produces active enzyme and in large amounts: 250mg was obtained from 160g cells. The purification procedure was not optimised, and a large amount of epimerase, present in slightly less pure side fractions from each step, was not taken on to the next and fully purified. With an improved purification, omitting the DE column and possibly adding the poly(ethylene glycol) fractionation used by Leadlay (1981), over 0.5g of pure epimerase should be readily obtainable from a 15l culture of TK24/pND2.

4.3 Discussion

The plasmid pND1 was shown to carry the epimerase gene. The DNA directed the synthesis of a protein that was the same size as epimerase, and was precipitable by anti-epimerase antibodies, in a S. lividans cell-free transcription-translation system. Mutase and transcarboxylase were not present on this clone. A 2.3kb XhoI fragment from the insert also gave expression, showing that the P. shermanii promoter was recognised by S. lividans. The expression from this promoter was in fact very strong.

Part of the original insert was subcloned into S. lividans in a multicopy plasmid vector. This gave overexpression of a protein that was confirmed to be epimerase by the enzymic activity and N-terminal amino acid sequence of

the purified protein. The epimerase was expressed as a substantial proportion of the total cell protein (over 10%), confirming that the P. shermanii promoter is strong in S. lividans in vivo. This clone should provide a good source of epimerase free of transcarboxylase and mutase for enzymological studies on these enzymes, and the high level of expression gives large amounts of pure enzyme which may allow crystallisation of the protein and elucidation of the full tertiary structure.

CHAPTER 5

NUCLEOTIDE SEQUENCE OF THE EPIMERASE GENE

5. NUCLEOTIDE SEQUENCE OF THE EPIMERASE GENE

The *P. shermanii* methylmalonyl-CoA epimerase gene had been cloned into *E. coli*, as shown by the *in vitro* expression of the DNA by an *S. lividans* cell-free system. This also showed that the genes for mutase and transcarboxylase were not present on the same fragment. There was no need, therefore, to sequence the entire 10.4kb insert, but only that portion carrying the epimerase gene. The *in vitro* transcription-translation with pND1 DNA digested with the restriction enzyme XhoI showed that all the information required to express the epimerase gene was present on a XhoI fragment. The 2.3kb XhoI fragment that hybridised with the oligonucleotide probe was, therefore, sequenced.

5.1 Nucleotide sequence of the 2.3kb XhoI fragment

The 2.3kb XhoI fragment of pND1, carrying the epimerase gene, was isolated from an agarose gel using DE81 paper. The fragment was ligated to itself, to prevent overrepresentation of the ends, and sonicated. The sheared fragments were fractionated on an agarose gel and pieces of length 350-1 000bp were isolated using DE81 paper. These fragments were then end-repaired, to give blunt ends, and ligated into M13mp10 that had been cut with SmaI and phosphatased. The ligation mix was transformed into *E. coli* strain TG1. White plaques were picked (recombinant phage have an inactivated β-galactosidase gene so remain white on plates containing IPTG and X-Gal) and single-stranded bacteriophage DNA was prepared.

Seventy-two clones were sequenced by the Sanger dideoxy chain-termination method (Sanger *et al.*, 1977 and 1980). The primer used (3' TGTTGCAGCACTGACCC) was synthesised on a Kieselguhr resin support and purified by gel-filtration, gel-electrophoresis and an alkaline salt gradient run on a NACS-20 FPLC column. The DNA sequence was read from the autoradiograms (one of which is shown in Figure 5.1) into a microcomputer using a sonic digitiser. The individual gel readings were then aligned on a main-frame computer.

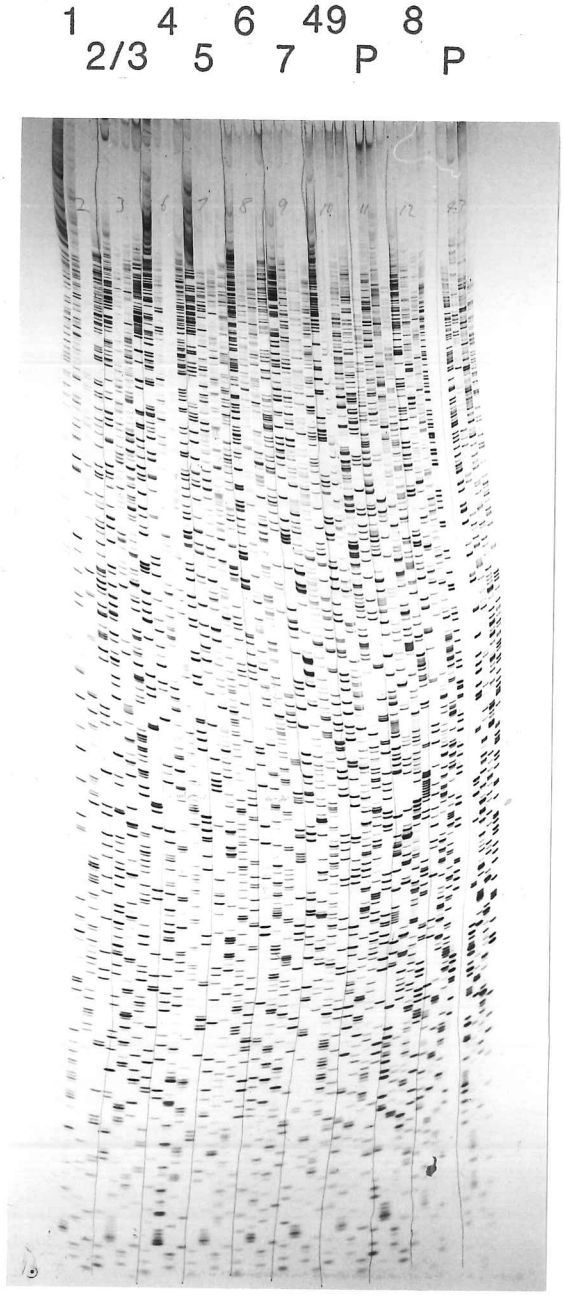


Figure 5.1 A DNA sequencing gel

The sequence ladders of ten clones are shown. The four tracks for each are A, C, G and T. The clone numbers correspond to gel readings in the Appendix. Clones (P) are the two clones derived from some pUC12 present in the preparation of the fragment to be sequenced.

The full sequence of the 2.3kb XhoI fragment is given in the Appendix, together with the gel readings. Of the 72 clones sequenced 57 are included in the sequence. Five were vector with no insert, one had deleted part of the M13 vector and another gave no sequence, probably due to the priming site being deleted. Two gave pUC12 sequence, contaminants from the adjacent band on the gel from which the fragment was isolated. The remaining six had very small inserts (16-101bp) with as little as 34% G+C content. These may have derived from the tRNA that was used in the elution of the DNA from the DE81 paper. This was found to be unnecessary for elution and was omitted in other experiments.

The total sequence is 2 263bp long and is covered by at least four, and on average by 9.5, gel readings at each nucleotide. Each position is covered on both strands except for 1 723-1 799 which is well outside the epimerase coding region. On some gels certain bases are unresolved. These compressions are caused by the DNA forming secondary structure despite the 8M urea in the gel. In many cases other gels covering the same region in the same direction are readable because the problem sequence is running in a hotter, more denaturing, part of the gel. Occasionally the sequence had to be inferred from that on the opposite strand, the gels for which are not usually compressed at the same point in the sequence. Compressions may often be resolved by using 25-40% formamide as well as urea to give even more denaturing conditions (Sanger et al., 1982). Alternatively, inosine can be used in place of guanine in the reactions. Inosine does not base-pair well and prevents the formation of secondary structure (Mills and Kramer, 1979; Sanger et al., 1982). These methods were not employed because the only unresolvable compression was in the region 320-350 which was not part of the epimerase gene. The number of bases in this region appears to be correct, as described below, but there may be an error in the order given.

5.2 Reading frames in the 2.3kb sequence

The high G+C content of P. shermanii DNA means that initiation and termination codons for proteins are rare since they all contain two or three A or T bases (ATG, TGA, TAA, TAG). Stop codons should occur at only 1.9% triplets of bases, instead of the usual 4.7%, making the possibility of long

open reading frames occurring by chance greater than normal. Initiation codons should occur at a frequency of 0.8%, instead of the expected 1.6% for 50% G+C DNA, which might reduce the amount of ambiguity about possible start sites. However, many Streptomycete proteins (approx. 30%) appear to start at GTG (Valine) instead of ATG (Methionine), and the similar G+C content of P. shermanii and Streptomycete DNA may mean that some P. shermanii proteins are also initiated by valine (Bibb et al., 1984). The sequence of the epimerase containing fragment does, in fact, contain even less stop codons among the triplets of bases than expected (1.6%), but has more methionine codons (1.7%). Also, even when part of the protein sequence is known, in this case the N-terminal 32 amino acids, it is possible that an error in the reading might give a frame-shift that might not be noticed if the resultant protein was of approximately the correct size.

Many researchers compare the codon usage of each possible reading frame with the known codon usage for that organism to determine which sequences are actually protein coding. No genes had been sequenced from P. shermanii before this sequence was obtained so the usual codon usage analysis could not be used. However, the high G+C content of the DNA allows the use of a similar method. As was argued for the choice of oligonucleotide as probe (see Section 3.2), in a coding region the second position in each codon is determined by the amino acid to be used and is 35-45% G+C. To keep the overall G+C content high, the third codon position has a very high G+C content (Bibb et al., 1984). This analysis works well for predicting Streptomycete protein coding sequences, and the recently published DNA sequence of the biotinyl subunit of transcarboxylase from P. shermanii also follows this pattern (Murtif et al., 1985).

A computer program was written based on this idea, plotting the G+C content of each of the three codon positions on three superimposed graphs. The G+C content is averaged over a number of codons, and plotted as a funtion of the position in the sequence. Where the sequence is not coding for a protein all three positions should have a similar G+C content and the lines will be close together, whereas where there is a protein coding sequence the three lines should separate, one dropping to approx. 40% and another rising to 85-95%.

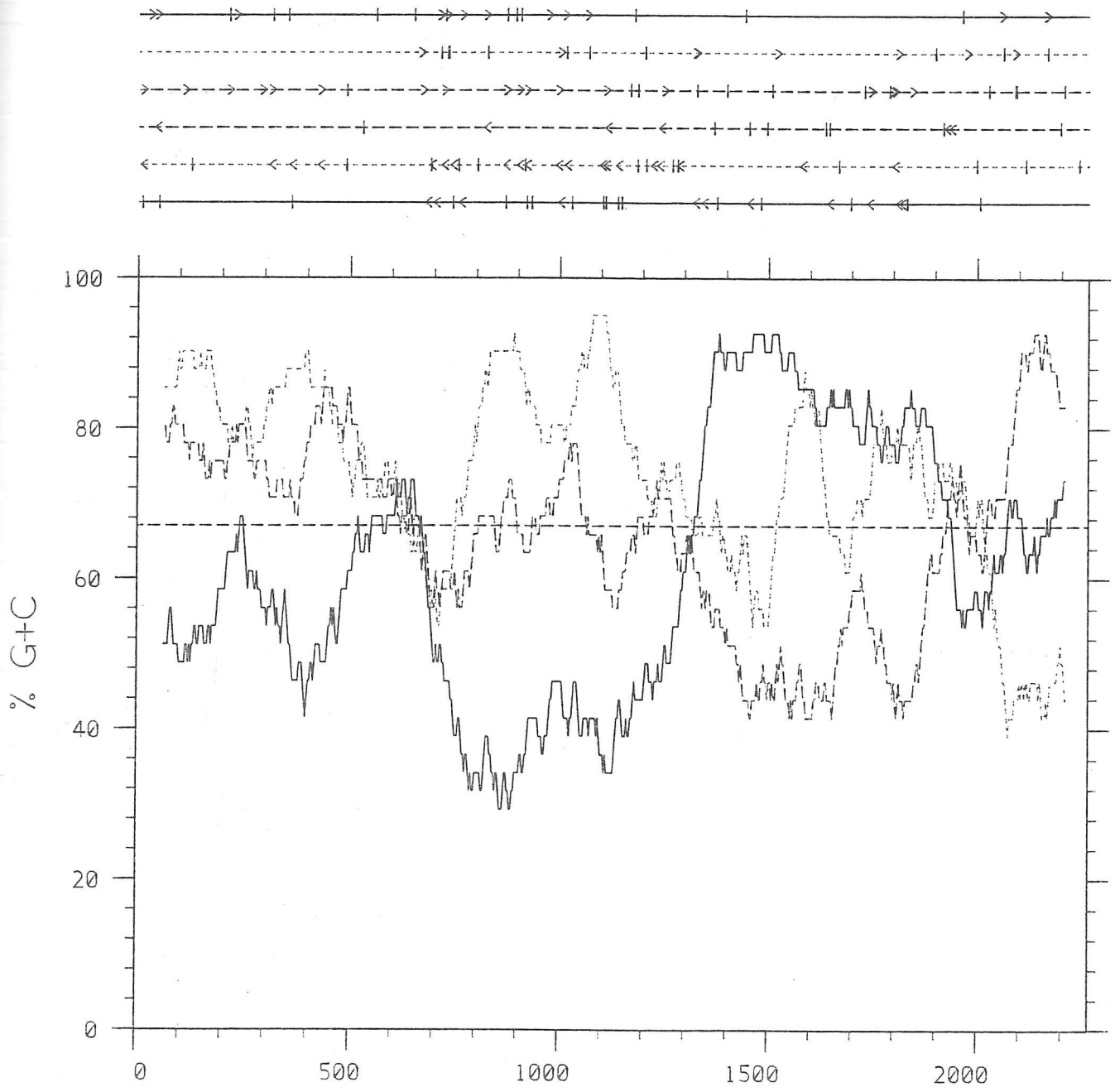


Figure 5.2 Graph of G+C content of each codon position of the 2.3kb XhoI restriction fragment carrying epimerase

The G+C content of each codon position, averaged over 41 codons, are plotted: position 1 (——), position 2 (·····), position 3 (– – –). The sequence is in the orientation of the coding strand of the epimerase gene and the sequence in the Appendix. Stop codons (|) and methionine codons (> or <) for all six possible reading frames are plotted above the graphs. The overall G+C content of 63.5% is also plotted.

The direction in which the protein is read is also shown by this. The second position is always the one that drops, but if the third position rises the protein is read in the same sense as the DNA sequence was read, whereas if the first position rises (which would be the third on the complementary strand read the other way) then the protein is read off the negative strand in the opposite direction. The start and stop codons are also plotted on all reading frames in each direction to allow the correlation between these and the G+C content to be seen. The graphs also show up any frame-shift errors in the sequence by the three lines suddenly crossing over as each position gains the characteristic G+C content of another.

Figure 5.2 shows the G+C content graph for the whole 2.3kb XhoI fragment and suggests that there are two complete reading frames and two partial reading frames. All four are read in the same direction. One starts outside the fragment and probably ends at position 491 (assuming that the correct number of characters are present in the unresolved compression around 340, which seems likely because the reading frame is maintained on the G+C content graph). The second runs from 723 to 1169 (the G+C content suggests that this does not start at the earlier, in frame, ATG at position 672). A third runs from 1325 or 1328 to 1900, and the fourth begins at 2059 and runs off the end of the sequenced fragment. A long possible reading frame on the opposite strand runs from 1244 to 528 but the graph shows that this is fortuitous and it is non-coding.

5.3 The epimerase gene

The reading frame from position 723 to 1169 contains the sequence which hybridises with the oligonucleotide (795-811), and when translated the amino acid sequence (Figure 5.3) shows that the protein begins with the N-terminal sequence of epimerase obtained previously, preceded by a methionine which is removed to produce the mature protein. The complete sequence encodes a protein of $\underline{M}_r$ 16 698, from which the N-terminal methionine residue is removed to yield the mature protein of $\underline{M}_r$ 16 549. This agrees well with $\underline{M}_r$ 16 500 obtained from the protein itself (Leadlay, 1981). The amino acid composition deduced from the sequence agrees fairly well with that obtained from the protein (Table 5.1). In particular, there are two cysteine residues as

```
                    5                   10                  15
Met Ser Asn Glu Asp Leu Phe Ile Cys Ile Asp His Val Ala Tyr

                   20                  25                  30
Ala Cys Pro Asp Ala Asp Glu Ala Ser Lys Tyr Tyr Gln Glu Thr

                   35                  40                  45
Phe Gly Trp His Glu Leu His Arg Glu Glu Asn Pro Glu Gln Gly

                   50                  55                  60
Val Val Glu Ile Met Met Ala Pro Ala Ala Lys Leu Thr Glu His

                   65                  70                  75
Met Thr Gln Val Gln Val Met Ala Pro Leu Asn Asp Glu Ser Thr

                   80                  85                  90
Val Ala Lys Trp Leu Ala Lys His Asn Gly Arg Ala Gly Leu His

                   95                 100                 105
His Met Ala Trp Arg Val Asp Asp Ile Asp Ala Val Ser Ala Thr

                  110                 115                 120
Leu Arg Glu Arg Gly Val Gln Leu Leu Tyr Asp Glu Pro Lys Leu

                  125                 130                 135
Gly Thr Gly Gly Asn Arg Ile Asn Phe Met His Pro Lys Ser Gly

                  140                 145
Lys Gly Val Leu Ile Glu Leu Thr Gln Tyr Pro Lys Asn
```

Figure 5.3 The amino acid sequence of epimerase

Table 5.1 Amino acid composition of epimerase

The amino acid compositions are expressed as residues per subunit. The composition from epimerase protein is taken from Leadlay (1981).

Amino acid	Composition From DNA sequence	Composition From protein
Lys	8	7.7
His	8	11.2
Arg	6	7.8
Cys	2	2.0
Asn	7	} 8.8
Asp	9	
Thr	7	7.8
Ser	5	5.9
Gln	6	}17.1
Glu	13	
Pro	7	8.8
Gly	10	11.3
Ala	14	14.1
Val	10	12.4
Met	7	5.0
Ile	6	6.0
Leu	12	12.4
Tyr	5	2.9
Phe	3	3.1
Trp	3	3.3
	148	147.6

expected, and a total of 14 arginine and lysine residues agreeing with the 14-15 peptides obtained when the protein is digested with trypsin (Leadlay, 1981).

The DNA sequence was used to search the EMBL and Genbank DNA libraries for similar sequences, and epimerase and the other putative protein sequences were used to search the Doolittle, Protein Identification Resource and Claverie protein sequence banks on the University of Cambridge IBM3081 computer. No significant homologies were found in any of the searches.

As expected from the lack of expression in E. coli there are no sequences homologous to E. coli promoter regions upstream of the epimerase gene. The expression in S. lividans shows that a Streptomycete like promoter should be present. There are at least two classes of Streptomycete promoters recognised by different RNA polymerases (Westpheling et al., 1985). One of these types is similar to E. coli promoters, and E. coli promoters are recognised by Streptomycetes but the stringency of some of the requirements appears to be relaxed. Insufficient Streptomycete promoters have been rigorously identified to determine the sequences and spacings that are important or necessary. Even if this information was available the epimerase promoter could only be unambiguously determined by mapping the RNA start position. This should be done both in P. shermanii and in the S. lividans clone.

The M13 clones prepared for sequencing may now be used to prepare highly-labelled single-stranded probes to use in analysing the transcription of the 2.3kb XhoI fragment (Burke, 1984; Hudson and Davidson, 1984). This analysis would also show whether there are any other promoter activities on the fragment. If there are they are not well-recognised by S. lividans because epimerase was the major product. Comparison of any such promoters that may be less active in S. lividans than in P. shermanii might help to identify the important components of Streptomycete promoters.

The DNA encoding the biotin-containing subunit of transcarboxylase has been cloned recently (Murtif et al., 1985). Unfortunately, this was cloned without a promoter and the sequence after the coding region was not published so no comparisons can yet be made. Although the epimerase promoter cannot be

identified yet, the DNA sequence upstream of the epimerase gene is relatively A+T rich suggesting that a promoter sequence is present in this region.

Bibb and Cohen (1982) have determined the sequence of the 3' end of 16S rRNA from S. lividans (5' GAUCACCUCCUUUCU). This part of the 16S rRNA is thought to bind to the mRNA during initiation of protein synthesis (Shine and Dalgarno, 1974), and each of the three reading-frames which start in the sequenced fragment have a sequence upstream of the putative start-codons with some homology to this.

A 13bp inverted repeat with a loop of six bases is present at 60-90bp downstream of the epimerase terminus. This may be a transcription terminator or a processing site but, again, this requires mapping of the mRNA endpoint.

5.4 Discussion

The nucleotide sequence of the 2 263bp XhoI fragment carrying the epimerase gene was determined. Open reading frames in this sequence were identified by the different G+C content in each codon position. One of these was the epimerase gene. The N-terminal protein sequence matched the DNA sequence and the amino acid compositions also agreed. The complete protein encoded consisted of 148 amino acids giving an M_r of 16 698.

Other open reading frames were also present on the fragment but no homologies were found by searching protein-sequence data-bases. No meaningful comparisons of the presumed promoter region with known promoters could be made without identifying the mRNA endpoints.

CHAPTER 6

METHYLMALONYL-CoA MUTASE

6. METHYLMALONYL-CoA MUTASE

Methylmalonyl-CoA mutase has been purified from several sources including human and sheep liver and an intestinal worm. The P. shermanii enzyme purified by Zagalak et al. (1974) was apparently smaller than the others (M_r 124 000) and consisted of two different subunits (M_r 66 000 and 61 000), whereas the eukaryotic enzymes appeared to be dimers of one type of subunit. No protein sequence had been obtained for any of these enzymes so, before the gene for this enzyme could be cloned using oligonucleotide probes, some partial peptide sequence had to be obtained. It was also necessary to prove that there were two different subunits, and that the smaller sizes were not due to proteolysis, to determine the number and length of protein coding sequences that should be expected. A good purification method would also be required to obtain native enzyme for enzymological and protein chemical studies.

6.1 Purification and properties of methylmalonyl-CoA mutase

Mutase was purified from P. shermanii in collaboration with P.F. Leadlay and F. Francalanci (Francalanci et al., 1986). The procedure is also described in detail in Section 2.6.1. In brief, the supernatant from cell paste ground in a bead mill was adsorbed to DEAE cellulose and proteins eluted by salt washes, an ammonium sulphate cut was taken, and further purification obtained by gel-filtration and ion-exchange chromatography.

The pure enzyme ran as two bands in SDS-polyacrylamide gels with M_r 79 000 and 67 000 (Figure 6.1). The ratio of the two bands was consistently 1:1, or 0.85:1 when allowance was made for the relative sizes of the two bands, determined by densitometric scanning of stained gels. The amino acid compositions of the isolated bands were extremely similar (Table 6.1), suggesting that the dye binding characteristics of the two species should also be similar, but the consistency of the ratio obtained in different preparations suggests that the smaller band is not derived from the larger by proteolysis during the preparation.

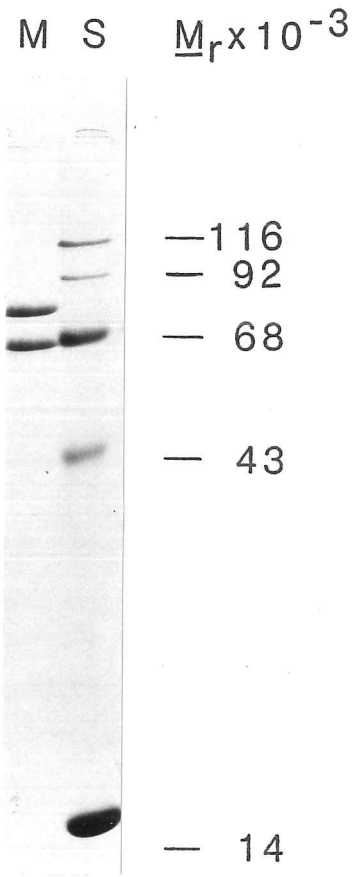


Figure 6.1 SDS-polyacrylamide gel of purified mutase

Mutase, purified from P. shermanii, was electrophoresed on a 15% SDS-polyacrylamide gel.

Table 6.1 Amino acid compositions of the mutase subunits

The compositions, expressed as residues per hundred residues, were obtained from duplicate analyses of 24, 48 and 72hr hydrolysates. Values for threonine and serine were obtained by extrapolation to zero time. Isoleucine and valine were from 72hr hydrolysates. Cysteine was determined as carboxymethylcysteine. Tryptophan was not determined.

Amino acid	Content (mol %)	
	α(M_r 79 000)	β(M_r 67 000)
Asx	8.1 ± 0.2 (6)	8.6 ± 0.3 (6)
Thr	4.8	4.9
Ser	13.1	11.7
Glx	11.4 ± 0.4 (6)	11.9 ± 0.3 (6)
Pro	3.5 ± 0.3 (6)	3.6 ± 0.5 (6)
Gly	16.7 ± 1.4 (6)	16.2 ± 0.9 (6)
Ala	8.6 ± 0.3 (6)	8.9 ± 0.2 (6)
Cys	0.8 ± 0.3 (6)	0.6 ± 0.3 (6)
Val	5.2	5.7
Met	1.5 ± 0.3 (6)	1.4 ± 0.4 (6)
Ile	3.9	4.0
Leu	6.6 ± 0.7 (6)	6.9 ± 0.2 (6)
Tyr	2.5 ± 0.2 (6)	2.4 ± 0.3 (6)
Phe	2.7 ± 0.2 (6)	2.9 ± 0.2 (6)
His	2.8 ± 0.4 (6)	2.6 ± 0.5 (6)
Trp	N.D.	N.D.
Lys	4.4 ± 0.2 (6)	4.2 ± 0.3 (6)
Arg	4.2 ± 0.2 (6)	4.3 ± 0.2 (6)

The native mutase was shown by sedimentation equilibrium analysis to have an M_r of 165 000 ± 8 000. These results suggest that mutase from P. shermanii is an αβ-dimer. M_r estimates were also obtained from sedimentation velocity experiments, 165 000 ($S^o_{20,w}$ of 7.7S), and from gel filtration on a TSK 3000SWG HPLC column, 163 000, by P.F. Leadlay (Francalanci et al., 1986).

These molecular weights are greater than those obtained by Zagalak et al. (1974) who reported a native M_r of 124 000 and subunits of M_r 66 000 and 61 000, which may have been the result of proteolytic degradation. Protease inhibitors were used throughout this preparation, a gel-filtration step was included, and the procedure was simpler than that of Zagalak et al. (1974). Mutase stored in solution at 4°C gradually lost activity and a band of lower M_r material (60 000) appeared. P.F. Leadlay separated three enzyme forms from this stored mutase by anion-exchange FPLC (Francalanci et al., 1986). One species was highly active, colourless and had the M_r 79 000 and 67 000 subunits (in the usual ratio). A second species had very little activity and was pink, binding photoinactivated B_{12}, but was indistinguishable from the first by SDS-polyacrylamide gel electophoresis. The third species was also inactive and contained material with M_r 67 000 and 60 000 in approx. 1:1 ratio. The active fraction had a specific activity of 120-150nkat mg^{-1}.

Further evidence that the true M_r of the subunits was 79 000 and 67 000 was shown by performing a Western blot of total soluble protein from freshly sonicated P. shermanii. Mutase antibody bound to bands of the same size in the crude cell extract as in the purified enzyme preparation, suggesting that the smaller subunit is not derived from the larger by proteolysis during the purification.

6.2 Partial proteolysis of the mutase subunits

The mutase from P. shermanii appears to be a dimeric enzyme. The close similarity of the amino acid compositions of the two subunits might suggest that the smaller is derived from the larger by proteolysis but the other evidence described above suggests that there are two different subunits. An alternative explanation for the similarity in the amino acid compositions is that the two peptide sequences are closely related, although non-identical.

To test this hypothesis the two subunits were each partially digested with various proteinases to look for similar peptides by the method of Cleveland et al. (1977). Mutase subunits were separated on an SDS-polyacrylamide gel. Pieces of gel containing one or other subunit were then inserted in the wells of a new gel and proteinase solutions added. All the proteins were electrophoresed into the stacking gel where digestion was allowed to proceed for 30min before the current was reapplied and the resulting peptides separated. If the smaller subunit were a part of the larger then most of the peptides should be identical from each subunit, whereas if the subunits were unrelated there should be no identical peptides. A close relation between the two subunits might give rise to some peptides occuring in digests of both subunits.

The results for digestion with several concentrations of each of four different proteinases are shown in Figure 6.2. (Increasing amounts of digestion are shown from right to left.) This shows that the smaller subunit is not a portion of the larger because the majority of peptides from each subunit are different, although the two may be related because there are some peptides of apparently identical sizes in both subunits and some similar patterns of cleavage.

The large subunit, in particular, does not run as a single species on the second gel, even though a narrow band was cut from the original gel. Presumably there has been some dimerisation of this subunit between the running of the two gels. This would have occurred during the brief staining in acetic acid/methanol/Coomassie blue. It is not known whether this is due to the acid or the removal of the SDS allowing association of the peptide chains. Mutase sometimes runs on gels with part of the larger subunit precipitated at the top of the separating gel or running at a higher than usual molecular weight. This does not, however, explain the 0.85:1 ratio of the two subunits as this was obtained from gels on which all the material was present in just the two bands.

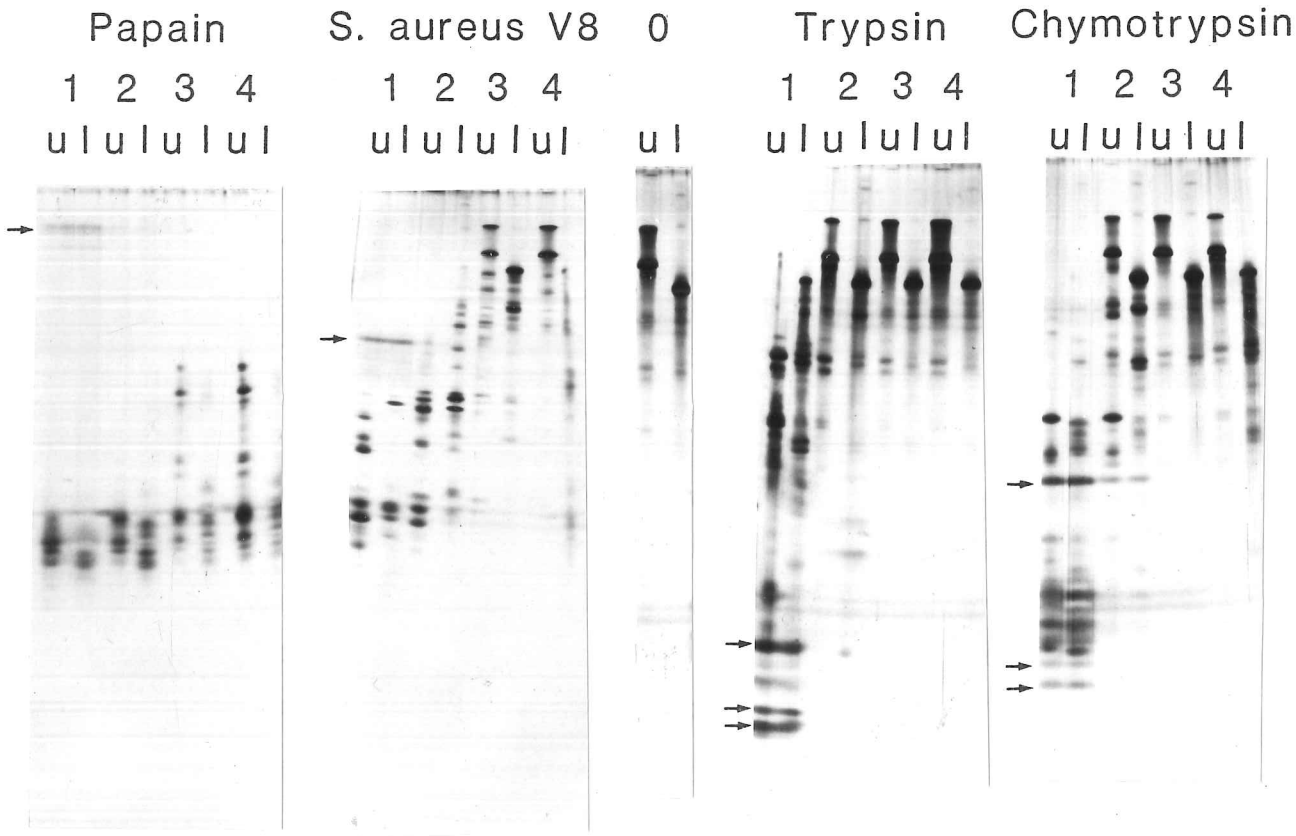


Figure 6.2 Partial proteolysis of the subunits of mutase

Gel slices containing the subunits of mutase were inserted in the wells of SDS-polyacrylamide gels (20% for trypsin and chymotrypsin, 24% for papain and S. aureus V8 proteinase). Solutions of each proteinase were added and allowed to digest mutase in the stacking gel for 30min before electrophoresis was continued. Increasing amounts of digestion are shown from right to left. Tracks 1, 2, 3, and 4 refer to 10, 100, 1 000 and 10 000 fold dilutions of a stock solution of each proteinase at 0.5mg ml^{-1}. Arrows indicate protein bands derived from the proteinase solutions. The peptides were visualised by silver-staining.

u- Upper subunit
l- lower subunit

6.3 Purification of peptide fragments of mutase

To be able to clone the mutase gene using oligonucleotide probes it was necessary to obtain some protein sequence. The amino acid compositions of the separated subunits were obtained from the two peptides eluted from an SDS-polyacrylamide gel. However, this material gave no sequence on an automated solid-phase protein sequenator. Therefore, smaller fragments were sequenced.

It was not possible to separate the subunits other than by gel electrophoresis, perhaps due to their similar compositions and also because once denatured mutase is virtually insoluble. (This may be due to hydrophobic portions, that normally bind the coenzyme, and perhaps provide contacts between the two subunits, folding and binding incorrectly after denaturation.) Because of this problem, fragments were derived from the whole protein rather than the individual subunits.

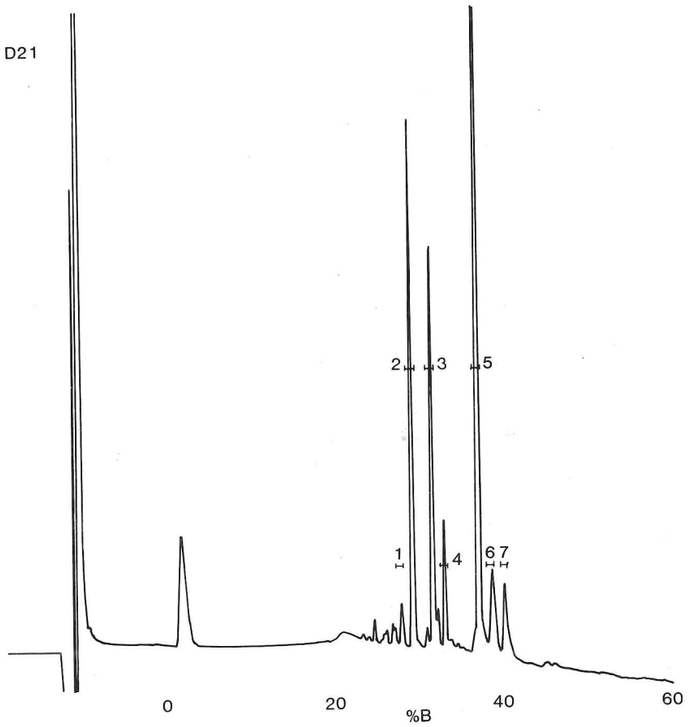
Cyanogen bromide cleavage, which often leaves an insoluble core, released very little soluble material. This is not surprising in view of the overall lack of solubility of mutase and the fact that the subunits have only approx. 10 and 8 methionines according to the amino acid analyses. This would result in peptides of average length approx. 65, and large peptides are often poorly soluble.

Although this cleavage means that all peptides may be assumed to start with methionine, which is useful for making oligonucleotides because it only has one codon, particularly if the N-terminus was blocked and so would give no sequence, it would be best to have several peptide sequences from which to choose for making the probes. Access to a gas-phase sequenator would have allowed large cyanogen bromide fragments to be sequenced. They could be separated by SDS-polyacrylamide gel electrophoresis and electroblotted to derivatised glass-fibre paper. Pieces of this could then be inserted in the sequenator (Vandekerckhove *et al.*, 1985). The small quantities of peptide that could be purified in this manner would be difficult to sequence manually but an automated gas-phase sequenator should give 20 or more residues, allowing a choice of the sequence to use for making oligonucleotides.

Instead, endoproteinase Arg-C from mouse submaxillary gland (Levy et al., 1969; Schenkein et al., 1977) was chosen to produce the peptides to be sequenced. This enzyme cuts on the carboxyl side of arginine residues, which occur four times as often as methionines in mutase and should give many more soluble peptides. Several of these peptides could be sequenced manually by the microscale DABITC technique (Chang, 1983). Although it is not practical to obtain many residues from each peptide by manual methods, sequencing several should give a reasonable choice for making oligonucleotide probes. Arginine is also one of the worst amino acids for oligonucleotide probe making because it is coded for by six codons. Cutting at arginines should ensure that this residue is not encountered in the middle of a peptide. Trypsin, which is more commonly used, would also cut at lysine residues which have only two codons and so are quite suitable for oligonucleotide probes.

Mutase was reduced and reacted with low specific activity [^{14}C]-iodoacetic acid to protect the cysteine residues, and the protein, now rather poorly soluble, was digested with the Arg-C proteinase. Over 80% of the radioactivity incorporated in the protein was solubilised during the digestion and the pink colour due to bound coenzyme-B_{12} was also released.

Soluble peptides were fractionated by gel filtration and five pools collected. The smallest molecular weight pool was not used further. Each of the other pools was then fractionated on an HPLC RP8 reverse-phase column and many of the peaks collected were further chromatographed on the same column with a different buffer system. Examples of each of these separations are shown in Figure 6.3. A different reverse-phase column was also tried for the second separation, using the same buffer system as for the first column, but this gave less good resolution. Most of the peptides eluting early from the first reverse-phase column were positively charged, as shown by thin layer electrophoresis at pH6.8, so some of these were separated by FPLC cation-exchange chromatography. Samples of each peak collected were analysed by thin-layer electrophoresis to determine their purity and concentration. Some of the TLE plates were tested for the presence of tryptophan. Peptides containing tryptophan might be good for deducing oligonucleotide sequences from, because tryptophan is coded for by just one codon.



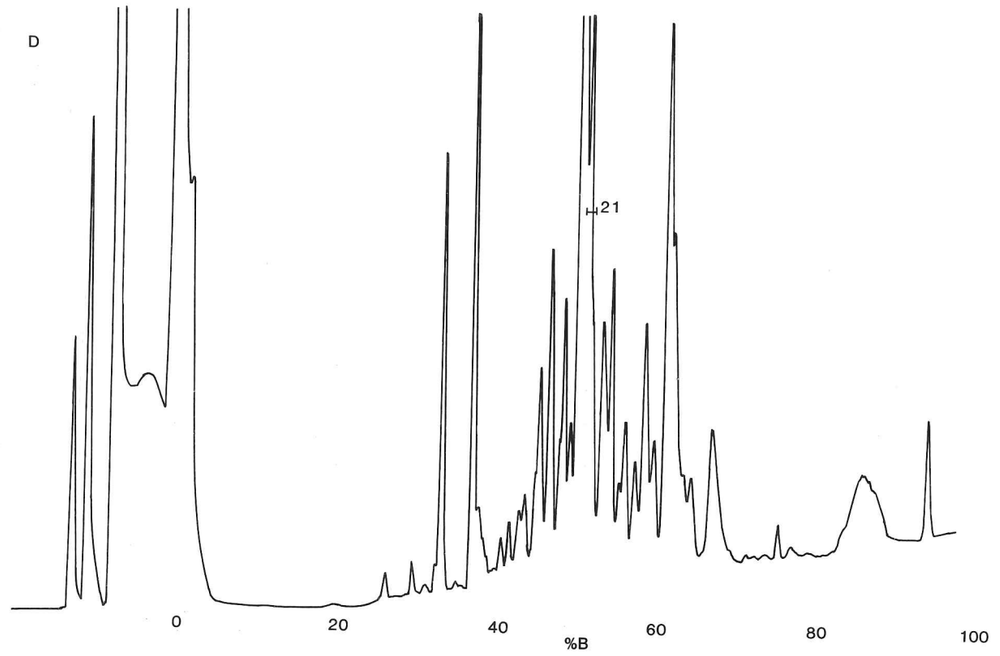


Figure 6.3 Purification of mutase peptides by HPLC

Mutase peptides in pool D from the gel-filtration step were chromatographed on an RP8 reverse-phase column. They were loaded in 0.1% trifluoroacetic acid and eluted using a gradient of 0-100% 0.1% trifluoroacetic acid/acetonitrile (B). Peak 21 is shown chromatographed on the same column using 10mM ammonium acetate, pH5.8 and a gradient of 0-60% acetonitrile (B).

6.4 Manual peptide sequencing

Many of the peptides obtained were sequenced manually by the microscale DABITC method (Chang, 1983). The DABTH-amino acid derivatives released were identified by 2-dimensional thin layer chromatography. Table 6.2 lists the sequences obtained. Unfortunately, there is no reliable buffer system for separating leucine from isoleucine, increasing the codon degeneracy at these positions to 9 (6 for leucine and 3 for isoleucine), and these amino acids between them comprise over 10% of the protein.

A few peptides were impure despite giving a single spot on thin layer electophoresis. Some peptides contained arginine before the C-terminus suggesting that the digestion was incomplete, despite the excess of enzyme used, or that it does not cut in specific circumstances, e.g. after an N-terminal arginine obtained by cleavage between two in a row, or when followed by proline. Schenkein *et al.* (1977) also reported that the enzyme does not give complete digestion.

Sequence was obtained from 24 peptides (Table 6.2), containing a total of 112 amino acids, approx. 9% of the protein (approx. 1 270 total amino acids).

Only once did two fractions give the same peptide sequence, but, interestingly, there are several pairs of peptides that are very similar to each other. In some cases, but not all, this is probably due to the partial digestion. The rest might be due to the purification procedure favouring particular sorts of sequence, but hints at there being related sequences in the two subunits.

None of the sequences obtained were present in the reading frames in the DNA sequence of the 2.3kb epimerase-carrying fragment, as expected from the lack of any detectable expression of mutase by the epimerase clone.

Table 6.2 Mutase peptide sequences

The sequences of 24 peptides were determined by a manual DABITC method. A question mark indicates the most probable amino acid at uncertain positions. Xxx is used were the amino acid could not be identified. Peptides that were not fully sequenced end with three dots to show that they continue.

A13-14/5	Phe Ser Gly Ala Gln Val ...
A13-14/8	Ser Leu/Ile Glu Thr Lys Pro ...
B7-8/14	Leu/Ile Arg
B7-8/17	Arg Asn Leu/Ile Ala Ala Gly Gln ...
B7-8/20	Arg Gly Thr Thr Val Arg
C16	Leu/Ile Arg Pro Pro ?Arg
C20/1	(Phe+Ala) (Gly+Gln) (Pro+Phe) Asp Gly Arg
C20/5	Gly Val Asn Pro ?Tyr ...
C20/14	Lys Leu/Ile Arg
C22/1	Val Leu/Ile Ala Gly Val Tyr ?Arg
C22/3	Arg Asp Phe Gly Gly ?Gln ...
C22/9	Cys Tyr Pro Pro ?Arg
C23/5	Glu Ala Trp Ala Arg
C23/9	Leu/Ile Gly Leu/Ile Asn Lys ?Arg
C24/6	Gly Xxx Leu/Ile Ala ...
C26/4	Phe ?Leu/Ile ... (Contains Trp)
C26/12	?Glu ?His Ala Leu/Ile ...
C30/5	Ala Leu/Ile Tyr ...
D16/1	Asp Phe Gly Gly Arg
D20/5	Glu Ala Trp Ala Arg
D20/7	Arg Asp Phe Gly ?Ser ?Gln ...
D21/2	Glu ?Leu/Ile ?Arg Ala ...
D21/3	Glu ?Leu/Ile Asn Ala Phe Tyr Arg
D23/4	Gln Xxx Ala Leu/Ile ... (Contains Trp)

6.5 Discussion

Methylmalonyl-CoA mutase was purified from P. shermanii. An improved preparation procedure yielded a larger protein than obtained previously. This enzyme was shown to be a dimer of two non-identical subunits.

Several peptides produced by proteolytic digestion of the enzyme were purified and sequenced manually to obtain peptide sequences for which oligonucleotide probes could be made. The availability of automated peptide sequencing facilities would have greatly speeded up this part of the project.

6.5 Discussion

Methylmalonyl-CoA mutase was purified from P. shermanii. An improved preparation procedure yielded a larger protein than obtained previously. This enzyme was shown to be a dimer of two non-identical subunits.

Several peptides produced by proteolytic digestion of the enzyme were purified and sequenced manually to obtain peptide sequences for which oligonucleotide probes could be made. The availability of automated peptide sequencing facilities would have greatly speeded up this part of the project.

CHAPTER 7

CLONING AND EXPRESSION OF PART OF THE MUTASE GENE

7. CLONING AND EXPRESSION OF PART OF THE MUTASE GENE

Several peptides from a proteolytic digestion of mutase had been sequenced, so oligonucleotide probes could now be made and used to clone the genes for this enzyme.

None of the sequences obtained were very good for making oligonucleotides. The one requiring fewest different sequences was obtained twice (C23/5 and D20/5). This would require 64 different 14-mers but preferably 17-mers should be used for better specificity. The best of these would be C22/3 (without the first arginine), requiring 256 sequences.

As explained previously (Section 3.2) the high G+C content of *P. shermanii* DNA means that G or C is usually used in the third position of codons in preference to A or T, and also in the first position where that choice exists (arginine and leucine). Thus, only two sequences are required for arginine instead of six, and only three for leucine/isoleucine instead of nine. Using only G or C in all degenerate positions dramatically reduces the number of oligonucleotides required to 4 14-mers for D20/5 and 8 17-mers for C22/3. However, some A and T is found in the third position, 15% in the epimerase gene, so there would only be a 61% chance of one of these 14-mers being correct and a 44% chance for the 17-mers. These are not good odds, but fewer oligonucleotides have to be made. If these more likely sequences were made for several peptides the chance of having a correct probe would be greatly increased. There would also be the possibility of having probes to peptides from both subunits. For these reasons, instead of making all possible oligonucleotides for peptide D20/5, a total of 90 oligonucleotides were made representing the 'high G+C' sequences for 8 of the peptides, giving a 99.6% chance of having at least one correct sequence present. The sequences made are listed in Table 7.1.

7.1 Synthesis of oligonucleotides

The oligonucleotides were synthesised on filter paper disc supports over two days by the method of Matthes *et al.* (1984). Each sequence can be kept

Table 7.1 Oligonucleotides synthesised as probes for mutase

	Peptides and DNA (5'–3')	Oligonucleotides (3'–5')
A13-14/8	Ser Leu/Ile Glu Thr Lys Pro TCX CTX T AG$^{C}_{T}$ TT$^{G}_{A}$/AT$^{C}_{A}$ GA$^{G}_{A}$ ACX AA$^{G}_{A}$ CC	1 AGC GAC CTC TGC TTC GG 2 AGC GAC CTC TGC TTG GG 3 AGC GAG CTC TGC TTC GG 4 AGC GAG CTC TGG TTC GG 5 AGC TAG CTC TGC TTC GG 6 AGC TAG CTC TGG TTC GG 7 AGG GAC CTC TGC TTC GG 8 AGG GAC CTC TGG TTC GG 9 AGG GAG CTC TGC TTC GG 10 AGG GAG CTC TGG TTC GG 11 AGG TAG CTC TGC TTC GG 12 AGG TAG CTC TGG TTC GG 13 TCG GAC CTC TGC TTC GG 14 TCG GAC CTC TGG TTC GG 15 TCG GAG CTC TGC TTC GG 16 TCG GAG CTC TGG TTC GG 17 TCG TAG CTC TGC TTC GG 18 TCG TAG CTC TGG TTC GG
B7-8/17	Asn Leu/Ile Ala Ala Gly Gln AA$^{C}_{T}$ CTX T TT$^{G}_{A}$/AT$^{C}_{A}$ GCX GCX GGX CA	19 TTG GAC CGC CGC CCC GT 20 TTG GAC CGC CGC CCG GT 21 TTG GAC CGC CGG CCC GT 22 TTG GAC CGC CGG CCG GT 23 TTG GAC CGG CGC CCC GT 24 TTG GAC CGG CGC CCG GT 25 TTG GAC CGG CGG CCC GT 26 TTG GAC CGG CGG CCG GT 27 TTG GAG CGC CGC CCC GT 28 TTG GAG CGC CGC CCG GT 29 TTG GAG CGC CGG CCC GT 30 TTG GAG CGC CGG CCG GT 31 TTG GAG CGG CGC CCC GT 32 TTG GAG CGG CGC CCG GT 33 TTG GAG CGG CGG CCC GT 34 TTG GAG CGG CGG CCG GT 35 TTG TAG CGC CGC CCC GT 36 TTG TAG CGC CGC CCG GT 37 TTG TAG CGC CGG CCC GT 38 TTG TAG CGC CGG CCG GT 39 TTG TAG CGG CGC CCC GT 40 TTG TAG CGG CGC CCG GT 41 TTG TAG CGG CGG CCC GT 42 TTG TAG CGG CGG CCG GT
D16/1	Asp Phe Gly Gly Arg GA$^{C}_{T}$ TT$^{C}_{T}$ GGX GGX $^{C}_{A}$G	43 CTG AAG CCC CCC GC 44 CTG AAG CCC CCG GC 45 CTG AAG CCG CCC GC 46 CTG AAG CCG CCG GC

	Peptides and DNA (5'–3')	Oligonucleotides (3'–5')
A13-14/5	Phe Ser Gly Ala Gln Val TT$^{C}_{T}$ TCX AG$^{C}_{T}$ GGX GCX CA$^{G}_{A}$ GT	47 AAG AGC CCC CGC GTC CA 48 AAG AGC CCC CGG GTC CA 49 AAG AGC CCG CGC GTC CA 50 AAG AGC CCG CGG GTC CA 51 AAG AGG CCC CGC GTC CA 52 AAG AGG CCC CGG GTC CA 53 AAG AGG CCG CGC GTC CA 54 AAG AGG CCG CGG GTC CA 55 AAG TCG CCC CGC GTC CA 56 AAG TCG CCC CGG GTC CA 57 AAG TCG CCG CGC GTC CA 58 AAG TCG CCG CGG GTC CA
C22/3	Arg Asp Phe Gly Gly Gln CGX AG$^{G}_{A}$ GA$^{C}_{T}$ TT$^{C}_{T}$ GGX GGX CA	59 GCC CTG AAG CCC CCC GT 60 GCC CTG AAG CCC CCG GT 61 GCC CTG AAG CCG CCC GT 62 GCC CTG AAG CCG CCG GT 63 GCG CTG AAG CCC CCC GT 64 GCG CTG AAG CCC CCG GT 65 GCG CTG AAG CCG CCC GT 66 GCG CTG AAG CCG CCG GT
C23/9	Leu/Ile Gly Leu/Ile Asn Lys CTX T CTX T TT$^{G}_{A}$/AT$^{C}_{A}$ GGX TT$^{G}_{A}$/AT$^{C}_{A}$ AA$^{C}_{T}$ AA	67 GAC CCC GAC TTG TT 68 GAC CCC GAG TTG TT 69 GAC CCC TAG TTG TT 70 GAC CCG GAC TTG TT 71 GAC CCG GAG TTG TT 72 GAC CCG TAG TTG TT 73 GAG CCC GAC TTG TT 74 GAG CCC GAG TTG TT 75 GAG CCC TAG TTG TT 76 GAG CCG GAC TTG TT 77 GAG CCG GAG TTG TT 78 GAG CCG TAG TTG TT 79 TAG CCC GAC TTG TT 80 TAG CCC GAG TTG TT 81 TAG CCC TAG TTG TT 82 TAG CCG GAC TTG TT 83 TAG CCG GAG TTG TT 84 TAG CCG TAG TTG TT
D20/5	Glu Ala Trp Ala Arg GA$^{G}_{A}$ GCX TGG GCX $^{C}_{A}$G	85 CTC CGC ACC CGC GC 86 CTC CGC ACC CGG GC 87 CTC CGG ACC CGC GC 88 CTC CGG ACC CGG GC
D21/3	Asn Ala Phe Tyr Arg AA$^{C}_{T}$ GCX TT$^{C}_{T}$ TA$^{C}_{T}$ $^{C}_{A}$G	89 TTG CGC AAG ATG GC 90 TTG CGG AAG ATG GC

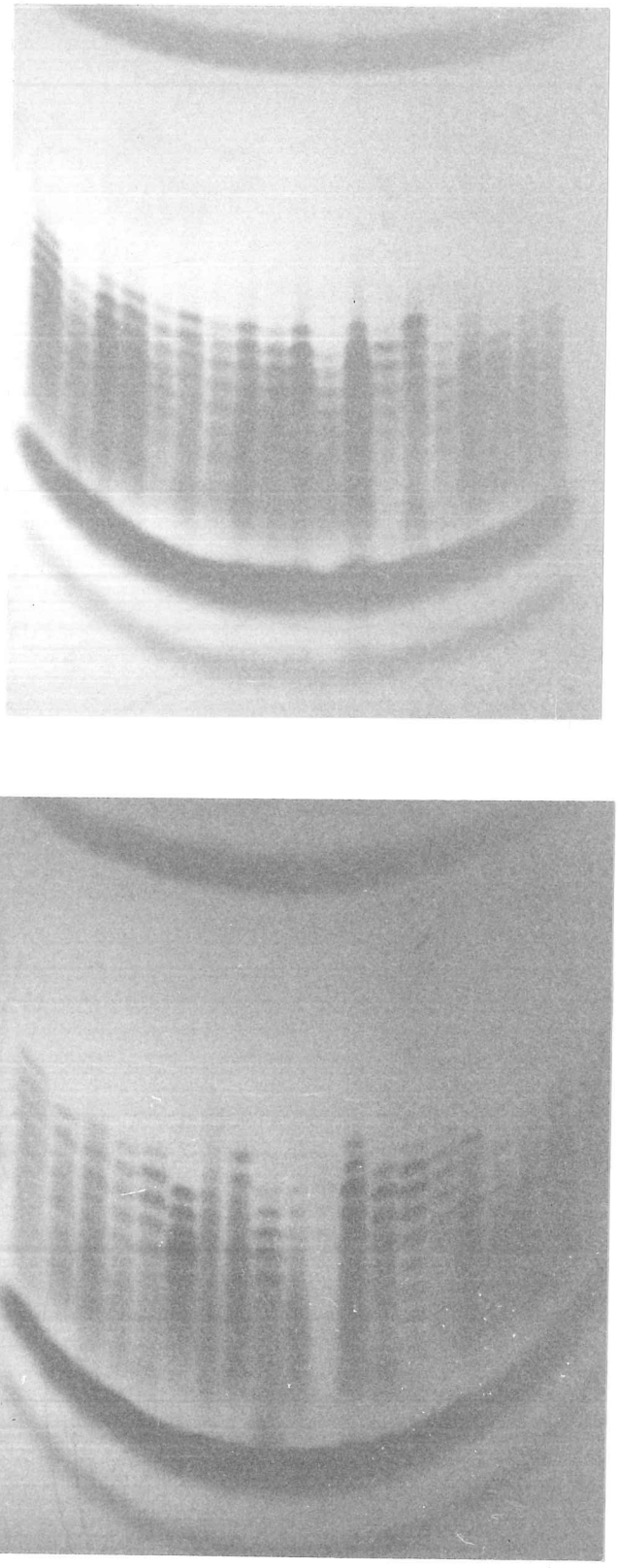


Figure 7.1 Purification of oligonucleotides by gel electrophoresis

Two of the oligonucleotide purification gels are shown. The oligonucleotides absorb UV light, casting shadows on a fluorescent TLC plate. The upper band in each track is the longest, and so full length, oligonucleotide. Variations in mobility between tracks are caused by the different compositions and sequences of the oligonucleotides.

separate, so avoiding the use of mixtures, but they are all made at the same time. Several other oligonucleotides were also synthesised making a total of 131. An extra column was used in some rounds (when more than 60 sequences required the same base). This was particularly prone to happen because most of the oligonucleotides required either G or C for every third base, but was largely avoided by starting some oligonucleotides after the first base had been added to the others.

After the synthesis the oligonucleotides were individually deprotected and removed from the supports and then purified by gel electrophoresis. The initial coupling loaded approx. 100nmol of nucleotide succinate onto each paper disc. On average 800ng of oligonucleotide was obtained which would be sufficient for 80 experiments. This yield is approx. 200µmol or 0.2% and corresponds to a repetitive yield of 68% at each cycle. Figure 7.1 shows two of the purification gels.

7.2 Hybridisation of the oligonucleotides to P. shermanii DNA

7.2.1 Selection of the best oligonucleotide probes

To discover which, if any, of the oligonucleotides made for a particular peptide was the correct sequence they were all hybridised to P. shermanii chromosomal DNA. An EcoRI digest of the chromosomal DNA was run in wide tracks on an agarose gel. This gel was dried and denatured in the usual way and then cut into strips. Each oligonucleotide was individually labelled using γ-[^{32}P]-ATP and T4-polynucleotide kinase and used without purification from unincorporated label. A gel slice was rolled and inserted into each tube and hybridisation buffer was added. After hybridisation they were washed three times in the same tubes. This procedure avoided mixing up the strips.

The oligonucleotides had a range of T_m from 40 to 60°C, and ideally should have all been hybridised at 5-7°C below their T_m to give the same stringency to each. However, only two water baths were available, so oligonucleotides were hybridised at 5-15°C below the T_m, which might allow two, or more, mismatches in the worst cases. The range of T_m of oligonucleotides for any one peptide was at most 4°C so that any binding differences seen between them

should be due to the number of mismatches in the hybrids. The autoradiogram of all the gel slices is given in Figure 7.2 and shows that, as expected, most of the oligonucleotides do not hybridise specifically to any sequence.

Probes 1-18 for peptide A13/8 were hybridised at 7-9°C below the T_m, sufficient to allow one mismatch. Probe 18 appears to be correct. Probe 16 is identical except for a G in place of a T, probe 12 has two mismatches at one end (less destabilising than in the middle) and probe 8, which binds less well, has both of these combined. Probe 6, which hybridises very weakly, has three mismatches at the end. Probe 2 binds to a completely different size band to the rest. The slice was, unfortunately, inverted relative to the rest on the autoradiogram and the hybridising band is, in fact, of higher molecular weight. This sequence is probably not part of mutase because, if correct, some of the other probes 1-18 would bind weakly to this size DNA.

Probes 43-46 for peptide D16/1 were hybridised well below the T_m (15°C), so mismatched sequences should bind. None of them bind strongly to any particular size DNA but, interestingly, probe 46 does bind weakly to the same size DNA as probe 18.

Probes for peptide D7-8/17 (19-42) do not seem to bind stringently to particular sequences except, perhaps 27, 42 and 38. However, if these are correct then probes with one difference from them should bind to the same sizes of DNA, as for peptide A13/8. The absence of these for 27, which binds most strongly, suggests that, as for probe 2, this is not correct.

For peptide C23/9 (67-84) probe 78 binds strongly to two bands at 7°C below the T_m. Again, none of the others bind to one of these sizes of fragment, but probe 72 binds well to the larger fragment, and probes 69, 71, 74, and 75 also hybridise to this size of fragment. Probes 69,71 and 78 each have one difference from 72, 75 has two and 74 has three (two of which are together). Probe 72 does not bind very strongly so may itself have a mismatch from the correct sequence.

Peptides A13/5, C22/3, D20/5 and D21/3 have no strongly binding probes (47-66, 85-90), suggesting that the correct sequences for these have at least one A or T in place of the G or C used.

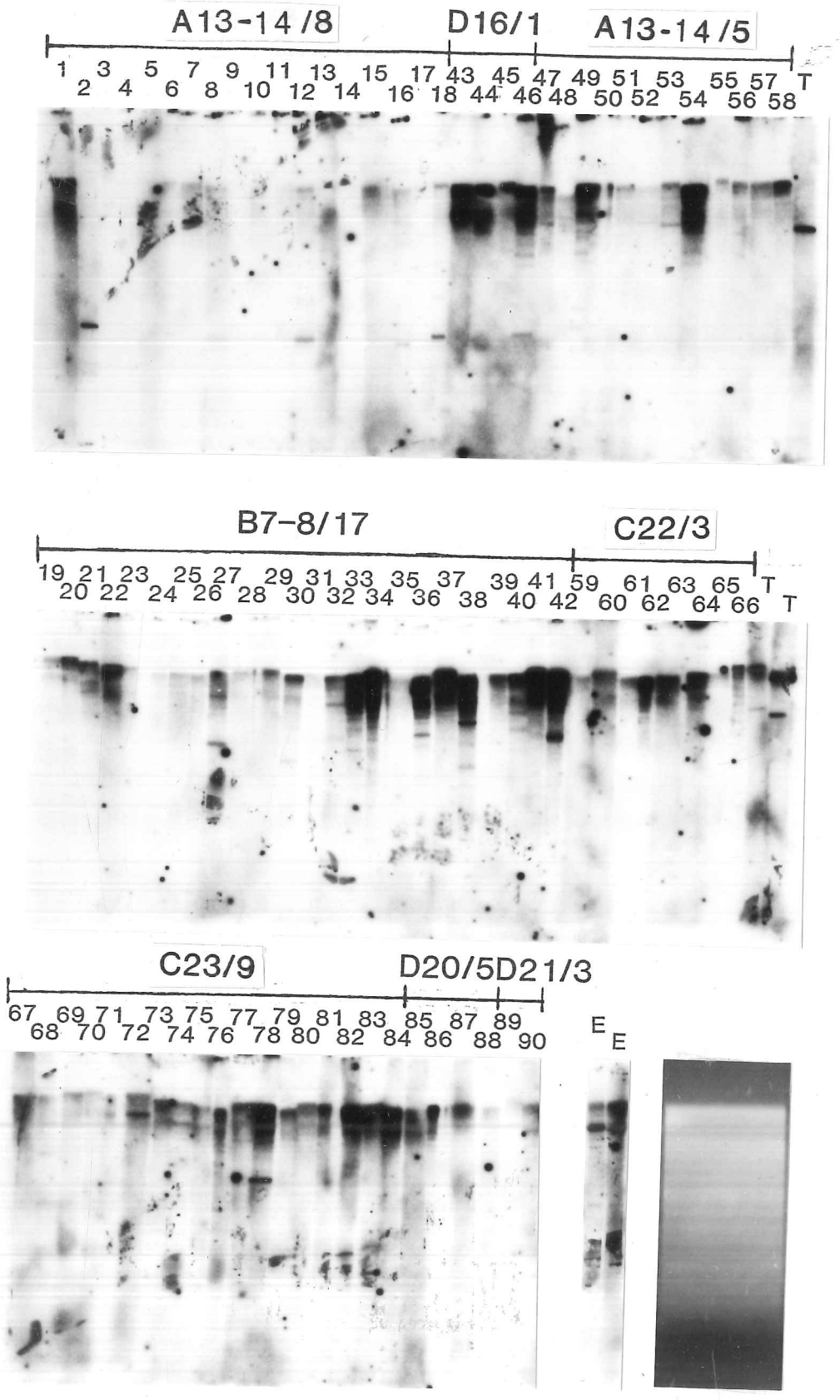


Figure 7.2 Hybridisation of 90 mutase oligonucleotides to P. shermanii DNA

P. shermanii DNA, digested with EcoRI, was electrophoresed in wide tracks on 0.7% agarose gels. The gels were dried and slices hybridised with each oligonucleotide separately. Unincorporated label was not removed from the probes. The probes are grouped according to the peptides for which they are probes. Also shown are hybridisations with epimerase (E) and transcarboxylase (T) probes. A photograph of part of a gel is also shown.

The result from this multiple probing experiment was that one oligonucleotide almost certainly bound to the mutase gene, probe 18, and one for another peptide, probe 46, bound with a mismatch to the same size fragment of chromosomal DNA. Probe 72 was possibly binding to a different fragment carrying another part of mutase, but might have a mismatch to the correct sequence, and for another peptide either probe 38 or 42 might be correct or nearly so. Two probes that bound strongly, but were probably not binding to the mutase gene, were 2 and 27. Probe 78 hybridised to the same size of fragment as probe 72 but also to another band that was probably not the mutase gene.

7.2.2 Hybridisation of the selected probes to several DNA digests

Probe 18 hybridised quite cleanly to a fragment of P. shermanii DNA, and was likely to be binding to the mutase gene. This probe and a selection of others that bound to specific bands in the EcoRI digest (2, 46, 27, 42 and 78) were used to probe several restriction digests of the P. shermanii chromosomal DNA, although most of them were not expected to be binding to the mutase gene (Figure 7.3). Probes 72 and 38 were not used, but perhaps 72 in particular should have been. It was hoped that if two or more of these probes were specific for the mutase gene that a restriction digest might be found in which they bound to the same size fragment.

These gels were hybridised at nominally the same two temperatures as used in the previous section (7.2.1), but may have been slightly different because the EcoRI band labelled by probe 46 was now very faint, confirming that it was not a perfect match. Probe 18 labelled an extra EcoRI band that was extremely faint before. This is probably due to a partial digestion, although on the gel the digest looked no different from before. It is possible that restriction sites in some environments of local sequence may not be cut as fast as others and differential digestion rates have been reported (Thomas and Davis, 1975).

The amount of non-specific binding with each probe can be seen from the λHindIII standards as well as the parts of the digests with most DNA. The two 14-mers give most, as expected, and, although more stringent conditions could be used for probe 46, this probe gave no particularly strong bands. Therefore,

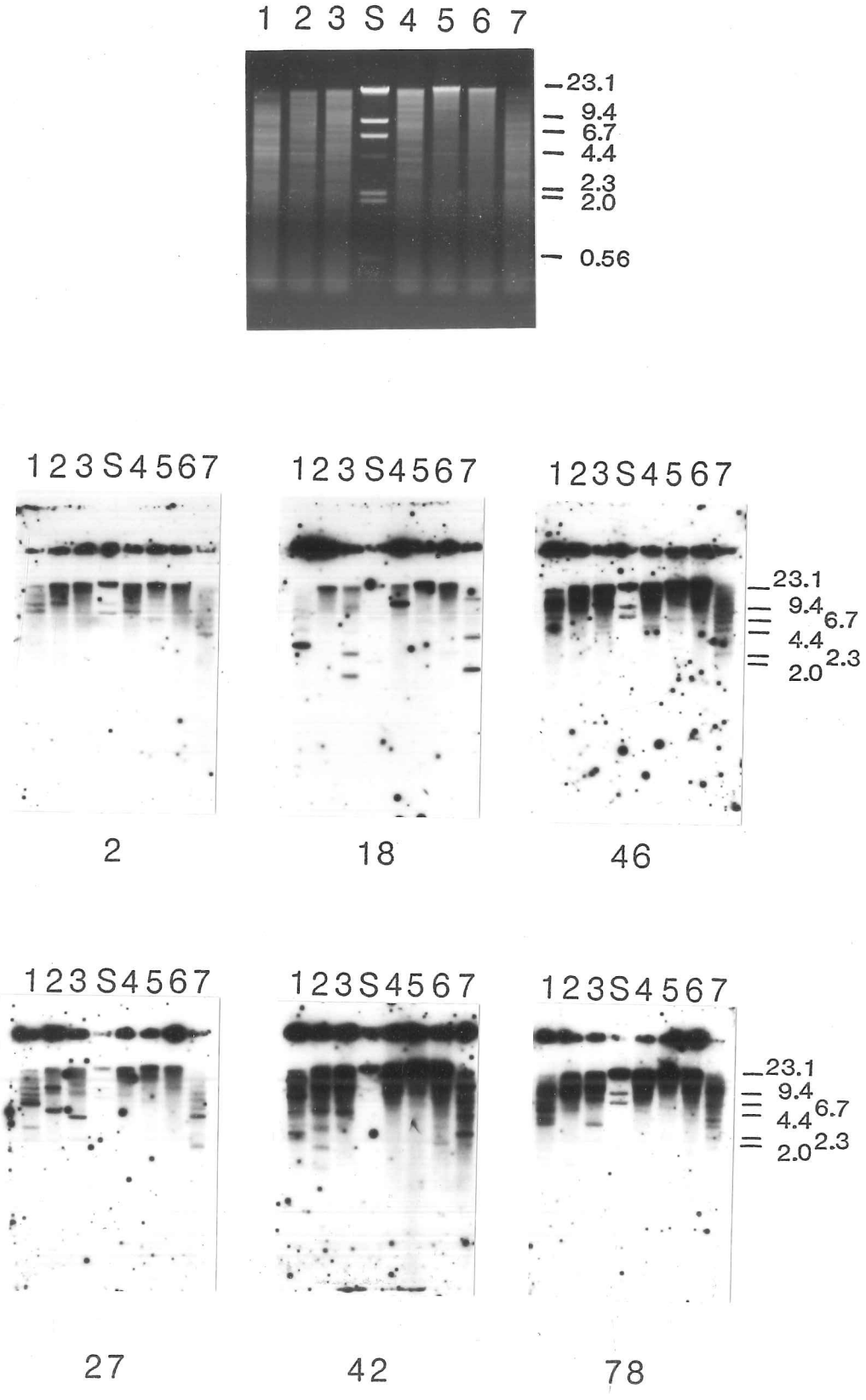


Figure 7.3 Hybridisation of six mutase oligonucleotides to several restriction digests of P. shermanii DNA

Six sets of seven DNA digests were run on a 0.7% agarose gel using two lines of wells. Six oligonucleotides (2, 18, 46, 27, 42 and 78) were each hybridised to one section of the dried gel. The digests used were: 1, BamHI; 2, BglII; 3, EcoRI; 4, EcoRV; 5, HindIII; 6, PstI; 7, XhoI. The PstI digestion did not go to completion. A photograph of one section of the gel is also shown.

neither 14-mer is suitable to use as a probe for cloning. The 17-mers were hybridised at 47°C, which was 7 and 9°C below the T_m of probes 18 and 2 respectively. Probe 18 gives little background and binds strongly to just one or two bands in each digest, but probe 2 does not give very strong bands or bind to any of the same size bands as probe 18. This was further evidence that probe 2, made as a probe to the same peptide as probe 18, was not binding to the mutase gene but, with a mismatch, to another DNA sequence. Probes 27 and 42 have higher T_m, accounting for some of the background in 42, but 27 gives little background anyway making it more suitable as a probe, although, the evidence presented earlier suggested that it might not be binding to the mutase gene.

None of the oligonucleotides, other than 18 and 46, hybridised to the same restriction fragments as each other, reinforcing the suspicion that most of these six probes were not binding to the mutase gene. The hybridisation conditions for the other probes could have been improved to allow only specific binding, and oligonucleotides 38 and 72 could also have been tried, but at this stage it was better to prove that at least one oligonucleotide bound to the mutase gene. Therefore, probe 18 was used to clone DNA fragments to which it bound.

7.3 Cloning restriction fragments that hybridise to the mutase probes

Two or more oligonucleotides binding to a restriction fragment large enough to carry the whole gene would make that a good fragment to clone but no such fragments were found. Probe 18 only bound to one fragment that was large enough to carry all of the mutase gene (at least 4.5kb). This was a 10.5kb EcoRV fragment, but this is a poor enzyme for use in cloning. Also, the position within the protein of the peptide to which this oligonucleotide was a probe was not known, so it was not worth using Bal31 to locate the binding site within a fragment to ensure that the whole gene was present, as was done for epimerase (see Section 3.4).

Instead, smaller hybridising fragments were cloned and then analysed by transcription-translation. The antibody to mutase should still pick out any partial peptide sequence of the protein, even if only a portion of the gene

was present. The sensitivity of this technique, as shown by the detection of epimerase synthesis directed by chromosomal DNA (Section 4.1.2), might show the presence of mutase even without its promoter by read through from vector promoters which the extract also recognises. If such a fragment were found it could be used as a very specific probe to clone the entire gene from a different restriction digest or a library.

Two different restriction digests were used for the cloning, EcoRI and BamHI. DNA fragments were isolated with DE81 paper from the digests size-fractionated by agarose gel electrophoresis. They were then ligated into pUC13, which had previously been cut with the appropriate enzyme and treated with phosphatase. EcoRI fragments of 1-1.8kb and 2.3-3kb, and BamHI fragments of 2.4-3kb were used, as well as unfractionated EcoRI digested DNA.

The ligated DNA was used to transform E. coli strain TG1, and the transformants screened using the 541 paper method. Positive colonies were obtained from each transformation with ligations of EcoRI fragments, but not of BamHI fragments. The G+C content of the BamHI recognition site is higher than that of the EcoRI site, so there are more fragments in this fraction and insufficient numbers of recombinants were screened.

The positive strains were streaked out to single colonies, numbered colonies were patched out to store, and the streak plates were then screened. This shortened the screening procedure by one day, because the patched out colonies did not have to be grown up before screening.

7.4 Analysis of the cloned DNA fragments

Plasmid DNA was prepared from the recombinants and run on an agarose gel to analyse the inserts obtained (Figure 7.4). The inserts were excised with EcoRI to show their size and were also digested with BglII. The gel of the recombinant plasmids was hybridised with the oligonucleotide. A stringently binding 1.4kb EcoRI insert was obtained in plasmids pND4, pND5 and pND6. This was the same fragment in each case, with a BglII site very close to one end, and was the same size as the strongly-hybridising band in the EcoRI chromosomal digest.

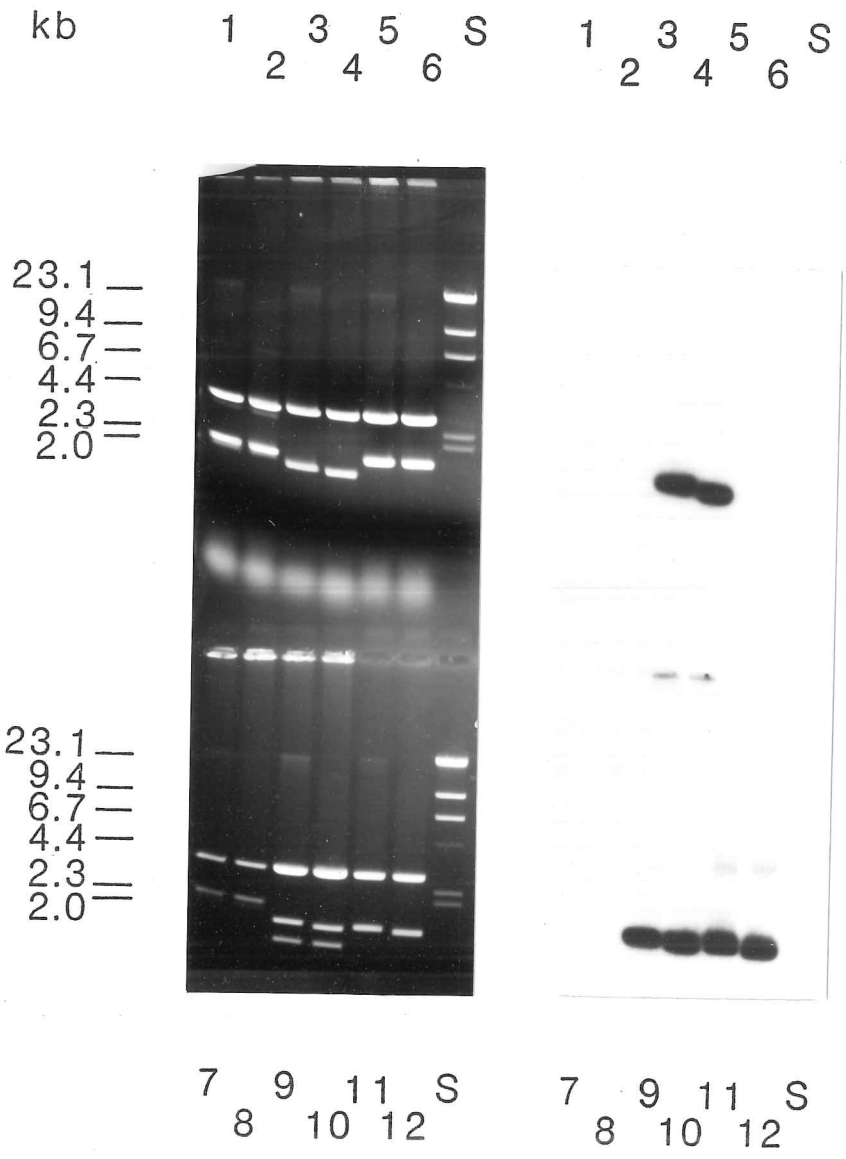


Figure 7.4 Recombinant plasmids obtained using mutase probe 18

All tracks contain recombinant plasmids digested with EcoRI to excise the inserts. DNA in even numbered tracks was digested with BglII as well as EcoRI. The agarose gel was dried and probed with oligonucleotide 18. The resulting autoradiograph is shown to the right. Tracks 3 and 4 contain pND4, tracks 9 and 10 contain pND5, and tracks 11 and 12 contain pND6. Tracks 1, 2, 5, 6, 7, and 8 contain other plasmids with non hybridising inserts.

Plasmid pND5 has two insert bands of equal intensity. The vector band is much more intense than these so the plasmid is either an equal mixture of two recombinant plasmids, or one recombinant with an EcoRI site in the insert, derived by partial digestion, and some contaminating vector. The latter case is more likely because pND5 derived from the ligation with inserts of 2.3-3kb and it seems unlikely that two inserts which should be in the size fraction 1-1.8kb would both be cloned from the 2.3-3kb fraction in adjacent transformants on the selection plate. (The 1.8-2.3kb size fraction was also taken but not used in these experiments, so there should not be any significant contamination between the other two fractions). A larger excess of enzyme, which was also a different batch, was used in the digestions for this gel, so the insert is fully digested. The complete insert in pND5 (1.4 + 1.1kb) corresponds to the 2.5kb fainter band seen in the hybridisations to chromosomal DNA digested with EcoRI.

7.5 Transcription-translation of the cloned fragments

The sequence in *P. shermanii* DNA complementary to probe 18 had been cloned on a small restriction fragment. In order to show whether or not this fragment contained part of the mutase gene the recombinant plasmids were used to direct protein synthesis in the Streptomycete cell-free system (Thompson *et al.*, 1984). The extract can recognise *E. coli* promoters (e.g. the ampicillin resistance gene on pUC12) and also at least one *P. shermanii* promoter (the epimerase gene, see Section 4.1) so it was hoped that, although only part of the gene would be present on the cloned fragments, a partial protein product might be detected. This might be an N-terminal portion read from the *P. shermanii* promoter, or another fragment expressed by fusion to the β-galactosidase gene of pUC13 by the insertion at the polylinker. (The *lac* repressor would not be present to prevent this expression.) The extract might also be able to initiate protein synthesis, at a slower rate, from points other than the normal N-terminus giving a little expression without protein fusion, because the *in vitro* system has a slightly lowered specificity.

Figure 7.5 shows the products of transcription-translation of the recombinant plasmids immunoprecipitated by anti-mutase antiserum, and pND1 immunoprecipitated by anti-epimerase antiserum as a control. Mutase

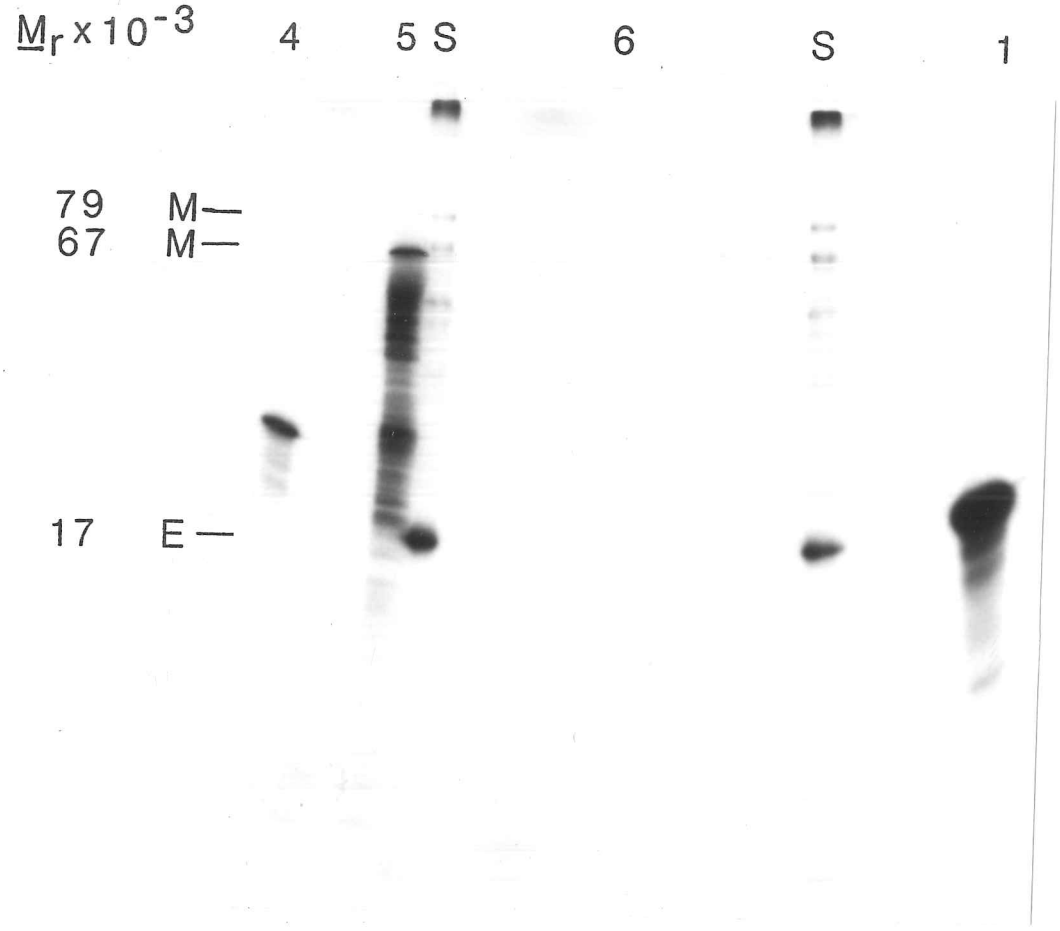


Figure 7.5 Transcription-translation of plasmids obtained using mutase probe 18

Products of transcription-translation of various plasmids using the S. lividans cell-free extract were immunoprecipitated by an anti-mutase antiserum and electrophoresed on a 20% SDS-polyacrylamide gel. An anti-epimerase antiserum was used to immunoprecipitate epimerase produced from pND1 as a control. The tracks containing products produced from pND4, pND5, pND6 and pND1 are labelled 4, 5, 6 and 1 respectively. Standards (S) are [^{14}C]-carboxymethylated mutase (M) and [^{14}C]-NEM-labelled epimerase (E).

polypeptide was detected from each clone that hybridised with oligonucleotide 18.

Plasmid pND4 directed the synthesis of an immunoprecipitable polypeptide of M_r 23 500, and a trace of a larger peptide of M_r 27 000. Plasmid pND6, which has the same insert, only shows a faint polypeptide of M_r 23 500. Further restriction digests of the two plasmids showed that the 1.4kb EcoRI insert was in the same orientation in each, with the BglII site in the insert close to the HindIII site in the vector polylinker as shown in Figure 7.6. The expression from each plasmid should, therefore, be the same, but the preparation of plasmid pND6 was poor containing much more RNA. This might reduce the specific expression from the plasmid by competition for ribosomes in the transcription-translation extract.

The insert in plasmid pND5 contains an extra 1.1kb EcoRI fragment. This extra DNA could encode up to approx. M_r 42 000 more of mutase and a peptide of M_r 65 000 is observed (almost the size of the smaller of the two mutase subunits). The 1.4kb fragment is in the opposite orientation in plasmid pND5 compared with pND4 and pND6 (see Figure 7.6). The plasmid preparation, like that of pND4, contains little RNA, and the similar levels of expression of immunoprecipitable material from the insert in each orientation suggest that the gene is being expressed from a *P. shermanii* promoter on the 1.4kb fragment, rather than from different *E. coli* vector promoters in each clone.

There are many partial products or degradation products from the expression of pND5, which could be due to *E. coli* derived sequences at the C-terminal end of the peptide preventing the polypeptide from folding properly, and so rendering it more susceptible to proteolysis. *E. coli* proteins are degraded by this system (Thompson *et al.*, 1984), so some *P. shermanii* proteins may also be degraded because they are foreign. Alternatively, mutase polypeptides may require coenzyme-B_{12} to fold correctly. Mutase from which the coenzyme is removed is generally more susceptible to chemical and enzymic attack. Also, the second subunit is not present so the peptide synthesised here will not be able to take up its normal three-dimensional structure, and, as reported above (Chapter 6), the subunits precipitate once separated rather than reforming the correct

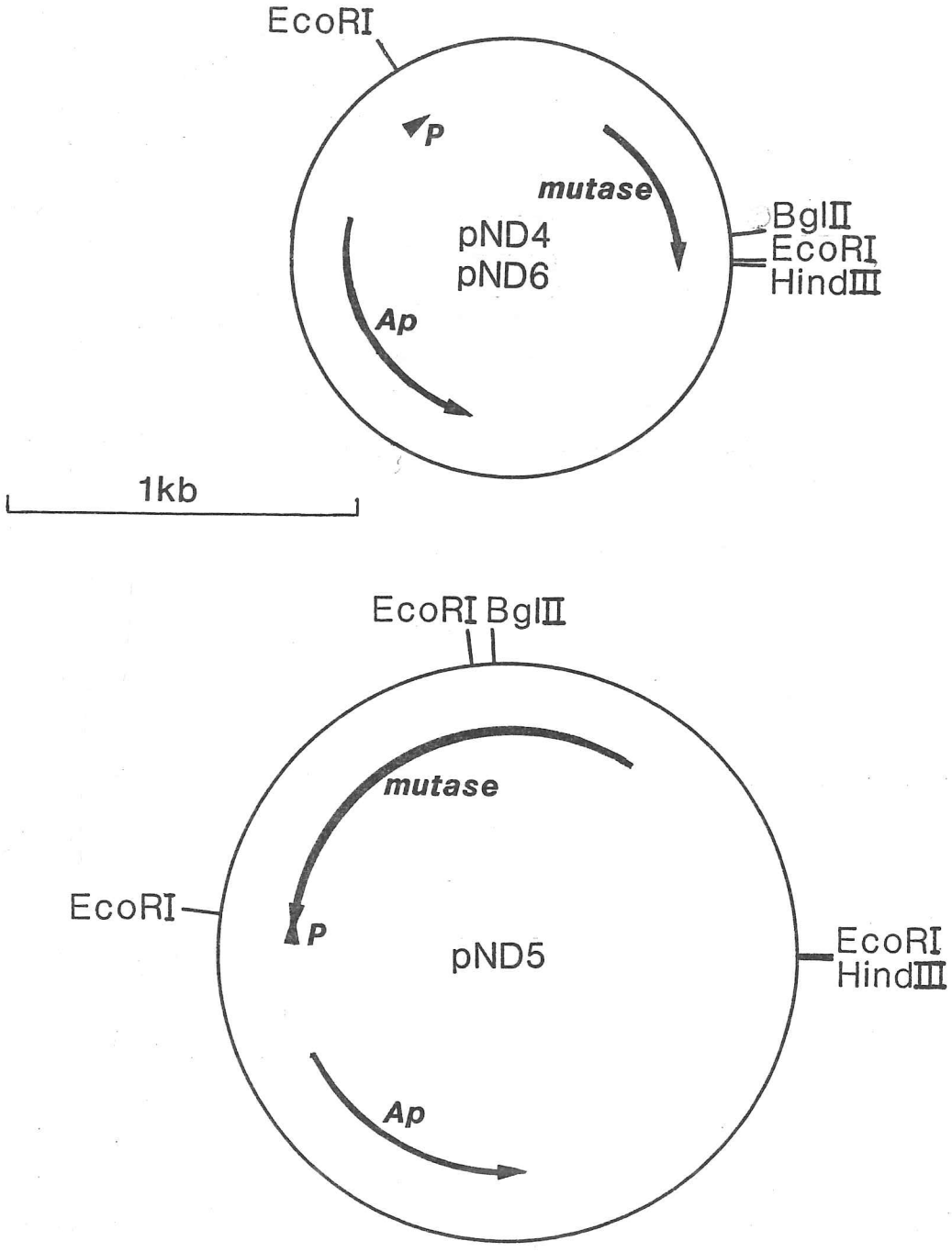


Figure 7.6 Clones containing parts of the mutase gene

Fragments of P. shermanii DNA carrying part of the mutase gene were introduced into the EcoRI site of pUC13. This insertion disrupts the β-galactosidase gene leaving the promoter (p). Plasmid-containing cells were selected by the ampicllin resistance gene (Ap) on pUC13. Plasmid pND5 contains the same insert as plasmids pND4 and pND6, but in the opposite orientation, as well as an additional fragment.

structure.

7.6 Discussion

Ninety oligonucleotides were made representing all possible DNA coding sequences for eight of the mutase peptides with G and C used in place of A and T at every ambiguity. As expected very few of these bound stringently to *P. shermanii* DNA. Of those that did one was chosen and used as a probe to clone restriction fragments to which it bound.

The recombinant plasmids obtained were used to programme protein synthesis in the *S. lividans* transcription-translation system. Three clones encoded polypeptide precipitable by the anti-mutase antiserum. Two contained the same 1.4kb insert and the other contained an additional 1.1kb fragment and was derived by partial digestion. This larger insert encoded a larger polypeptide than the others, in proportion to its increased size, and carried almost the entire gene for one of the subunits of mutase.

These clones may now be used to probe the chromosomal DNA to find a restriction fragment that will carry the entire gene. They could also be used as very specific probes to clone the complete gene from a bank generated by partial digestion of the whole chromosome.

CHAPTER 8

OVERVIEW

8. OVERVIEW

Three of the enzymes of the propionate pathway of P. shermanii have methylmalonyl-CoA as substrate or product, and all three have been studied from a mechanistic point of view. Transcarboxylase, a large multi-subunit complex, has also been studied structurally.

Carboxylations involving transcarboxylase have been shown to be non-concerted (O'Keefe and Knowles, 1986). The structures of the three subunit types and their roles in the catalysis have been studied in detail (Wood and Zwolinski, 1976; Zwolinski et al., 1977). The peptide sequence of the small, biotin-carrying peptide has been determined and the biotin binding site shows homology with other biotin containing enzymes (Maloy et al., 1979).

Methylmalonyl-CoA epimerase is metal-activated, but contains no other cofactor, and is a dimer of one subunit type (Leadlay, 1981). The active site contains two bases but their identity is unknown (Fuller and Leadlay, 1983; Leadlay and Fuller, 1983). The two thiols in each subunit show unusual reactions with thiol-directed reagents (P.F. Leadlay, personal communication).

Methylmalonyl-CoA mutase is a coenzyme-B_{12} dependent enzyme. The mechanism of action of the coenzyme has been studied in detail because of the unusual carbon-skeleton rearrangement catalysed and the involvement of radicals (Halpern, 1985). However, nothing is known about the protein.

The understanding of mechanistic and protein chemical studies of all these enzymes is limited by the lack of primary structure information. For this reason it would be desirable to clone the genes and sequence them. The approach adopted for cloning these enzymes was to make oligonucleotide probes complementary to DNA sequences deduced from peptide sequences. This project concentrated on cloning the genes for epimerase and mutase.

Samples of epimerase were sequenced on a solid-phase sequenator and the N-terminal 31 amino acids were obtained. The high G+C content of P. shermanii DNA allowed choices to be made at ambiguous positions in the sequence caused

by the degeneracy of the genetic code. Part of the peptide sequence was favourable for oligonucleotide synthesis and allowed a single 17-mer to be synthesised. This hybridised stringently to *P. shermanii* DNA. A new method of oligonucleotide synthesis allowed all the possible coding sequences to be made separately. Of these 32 oligonucleotides the sequence made previously hybridised most strongly. DNA restriction fragments of the size that the probe hybridised to were isolated from a gel and cloned into *E. coli* in pUC12. The recombinant clones were screened with the oligonucleotide and hybridising plasmids obtained. A restriction map was constructed, and the hybridising fragments corresponded to fragments which had hybridised in the chromosomal DNA digests, showing that no rearrangement had occured in the clone.

To confirm that the correct gene had been cloned the DNA was analysed by transcription-translation in a Streptomycete cell-free system. A peptide of the correct size was immunoprecipitated by antiserum raised against epimerase. Transcarboxylase and mutase were apparently not present on this fragment because their respective antisera did not precipitate any peptide from the products of transcription-translation.

A fragment of the clone was identified as carrying the epimerase gene by transcription-translation of the digested DNA. This also showed that the *P. shermanii* promoter was present on this fragment. The DNA sequence of this fragment contained an open reading frame that encoded a protein of M_r 16 698 which had the N-terminal peptide sequence and amino acid composition expected. Other open reading frames were present on this fragment, but they showed no homology with any previously sequenced proteins.

Methylmalonyl-CoA is a precursor of many antibiotics so it was hoped that cloning the methylmalonyl-CoA-using enzymes from *P. shermanii* into antibiotic producing organisms might affect production. A very large proportion of antibiotics are produced by Streptomycetes and the transcription-translation showed that an *S. lividans* extract could express a *P. shermanii* gene *in vitro*, so part of the original clone was subcloned into *S. lividans* using the multicopy plasmid pIJ702. The recombinant plasmid was stably maintained and epimerase was produced as over 10% of the total cell

protein. The purified protein had enzymatic activity, but on its own epimerase was not expected to affect antibiotic production.

The protein methylmalonyl-CoA mutase had been studied less than epimerase at the beginning of this project. An improved method of purification yielded a protein of M_r 165 000. This was larger than had been reported previously. The enzyme was an αβ-dimer (M_r 79 000 and 67 000). Partial proteolysis showed that these were distinct subunits, rather than the smaller being a proteolytic fragment of the larger, although the amino acid compositions of the two subunits were very similar and they could be related.

Several peptides from a proteolytic digestion of mutase were purified and then sequenced by a manual microscale method. None of the sequences obtained were very good for making oligonucleotides. All had a high proportion of amino acids with very degenerate codons. Instead of making all possible oligonucleotides for one of the peptides, G and C were used at all ambiguous positions and all the 'high G+C' oligonucleotides were made for eight peptides. One that hybridised stringently was used as a probe to clone restriction fragments into *E. coli*.

The recombinant plasmids were used to programme protein synthesis in the *S. lividans* transcription-translation system and caused the synthesis of peptides precipitable by the anti-mutase antiserum. Only small restriction fragments had been cloned in this experiment, but the largest clone encoded a peptide of almost the same size as the smaller subunit of mutase. It was also shown that the *P. shermanii* promoter was present.

The partial mutase clone that has been obtained can now be used as a probe for clones of the intact gene. When one has been obtained it may be sequenced and also cloned into *S. lividans*. Clones of both the epimerase and mutase genes in suitable vectors can then be introduced into other Streptomycetes and the effects on antibiotic production studied.

The overproduction of epimerase by the *S. lividans* clone has enabled large amounts to be purified and it should now be possible to obtain crystals of the enzyme.

REFERENCES

REFERENCES

Abeles, R.H., Brownstein, A.M. & Randles, C.H. (1960) Biochim. Biophys. Acta 41, 530-531

Allen, S.H.G., Kellermeyer, R., Stjernholm, R., Jacobson, B. & Wood, H.G. (1963) J. Biol. Chem. 238, 1637-1642

Anderson, S. & Kingston, I.B. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 6838-6842

Anderson, V.E. & Hammes, G.G. (1984) Biochemistry 23, 2088-2094

Bankier, A.T. & Barrell, B.G. (1983) in 'Techniques in the Life Sciences' B508, 'Nucleic Acid Biochemistry', (Flavell, R.A., ed.), pp1-34, Elsevier Scientific Publishers Ireland Ltd.

Barker, D.G. (1982) Eur. J. Biochem. 125, 357-360

Barker, H.A., Roose, V., Suzuki, F. & Iodice, A.A. (1964) J. Biol. Chem. 239, 3260-3266

Bell, G.I., Merryweather, J.P., Sanchez-Pescador, R., Stempien, M.M., Priestley, l., Scott, J. & Rall, L.B. (1984) Nature 310, 775-777

Bibb, M.J. & Cohen, S.N. (1982) Mol. Gen. Genet. 187, 265-277

Bibb, M.J., Chater, K.F. & Hopwood, D.A. (1983) in 'Experimental Manipulation of Gene Expression', (Inouye, M., ed.), pp53-82, Academic Press, New York

Bibb, M.J., Findlay, P.R. & Johnson, M.W. (1984) Gene 30, 157-166

Brown, C.E., Shemin, D. & Katz, J.J. (1973) J. Biol. Chem. 248, 8015-8021

Buckner, J.S. & Kolattukudy, P.E. (1976) Biochemistry 15, 1948-1957

Burke, J.F. (1984) Gene 30, 63-68

Burnette, W.N. (1981) Anal. Biochem. 112, 195-203

Cane, D.E., Hasler, H. & Liang, T.C. (1981) J. Am. Chem. Soc. 103, 5960-5962

Cannata, J.J.B., Foceci, A., Mazumder, R., Warner, R.C. & Ochoa, S. (1965) J. Biol. Chem. 240, 3249-3257

Cardinale, G.J. & Abeles, R.H. (1968) Biochemistry 7, 3970-3978

Carroll, M.C. & Porter, R.R. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 264-267

Chang, J-Y. (1983) Methods Enzymol. 91, 455-466

Chollet, A. & Kawashima, E.H. (1985) Nucleic Acids Res. 13, 1529-1541

Chou, P.Y. & Fasman, G.D. (1974) Biochemistry 13, 211-245

Clarke, L. & Carbon, J. (1976) Cell 9, 91-99

Clarke, L., Hitzeman, R. & Carbon, J. (1979) Methods Enzymol. 68, 436-442

Cleland, W.W. (1982) Crit. Rev. Biochem. 13, 385-429

Cleveland, D.W., Fischer, S.G., Kirschner, M.W. & Laemmli, U.K. (1977) J. Biol. Chem. 252, 1102-1106

Cohn, D.H., Ogden, R.C., Abelson, J.N., Baldwin, T.O., Nealson, K.H., Simon, M.I. & Mileham, A.J. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 120-123

Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.C. & Wallace, R.B. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 278-282

Corcoran, J.W. (1981) in 'Antibiotics 4: Biosynthesis', (Corcoran, J.W., ed.), pp132-174, Springer-Verlag

Dretzen, G., Bellard,M., Sassone-Corsi, P. & Chambon, P. (1981) Anal. Biochem. 112, 295-298

Efimov, V.A., Buryakova, A.A., Reverdatto, S.V., Chakhmakhcheva, O.G. & Ovchinnikov, Y.A. (1983) Nucleic Acids Res. 11, 8369-8387

Ehrlich, H.A., Cohen, S.N. & McDevitt, H.O. (1979) Methods Enzymol. 68, 443-453

Elmblad, A., Josephson, S. & Palm, G. (1982) Nucleic Acids Res. 10, 3291-3301

Fenton, W.A., Hack, A.M., Huntington, F.W., Gertler, A. & Rosenberg, L.E. (1981) Arch. Biochem. Biophys. 214, 815-823

Flores, M.E. & Sanchez, S. (1985) FEMS Microbiol. Lett. 26, 191-194

Francalanci, F., Davis, N.K., Fuller, J.Q., Murfitt, D. & Leadlay, P.F. (1986) Biochem. J. 236, in press

Friedman, S.M., Kaneda,T. & Corcoran, J.W. (1964) J. Biol. Chem. 239, 2386-2391

Fuller, J.Q. & Leadlay, P.F. (1983) Biochem. J. 213, 643-650

Gait, M.J., Matthes, H.W.D., Singh, M., Sproat, B.S. & Titmas, R.C. (1982a) in 'Chemical and Enzymatic Synthesis of Gene Fragments: A Laboratory Manual', (Gassen, H.G. & Lang, A., eds.), pp1-42, Verlag Chemie GmbH, Weinheim

Gait, M.J., Matthes, H.W.D., Singh, M., Sproat, B.S. & Titmas, R.C. (1982b) Nucleic Acids Res. 10, 6243-6252

Gergen, J.P., Stern, R.H. & Wensink, P.C. (1979) Nucleic Acids Res. 7, 2115-2136

Halpern, J. (1985) Science 227, 869-875

Han, Y.S., Bratt, J.M. & Hogenkamp, H.P.C. (1984) Comp. Biochem. Physiol. 78B, 41-45

Hanahan, D. (1983) J. Biol. Chem 166, 557-580

Hara, O. & Beppu, T. (1982) J. Antibiotics 35, 349-358

Hawkes, R., Niday, E. & Gordon, J. (1982) Anal. Biochem. **119**, 142-147

Hopwood, D.A. & Merrick, M.J. (1977) Bacteriol. Rev. **41**, 595-635

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. & Schrempf, H. (1985) 'Genetic Manipulation of Streptomyces: A Laboratory Manual', The John Innes Foundation, Norwich

Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A.M., Floss, H.G. & Omura, S. (1985) Nature **314**, 642-644

Horinouchi, S., Hara, O. & Beppu, T. (1983) J. Bacteriol. **155**, 1238-1248

Hudson, G.S. & Davidson, B.E. (1984) J. Biol. Chem **180**, 1023-1051

Hudson, L. & Hay, F.C. (1976) 'Practical Immunology', Blackwell Scientific Publications

Hunaiti, A.R. & Kolattukudy, P.E. (1982) Arch. Biochem. Biophys. **216**, 362-371

Hunaiti, A.R. & Kolattukudy, P.E. (1984a) Antimicrob. Agents Chemother. **25**, 173-178

Hunaiti, A.R. & Kolattukudy, P.E. (1984b) Arch. Biochem. Biophys. **229**, 426-439

Ike, Y., Ikuta, S., Sato, M., Huang, T. & Itakura, K. (1983) Nucleic Acids Res. **11**, 477-488

Itakura, K., Rossi, J.J. & Wallace, R.B. (1984) Annu. Rev. Biochem. **53**, 323-356

Jaye, M., de la Salle, H., Balland, A., Kohli, V., Findeli, A., Tolstoshev, P. & Lecocq, J-P. (1983) Nucleic Acids Res. **11**, 2325-2335

Johnson, J.L. & Cummins, C.S. (1972) J. Bacteriol. **109**, 1047-1066

Kaplan, B.H. & Stadtman, E.R. (1968) J. Biol. Chem. **243**, 1794-1803

Kaplan, B.H., Greenfield, L. & Collier, R.J. (1983) Methods Enzymol. **100**, 342-368

Kellermeyer, R.W. & Wood, H.G. (1969) Methods Enzymol. **13**, 207-215

Kempe, T., Sundquist, W.I., Chow, F. & Hu, S-L. (1985) Nucleic Acids Res. **13**, 45-57

Kieser, T. (1984) Plasmid **12**, 19-36

Kinnaird, J.H., Keighren, M.A., Kinsey, J.A., Eaton, M. & Fincham, J.R.S. (1982) Gene 20, 387-396

Kumar, G.K., Bahler, C.R., Wood, H.G. & Merrifield, R.B. (1982) J. Biol. Chem. **257**, 13828-13834

Kung, H.F. & Tsai, L. (1971) J. Biol. Chem. 246, 6444-6451

Kuo, D.J. & Rose, I.A. (1982) J. Am. Chem. Soc. **104**, 3235-3236

Laemmli, U.K. (1970) Nature 227, 680-685

Lancini, G. & Grandi, M. (1981) in 'Antibiotics 4: Biosynthesis', (Corcoran, J.W., ed.), pp12-27, Springer-Verlag, Berlin

Lathe, R. (1985) J. Biol. Chem 183, 1-12

Leadlay, P.F. (1981) Biochem. J. 197, 413-419

Leadlay, P.F. & Fuller, J.Q. (1983) Biochem. J. 213, 635-642

Levy, M., Fishman, L. & Schenkein, I. (1969) Methods Enzymol. 19, 672-681

Maas, R. (1983) Plasmid 10, 296-298

Maloy, W.L., Bowien, B.U., Zwolinski, G.K., Kumar, K.G., Wood, H.G., Ericsson, L.H. & Walsh, K.A. (1979) J. Biol. Chem. 254, 11615-11622

Malpartida, F. & Hopwood, D.A. (1984) Nature 309, 462-464

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) 'Molecular Cloning: A Laboratory Manual', Cold Spring Harbor Lab., CSH, New York

Marshall, V.P., Cialdella, J.I., Fox, J.A. & Laborde, A.L. (1984) J. Antibiotics 37, 923-925

Martin, J-F. (1977) Annu. Rev. Microbiol. 31, 13-38

Matthes, H.W.D., Zenke, W.M., Grundstrom, T., Staub, A., Wintzerith, M. & Chambon, P. (1984) EMBO J. 3, 801-805

Mazumder, R., Sasakawa, S., Kaziro, T. & Ochoa, S. (1962) J. Biol. Chem. 237, 3065-3068

McFarland, G.D., Cohen, L.W., Riggs, A.D., Morris, C., Itakura, K. & Richards, J.H. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6409-6413

Meinkoth, J. & Wahl, G. (1984) Anal. Biochem. 138, 267-284

Mills D.R. & Kramer, F.R. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 2232-2235

Murtif, V.L. Bahler, C.R. & Samols, D. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 5617-5621

Nureddin, A. & Johnson, P. (1977) Biochemistry 16, 1730-1737

Ohlrogge, J.B. & Kuo, T.M. (1985) J. Biol. Chem. 260, 8032-8037

Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. & Matsubara, K. (1985) J. Biol. Chem. 260, 2605-2608

O'Brien, R.J., Fox, J.A., Kopczynski, M.G. & Babior, B.M. (1985) J. Biol. Chem. 260, 16131-16136

O'Keefe, S.J. & Knowles, J.R. (1986) J. Am. Chem. Soc. 108, 328-329

Omura, S. (1981) Methods Enzymol. 72, 520-532

Omura, S. & Tanaka, Y. (1983) in 'Biochemistry and Genetic Regulation of Commercially Important Antibiotics', (Vining, L.C., ed.), pp179-206, Addison-

Wesley Publishing Co.

Omura, S., Takeshima, H., Nakayama, A., Kanemoto, N. & Lukacs, G. (1976) Bioorg. Chem. 5, 451-455

Perham, R.N. (1978) in 'Techniques in the Life Sciences', B110, 'Techniques in Protein and Enzyme Biochemistry', (Kornberg, H.L., Metcalfe, J.C., Northcote, D.H., Pogson, C.I. & Tipton, K.F., eds.), pp1-39, Elsevier/North-Holland, Amsterdam

Rudnick, G. & Abeles, R.H. (1975) Biochemistry **14**, 4515-4522

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B.A. (1980) J. Biol. Chem **143**, 161-178

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. & Peterson, G.B. (1982) J. Biol. Chem **162**, 729-773

Schenkein, I., Levy, M., Franklin, E.C. & Frangione, B. (1977) Arch. Biochem. Biophys. **182**, 64-70

Schleifer, K.H., Plapp, R. & Kandler, O. (1968) Fed. Eur. Biochem. Soc. Lett. **1**, 287-289

Scott, A.I., Beadling, L.C., Georgopapadakou, N.H. & Subbarayan, C.R. (1974) Bioorg. Chem. **3**, 238-248

Shine, J. & Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 1342-1346

Singer-Sam, J., Simmer, R.L., Keith, D.H., Shively, L., Teplitz, M., Itakura, K., Gartler, S.M. & Riggs, A.D. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 802-806

Southern, E. (1975) J. Biol. Chem **98**, 503-517

Sproat, B.S. & Bannwarth, W. (1983) Tetrahedron Lett. **24**, 5771-5774

Sproat, B.S. & Gait, M.J. (1984) in 'Oligonucleotide Synthesis: A Practical Approach', (Gait, M.J., ed.), pp83-115, IRL Press, Oxford

Stabler, S.P., Marcell, P.D. & Allen, R.H. (1985) Arch. Biochem. Biophys. **241**, 252-264

Staden, R. (1982) Nucleic Acids Res. **10**, 4731-4751

Stubbe, I., Fish, S. & Abeles, R.H. (1980) J. Biol. Chem. **255**, 236-242

Thomas, M. & Davis, R.W. (1975) J. Biol. Chem **91**, 315-328

Thompson, J., Rae, S. & Cundliffe, E. (1984) Mol. Gen. Genet. **195**, 39-43

Towbin, H., Staehelin, T. & Gordon, J. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4350-4354

Tsao, S.G.S., Brunk, C.F. & Pearlman, R.E. (1983) Anal. Biochem. **131**, 365-372

Vandekerckhove, J., Bauw, G., Puype, M., Van Damme, J. & Van Montagu, M. (1985) Eur. J. Biochem. **152**, 9-19

Viera, J. & Messing, J. (1982) Gene **19**, 259-268

Vu-Trong, K., Bhuwapathanapun, S. & Gray, P.P. (1980) Antimicrob. Agents Chemother. **17**, 519-525

Vu-Trong, K., Bhuwapathanapun, S. & Gray, P.P. (1981) Antimicrob. Agents Chemother. **19**, 209-212

Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. & Itakura, K. (1981) Nucleic Acids Res. **9**, 879-895

Wallace, R.B., Shaffer, W.J., Murphy, R.F., Bonner, J., Hirose, T. & Itakura, K. (1979) Nucleic Acids Res. **6**, 3543-3557

Westpheling, J., Ranes, M. & Losick, R. (1985) Nature **313**, 22-37

Winter, G., Fersht, A.R., Wilkinson, A.J., Zoller, M. & Smith, M. (1982) Nature **299**, 756-758

Wood, H.G. & Barden, R.E. (1977) Annu. Rev. Biochem. **46**, 385-413

Wood, H.G. & Zwolinski, G.K. (1976) Crit. Rev. Biochem. **4**, 47-122

Wood, W.I., Gitschier, J., Laskey, L.A. & Lawn, R.M. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 1585-1588

Wray, W., Boulikas, T., Wray, V.P. & Hancock, R. (1981) Anal. Biochem. **118**, 197-203

Zagalak, B. & Friedrich, W. (1979) Vitamin B_{12} (eds.), Walter de Gruyter, Berlin

Zagalak, B., Retey, J. & Sund, H. (1974) Eur. J. Biochem. **44**, 529-535

Zwolinski, G.K., Bowien, B.U., Harmon, F. & Wood, H.G. (1977) Biochemistry **16**, 4627-4637

APPENDIX

APPENDIX

The sequence of the 2.3kb fragment carrying the epimerase gene is listed (in bold type), together with the individual gel readings. A negative number for the gel reading indicates that the sequence read was complementary to that shown. An asterisk in a gel reading is a pad inserted to maintain the alignment with other readings. A dash indicates a position where the base could not be identified.

```
                10        20        30        40        50        60
-15  TCGAG*AGGTCATGGCCTCCGCCGTGGCAAATGGCGTGGCAGGCCATGTCAAG
 28  TCGAGCAGGTCATGGCCTCCGCCGTGGCAAATGGCGTGGCAGGCCATGTCAAGCCCGTCT
  3  T*GAGCAGGTCATGGCCTCCGCCGTGGCAAATGGCGTGGCAGGCCATGTCAAGCCCGTCT
 20  T*GAGCAGGTCATGGCCTC*G*CGTGGCAAATGGCGTGGCAGGCCATGT*AA
-41  TCGAG*AGGTCATGGCCTCCGCCGTGGCAAATGGCGTGGCAGGCCATGTCAAGCCCGTC*
-10  TC-AGCAGGTCATGGCCTCCGCC*TGGCAAATGGC*TGGCAGGCCATGTCAAGCCCGTCT
 59  TCGAGCAGGTCATGGCCTCCGCCGTGGCAAATGGCGTGGCAGGCCATGTCAAGCCCGTCT
-26  TCGAGCAGGTCATGGCCTCCGCCGTGGCAAATGGCGTGGCAGGCCATGTCAAGCCCGTCT
 24                                                    TGTCAAGCCCGTCT
     TCGAGCAGGTCATGGCCTCCGCCGTGGCAAATGGCGTGGCAGGCCATGTCAAGCCCGTCT

                70        80        90        100       110       120
 28  CCCACGGGGAGGCCGGCCCCGATCCGCGCTTCGCCGCAGCCGACACGGCGATGGAGGCCG
  3  CC*A*GGGGAGGCCGGCCC*GATCCGCGCTTCGCCGCAGCCGACACGGCGATGGAGGCCG
-41  *CCACGGGGA
-10  CCCACGGGGAGGCCGGCCCCGATCCGCGCTTCGCCGCAGCCGACACGGCGATGGAGGCC-
 59  CCCACGGGGAGGCCGGCCCCGATCCGCGCTTCGCCGCAGCCGACA*GGCGATGGAGGCCG
-26  CCCACGGGGAGGCCGGCCCCGATCCGCG*TTCGCCGCACGCGACACGGC
 24  CCCACGGGGAGGCCGGCCCCGATCCGCGCTTCGCCGCAGCCGACA*GGCGATGGAGGCCG
-13                                TTCGCCGCAGCCGACACGGCGATGGAGGCCG
     CCCACGGGGAGGCCGGCCCCGATCCGCGCTTCGCCGCAGCCGACACGGCGATGGAGGCCG

                130       140       150       160       170       180
 28  AGGACTACGACCGGGCGGCCGACGAGTTCGGCAAGCTTCTGGCGGCCAATCCCAAGGACA
  3  AGGACTACGACCGGG
-10  AGGACTACGACCGGGCGGCCGACGAGTTCGGCAAGCTTCTGGCGGCCAATCCCAAGGACA
 59  AGGACTACGACCGGGCGGC
 24  AGGACTACGACCGGGCGGCCGACGAGTTCGGCAAGCTTCTGGCGGCCAATCCCAAGGACA
-13  AGGACTACGACCGGGCGGCCGAC*A*TTCGGCAAGCTTCTGGCGGCCAATCCCAAGGACA
 30       TACGACCGGGCGGCCGACGAGTTCGCGAAGCTTCTGGCGGCCAATCCCAAGGACA
     AGGACTACGACCGGGCGGCCGACGAGTTCGGCAAGCTTCTGGCGGCCAATCCCAAGGACA

                190       200       210       220       230       240
 28  GCGAGGCCGCTGCCGGGCAGGCCACCGCA*GGCTGATGTCGCGCGCCG*GAATGCCGATC
-10  GCGAGGCCGCTGCCGGGCAGGCCA-CGCACGGCTGATGTCGC*CGCCGC*AATGCCGATC
 24  GCGAGGCCGCTGCCGGGCAGGCCACCGCACGGCTGATGTCGCGCGCCGCGAATGCCGATC
-13  GCGAGGCCGCTGCCGGGCAGGCCACCGCACGGCTGATGTCGCGCGCCGCGAATGCCGATC
 30  GCGAGGCCGCTGCCGGGCAGGCCACCGCACGGCTGATGTCGCGCGCCGCGAATGCCGATC
     GCGAGGCCGCTGCCGGGCAGGCCACCGCACGGCTGATGTCGCGCGCCGCGAATGCCGATC
```

```
          250        260        270        280        290        300
 28 CGGAAGCGACG*TTGCCGCCGCCAAGGCG*CACC
-10 C
 24 CGGAAGCGACGCTTGCCGCCGCCAAGGCG**ACCCGACGACGTACCGGCGGCGATGGCCG
-13 CGGAAGCGACGCTTGCC**CGCCAAGGCGGCACCCGACGACGTA-CGGCGGCGATGGCCG
 30 CGGAAGCGACGCTTGCCGCCGCCAAGGCG**ACCCGACGACGTACCGGCGGCGATGGCCG
-45         ACG*TTG*CG**GCCAAGGCGGCA*CCGACGACGTACC*GCGGCGATGGCCG
-57                                                    ATGGCCG
-60                                                      GGCCG
    CGGAAGCGACGCTTGCCGCCGCCAAGGCGGCACCCGACGACGTACCGGCGGCGATGGCCG

          310        320        330        340        350        360
 24 CCTCCGACGTCGACATGATTGCCGGGCGTCCCAAGGACCGCTTCGGCCGCCTGATCGGCC
-13 CCTCCGACGTCGACATGATTGCCG*CGGTCCCAAGGACCG*TTCGGCCGCCTGATCGGCC
 30 CCTCCGACGTCGACATGATTGCCGGGCGTCCCAAGGACCGCTTCGGCCGCCTGATCGGCC
-45 CCTCCGACGTCGACATGATTGCCGGCGGTCCCAAGGACCG*TTCGGCCGCCTGATCGGCC
-57 CCTCCGACGTCGACATGATTGCCGG*C*TCCCAAGGACCG*TTCGGCCGCCTGATCGGCC
-60 CCTCCGACGTCGACATGATTGCCGGGC*TCCCAAGGACCG*TTCGGCCGCCTGATCGGCC
-65               ATGA*TGC*GGGC*TCCCAAGGAC*G*TTCGG*CGC*TGATCGGCC
    CCTCCGACGTCGACATGATTGCCGG--GTCCCAAGGACCGCTTCGGCCGCCTGATCGGCC

          370        380        390        400        410        420
 24 TCATC*GCAC*A*TGCCGGCGACGAGCGCGACGCAGTCCGCACGCG*CT*CTGGAGCTGT
-13 TCATCCGCACCACTGCCGGCGACGAGCGCGACGCAGTCCG
 30 TCATCCGCA*CACTGCCGGCGACGAGCGCGACG*AGTCCGCACGCGGCTCCTGGAGCTGT
-45 TCATCCGCACCACTGCCGGCGACGAGCGCGACGCAGTC*GCACGCGGCTCCTGGAGCTGT
-57 TCATCCGCACCACTGCCGGCGACGAGCGCGACGCAGTCCGCACGCGGCTCCTGGAGCTGT
-60 TCATCCGCACCACTGCCGGCGACGAGCGCGACGCAGTCCGCACGCGGCTCCTGGAGCTGT
-65 TCATCCGCACCACTGCCGGCGACGAGCGCGACGCAGTCCGCACGCGGCTCCTGGAGCTGT
  6                                  ACGCAGTCCGCACGCG*CTCCTGGAGCTGT
    TCATCCGCACCACTGCCGGCGACGAGCGCGACGCAGTCCGCACGCGGCTCCTGGAGCTGT

          430        440        450        460        470        480
 24 *CGAGAC*ATGGAC*AGGC*GATCC*GA
 30 TCGAGACCATGGACCAGGCCGATCCCGAACTGCTGGCC*CA*GTCGCGCCCT*GGG
-45 TCGAGACCATGGACCAGGCCGATCCCGAACTGCTGGCCGCACGTCGCGCCCTCGGGGCGG
-57 TCGAGACCATGGACCAGGCCGATCCCGAACTGCTGGCCGCACGTCG*GCCCTCGGGGCGG
-60 TCGAGACCATGGACCAGGCCGATCCCGAACTGCTGGCCGCACGTCGCGCCCTCGGGGCGG
-65 TCGAGACCATGGACCAGGCCGATCCCGAACTGCTGGCCGCACGTCGCGCCCTCGGGGCGG
  6 TCGAGACCATGGACCAGGCCGATCCCGAACTGCTGGCCGCACGTCGCGCCCTC*GGGCGG
-38                                                  CCCTCGGGGCGG
-47                                                    CTCGGG*C*G
    TCGAGACCATGGACCAGGCCGATCCCGAACTGCTGGCCGCACGTCGCGCCCTCGGGGCGG
```

```
          490        500        510        520        530        540
-45  CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTTCGGCACTTCTCACCGCGGCCCT
-57  CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTTCGGCACTTCTCACCGCGGCCCT
-60  CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTTCGGCACTTCTCACCGCGGCCCT
-65  CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTTCGGCACTTCTCACCGCGGCCCT
  6  CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTT*GGCACTTCT*ACCGCGGCCCT
-38  CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTTCGGCACTTCTCACCGCGGCCCT
-47  CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTTCGGCACTTCTCACCGCGGCCCT
-31                                AGCGCCTTTTCGGCACTTCTCACCGCGGCCCT
     CCCTGTACTAGGCCCCCGCTCCCGGCCGAGCGCCTTTTCGGCACTTCTCACCGCGGCCCT

          550        560        570        580        590        600
-45  TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
-57  TTCTCCAGGGCCCGGATTTG*GACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
-60  TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
-65  TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
  6  TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
-38  TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
-47  TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
-31  TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC
     TTCTCCAGGGCCCGGATTTGAGACCGCGCGGCGTGTGGTTCCTCCGTTCCCGACGACCAC

          610        620        630        640        650        660
-45  CGGACCGGAGGAACCCCCG*CGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
-57  CGGACCGGAGGAACCCC*GCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
-60  CGGACCGGAGGAACCCCCGCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
-65  CGGACCGGAGGAACCCCCGCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
  6  CGGACCGGAGGAACCCCCGCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
-38  CGGACCGGAGGAACCCCCGCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
-47  CGGACCGGAGGAACCCCCGCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
-31  CGGACCGGAGGAACCCCCGCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC
     CGGACCGGAGGAACCCCCGCCGCGGTGCCGGCGGCAATATCCGCGCGGTAACACCGACCC

          670        680        690        700        710        720
-45  GGGACGGATGGATGCGAATACCATCTGGTCACA
-57  GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATG
-60  GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATGATGGCC
-65  GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATGATGGCC
  6  GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATGATGGCC
-38  GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATGATGGCC
-47  GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATGATGGCC
-31  GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATGATGGCC
-14                                CGGCGGCAT-*T*GGATGATGGCC
     GGGACGGATGGATGCGAATACCATCTGGTCACATCGCGGCGGCATTGTGGGATGATGGCC
```

```
          730        740        750        760        770        780
-60  CCATGAGTAATGAGGATCTTTTCATCTGTATCGAT
-65  CCATGAGTAATGAGGATCTTTTCATCTGTATCGATCACGTGGCATATGCGTGCC**GACG
  6  CCATGAGTAATGAGGATCTTTTCATCTGTATCGATCA
-38  CCATGAGTAATGAGGATCTTTTCATCTGTATCGATCACGTGGCATATGCGTGCCCCGACG
-47  CCATGAGTAATGAGGATCTTTTCATCTGTATCGATCACGTGGCATATGCGTGCC
-31  CCATGAGTAATGAGGATCTTTTCATCTGTATCGATCACGTGGCATATGCGTGCCCCGACG
-14  CCATGAGTAAT*AGGATCTTT*CATCT-TATCGATCAC*T*GCATAT*C*TGCCCCGACG
-56  CCATGAGTAATGAGGATCTTT*CATCTGTATCGATCACGTGGCATATGCGTGCCCCGACG
-55        TAATGAGGATCTTT*CATCTGTATCGATCAC*T*GCATATGC*TGCCCCGACG
 21                          TATCGATCACGTGGCATATGCGTGCCCCGACG
 54                                                ATATGCGTGCCCCGACG
     CCATGAGTAATGAGGATCTTTTCATCTGTATCGATCACGTGGCATATGCGTGCCCCGACG

          790        800        810        820        830        840
-38  CCGACGAGGCTTCCAAGTACTACCAGGAGAC*TTCGGCTGGCATGA
-31  CCGACGAGGCTTCCAA
-14  CCGACGAGGCTTCCAAGTACTACCAGGAGACCTT*GGCTGGCATGAGCTCCACCGCGAGG
-56  CCGACGAGGCTTCCAAGTACTACCAGGAGACCTTCGGCTGGCATGAGCTCCACCGCGAGG
-55  CCGACGAGGCTTCCA*GTACTACCAGGAGACCTTCGGCTGGCATGAGCTCCACCGCGAGG
 21  CCGACGAGGCTTCCAAGTACTACCAGGAGACCTTCGGCTGGCATGAGCTCCACCGCGAGG
 54  CCGACGAGGCTTCCAAGTACTACCAGGAGACCTTCGGCTGGCATGAGCTCCACCGCGAGG
 -5        AGGC*TCCA*GTACTACCAGGAGACCTTCGGCTGGCATGAGCTCCACCGCGAGG
-63        AGGC*TCCA*GTACTACCAGGAGACC*TCGGCTGGCATGAGCTCCACCGCGAGG
-40             TCCA*GTACTACCAGGAGACCTTCGGCTGGCATGAGCTCCACCGCGAGG
-43             TCCAAGTACTACCAGGAGAC*TTCGGCTGGCATGAGCTCCACCGCGAGG
-17                   AC*ACCA*GAGAC*TTCGGCTGGCAT*AG*TCCACCGC*A*G
-35                                                TGAGCTCCACCGC*AGG
 51                                                      ACCGCGA*G
     CCGACGAGGCTTCCAAGTACTACCAGGAGACCTTCGGCTGGCATGAGCTCCACCGCGAGG

          850        860        870        880        890        900
-14  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-56  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-55  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
 21  AGAACCCGGAGC
 54  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
 -5  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-63  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-40  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-43  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-17  AGAACCCG*AGCAGG*A*TCGTC*AGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-35  AGAACCCGGAGCAGGGAGTCGTC*AGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
 51  AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAACGTGACCGAGC
-52                 AGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
-33                           G*TGG*CCCGGCTGCGAAGCTGACCGAGC
-34                             TGG*CCCGGCTGCGA*GCTGACCGAG*
    AGAACCCGGAGCAGGGAGTCGTCGAGATCATGATGGCCCCGGCTGCGAAGCTGACCGAGC
```

```
          910        920        930        940        950        960
-14  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-56  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-55  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
 54  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
 -5  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-63  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-40  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-43  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-17  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAA*TGGC
-35  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
 51  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-52  ACATGACCCAG*T*CAG*TCATGGCCCCGCTCAACGACGAGTCGACC*T*G*CAA*TGGC
-33  ACATGACCCAGGT*CAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC
-34  ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGG*
 -7            GTT*AGGTCATGG*CCCGCTCAACGACGAGTCGACCGTTG*CAAGTGGC
     ACATGACCCAGGTTCAGGTCATGGCCCCGCTCAACGACGAGTCGACCGTTGCCAAGTGGC

          970        980        990       1000       1010       1020
-14  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-56  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-55  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
 54  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCA-ATGGCATGGCGTGTCGATGACATCG
 -5  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-63  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-40  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-43  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-17  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGC*TGTCGATGACATCG
-35  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
 51  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-52  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-33  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
-34  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
 -7  TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATG*CAT*GCGTGTCGATGACATCG
 62     CCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
 22                                          GCATGGCGTGTCGATGACATCG
     TTGCCAAGCACAATGGTCGCGCCGGACTGCACCACATGGCATGGCGTGTCGATGACATCG
```

```
          1030        1040        1050        1060        1070        1080
-14  ACGCC
-56  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-55  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 54  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 -5  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-63  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-40  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-43  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-17  ACGCCGTCA*CGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-35  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 51  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-52  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-33  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
-34  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 -7  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 62  ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 22  A*GCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 61            CA*CCTGCGCGA*CGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC
 -8                                    G*TGTATGACGAG*CCAAGC
     ACGCCGTCAGCGCCACCCTGCGCGAGCGCGGCGTGCAGCTGCTGTATGACGAGCCCAAGC

          1090        1100        1110        1120        1130        1140
-56  TCGGCACCGGCGGCAACCGCATCAACTTCATG
-55  TCGGCACCGGCGGCAACCGCA
 54  T*GGCAC*GGCGGCAAC*GCATCAACTTCATGCATCC*AAGT*GG*CA
 -5  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
-63  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
-40  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
-43  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
-17  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
-35  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
 51  TCGGCACCGGCGGCAACCG*ATCA*CTTCATGCATCCCAAGTCGGGCA*GGGCGTGCTCA
-52  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
-33  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
-34  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
 -7  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
 62  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
 22  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
 61  TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
 -8  TCGGCACC*GCGGCAACCGCATCAACT*CATGCATCC*AAGTCGG*CAA*GGCGTGCTCA
-29            AACCGCATCAACTTCATGCAT*CCAAGTCGGGCAAGGGCGTGCTCA
 49                  TCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
 37                              ATCCCAAGTCGGGCAAGGGCGTGCTCA
     TCGGCACCGGCGGCAACCGCATCAACTTCATGCATCCCAAGTCGGGCAAGGGCGTGCTCA
```

```
          1150       1160       1170       1180       1190       1200
 -5 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
-63 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCG
-40 TCGAGCTCACCCAGT
-43 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTA
-17 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGT
-35 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 51 TCGAGCTCACC*AGTACC*GAAGAACTGACTGCGAGTAGTCGATTAGTCGCC*GCGGGTC
-52 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
-33 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
-34 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 -7 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 62 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 22 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 61 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 -8 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
-29 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 49 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 37 TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
 39    AGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC
    TCGAGCTCACCCAGTACCCGAAGAACTGACTGCGAGTAGTCGATTAGTCGCCCGCGGGTC

          1210       1220       1230       1240       1250       1260
 -5 CTGATCACCACCGACCCG
-35 CTGATCACCACC
 51 CTGATCAC*A*CGACC*G*CGA
-52 CTGATCACCACCGACCCGGCGACATCTGCCT*T*CCGGATCCATCGTCGGATGCGCCACA
-33 CTGATCACCACCGACCCGGCGACATCTGCCTTCG*CGGATCCATCGTCGGATGCGCCA-A
-34 CTGATCACCACCGACCCGGCGACATCTGCCGTTAG*G**TCCATCGTCGGATGCGC
 -7 CTGATCACCACCGACCCGGCGACATCTGCCTTCG**GCATCCATCGTCGGATGCGCCACA
 62 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
 22 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
 61 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
 -8 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCG*CA*A
-29 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
 49 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
 37 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
 39 CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
  1                                                GGATGCGCCACA
-18                                                        CA
    CTGATCACCACCGACCCGGCGACATCTGCCTGTGGCGCATCCATCGTCGGATGCGCCACA
```

```
          1270       1280       1290       1300       1310       1320
-52  GCTGCGTTCACGACCACTTTCA
-33  GCTGCGTTCACGACCACTTT
 -7  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
 62  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
 22  GCTGCGTTCACGACCA*TTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
 61  GCTGCGTTCACGACCACTTTCA*CCCATCCCCACCCACCAGCTTCCGAGAAGGGAACG*C
 -8  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
-29  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
 49  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
 37  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
 39  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
  1  GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC
-18  GCTGCGT*CACGACCAC*TTCACCCCATCCCCACCCACCAGCT*CC*AGAAGGGAACGCC
 -4                    TT*A*CCCATCCC*ACCCA-CAGCTTCCGAGAAGGGAACGCC
 48                                                TTCCGAGAAGGGAACGCC
     GCTGCGTTCACGACCACTTTCACCCCATCCCCACCCACCAGCTTCCGAGAAGGGAACGCC

          1330       1340       1350       1360       1370       1380
 -7  GCCCATGATGCGAGT
 62  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATC
 22  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCC*TACCA
 61  GCCCATGATGCGAGTGCGC*AGGC*ATCGC*AGGATCGGGTGGGCGATCAGTCCCTACCA
 -8  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
-29  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
 49  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
 37  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
 39  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
  1  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
-18  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
 -4  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
 48  GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
 36        GATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
-53                                          ATCGGGTGGGCGATCAGTCCCTACCA
     GCCCATGATGCGAGTGCGCCAGGCCATCGCCAGGATCGGGTGGGCGATCAGTCCCTACCA
```

```
          1390       1400       1410       1420       1430       1440
 22 GCTGGTCGGCCGGGAA
 61 GCTGGTCGGCCGGGA
 -8 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGG*CGCGCCCCACACCAG
-29 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGG*CGCGCCCCACACCAG
 49 *CTGGTCGGCCGGGAA*TGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCC*ACAC*A*
 37 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCC*ACACCAG
 39 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCCCACACCAG
  1 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCCCACACCAG
-18 GCTG*TCGGCCGGGAACTGACCAGCGAGCCGGGC*TGGTCCTGGGCGCGCCCCACACCAG
 -4 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCCCACACCAG
 48 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCCCACACCAG
 36 GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCCCACACCAG
-53 GCTG*TCGGCCGG*AACTGACCAGCGAGCCGGGC*TGGTCCTGG*CGCGCCCCACACCA*
 25                                   GTGGT*CTGGGCGCGCC**ACAC*AG
 44                                                         AG
    GCTGGTCGGCCGGGAACTGACCAGCGAGCCGGGCGTGGTCCTGGGCGCGCCCCACACCAG

          1450       1460       1470       1480       1490       1500
 -8 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAA
-29 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCT*GTATTACC*GGTGCCGCTCAA
 49 TAA*TGGGATTT*AT*G*CTT*CTGG*CGTGTC*TGGTAT*A*CGG*TG*CG*T*A*G*T
 37 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCG*T
 39 TAA*TGGGATTT*AT*GCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCG*T
  1 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAAGGT
-18 TAACTGGGATTTCATCGCCTTCCTGGGC*TGTCCTGGTATTACCGGGTGCCGCTCAAGGT
 -4 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAAGGT
 48 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAAGGT
 36 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAAGGT
-53 TAACTGGGATTTCATCGCCTTCCTGGGC*T-TCCTGGTATTACCGGGTGCCGCTCAAG-T
 25 TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAAGGT
 44 TAACT*GGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAAGGT
    TAACTGGGATTTCATCGCCTTCCTGGGCGTGTCCTGGTATTACCGGGTGCCGCTCAAGGT

          1510       1520       1530       1540       1550       1560
 49 GCTGGTGA*GA*GT*GTG*ATG
  1 GCTGGTGAAGAAGTCGTGGATGCGGGGGCCCCTGTGG**CCTCGGAAAGGCCCTCGGCGC
-18 GCTGGTGAAGAAGTCGTGGATGCGG*GGCCCCTGTGGGCCCTCGGAAAGGCCCTCGGCGC
 -4 GCTGGTGAAGAAGTCGTGGATGCGGCGG*CCCTGTGGGCCCTCGGAAAGGCCCTCGGCGC
 48 GCTGGTGAAGAAGTCGTGGATGCGGGGGCCCCTGTGGG*CCTCGGAAAGGCCCTCGGCGC
 36 GCTGGTGAAGAAGTCGTGGATGCGGGGGCCCCTGTGGGCCCTCGGAAAGGCCCTCGGCGC
-53 GCTGGTGAAGAAGTCGTGGATGCGGGGGCCCCTGTGGGCCCTCGGAAAGGCCCTCGGCGC
 25 GCTGGTGAAGAAGTCGTGGATGCGGGGGCCCCTGTGGG*CCTCGGAAAGGCCCTCGGCGC
 44 GCTGGTGAAGAAGTCGTGGATGCGGGGGCCCCTGTGGG*CCTCGGAAAGGCCCTCGGCGC
 32                    GATGCGG*G*CCCCTGTGG***CTCGGAAAGGCCCTCGGCGC
    GCTGGTGAAGAAGTCGTGGATGCGGGGGCCCCTGTGGGCCCTCGGAAAGGCCCTCGGCGC
```

```
          1570       1580       1590       1600       1610       1620
  1 CGTGGCG*TG*ATCGCGCCCATCCC
-18 CGTGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCACCT
 -4 CGTGGCGGTGGATCG
 48 CGTGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCAC*TGGTGGCGCAGGCAGA
 36 CGTGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCAC*TGGTGGCGCAGGCAGA
-53 C*TGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCACCTGGTGGCGCAGGCAGA
 25 CGTGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCACCTGGTGGCGCAGGCAGA
 44 CGTGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCACCTGGTGGCGCAGGCAGA
 32 CGTGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCACCTGGTGGCGCAGGCAGA
    CGTGGCGGTGGATCGCGCCCATCCCGGACAGGTGGTGGACCACCTGGTGGCGCAGGCAGA

          1630       1640       1650       1660       1670       1680
 48 GCAGGGACACTCCTTCAAGCTCGTCATCGCCCCCAAGGG*ACCCGATCA*CTCGCCAGTA
 36 GCAGGGACACTCCTTCAAGCTCGTCATCGCCCCCAAGGG*ACCCGATCACCTCGCCAGTA
-53 GCAGGGACACTCCTTCAAGCTCGTCATCGCCCC*AAGGGGACCCGATCACCTCGCCAGTA
 25 GCAGGGACACTCCTTCAAGCTCGTCATCGCCCCCAAGGGGACCCGATCACCTCGCCAGTA
 44 GCAGGGACA-TCCTTCAAGCTCGTCATCGCCCCCAAGGGGACCCGATCA*CTCGCCAGTA
 32 GCAGGGACA-TCCTTCAAGCTCGTCATCGCCCCCAAGGGGACCCGATCACCTCGCCAGTA
    GCAGGGACACTCCTTCAAGCTCGTCATCGCCCCCAAGGGGACCCGATCACCTCGCCAGTA

          1690       1700       1710       1720       1730       1740
 48 CTG*A
 36 CTGGAAGTCCGGCTTCTACCG*ATCGCACTGG*T
-53 CTGGAAGTCCGGCTTCTACCGGATCCGACTGGGTGCCGGGCT
 25 CTGGAAGTCCGGCTTCTACCGGATCGCACTGGGTGCCGGGCTGCCGGTGACGCTGGCGGG
 44 CTGGAAGTCCGGCTTCTACCGGATCGCACTGGGTGCCGGGCTGCCGGTGACGCTGGCGGG
 32 CTGGAAGTCCGGCTTCTACCGGATCGCACTGGGTGCCGGGCTGCCGGTGACGCTGGCGGG
 64               CTACCGGATCGCACTGGGTGCCGGGCTGCCGGTGACGCTGGCGGG
    CTGGAAGTCCGGCTTCTACCGGATCGCACTGGGTGCCGGGCTGCCGGTGACGCTGGCGGG

          1750       1760       1770       1780       1790       1800
 25 CATCGATGCCGGTCGCCGACA*GTCGAGGTCGGACCCACGATCCGCCTGACCGGCGATGT
 44 CATCGATGCCGGTCGCCGACAGGTCGAGGTCGGACCCACGATCCGCCTGACCGGCGATGT
 32 CATCGATGCCGGTCGCCGACAGGTCGAGGTCGGACCCACGATCCGCCTGACCGGCGATGT
 64 CATCGATGCCGGTCGCCGACAGGTCGAGGTCGGACCCACGATCCGCCTGACCGGCGATGT
-12                                                           T
    CATCGATGCCGGTCGCCGACAGGTCGAGGTCGGACCCACGATCCGCCTGACCGGCGATGT

          1810       1820       1830       1840       1850       1860
 25 CCATGCCGACATGGATCGC*TCCGTGCC*TCTA*GAC
 44 CCATGCCGACATGGATCGCATCCGTGCCTTCTACGACCGCTTCGATGGCGTCCACC*GCA
 32 CCATGCCGACATGGATCG*ATCCGTGCCTTCTACGACCGCTTCGATGGCGTCCACCCGCA
 64 CCATGCCGACATGGATCGCATCCGTGCCTTCTACGACCGCTTCGATGGCGTCCACCCGCA
-12 CCATGCCGACAT*GATCGCATCC*TGCCTTCTACGACCGCTTCGAT*GCGTCCACCCGCA
    CCATGCCGACATGGATCGCATCCGTGCCTTCTACGACCGCTTCGATGGCGTCCACCCGCA
```

```
          1870        1880        1890        1900        1910        1920
 44 ACTGCGCTCCGACCCGCGG
 32 A-TGCGCTCCGACCCG*GGCTGCGCGAGGAGGACTC*TGACC*GGCCGTCGCGCGGTCA
 64 ACTGCGCTCCGACCCGCGGCTGCGCGAGGAGGACTCCTGACCGGGCCGTCGCGCGGTCAC
-12 ACTGCGCTCCGACCCGCGGCTGCGCGAGGAGGACTCCTGACCGGGCCGTCGCGCGGTCAC
 46                    GCTG*GC*AGGAGGACTCCTGACCGGGCCGTCGCGCG*TCAC
 -9                                                  CGCGCGGTCAC
-16                                                      CGGTCAC
    ACTGCGCTCCGACCCGCGGCTGCGCGAGGAGGACTCCTGACCGGGCCGTCGCGCGGTCAC

          1930        1940        1950        1960        1970        1980
 64 GGCGACATCCAGATCCCCATCCACC*GACAAGTACAC*TGCCTGAGCCTGGTCGGATGCC
-12 GGCGACATCCAGATCCCCATCCACCCGACAAGTACACCTGCCTGAGCCTGGTCGGATGCC
 46 GGCGACATCCAGATCCCCATCCACCCGACAAGTACACCTGCCTGAGCCTGGTCGGATGCC
 -9 GGCGACATCCAGATCCCCATCCACCCGACAAGTACACCTGCCTGAGCCTGGTCGGATGCC
-16 GGCGACATCCAGAT*CCCATCCACCCGACAAGTACACCTGCCTGAGCCTGGTCGGATGCC
-50                              ACA**TACACCTGCCTGAGCCTGGTCGGATGCC
 58                                  TACACCTGCCTGAGCCTGGTCGGATGCC
  2                                                           C
    GGCGACATCCAGATCCCCATCCACCCGACAAGTACACCTGCCTGAGCCTGGTCGGATGCC

          1990        2000        2010        2020        2030        2040
 64 CCCACGTCCGGGCGGTTCAC*AGTTCA*TCCCGCCCGGAATTGGTAGCCT*GAT*CAGAC
-12 CCCACGTCCGGGCGGTTCACAAGTTCAATCCCGCCCGGAATTGGTAGCCTCGATCCAGAC
 46 CCCACGTCCGGGCGGTTCACAAGTTCAATCCCGCCCGGAATTGGTAGCCTCGATCCAGAC
 -9 CCCACGTCCGGGCGGTTCACAAGTTCAATCCCGCCCGGAATTGGTAGCCTCGATCCAGAC
-16 CCCACGTCCGGGCGGTTCACAAGTTCAATCCCGCCCGGAATTGGTAGCCTCGATCCAGAC
-50 CCCACGTCCGG*CG*TTCACAAGTTCAATCCCGCCC*GAAT*GGTAGCCTCGATCCAGAC
 58 CCCACGTCCGGGCGGTTCACAAGTTCAATCCCGCCCGGAATTGGTAGCCTCGATCCAGAC
  2 CCCACGTCCGGGCGGTTCACAAGTTCAATCCCGCCC*GAATTGGTAGCCTCGATCCAGAC
-42       CCGGGCGGT*CACAAGT*CAA*CCCGCCCGGAATTGGTAGCCTCGATCCAGAC
-23                                 CCCGCCCGGAATTGGTAGCCTCGATCCAGAC
    CCCACGTCCGGGCGGTTCACAAGTTCAATCCCGCCCGGAATTGGTAGCCTCGATCCAGAC

          2050        2060        2070        2080        2090        2100
 64 *TG*CCGAGGAGC*TTCGATGAGCC*ATA*GA*GA*AGC*A*A*CGGTGATGACCAGAGC
-12 CTGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
 46 *TGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
 -9 CTGCCCGAGGAGCCTTCGAT*AGCCAATACGACGACAGCCACACC*GTGATGACCAGAGC
-16 CTGCCC*AGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
-50 CTGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
 58 CTGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
  2 CTGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
-42 CTGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
-23 CTGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
 19                                 TACGACGACAGCCACACCGGTGATGACCAGAGC
    CTGCCCGAGGAGCCTTCGATGAGCCAATACGACGACAGCCACACCGGTGATGACCAGAGC
```

```
          2110       2120       2130       2140       2150       2160
 64  GAGGAGA*CGG*CT
-12  GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
 46  GAGGAGA**GGCCT*AAC*TCTTCGACGA*CGGGCCAGCGCGGCCGGGAGCTTCCCC*AC
 -9  GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
-16  GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
-50  GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
 58  GAGGAGACCGGCCT*AACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
  2  GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
-42  GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
-23  GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
 19  GAGGAGACCGGCCT*AACCTCTT*GACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC
 27                                  GGCCAGCGCGGC**GGAGCTTCCCCCAC
     GAGGAGACCGGCCTCAACCTCTTCGACGACCGGGCCAGCGCGGCCGGGAGCTTCCCCCAC

          2170       2180       2190       2200       2210       2220
-12  GCGATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT
 46  GCGATGATGGGATACGACCGGAGCA*GGTCGACA**TACGTACGTGACCTGGAGCAGCGT
 -9  GCGATGATGGGATAC*ACCGGAGCACGGTCGACAACT
-16  GCGATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT
-50  GCGATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT
 58  GCGATGATGGGATA*GACCGGAGCA*GGTCGACAACTACGTACGTGACCTGGAGCAGCGT
  2  GCGATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT
-42  GCGATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT
-23  GCGATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT
 19  GCGATGATGGGATACGACCGGAGCA*GGTCGACAACTACGTACGTGACCTGGAGCAGCGT
 27  G-GATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT
 11                                                    GGAGCAGCGT
     GCGATGATGGGATACGACCGGAGCACGGTCGACAACTACGTACGTGACCTGGAGCAGCGT

          2230       2240       2250       2260       2270       2280
-12  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
 46  CTCTCGGCAGCC*GC*AGCTC*ATCGCGATCGGCTGCGCGACC
-16  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGG*TGCGCGACC
-50  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
 58  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
  2  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
-42  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
-23  CTCTCG*CAG*CCGCCAGCTCAATCGCG
 19  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCG-GACC
 27  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
 11  CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
     CTCTCGGCAGCCCGCCAGCTCAATCGCGATCGGCTGCGCGACC
```