Nucleotide Sequence and Mutagenesis of the Gene Encoding Isocitrate Lyase From *Escherichia coli* ML308

by

Alan George Simpson Robertson

A Thesis submitted for the degree of Doctor of Philosophy in the Faculty of Science, University of Glasgow

February 1992

For Mum and Dad

and

In Loving Memory Of My Gran

CONTENTS

Page

List Of Contents

List Of Figures

List Of Tables

Abbreviations

Acknowledgements

Summary

Chapter 1: Introduction

1.1	**The Glyoxylate Bypass:**	1
1.1.1	Introduction:	1
1.1.2	Growth Of *Escherichia coli* On Fatty Acids And Acetate:	1
1.1.3	The Glyoxylate Bypass:	2
1.2	**Control Of The Branchpoint Between ICL and ICDH In *E. coli*:**	4
1.2.1	Reversible Inactivation Of ICDH:	4
1.2.2	Phosphorylation Of ICDH *in vivo*:	10
1.2.3.	ICDH Kinase/Phosphatase:	11
1.2.4	Control Of ICDH Phosphorylation:	13
1.2.5.	Mechanism Of Inactivation Of ICDH:	15
1.3	**The Glyoxylate Bypass (*ace*) Operon:**	18
1.3.1	Genetic Organisation Of The Glyoxylate Bypass:	18
1.3.2	Evidence For A Fourth *ace* Gene:	19
1.3.3	Genetic Control Of The *ace* Operon:	20
1.3.4	Possible Advantages Of A Dual Repressor System:	22

1.3.5	Metabolic Control Of The *ace* Operon:	22
1.3.6	Downshift In Expression From *aceK*:	24
1.4	**Isocitrate Lyase:**	26
1.4.1	Physical Properties:	26
1.4.2	Requirement Of ICL for Magnesium:	27
1.4.3	K_m Of ICL From *E. coli* For Isocitrate:	29
1.4.4	Effect OF pH On ICL:	31
1.4.5	Effect Of Anions On ICL:	33
1.4.6	Potential Metabolic Regulators Of ICL:	33
1.4.7	K_m Of ICL For Isocitrate *in vivo*:	35
1.4.8	Phosphorylation Of ICL:	35
1.4.9	Kinetic Mechanism Of ICL From *E. coli* ML308:	37
1.4.10	Active Site Cysteine Residue:	41
1.4.11	Other Active Site Groups:	43
1.4.12	Catalysis:	44
1.5	**Aims Of This Thesis:**	45

Chapter 2: Materials And Methods

2.1	**Materials:**	47
2.2	**General Biochemical Methods:**	48
2.2.1	pH Measurements:	48
2.2.2	Conductivity Measurements:	48
2.2.3	Glassware And Plastics:	48
2.2.4	Distilled Water:	48
2.2.5	Protein Estimations:	49
2.2.6	Dialysis:	49
2.2.7	Preparation Of Chromatographic Media:	49
2.2.8	Spectrophotometric Determination Of Nucleic Acid Concentrations:	49
2.2.9	Sterilisation:	50
2.3	**Microbiological Techniques:**	50
2.3.1	Bacterial Strains, Plasmids And Phage:	50
2.3.2	Storage Of Bacteria:	51
2.3.3	Growth Media:	52
	a) Rich Media:	52
	b) Minimal Media:	53
	c) Specific Supplements To Defined Media:	53
2.3.4	Antibiotics:	53
2.3.5	Growth Of Bacteria:	54
2.3.6	Measurement Of Growth:	54
2.3.7	Harvesting Of Bacteria:	54
2.3.8	Disruption Of Bacteria:	54
	a) Ultrasonic Disruption:	54
	b) French Pressure Cell Disruption:	54

2.4	**General Recombinant DNA Techniques:**	55
2.4.1	Large Scale Plasmid Purification:	55
	a) Using CsCl Gradients:	55
	b) Using QIAGEN-Tips:	56
2.4.2	Small Scale Plasmid Purification:	57
2.4.3	Digestion Of DNA With Restriction Enzymes:	57
2.4.5	Filling Recessed Termini Using Klenow Polymerase:	57
2.4.5	Agarose Gel Electrophoresis:	58
2.4.6	Recovery Of DNA From Agarose Gels:	58
2.4.7	Phenol/Chloroform Extraction of Nucleic Acids:	59
2.4.8	Extraction Of DNA With Butanol:	59
2.4.9	Ethanol Precipitation:	59
2.4.10	Ligations:	60
2.4.11	Transformation Of *E. coli* With DNA:	60
	a) Preparation Of Competent Cells:	60
	b) Transformation Of Competent Cells:	61
2.5	**Cloning In M13:**	61
2.5.1	Maintenance Of Host Cells:	61
2.5.2	Preparation Of Phage Stock:	62
2.5.3	Preparation Of RF M13 DNA:	62
	a) Small Scale:	62
	b)Large Scale:	63
2.5.4	Preparation Of Single Stranded DNA:	63
2.6	**Generation Of Nested Deletions Using Exonuclease III:**	63
2.6.1	Digestion With PstI And BamHI:	63
2.6.2	Digestion With Exonuclease III:	64
2.6.3	Removal Of Single Stranded Overhangs:	64

2.6.4	Ligation And Transformation:	64
2.6.5	Analysis Of Deletion Clones:	64
2.7	**Sequencing By The Chain Termination Method:**	65
2.7.1	Sequencing With The Klenow Fragment Of *E. coli* DNA Polymerase I:	65
	a) Annealing Primer To Template:	65
	b) Sequencing Reactions:	66
	c) Composition Of Reaction Mixes:	66
2.7.2	Sequencing With Sequenase Version 2.0:	67
	a) Annealing Primer To Template:	67
	b) Sequencing Reactions:	67
2.7.3	Polyacrylamide Gel Electrophoresis:	68
2.8	**Computer Programs Used In The Analysis Of DNA And Amino Acid Sequences:**	69
2.9	**Site Directed Mutagenesis:**	71
2.9.1	Preparation Of Oligonucleotides:	71
2.9.2	Phosphorylation Of Oligonucleotides:	71
2.9.3	Annealing Oligonucleotides To Single Stranded Templates:	71
2.9.4	Synthesis And Ligation Of Mutant DNA Strand:	72
2.9.5	Digestion Of The Non-Mutant Strand:	72
2.9.6	Repolymerisation And Ligation Of The Gapped DNA:	72
2.10	**Polyacrylamide Gel Electrophoresis Of Proteins:**	73
2.10.1	Denaturing Electrophoresis:	73
2.10.2	Non-denaturing Electrophoresis:	73
2.10.3	Staining Gels For Protein:	75
2.10.4	M_r Determination By SDS-PAGE:	75

2.11 Purification Of ICL From *E. coli* KAT-1/pAR9023, KAT-1/pAR654 And KAT-1/pAR655: 76

2.11.1 Purification Of Wild Type ICL From *E. coli* KAT-1/pAR9023: 76

a) Purification Buffers: 76

b) Preparation Of Crude Extract: 76

c) Ammonium Sulphate Fractionation: 77

d) Gel Filtration On Sephacryl S-300 Superfine: 77

e) Phenyl-Sepharose Chromotography: 77

f) FPLC Ion-exchange Chromotography On Mono Q: 77

g) Storage: 77

2.11.2 Purification Of ICL, ICLSer195 And ICLAla195 From *E. coli* KAT-1/pAR9023, KAT-1/pAR654 And KAT-1/pAR655 Respectively: 78

a) Purification Buffers: 78

b) Preparation Of Crude Extract: 78

c) Ammonium Sulphate Fractionation: 78

e) FPLC Ion-exchange Chromotography On Mono Q: 78

g) Storage: 78

2.12 Isocitrate Lyase Assays: 79

a) Method 1: 79

b) Method 2: 79

2.13 Reaction Of ICL, ICLSer195 and ICLAla195 With Iodoacetate: 79

2.14 Fluorimetry: 80

2.15 Circular Dichroism: 80

Chapter 3: Cloning And Sequencing Of *aceA*

3.1	**Introduction:**	81
3.2	**Cloning Of The *aceA* Gene From *E. coli* ML308:**	82
3.2.1	Subcloning *aceA* From pEM9:	82
3.2.2	Expression Of ICL From *E. coli* KAT-1/pAR9023:	85
3.3	**Sequence Of *aceA*:**	89
3.3.1	Sequencing Strategy:	89
3.3.2	Determination Of Rate Of Digestion By Exonuclease III:	94
3.3.3	Generation Of A Set Of Nested Deletions In Both Strands To Sequence *aceA*:	94
3.3.4	Nucleotide Sequence Of The Insert Encoding The *aceA* Gene:	97
3.4	**Sequence Analysis:**	103
3.4.1	Translation:	103
3.4.2	Comparison Of The *E. coli* ICL Sequences With Each Other:	107
3.4.3	Comparison Of The Amino Acid Sequences Of *E. coli* ICL With ICL From Other Organisms:	110
3.5	**A Consensus Isocitrate Binding Site Between ICL And ICDH:**	114
3.5.1	Comparison Of ICL And ICDH From *E. coli*:	114
3.5.2	Secondary Structure Analysis Of ICL And ICDH:	115
3.5.3	Comparison Of The Predicted ICL Secondary Structure With The Actual Structure Of ICDH:	117
3.6	**Discussion:**	120

Chapter 4: Site Directed Mutagenesis Of ICL

4.1 Introduction: 126

4.2 Site Directed Mutagenesis Of ICL: 128

4.2.1 Strategy: 128

4.2.2 Design Of Oligonucleotides For The Mutagenesis Reaction: 128

4.2.3 The Mutagenesis Reaction: 130

4.2.4 Nucleotide Sequence Of Mutants: 133

4.3 Expression Of Mutant *aceA* Gene Produts, ICLSer195 And ICLAla195: 138

4.3.1 Cloning Of The Inserts Of pAR24-654 And pAR24-655 Into pGLW11: 138

4.3.2 Expression Of ICLSer195 And ICLAla195 From pAR654 And pAR655 Respectively: 139

4.4 Purification Of Mutant Enzymes, ICLSer195 And ICLAla195: 143

4.4.1 Introduction: 143

4.4.2 Purification Of w.t. ICL From *E. coli* KAT-1/pAR9023: 143

4.4.3 Purification Of ICLSer195 and ICLAla195: 144

4.5 Characterisation Of ICLSer195 And ICLAla195: 151

4.5.1 Conformation Of ICLSer195, ICLAla195 And Wild Type ICL: 151

4.5.2 Sensitivity To Inhibition By Iodoacetate : 156

4.5.3 Kinetic Properties Of ICLSer195, ICLAla195 And Wild Type ICL: 158

4.5.4 Effect Of pH On ICLSer195, ICLAla195 And Wild Type ICL: 163

4.6 Discussion: 170

Chapter 5: General Discussion And Future Prospects

5.1	**Introduction:**	175
5.2	**Comparison Of ICL From Different Sources:**	176
5.3	**The Active Site Of ICL:**	179
5.4	**Importance Of Sequence Downstream Of *aceK*:**	183
5.5	**Future Work:**	184

References:

List Of Figures:

Chapter 1:

1.1	The Glyoxylate Bypass And The TCA Cycle	3
1.2	Reactions Catalysed By Isocitrate Lyase And Malate Synthase-A	5
1.3	Control Of ICDH Kinase/Phosphatase	14
1.4	Possible Secondary Structure In The *aceA-aceK* Intercistronic Region	25
1.5	Model For Binding Of Magnesium By Isocitrate Lyase	30
1.6	Postulated Mechanism Of Isocitrate Lyase From *E. coli* ML308	39

Chapter 3:

3.1	Comparison Of Restriction Maps Of Clones Encoding The Complete *ace* Operon	83
3.2	Restriction Maps Of pAR9023 And pAR9024	86
3.3	Effect Of IPTG Concentration On Growth And Expression Of ICL Activity From *E. coli* KAT-1/pAR9023	87
3.4	Effect Of IPTG Concentration On The Rate Of Increase In Specific Activity Of ICL In *E. coli* KAT-1/pAR9023	88
3.5	Overexpression Of ICL From *E. coli* KAT-1/pAR9023	90
3.6	Strategy Used To Create A Series Of Nested Deletions	93
3.7	Determination Of Rate Of Digestion By Exonuclease III	95
3.8	Restriction Maps Of pAR23 And pAR24	96
3.9	Exonuclease III Deletion Of pAR23 and pAR24	98
3.10	Examples Of Exo III Deletion Of pAR23 and pAR24	99
3.11	The Sequencing Scheme Of *aceA*	101
3.12	Nucleotide Sequence Of The Insert Of pAR24	102
3.13	CODONPREFERENCE Table For Nucleotides 1-1642 Of pAR24	104

3.14	Nucleotide Sequence And Derived Amino Acid Sequence Of The Insert Of pAR24	106
3.15	Comparison Of The *E. coli* ICL Derived Amino Acid Sequences	108
3.16	Comparison Of The Nucleotide Sequences Of The *aceA* Genes From *E. coli*	109
3.17	LINEUP Of ICL Sequences	111
3.18	Comparison Of The Derived Amino Acid Sequence And Secondary Structure Prediction Of ICL With ICDH	116
3.19	Three Dimensional Structure Of ICDH	118
3.20	Comparison Of The Derived Amino Acid Sequence And Secondary Structure Prediction Of ICL With The Actual Structure Of ICDH	119

Chapter 4:

4.1	Site Directed Mutagenesis Scheme	129
4.2	Oligonucleotide Sequences	131
4.3	Analysis Of Mutagenesis	132
4.4	Sequencing Of pAR654 and pAR655	136
4.5	Sequencing Strategy For Mutant *aceA* Genes	137
4.6	Strategy For Constructing Expression Cassettes For Mutant *aceA* Genes	140
4.7	Expression Of ICLSer195 and ICLAla195 From *E. coli* KAT-1/pAR654 and KAT-1/pAR655 Respectively	141
4.8	Expression Of Mutant ICL Enzymes	142
4.9	Elution Profile Of Wild Type ICL From Mono Q	145
4.10	Elution Profile Of ICLSer195 From Mono Q	147
4.11	Elution Profile Of ICLAla195 From Mono Q	148
4.12	Purification Gel Of ICL, ICLSer195 and ICLAla195	149

4.13	Non-Denaturing Polyacrylamide Gel Electrophoresis Of ICL, ICLSer195 and ICLAla195	150
4.14	Circular Dichroism Of w.t.ICL, ICLSer195 And ICLAla195, Far U.V. Spectrum	153
4.15	Circular Dichroism Of w.t.ICL, ICLSer195 And ICLAla195, Near U.V. Spectrum	154
4.16	Fluorimetric Analysis Of w.t.ICL, ICLSer195 and ICLAla195	155
4.17	Inhibition Of w.t.ICL, ICLSer195 and ICLAla195 With Iodoacetate	157
4.18	Isocitrate Protection of w.t.ICL, ICLSer195 and ICLAla195 Against Inactivation By Iodoacetate	159
4.19	Mg^{2+} Protection of w.t.ICL, ICLSer195 and ICLAla195 Against Inactivation By Iodoacetate	160
4.20	Mg^{2+}-Isocitrate Protection of w.t.ICL, ICLSer195 and ICLAla195 Against Inactivation By Iodoacetate	161
4.21	Effect of pH On log K_m Of w.t.ICL, ICLSer195 and ICLAla195	165
4.22	Effect of pH On log $^1/_{K_m}$ Of w.t.ICL, ICLSer195 and ICLAla195	166
4.23	Effect of pH On log k_{cat} Of w.t.ICL, ICLSer195 and ICLAla195	167
4.24	Effect of pH On log $^{k_{cat}}/_{K_m}$ Of w.t.ICL, ICLSer195 and ICLAla195	168
4.25	Postulated Roles Of Cys195 In The Catalytic Mechanism Of ICL	173

Chapter 5:

5.1	Binding Of Mg^{2+} Isocitrate In The Active Site Of ICDH	182

List Of Tables:

Chapter 1:

1.1	Properties Of ICL And ICDH	7
1.2	Amino Acid Compositions Of ICL From Various Organisms	28
1.3	K_m Values Of ICL From Various Organisms	32
1.4	Kinetic Properties Of ICL From *E. coli* ML308	38

Chapter 2:

2.1	Preparation Of Polyacrylamide Gels	74

Chapter 4:

4.1	Results Of Site Directed Mutagenesis Reactions	134
4.2	Oligonucleotide Primers For Sequencing *aceA* Mutants	135
4.3	Purification Of ICL, ICLSer195 And ICLAla195	146
4.4	Kinetic Properties Of ICLSer195, ICLAla195 And Wild Type ICL	162

ABBREVIATIONS

All abbreviations used in this thesis are those recommended by the Biochemical Society, London, except for those listed below:

Amp	Ampicillin
3-BrP	3-Bromopyruvate
BSA	Bovine serum albumin
CD	Circular dichroism
DEPC	Diethylpyrocarbonate
DNase I	Deoxyribonuclease I
DTE	1,4-dithioerythritol
DTT	Dithiothreitol
Exo III	Exonuclease III
FPLC	Fast Protein Liquid Chromatography
HPLC	High Pressure Liquid Chromatography
ICDH	Isocitrate dehydrogenase
ICL	Isocitrate Lyase
IPTG	Isopropyl-β-thiogalactoside
Mops	Morpholino propane sulphonic acid
MS-A	Malate synthase-A
MSB	Malate synthase-B
NADP	Nicotinamide-adenine dinucleotide phosphate
OAA	Oxaloacetate
PEP	Phosphoenolpyruvate
3-PG	3-phosphoglycerate
PMSF	Phenylmethylsulphonyl fluoride
PTPase	Protein tyrosine phosphatase
REP	Repetitive extragenic palindromic

RNase	Ribonuclease A
SDS	Sodium dodecyl sulphate
Taps	N-tris[Hydroxymethyl]methyl-3-aminopropane sulphonic acid
TEMED	NNN'N'-Tetramethylethylenediamine
Tet	Tetracycline
Tris	Tris (hydroxymethyl) aminomethane
X-gal	5-Bromo-4-chloro-3-indoyl-β-galactoside

AKNOWLEDGEMENTS

I would like to thank my supervisor Dr. H.G. Nimmo for his constant help and encouragement throught this project. I would also like to thank the late Professor R.M.S. Smellie and Professor M.D. Houslay for making available the facilities of the Biochemistry Department and the Science and Engineering Research Council for a research studentship.

I would like to thank Professor J.R. Coggins, Dr G.A. Nimmo, Dr L. Graham for their encouragement, support and advice.

Thanks to all those in the D-floor dungeon; Mr. J. Greene for keeping D-floor in good running order, Mrs. Margaret Fairbairn for keeping us supplied with clean glassware, Miss Fiona Douglas, fellow *E. coli* grower and all other D-floor and B4a members.

Thanks too to the "Lads" from 12 Kelvingrove St., Gordy, Bones, Greg and all the weekend visitors.

I am grateful to Dr. F. Rinaldi for proof-reading this thesis and for the workouts with Jools Galbraith.

Thanks are also due to Professor C.A. Fewson and the members of C24 and C26 for their encouragement over the last year.

Finally, I wish to thank Christine Harris for her support and encouragement, and for putting up with my writing-up blues.

SUMMARY

The glyoxylate bypass allows *E. coli* to generate precursors for biosynthesis during growth on acetate as the sole carbon source. During growth on acetate, there is competition between isocitrate dehydrogenase (ICDH) and isocitrate lyase (ICL) for the common substrate, isocitrate. Previous studies have shown that ICDH is regulated by reversible phosphorylation during growth on acetate. Although much is known about ICDH, comparatively little is known about the structure or catalytic mechanism and regulation of ICL.

Comparison of the restriction map of pEM9, a clone containing the *ace* operon from *E. coli* ML308 with the restriction/functional map of pCL1000, a clone encoding the *ace* operon of *E. coli* K12, indicated that the gene encoding ICL, *aceA*, was likely to be found within a 2.3kb *Bam*HI-*Ava*I fragment. This 2.3kb fragment was subcloned from pEM9 into the expresion vector pGLW11 to give pAR9023. *E. coli* KAT-1, contains a transposon Tn10 insertion in the gene encoding *aceA*. After transformation of this strain with pAR9023 ICL activity and protein were overexpressed after addition of isopropyl-β-thiogalactoside (IPTG) to the growth medium.

The 2.3kb insert of pAR9023 was then subcloned into M13mp8 and a series of nested deletions were generated using a combination of exonuclease III and S1 nuclease. Using this method, the insert was sequenced using the universal primer or the -40 primer, without further subcloning or synthesis of other oligonucleotide primers.

During the course of this work, the sequence of *aceA* from *E. coli* K12 was determined by two independent groups. Comparison of the *E. coli* ML308 sequence with the K12 sequences revealed several differences which are probably due to sequencing errors or spontaneous mutations in the other sequences rather than strain variation. In support of my sequence, a third group sequenced *aceA* from *E. coli* and found it was identical to that from *E. coli* ML308, although the authors did not indicate the strain from which the *aceA* gene was cloned. Comparison of the derived amino acid sequence of ICL from *E. coli* ML308 with that of three plant ICLs and one yeast ICL revealed a very high degree of sequence identity.

Comparison of the derived amino acid sequence of ICL from *E. coli* ML308 with

that of ICDH, revealed a region of possible sequence and secondary structure homology. In addition, the three dimensional structure of ICDH is known and several residues which bind isocitrate in ICDH are conserved in ICL.

During the course of this project, an active site Cys residue was identified as Cys195. Site directed mutagenesis was employed to replace this residue with either an Ala or a Ser. The mutated genes were sequenced to ensure that no other residues were replaced and then subcloned into pGLW11. The mutant ICLs, ICLSer195 and ICLAla195, were then overexpressed from transformed into KAT-1 strains and purified.

The conformation of both mutants were compared to that of the wild type enzyme by both near U.V. and far U.V. circular dichroism and fluorimetry. These studies indicated that all three ICLs had a very similar conformation and so replacment of Cys195 with either a Ser or an Ala did not result in any great change in the conformation of the enzyme.

Both ICLSer195 and ICLAla195 showed a 3.6-fold and a 8.4-fold increase K_m for isocitrate respectively and a 100-fold and 30-fold decrease in k_{cat} values respectively, indicating that Cys195 is an important residue. pH studies of the mutant enzymes indicated that Cys195 probably has a pK_a value of around 7.1, which is one pH unit below that usually found for sulphydryl groups. These studies also revealed that K_m and k_{cat} are dependent on ionisation of another group with a pK_a of 7.8. The identity of this group is unknown. Based on these results, two possible roles for Cys195 are postulated.

Chapter 1
Introduction

1.1 The Glyoxylate Bypass:

1.1.1 Introduction:

Cells contain a number of central metabolic pathways into which all carbon sources are directed and out of which all cell constituents are derived. The central metabolic pathways not only produce the precursors for biosynthesis but also generate the energy in the form of ATP necessary for biosynthesis and all necessary reducing power. This is achieved in aerobic organisms by complete combustion of the nutrients to CO_2 and H_2O.

The Tricarboxylic Acid Cycle, TCA cycle or Krebs cycle serves as the central pathway in most organisms. This pathway condenses a C_2 unit (acetyl-CoA) with an oxaloacetate (OAA) molecule to form citrate. This then undergoes a series of reactions in which coenzymes are reduced and two carbon atoms are lost as CO_2 and OAA is regenerated. The reduced coenzymes, NADH, NADPH and $FADH_2$, can then be used in either reductive biosynthesis or to regenerate ATP by oxidative phosphorylation.

Use of intermediates for biosynthesis would prevent regeneration of OAA and thus cause breakdown of the cycle. Hence there is an absolute requirement for regeneration of TCA cycle intermediates. Thus there are anaplerotic pathways which operate to maintain the levels of intermediates of the central pathways (Kornberg, 1966).

1.1.2 Growth Of *Escherichia coli* On Fatty Acids And Acetate:

E. coli, like many other organisms can grow on acetate or fatty acids as the sole carbon source. Both provide C_2 units, in the form of acetyl-CoA, for entry into the TCA cycle.

When fatty acids enter the cell, they are either incorporated into more complex lipids or catabolised to acetyl-CoA.The pathway by which *E. coli* degrades fatty acids is very similar to the β-oxidative pathways of mammals and other eukaryotic organisms. The first step in the degradation of fatty acids is their activation to fatty-acyl-CoA, catalysed by an acyl-CoA synthetase. The fatty-acyl-CoA then undergoes two oxidation steps, yielding one molecule of $FADH_2$ and one molecule of NADH, followed by thiolytic cleavage to give a fatty-acyl-CoA, which is shorter than the original by two carbon atoms, and acetyl-CoA. The shortened fatty-

acyl-CoA then re-enters the cycle thus yielding several molecules of acetyl-CoA for entry into the TCA cycle (Overath *et al.*, 1969).

Acetate can be activated to acetyl-CoA by two mechanisms. One mechanism involves the acetylation of CoA and cleavage of ATP to AMP and inorganic pyrophosphate by acetyl-CoA synthetase (acetate: CoA ligase [AMP-forming]; EC 6.2.1.1.) (Chou and Lipmann, 1952; Berg, 1956). The other mechanism uses two enzymes. Acetate kinase (ATP acetate phosphotransferase; EC 2.7.2.1.) converts acetate to acetyl phosphate, with cleavage of ATP to ADP, then phosphotransacetylase (acetyl-CoA: orthophosphate acyltransferase; EC .3.1.8.) transfers the acetyl moiety from acetyl phosphate to CoA liberating inorganic phosphate (Kaplan and Lipmann, 1948; Stadtman and Barker, 1950, Levine *et al.*, 1980). It has been suggested from studies with *ack* (acetate kinase) and *pta* (phosphotransacetylase) mutants that the acetate kinase and phosphotransacetylase are required for *E. coli* to grow optimally on acetate as the sole carbon source (Brown *et al.*, 1977). These authors also suggest that acetyl-CoA synthetase is used by these mutants as an alternative route for the conversion of acetate to acetyl-CoA.

1.1.3 The Glyoxylate Bypass:

Operation of the TCA cycle results in the loss of both carbon atoms of the acetyl unit as CO_2, and by itself would allow no net accumulation of carbon. However, *E. coli* is able to grow on acetate and fatty acids as the sole source of carbon. This is achieved by diverting flux from the two CO_2 evolving steps of the TCA cycle through the glyoxylate bypass (figure 1.1). In *E. coli* the two unique enzymes of the glyoxylate bypass, isocitrate lyase (ICL) and malate synthase A (MSA) are induced during growth on acetate or fatty acids as the only source of carbon (Kornberg, 1966; Vanderwinkel and DeVliegher, 1968). (There is a second malate synthase, MSB, which is induced during growth on glycolate and compounds which are metabolised to glyoxylate as the sole carbon source (Falmange *et al.*, 1965; Vanderwinkel *et al.*, 1963)).

The first enzyme of the glyoxylate bypass, ICL [Ds-isocitrate glyoxylate-lyase E.C. 4.1.3.1.], was first discovered in sonic extracts of *Pseudomonas aeroginosa* which catalysed the reversible formation of succinate and glyoxylate on addition of either citrate or

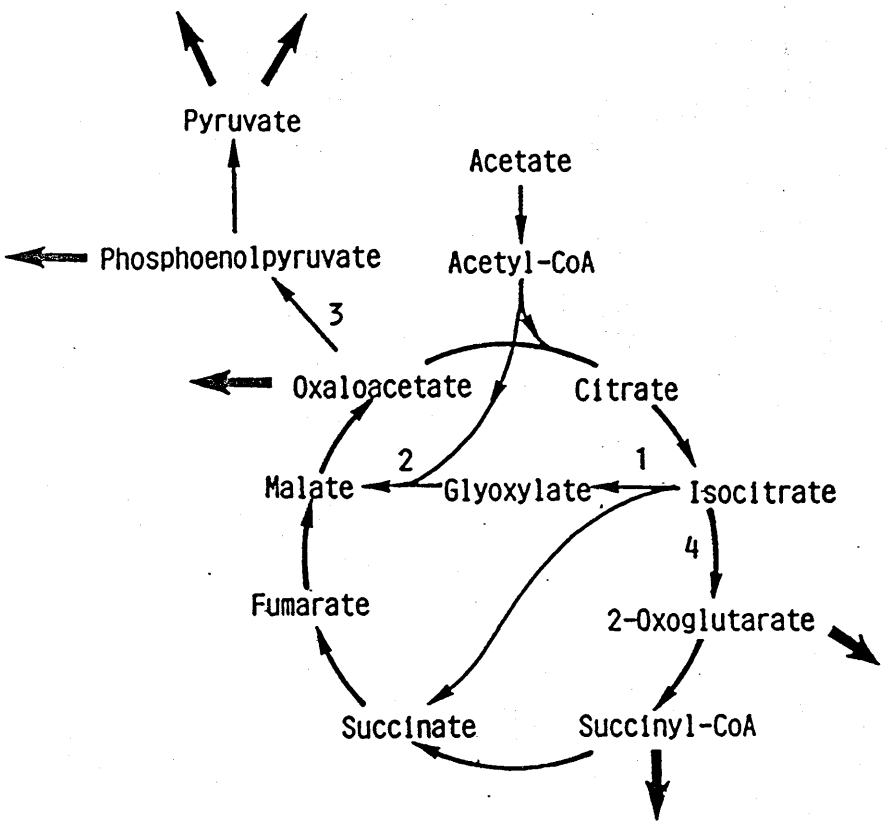


Figure 1.1: The Glyoxylate Bypass And The TCA Cycle

The reactions numbered 1-4 are catalysed by the enzymes isocitrate lyase, malate synthase-A, phosphoenolpyruvate carboxykinase, and isocitrate dehydrogenase respectively. The heavy arrows indicate fluxes to biosynthesis.

cis-aconitate (Campbell *et al.*, 1953). This result was quickly confirmed using fungi and several bacteria and it was established that Ds-(+)-isocitrate was the true substrate (Olsen, 1954; Saz, 1954; Smith and Gunsalus, 1954). (Figure 1.2a)

The second glyoxylate bypass enzyme, MSA [E.C. 4.1.3.2] catalyses the condensation of glyoxylate and a second molecule of acetyl CoA to form malate (Wong and Ajl, 1956) with the equilibrium strongly in favour of malate (Figure 1.2b).

It was later shown, using labelled acetate, that ICL and MSA act together to give the following overall reaction (Kornberg, 1958; Kornberg and Madsen, 1958; Kornberg and Quayle, 1958; Kornberg *et al.*, 1960):

Acetyl CoA + Isocitrate + H_2O = Succinate + Malate + CoA

Thus there is a net gain of one C_4-dicarboxylic acid, a TCA cycle intermediate, which can be converted to oxaloacetate (OAA). This can then be used to supply all the precursors required for biosynthesis since all the TCA cycle intermediates are mutually interconvertible and phosphoenolpyruvate carboxykinase can convert OAA to phophenolpyruvate (PEP) which can then be used to generate all the phosphorylated biosynthetic precursors.

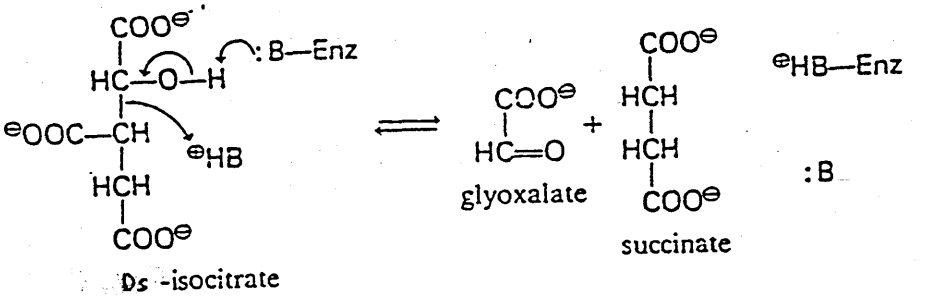
Since flux through ICDH is essential to produce 2-oxoglutarate and succinyl-CoA, as well as some of the energy and reducing power necessary to utilize the precursors for growth, there has to be flux through ICDH. ICDH and ICL compete for the common substrate, isocitrate, and one might therefore expect very tight regulation of the ICL and ICDH activities in order to ensure efficient use of the available acetate.

1.2 Control Of The Branchpoint Between ICL and ICDH In *E. coli*:

1.2.1 Reversible Inactivation Of ICDH:

Since there is no subcellular compartmentation in *E. coli*, ICL and ICDH compete directly with each other for the available isocitrate during growth on acetate or fatty acids as the sole carbon source. This raises the question of how fluxes at this branchpoint is regulated.

(A)



(B)

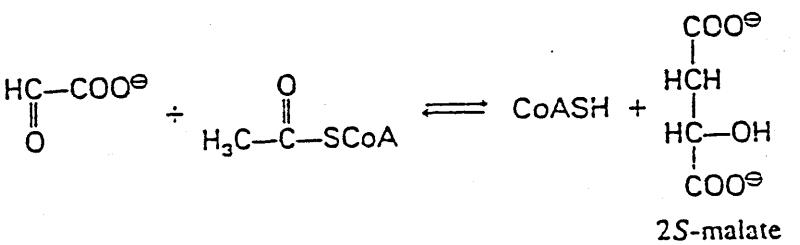


Figure 1.2: Reactions Catalysed By Isocitrate Lyase And Malate Synthase-A

(A) Isocitrate lyase catalyses the reversible aldol cleavage of isocitrate to give succinate and glyoxylate.

(B) Malate synthase-A catalyses the condensation of glyoxylate with acetyl-CoA to form malate.

There is no evidence to suggest that ICL from *E. coli* is allosterically controlled. Also, unlike the NAD-linked ICDH of eukaryotes, ICDH from *E. coli*, and most other bacteria, is NADP-linked and does not appear to be under the control of any allosteric inhibitors or activators (Ragland *et al.*, 1966). ICDH is however subject to concerted inhibition by OAA and glyoxylate, both of which are generated by the glyoxylate bypass. This inhibition is complex, partly caused by the condensation of glyoxylate and OAA to yield the unstable compound, oxalomalate, which is a very potent competitive inhibitor of ICDH and partly caused by the slow binding of OAA and glyoxylate to the isocitrate-binding site of ICDH (Nimmo, 1986). However, several groups have reported the condensation of OAA and glyoxylate is too slow to account for the inhibition of ICDH by oxalomalate (Johanson and Reeves, 1977; Marr and Weber, 1969; Shiio and Ozaki, 1968). Also, oxalomalate was found to decompose spontaneously to form 4-hydroxy-2-oxoglutarate, which is a much less potent inhibitor of ICDH (Nimmo, 1986). The second mechanism of inactivation of ICDH by OAA and glyoxylate is not thought to be physiologically significant because the intracellular concentrations of these metabolites are probably too low (Nimmo, 1986). Moreover, one might expect glyoxylate and OAA to activate ICDH and not inhibit it since these metabolites can be considered as end products of flux through the glyoxylate bypass during growth on acetate or fatty acids as the sole carbon source (see section 1.2.4).

Consideration of the kinetic properties of the two competing enzymes in isolation does not indicate how this branchpoint is regulated (Table 1.1). The K_m of ICL for isocitrate is 63μM in the absence of competing anions and is probably much higher *in vivo*, at least 260μM due to the presence of competing anions and inhibitors (see section 1.4.7) (MacKintosh and Nimmo, 1988) and the K_m of ICDH for the same substrate is around 5μM (Nimmo *et al.*, 1987). The intracellular concentration of isocitrate is approximately 4μM during growth on glucose (LaPorte *et al.*, 1984). Since the specific activities of both ICDH and ICL from cells grown on acetate are similar (0.9u.mg^{-1} for ICDH and 0.8u.mg^{-1} for ICL), if this were also true for acetate grown cells, flux through ICDH will be favoured, the acetate carbon lost as CO_2, no biosynthetic precursors will be accumulated and growth will cease. In order to allow flux through ICL, and hence the glyoxylate bypass, there must be a means of

Table 1.1: Properties Of ICL And ICDH

	ICDH	ICL
Specific activity in acetate grown cells	0.9u/mg	0.8u/mg
K_m for D_s-isocitrate in the absence of inhibitors	5μM	63μM
Significant inhibitors	none	3-phosphoglycerate (K_i=0.8mM: intracellular conc.=2.5mM). Sulphate (K_i=20mM: intracellular conc. unknown).
Estimated K_m for D_s-isocitrate *in vivo*	5μM	0.57mM

Taken from Nimmo *et al.* (1987). The K_m for D_s-isocitrate was measured at a $NADP^+$ concentration of 40μM. The intracellular concentrations of isocitrate in cells growing on glycerol and acetate are 20μM and 570μM respectively.

ICDH: Isocitrate Dehydrogenase
ICL: Isocitrate Lyase

either or both inhibiting ICDH or activating ICL *in vivo*.

During studies of the activities of different TCA cycle enzymes malate dehydrogenase, 2-oxoglutarate dehydrogenase and ICDH in *E. coli* growing on different carbon sources, the control of this branchpoint became clearer. Holms and Bennett (1971) found that after growth had stopped on limiting glycerol, the specific activities of all three of the above enzymes remained constant over a four hour period. However, after growth on limiting glucose, both malate dehydrogenase and 2-oxoglutarate dehydrogenase but not ICDH activity remained constant over the same period. The ICDH activity dropped to about 20% of its original level over 2 hours and then rose again to 75% at the end of the 4 hour period. This is explained by the fact that during growth on glucose but not glycerol, *E. coli* excretes acetate into the medium (Britten, 1954). Once the glucose is exhausted, growth stops and the enzymes of the glyoxylate bypass are induced and acetate is metabolised. During the period of adaption to acetate and its use as a carbon source, the ICDH activity declined but rose again after exhaustion of the acetate. Addition of chloramphenicol, which inhibits protein synthesis, at the end of growth on limiting glucose prevented ICL activity increasing and slowed but did not stop utilization of acetate. The activity of ICDH did not decrease under these conditions in the 4 hour period after cessation of growth. The specific activity of ICDH during growth of *E. coli* on acetate was approximately 30% of that during growth on glucose or glycerol (Bennett and Holms, 1975). Addition of 1mM pyruvate to cells growing on acetate resulted in a 4-fold activation of ICDH which persisted until the pyruvate was exhausted when the ICDH activity dropped to the level during growth on acetate. Similar results were obtained after addition of a number of other metabolites which make the glyoxylate bypass redundant. Thus, loss of ICDH activity appears to be peculiar to cells growing on acetate or fatty acids as the sole carbon source.

Bennett and Holms further investigated the inactivation of ICDH by trying to determine the mechanism of inactivation. Addition of chloramphenicol at the same time as pyruvate to *E. coli* growing on acetate had no effect on the change in activity of ICDH, indicating that protein synthesis is not involved. Degradation of an inhibitor protein could also be ruled out since ICDH activity returns to the level during growth on acetate after the pyruvate is exhausted. Experiments involving dialysis and mixing extracts from acetate and glucose grown cultures

indicated that changes in the intracellular concentration of a low molecular weight effector(s) was not responsible. Bennett and Holms concluded that the reversible inactivation may be the result of reversible covalent modification or reversible association with another macromolecule, and that the role of inactivation was to restrict flux through the TCA cycle and so permit isocitrate to be metabolised through the glyoxylate bypass (Holms and Bennett, 1971; Bennett and Holms, 1975).

The first evidence that ICDH from *E. coli* could be phosphorylated came from Garnak and Reeves (1979a,b). These workers added ^{32}P and acetate (5min later) to *E. coli* cultures at the beginning of stationary phase in a low phosphate medium with limiting glucose. From these cells, they found a 51,000 dalton radiolabelled protein which co-purified with ICDH. This protein was immunologically identical to ICDH and the radioactive band corresponded to an ICDH activity stain in native gels, suggesting that this protein was in fact ^{32}P-ICDH. Ion exchange chromatography followed by partial acid hydrolysis identified the phosphorylated residue as a serine. Borthwick *et al.* (1984a) purified active ICDH from glycerol grown cells, inactive and partially active ICDH from acetate grown cells. The active ICDH bound to an immobilised dye column, Procion red-Sepharose, while the inactive ICDH did not bind to this column. All three proteins had the same mobility on SDS gels corresponding to a subunit of 45,000 daltons. The amino acid composition of all three proteins was essentially the same as was the chymotryptic digestion pattern on SDS gels and the tryptic digestion pattern on 2-dimensional gels. Non-denaturing gel electrophoresis did however reveal differences in the three proteins. The partially active ICDH showed two bands after protein staining, of which the band with higher mobility did not appear on gels stained for ICDH activity. The inactive ICDH appeared as one band after protein staining with the same mobility as the higher mobility band of the partially active ICDH. Using non-denaturing gel electrophoresis in different acrylamide concentrations they showed that the difference in mobility was due to a charge difference and not a difference in molecular weight. The alkali-labile phosphate was also measured. This showed that active and inactive ICDH correspond to dephosphorylated and phosphorylated ICDH respectively differing in one phosphate per subunit .The ICDH purified by Garnak and Reeves (1979a,b), which contained ^{32}P but was active, was probably a monophosphorylated dimer.

1.2.2 Phosphorylation Of ICDH *in vivo*:

When studying phosphorylation systems, it is necessary to establish the occurrence of the phosphorylation *in vivo*.This was done for this system in *E. coli* ML308 by Borthwick *et al.* (1984b) When pyruvate was added to cells growing on acetate, the ICDH activity rose 4-fold over 20-30min, then declined to its basal level by 50min. Cells were labelled with $^{32}P_i$ during this experiment and denatured cell pellets were run on SDS gels. The ^{32}P-containing ICDH band was easily identified. On addition of the pyruvate, there was a loss of ^{32}P from this band followed by a rise after 40-60min. This effect was not due to changes in the amount of ICDH protein. Thus the result shows an inverse correlation between the phosphorylation state and the activity of ICDH in intact cells (Borthwick *et al.*, 1984b). Cell extracts from the same experiment were incubated with purified ICDH kinase/phosphatase (section 1.2.3) under conditions in which ICDH is dephosphorylated. This caused the ICDH activity to rise to the value seen after addition of pyruvate. It was therefore concluded that the reversible inactivation of ICDH *in vivo* is solely due to reversible phosphorylation. Borthwick *et al.* (1984b) also isolated and sequenced phosphorylated chymotryptic peptides from ICDH labelled *in vivo* and ICDH labelled *in vitro* by purified ICDH kinase/phosphatase. Both labelled proteins gave identical digestion patterns with one peptide containing over 80% of the recovered radioactivity. The sequences of these peptides were identical, showing that the serine residue phosphorylated *in vivo* is the same as the one phosphorylated *in vitro*.

It is believed that the role of phosphorylation and inactivation of ICDH is to render this enzyme rate limiting in the TCA cycle and thus allow the intracellular concentration of isocitrate to increase to a level high enough to sustain flux through ICL. In agreement with this hypothesis, El-Mansi *et al.* (1985) found that the intracellular concentration of isocitrate was much higher in cells growing on acetate (570μM) than on cells growing on glycerol or glucose as the sole carbon source (Table1.1). Hence this explains how phosphorylation of ICDH permits flux through the glyoxylate bypass. These authors also found that addition of pyruvate to cells growing on acetate, known to result in reactivation and dephosphorylation of ICDH, caused a transient two-fold increase in intracellular isocitrate concentration.

1.2.3. ICDH Kinase/Phosphatase:

LaPorte and Koshland (1982) devised a purification procedure for ICDH kinase, the enzyme that phosphorylates ICDH. Surprisingly, they also found that ICDH phosphatase could be purified in the same way and that the ratios of kinase:phosphatase remained constant throught the purification. The kinase and phosphatase activities eluted superimposably from ion exchange, gel filtration and affinity chromatography columns. Fractions from the affinity column (immobilised ICDH) which contained both kinase and phosphatase activities gave in a single band of 66,000 dalton on SDS gels indicating that both activities are probably associated with the same polypeptide. Similar results were obtained by Nimmo *et al.* (1984) who reported the native molecular weight to be 135,000 dalton suggesting that ICDH kinase/phosphatase is a dimer of identical subunits, M_r 66,000. Definite proof that both kinase and phosphatase activities are associated with the same polypeptide came from the isolation of a clone which restored the ability of an *aceK* mutant to grow on acetate (LaPorte and Chung, 1985). This clone expressed a 66,000 dalton protein with both ICDH kinase and ICDH phosphatase activities. Deletion mapping of the clone revealed a 1,800bp coding region, the length required to encode a 66,000 dalton polypeptide. These results were confirmed by determination of the nucleotide sequence (Cortay *et al.*, 1988; Klumpp *et al.*,1988). The deduced amino acid sequence revealed none of the consensus motifs characteristic of other protein kinases (Reed *et al.*,1985; Celenza and Carlson, 1986), with the exception of a concensus ATP binding site.

ICDH kinase transfers the γ-phosphate group of ATP to a serine residue of ICDH with a stoichiometry of one phosphate per subunit resulting in complete inactivation of ICDH. ICDH phosphatase catalyses the release of inorganic phosphate from phospho-ICDH, causing full reactivation of ICDH, in the presence of inhibitors of the kinase and either ADP or ATP. When ^{32}P-ICDH is incubated with ICDH kinase/phosphatase in the presence of ATP, ^{32}P is released from ICDH but there is only slow reactivation of ICDH. This can be explained by the rephosphorylation of ICDH by ICDH kinase which adds cold phosphate to the enzyme. Thus both the kinase and the phosphatase can be active similtaneously *in vitro* and so may also both be active *in vivo* (LaPorte and Koshland, 1982; Nimmo *et al.*, 1984).

An unusual feature of ICDH phosphatase activity is the absolute requirement for either ATP or ADP (LaPorte and Koshland, 1982; Nimmo *et al.*, 1984). Since ^{32}P released from [^{32}P]-ICDH appears as inorganic phosphate rather than [^{32}P]-ATP, the dephosphorylation cannot simply be the result of the back reaction of the kinase. Although both ICDH kinase and ICDH phosphatase require ATP, the deduced amino acid sequence shows only one region which matches the consensus ATP binding site (Klumpp *et al.*, 1988). Photoaffinity labelling of ICDH kinase/phosphatase using the ATP analogue 8-azido ATP resulted in parallel losses of kinase and phosphatase activities and in covalent attachment of the reagent at a single site on the enzyme. ATP or ADP protected against this inactivation, suggesting that this occurred at the active site (Valera and Nimmo, 1988). Using site directed mutagenesis, Stueland *et al.* (1989) changed the "invariant" lysine, which is found in every protein kinase to date and is involved in ATP binding (Hanks *et al.*, 1988), to a methionine. This mutation resulted in inhibition of both the kinase and phosphatase activities, indicating that they both use the same ATP binding site. Random mutagenesis of *aceK* resulted in several mutants which had retained their kinase activities but not their phosphatase activities (Stueland *et al.*, 1987). It was discovered that these mutants suffered a drastic decrease in their affinity for phospho-ICDH. This suggests that both the kinase and phosphatase reactions occur in the same site (LaPorte *et al.*, 1989). These mutations may be at the active site or they may cause a conformational change that inhibits the phosphatase and activates the kinase.

As well as a kinase and a phosphatase activity, ICDH kinase/phosphatase can also catalyse the hydrolysis of ATP to ADP, even in the absence of ICDH (Stueland *et al.*, 1987). In contrast to other kinases which can hydrolyse ATP, the ATPase of ICDH kinase/phosphatase has a higher V_{max} than either its kinase or its phosphatase activities. The ATPase was found to be active even when ICDH kinase/phosphatase is complexed with ICDH or phospho-ICDH. This ATPase activity is thought to be involved in the reaction of the phosphatase. The current hypothesis is that ICDH kinase/phosphatase has a single active site and that the phosphatase reaction represents the reverse of the kinase reaction coupled with ATP hydrolysis (LaPorte *et al.*, 1989).

1.2.4 Control Of ICDH Phosphorylation:

Since ICDH kinase and ICDH phosphatase can be active simultaneously *in vivo* (Borthwick *et al.*, 1984a), the phosphorylation state of ICDH is representitive of the steady state balance of ICDH kinase and phosphatase activities. Therefore, the activities of ICDH kinase and ICDH phosphatase have to be very tightly controlled.

ICDH kinase/phosphatase is controlled by a wide variety of metabolites (Figure 1.3). Isocitrate, 3-phosphoglycerate, pyruvate, AMP, ADP, oxaloacetate, 2–oxoglutarate and PEP activate ICDH phosphatase and inhibit ICDH kinase, while NADP, citrate, fructose-6-phosphate and glyoxylate inhibit ICDH kinase but have little effect on ICDH phosphatase (LaPorte and Koshland, 1983; Nimmo and Nimmo, 1984). Considering the intracellular concentrations of the above metabolites (Lowry *et al.*, 1971; Bautista *et al.*, 1979; Morikawa *et al.*, 1980; El-Mansi *et al.*, 1985), the effectors most likely to be significant *in vivo* are isocitrate, PEP, 3-PG, AMP, ADP and NADPH (Nimmo and Nimmo, 1984).

Perhaps the most important effector of ICDH kinase/phosphatase is isocitrate since a low concentration of this metabolite will prevent flux through the glyoxylate bypass. The other effectors can be divided into three catagories. PEP, 3-PG, pyruvate, 2-oxoglutarate and oxoaloacetate are all precursors for biosynthesis. High levels of these metabolites would be expected to favour flux through the TCA cycle instead of the glyoxylate bypass. A decrease in the energy charge of the cell will result in increased levels of AMP and ADP and might be expected to favour flux through the TCA cycle. As expected, all of the above metabolites inhibit ICDH kinase and activate ICDH phosphatase. NADPH is a product of the TCA cycle and its inhibition of ICDH phosphatase can be regarded as feedback inhibition. Thus ICDH phosphorylation is sensitive to a number of metabolic effectors and the fact that all of these effectors, with the exception of NADPH, act by affecting both opposing activities amplifies this sensitivity.

Further amplification of the sensitivity of the ICDH phosphorylation cycle is the result of "zero-order ultrasensitivity" (Goldbeter and Koshland, 1981), which refers to the situation where one or both converter enzymes are saturated with their substrates. Also, the phosphorylation system can respond to variations in the ICDH concentration *in vivo* (LaPorte *et al.*, 1985).

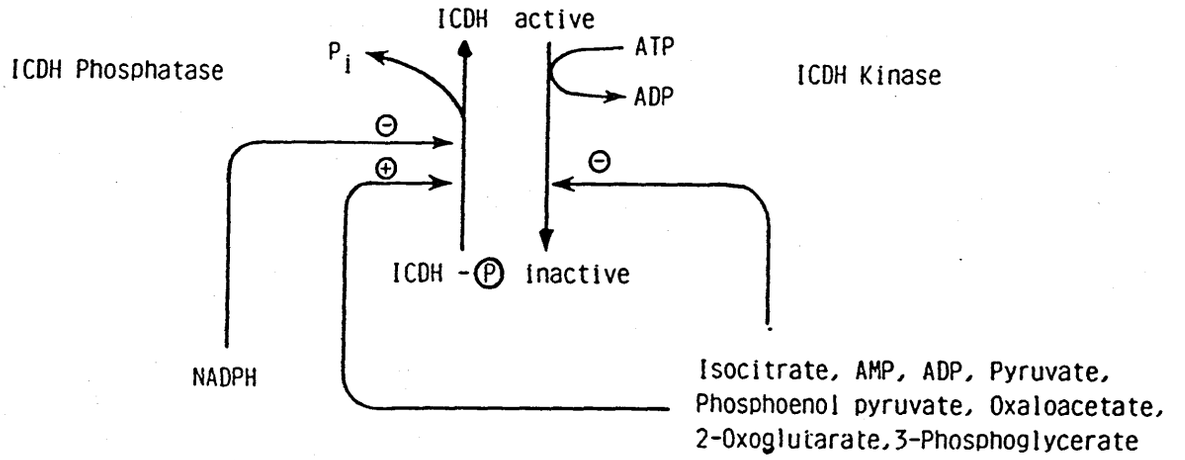


Figure 1.3: Control Of ICDH Kinase/Phosphatase

Taken from Nimmo (1984). + and - represent stimulatory and inhibitory effects respectively.

Therefore the system responds efficiently to a number of different effectors in order to maintain flux through the glyoxylate bypass during growth on acetate or fatty acids as the sole carbon source. La Porte *et al.* (1985), studied the levels of ICDH activity and phosphorylation in different *E. coli* strains with multiple copies of *icd*, the gene encoding ICDH. It was found that even when *icd* was expressed ten-fold over the level in the wild type, the ICDH activity remained constant as the proportion of phosphorylated ICDH increased.

1.2.5. Mechanism Of Inactivation Of ICDH:

Evidence has been presented which suggests that ICDH is phosphorylated at a number of sites. Cortay *et al.* (1986) resolved the ICDH from *E. coli* grown on acetate in the presence of $^{32}P_i$ into three distinct spots differing in charge and showed that ICDH is modified at both serine and threonine residues. These ^{32}P-laballed bands did not appear on control cells grown on glucose. Digestion of phospho-ICDH with several proteases revealed the presence of a number of phosphorylated peptides. The authors did not however, measure the stoichiometry of phosphorylation and inactivation of ICDH. Most of the evidence favours the conclusion that phosphorylation of ICDH at a single serine residue is responsible for inactivation.

Phosphorylation of ICDH on a single serine residue completely inactivates the enzyme and the difference between active and inactive ICDH is that inactive ICDH contains one phosphoserine group per subunit (LaPorte and Koshland, 1983; Borthwick *et al.*, 1984a,b,c; Nimmo *et al.*, 1984). Borthwick *et al.* (1984c) isolated a 22-residue phosphopeptide of ICDH generated by digestion of ICDH which had been phosphorylated either *in vivo* or *in vitro*. The first 14 amino acids of this peptide were sequenced. Malloy *et al.* (1984) independently determined the sequence of a phosphopeptide isolated from *in vivo* ^{32}P-labelled ICDH digested with trypsin. The same serine residue was phosphorylated in both cases and corresponds to Ser113 from the derived amino acid sequence of ICDH (Thorsness and Koshland, 1987):

Thr-Thr-Pro-Val-Gly-Gly-Gly-Ile-Arg-^{113}Ser(P)-Leu-Asn-Val-Ala

Since active but not inactive ICDH can bind to Procion Red Sepharose, which binds to the nucleotide binding site of many nucleotide binding proteins, Borthwick *et al.* (1984a) proposed that inactive ICDH was inactive because it was unable to bind $NADP^+$. Consistent with this, Garland and Nimmo (1984) found that inactive ICDH was resistant to limited proteolysis by a variety of proteases whereas active ICDH was not, but could be protected by $NADP^+$. Fluorescence titration experiments also indicated that $NADP^+$ bound to ICDH but not to phospho-ICDH. Also the sequence around the phosphorylated serine residue is similar to a region very close to the $NADP^+$ binding site of dihydrofolate reductase (DHFR) (Volz *et al.*, 1982): Arg-Ser-Leu-Asn. Nimmo (1984) proposed that phosphorylation of ICDH occurs at or near to the $NADP^+$ binding site of the enzyme so that the introduction of a negatively charged phosphate group could prevent $NADP^+$ binding by charge repulsion and also cause a conformational change similar to that caused by $NADP^+$ binding. Consistent with this hypothesis was the finding that there was a difference in the near UV CD spectra between ICDH and ICDH bound to $NADP^+$ (McKee *et al.*, 1989) This would explain why both $NADP^+$ and NADPH inhibit phosphorylation of ICDH and why phosphorylated ICDH cannot be reactivated by proteolytic removal of the phosphorylation site (Garland and Nimmo, 1984), unlike several eukaryotic enzymes regulated by phosphorylation (e.g. Huang and Cabib, 1974). Thorsness and Koshland (1987) introduced a negatively charged amino acid, aspartate, at position 113 of ICDH replacing the serine by site-directed mutagenesis. This substitution inactivated the enzyme suggesting that the inactivation by phosphorylation could be a result of the introduction of a negative charge. Replacement of Ser113 with a variety of amino acids, including threonine, completely prevented phosphorylation, indicating that Ser113 is necessary for recognition by ICDH kinase/phosphatase and confirming that ICDH kinase phosphorylates only one site in ICDH.

Contrary to this model, Dean *et al.* (1989), using binding studies, found that ICDH binds isocitrate, 2-oxoglutarate, $NADP^+$ and NADPH at a 1:1 ratio of substrate to enzyme monomer but that the phospho-ICDH cannot bind isocitrate although it retains the ability to bind $NADP^+$ and NADPH. This result is not compatible with the compulsory order, steady state kinetic mechanism of ICDH with $NADP^+$ binding first (Nimmo, 1986). Binding studies were also done for the Asp113 mutant generated by Thorsness and Koshland (1987). It was

found that like the inactive, phosphorylated ICDH, this mutant could not bind isocitrate but the binding of NADPH was unaffected. Thus these results suggest that the negative charge introduced by either the Asp at position 113 or phosphorylation of Ser113 causes inactivation by preventing binding of isocitrate. It should be noted however, that the kinetic and binding studies were carried out under different conditions and with enzymes from different *E. coli* strains.

The controversy over the mechanism of inactivation was not resolved until the 3-dimensional structure became available. Hurley *et al.* (1989) crystallised ICDH and determined the 3-dimensional structure by X-ray diffraction. Ser113 was found to lie at the edge of an inter-domain pocket which is formed between the two subunits. Because of its position, this pocket was suggested to be the active site. The same crystals were soaked in a buffer containing Mg^{2+} and isocitrate and the 3-dimensional structure determined. The results of this showed that isocitrate did in fact bind in this pocket (Hurley *et al.*, 1990a) and that the Ser113 actually forms a hydrogen bond to the γ-carboxyl group of isocitrate. Substitution of Ser113 with Asp or Glu, both negatively charged, did not cause the conformational change in ICDH observed in binding of isocitrate to the enzyme or on phosphorylation (Hurley *et al.*, 1990b), indicating that the introduction of a negative charge at Ser113 does not induce a conformational change. Instead, the conformation change observed on phosphorylation of ICDH is probably caused by steric effects of the phosphate group. Hurley *et al.* (1990a) calculated that the electrostatic potential change on phosphorylation could account for the decreased affinity of ICDH for isocitrate. Consistent with this is the finding that there is no significant decrease in activity of ICDH with 2R-malate as the substrate. 2R-malate is an anologue of isocitrate which lacks the γ-carboxyl group of isocitrate that interacts with Ser113 and would be repelled by phospho-serine on phosphorylation (Dean and Koshland, 1990). ICDH crystals were also soaked in $NADP^+$ and the structure determined (Hurley *et al.*, 1991). $NADP^+$ was found to bind to ICDH in the cleft between the α+β and α/β domains. The pyrophosphate moiety of $NADP^+$ binds near to the helical turn at amino acids 342-346 and might form hydrogen bonds with the nitrogens in the peptide bonds at amino acids 340-342. There is very little conformational change on binding $NADP^+$, which is not in agreement with McKee *et al.* (1989). However, the changes in the near UV CD spectra in binding $NADP^+$ to

ICDH may be accounted for by the electrostatic interactions between the 2' phosphate of the nucleotide with Tyr345 and Tyr391, by binding the adenine ring of $NADP^+$ to the enzyme, or both.

For many enzymes, it is clear that phosphorylation alters their allosteric properties rather than k_{cat}. The structural basis for this is known for glycogen phosphorylase (Goldsmith *et al.*, 1989). It is clear that the mechanism of inactivation of ICDH by phosphorylation is quite different.

1.3 The Glyoxylate Bypass (*ace*) Operon:

1.3.1 Genetic Organisation Of The Glyoxylate Bypass:

The structural genes for ICL, *aceA*, and MS-A, *aceB*, have been mapped by several groups independently and have been shown to be located at 90 min on the revised *E.coli* K12 linkage map (Bachmann and Low, 1980). Vanderwinkel and DeVlieghere (1968) mapped *aceA* and *aceB* with an adjacent gene, *metA*, and found the gene order to be *metA-aceB-aceA*. Brice and Kornberg (1968) also mapped this region and isolated mutants in an adjacent gene *iclR*,which caused constitutive expression of ICL, and showed close linkage with *aceA*. Maloy and Nunn (1982) confirmed the gene order as *metA-aceB-aceA-iclR*, although the exact distance between *aceA* and *iclR* was uncertain. Bachmann and Low placed *iclR* at approximately 0.4min from *aceA*.Similar results were obtained by Vinopal and Fraenkel (1974) who suggested that the very close linkage could be due to a bias in favour of co-inheritance of the *iclR* mutation and *aceA* since constitutive expression of the glyoxylate bypass enzymes enhances growth on plates containing acetate as the sole carbon source. Since both ICL and MS-A are coordinately expressed and because of the close genetic linkage between *aceA* and *aceB*, it was suggested that these genes may form an operon (Brice and Kornberg, 1968).

Using transposon Tn10 insertions, which prevent expression of genes downstream of the point of insertion in an operon, Maloy and Nunn (1982) found that *aceA*::Tn10 insertions eliminated ICL but not MS-A activity, and *aceB*::Tn10 insertions eliminated both ICL and

MS-A activities, in a *glc* background (*glc* encodes a second malate synthase, MSB, which is induced during growth on glyoxylate, or compounds such as glycolate which are metabolised to glyoxylate, as the sole carbon source). This confirms that *aceA* and *aceB* do constitute an operon, the *ace* operon, with transcription from *aceB* to *aceA*.

It was also found that the regulation of expression of *aceK*, the structural gene for ICDH kinase/phosphatase, paralleled that of *aceA* and *aceB*, although *aceK* was expressed 100-1000 times less than *aceA* or *aceB*.*aceK* was also shown to be closely linked with *metA*, suggesting that *aceK* may be part of the *ace* operon (LaPorte and Chung, 1985). This was confirmed by phage Mu insertion. Three phage Mu lysogens were isolated which had lost one or more of the activities of the glyoxylate bypass. Two of these had lost both ICL and ICDH kinase/phosphatase activities and had reduced MS-A activity. These were assigned as being *aceA*::Mu although they could have been *aceB*::Mu since this was not done in a *glc* background. Neverthless, this result confirms that *aceK* is part of the *ace* oeron. The third lysogen abolished only the ICDH kinase/phosphatase activity, therefore placing the *aceK* gene downstream of *aceA*.

The *ace* operon has now been cloned from *E.coli* by several groups. Chung *et al.* (1988) used BAL31 deletion in conjunction with enzyme assays and *in vitro* transcription/ translation using the maxicell labelling procedure to determine the 3'ends of *aceB* and *aceA*. *aceK* was mapped by comparing the restriction maps of the clone with the predicted restriction map from the DNA sequence of *aceK* (Klumpp *et al.*,1988). The promoter was identified by deletion mapping, nuclease S1 mapping and sequencing of the DNA immediately upstream of the startsite. These studies confirm the gene order as being *aceB-aceA-aceK* with transcription from a single promoter during growth on acetate and *aceB* proximal to the promoter.

1.3.2 Evidence For A Fourth *ace* Gene:

While sequencing the *aceB* gene, Byrne *et al.* (1988) found that DNA from the *aceB* gene hybridised to a 10kb messenger RNA. Since all three structural genes of the *ace* operon are encoded within 7kb of the promoter, it is possible that a fourth gene is encoded in the *ace* operon. An open reading frame has been found within the 1.9kb *Hin*dIII fragment immediately downstream of *aceK* (Cortay *et al.*, 1989), which the authors first claimed to be

iclR. However, *iclR* has since been cloned and sequenced by the same group and another group from a region further downstream from the 3' end of *aceK* than this open reading frame (Sunnarborg *et al.*, 1990; Negre *et al.*, 1991). Garlinier *et al.* (1991) have sequenced the entire intergenic region between *aceK* and *iclR* and found an open reading frame capable of encoding a 69,355 dalton polypeptide. This open reading frame is in the same orientation as *iclR* but in the opposite orientation from *aceK*. Although the authors did not identify any possible promoter sequences, there are regions upstream of the translation initiation startsite with considerable homology to the *E. coli* consensus promoter and the consensus ribosome binding site. The sequences CATTGACGC (at position 144-152) and TACACT (at position 171-176) are homologous to the -35 and -10 promoter consensus sequences, tcTTGACat and TAtAaT respectively (Rosenberg and Court, 1979; Siebenlist *et al.*, 1980), while two sequences GAGG and GGAG at positions 427-430 and 447-450 respectively are similar to the consensus ribosome binding sequence AGGAGG (Shine and Dalgarno, 1974). Thus it appears that there may be a gene immediately downstream from *aceK*, expressed in the same direction as *iclR*. Although not part of the ace operon, the possibility that this gene may be involved in the regulation of the ace operon or in growth on acetate cannot be ruled out.

1.3.3 Genetic Control Of The *ace* Operon:

As mentioned above, the *ace* operon appears to be controlled by the adjacent gene, *iclR*, since mutations in this gene lead to constitutive expression of ICL (Brice and Kornberg, 1968). In addition, the *fadR* gene, which is also involved in the synthesis of the enzymes involved in fatty acid degradation (Simons *et al.*, 1880a,b) and is required for maximal expression of unsaturated fatty acid biosynthesis (*fab*) (Nunn *et al.*, 1983), has also been shown to be involved in the regulation of the *ace* operon (Maloy *et al.*, 1980).The *fadR* gene is thought to encode a repressor of the *fad* regulon (Simons *et al.*, 1980a).

It was shown that ICL and MS-A activities were higher in *fadR* and *fadR*::Tn10 strains grown on succinate or malate as the sole carbon source, than in *fadR*+ strains (Maloy *et al.*, 1980). In support of this, a temperature sensitive mutant of *fadR* showed constitutive expression of the *fad* enzymes and also the *ace* operon at 42°C but low, inducible expression at 22°C. These mutants and Tn10 insertions were used to map the *fadR* gene at 25.5min on the

revised *E. coli* K12 linkage map. Similar results were obtained with strains which could not grow on long chain fatty acids due to mutations in different *fad* genes. Thus the increased expression of the glyoxylate bypass enzymes was not due to an increase in the acetate concentration from β-oxidation of fatty acids.*aceA*::Mu d(Ap *lac*) showed higher β-galactosidase and malate synthase activities in *iclR* or *fadR* mutants grown on succinate than *fadR*+,*iclR*+ grown on the same substrate but similar levels when grown on acetate (Maloy and Nunn, 1982). This result suggests that both *iclR* and *fadR* act at the level of transcription. Using F-primes which contained *iclR* or *fadR* in the host strains which were either *iclR* or *fadR*, it was shown that both *iclR* and *fadR* genes act in *trans* suggesting that both *iclR* and *fadR* genes encode repressors. Thus, the *ace* operon is regulated by two repressors encoded by the *iclR* and *fadR* genes.

The *iclR* gene has now been cloned and sequenced from *E. coli* (Sunnarborg *et al.*, 1990; Negre *et al.*, 1991) and *Salmonella typhimurium* (Garlinier *et al.*, 1990). This clone complemented a mutant *iclR* in *trans* and contained a 900base pair region which was both necessary and sufficient for complementation. This region was sequenced and was shown to be capable of encoding a 29,741 dalton protein (Sunnarborg *et al.*, 1990). Negre *et al.* (1991) also cloned and sequenced this region and obtained identical sequence except for a single nucleotide change in the putitive ribosome binding sequence. The deduced amino acid sequence data indicate a possible helix-turn-helix structure, which has been found to be present in the DNA-binding domains of several DNA-binding proteins. Negre *et al.* (1991) showed that the *iclR* gene product bound to the *ace* operator using extracts of *E. coli* which overexpress the *iclR* gene and [α^{32}P]dATP-labelled DNA containing the promoter/operator region of the ace operon. The same group have since purified the *iclR* gene product and confirmed the above result using this instead of the cell extracts. Interestingly, by incubating the purified *iclR* gene product with the *ace* promoter/operator fragment, in the presence of PEP, prevents the repressor from binding (Cortay *et al.*, 1991) (see section 1.3.5).

The *fadR* gene has also been cloned although the nucleotide sequence is not yet available (DiRusso and Nunn, 1985). This *fadR* clone exerted transcriptional control over β-galactosidase synthesis in a *fadR* strain which had a *fadE-lacZ*$^+$ operon fusion. The *fadR*

clone also suppressed an unsaturated fatty acid auxotrophy in a *fadR* mutation, *fadRfabA*(Ts) which could not grow without unsaturated fatty acids at permissive temperatures, suggesting that *fadR* is also involvd in unsaturated fatty acid biosynthesis. This 1.3kb clone produced a 29,000 dalton protein in maxicell analysis.

1.3.4 Possible Advantages Of A Dual Repressor System:

When *aceA*::Mu d(Ap *lac*)*fadR* mutants were grown on acetate as the sole carbon source, the β-galactosidase activity was higher than in the same strain grown on succinate. Also, *aceA*::Mu d(Ap *lac*)*iclRfadR* showed higher levels of β-galactosidase when grown on succinate than either *aceA*::Mu d(Ap *lac*)*iclR* or *aceA*::Mu d(Ap *lac*)*fadR* strains alone (Maloy and Nunn, 1982). These results suggest that the *iclR* and *fadR* repressors act independently causing partial repression of the ace operon, or together to cause full repression. Thus the operon could be activated to different levels according to different carbon sources. During growth on acetate as the sole source of carbon, the energy and reducing power required for growth can only be obtained by allowing flux through the TCA cycle. However, during growth on fatty acids as the sole carbon source, a substantial amount of energy and reducing power is generated during the β-oxidation of the fatty acid to acetyl-CoA, 1mol of $FADH_2$ and 1mol of NADH per turn of the β-oxidation cycle, thus reducing the requirement for flux through the TCA cycle. In support of this, higher ICL activities were observed in cells grown on oleate than the same cells grown on acetate (Maloy *et al.*, 1980).

1.3.5 Metabolic Control Of The *ace* Operon:

The *fad* regulon, which is under the control of the *fadR* gene, has been shown to be induced during growth on long chain fatty acids (>C_{11}). It therefore seems likely that *fadR* may regulate the *ace* operon in the same manner (Nunn *et al.*, 1979; Simons *et al.*, 1980a,b).

The rapid appearance of ICL activity after addition of acetate to *Achromobacter sp.* growing on succinate suggests that either (i) acetate, or acetyl CoA, acts directly as an inducer of the *ace* operon, (ii) acetate, or acetyl CoA, reacts with another metabolite to form a product which then induces the *ace* operon, or (iii) acetate, or acetyl CoA, reacts with a repressor metabolite to form a product, and in doing so reduces its concentration (Rosenberger, 1962).

Addition of acetate to *E. coli* growing on glucose, pyruvate or a usable intermediate of the TCA cycle did not induce ICL, suggesting that acetate, or acetyl-CoA, did not act directly as an inducer. This result also suggested that acetate did not react with another metabolite to produce an inducer, although during growth on glutamine, proline and γ-amino butyrate, the addition of acetate doubled the specific activity of ICL. However, no increase in the specific activity of ICL occurred in mutants which lack citrate synthase (Kornberg, 1963), which suggested that either citrate or a metabolite of citrate could be the inducer. Another possibility is that acetate, or acetyl CoA, reacts with oxaloacetate (OAA) to lower the concentration of an repressor such as oxaloacetate itself or a metabolite of it.

Supporting evidence exists for both of the above possibilities. Addition of pyruvate to an *E. coli* mutant devoid of the first enzyme of the pyruvate dehydrogenase complex (*pdc*-) growing on acetate and aspartate stopped ICL synthesis, suggesting that pyruvate acts as a repressor. Similar results were obtained for mutants which lacked phophoenolpyruvate carboxylase (PEP carboxylase), indicating that the inhibition was not due to the production of OAA from PEP. However, mutants devoid of both PEP carboxylase and PEP synthase showed inhibition of ICL synthesis on addition of glucose but not pyruvate. Therefore, pyruvate cannot be the repressor, instead, PEP or a metabolite close to PEP seems more likely (Kornberg, 1966). Supporting this is the evidence from gel retardation experiments with the promoter/operator region of the *ace* operon and the purified *iclR* repressor. In the presence of PEP, there was reduced binding of the operator region by the *iclR* repressor while acetate, acetyl-CoA, pyruvate and OAA had no effect (Cortay *et al.*, 1991). Contrary to the view that acetate relieves a repressor, it was shown using a series of mutants of the TCA cycle enzymes, that acetate had to be metabolised through the TCA cycle beyond succinate in order to induce ICL. Metabolism of acetate through both the TCA cycle and the glyoxylate bypass was necessary for full induction of ICL (Duckworth, 1981).

Thus, the mechanism of control of the *ace* operon is complex and the metabolites which regulate the expression have still to be identified. The regulation of the *ace* operon may be similar to that for the ICDH kinase/phosphatase in that it responds to a number of effectors. Perhaps now that *iclR* and *fadR* have been cloned and the promoter identified, the mechanism by which this operon is regulated can be elucidated.

1.3.6 Downshift In Expression From *aceK*:

Although *aceA* and *aceK* are part of the same operon, and as such are expressed from the same promoter, *aceA* is expressed 100-1000fold more than *aceK* (LaPorte and Chung, 1985).The reason for this downshift in *aceK* expression was unclear until the nucleotide sequence of the *aceA-aceK* intercistronic region became available (Cortay *et al.*, 1988; Klumpp *et al.*, 1988). This region was shown to contain several REP (repetitive extragenic palindromic) elements which have been identified in the untranslated regions of a number of operons in *E. coli*.

The *aceA-aceK* intercistronic region contains three complete REP elements and two partial elements which match the left half of the complete elements. One of these partial elements contains the Shine-Dalgarno ribosome-binding site. A number of stem-loop structures are possible within this region, several of which, including the most stable, contains the Shine-Dalgarno site and the translation initiation codon of *aceK*. Therefore, it was first thought that this stem-loop structure may contribute to the downshift in expression from *aceK* by masking the Shine-Dalgarno site and the initiation codon so preventing translation initiation in a similar manner to that observed between *uncE* and *uncF* in *E. coli* (Klionsky *et al.*, 1986) (figure 1.4). However, deletion of sequences contributing to the stem-loop between *aceA* and *aceK* had no effect on expression from *aceK* but instead halved expression from *aceA* (Klumpp *et al.*, 1988). Thus, the stem-loop does not inhibit translation initiation but instead, stabilizes the message upstream presumably in the same way as has been proposed for the *puf* operon of *Rhodobacter capsulatus* (Belasco and Chen, 1988) i.e. by preventing 3' exonucleolytic digestion of the message after an initial endonucleolytic cleavage downstream. It would therefore appear that the downshift in expression from *aceK* is due to fast degradation of the operon-length *aceBAK* transcript to *aceBA* relative to *aceBA* degradation. Another factor which may contribute to the downshift is that that the codon usage of *aceA* and *aceB* is similar to that in intermediately expressed genes whereas the codon usage of *aceK* is similar to that of genes which are expressed at a low level in *E.coli*, thus there may also be differential regulation at the level of translation during chain elongation (Cortay *et al.*, 1989).

Figure 1.4: Possible Secondary Structures In The *aceA-aceK* Intercistronic Region

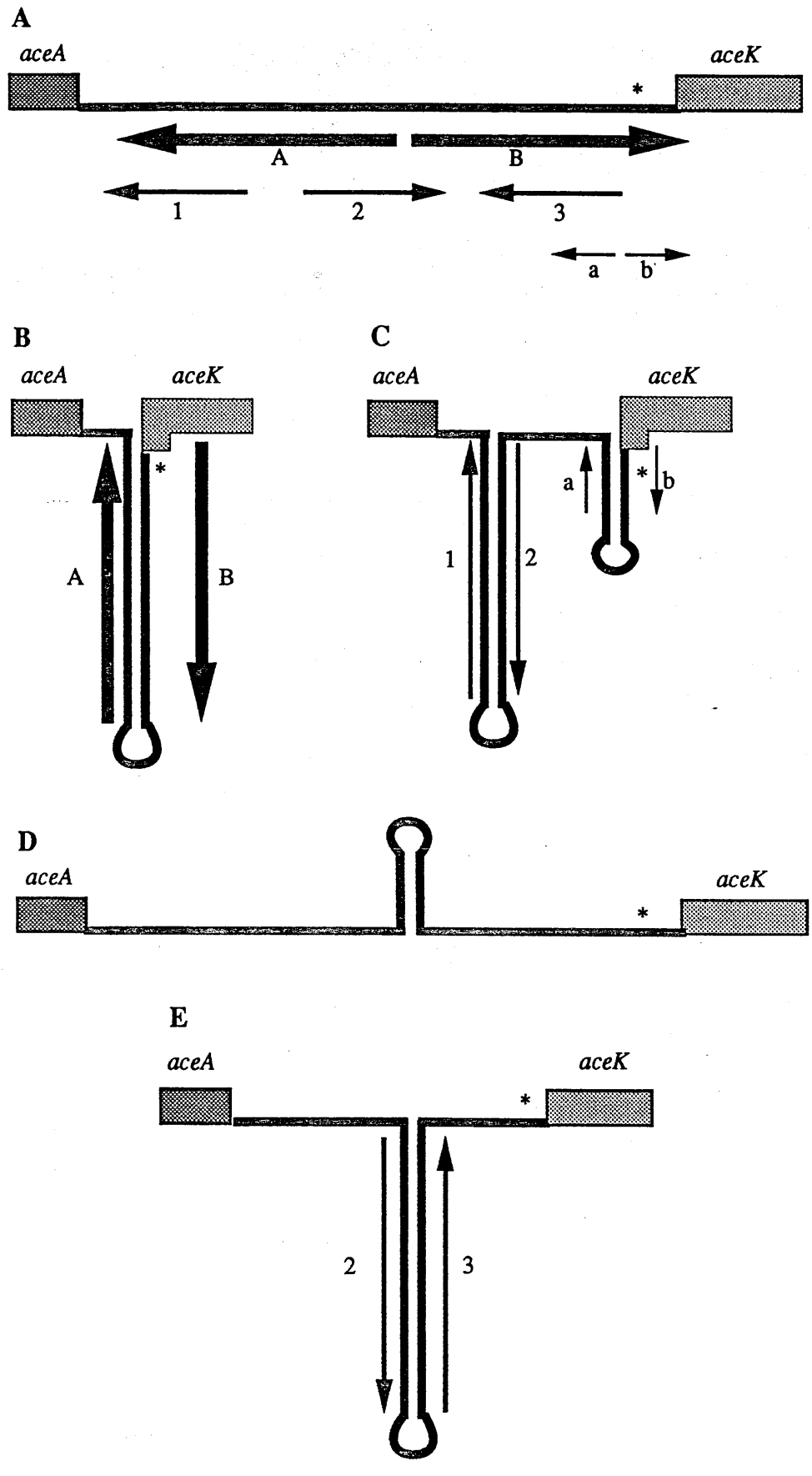
(A) Map of the intercistronic region between *aceA* and *aceK*, which are represented by the dark and light shaded boxes respectively. The heavy arrows, labelled A and B indicate a region of dyad symmetry, the lighter arrows, labelled 1, 2 and 3 represent the REP elements and arrows labelled a and b represent two partial REP elements. The asterisk indicates the Shine-Dalgarno, ribosome binding site.

(B) The most stable of the stem loops, formed by base pairing between regions A and B.

(C) Stem loops formed between REP elements 1 and 2 and also between the partial REP elements a and b.

(D) The stem loop formed REP element 2.

(E) The stem loop formed between REP elements 2 and 3.


A
aceA
aceK
*
A
B
1
2
3
a
b
B
aceA
aceK
*
A
B
C
aceA
aceK
a
*
b
1
2
D
aceA
aceK
*
E
aceA
aceK
*
2
3

1.4 Isocitrate Lyase:

1.4.1 Physical Properties:

E. coli ICL has a subunit M_r of approximately 45,000 (MacKintosh and Nimmo, 1988) which is similar to the values obtained for ICL from other prokaryotes, *Pseudomonas indigofera* (McFadden *et al.*, 1968) and a thermophilic *Bacillus* (Chell *et al.*, 1978). The plant enzymes however generally have a higher subunit M_r of between 62,000-68,000. The additional 17,000-18,000 in the plant ICL is thought to contain the signal sequence for sequestration into glyoxysomes, although neither the mechanism nor signal sequence are as yet known. The value for the subunit M_r of *Ricinus communis* ICL has been reported to be 62,000 (Roberts and Lord, 1981) and 35,000 (Malhorta and Srivastava, 1982). This discrepency could be due to proteolysis during isolation when the cells are ruptured, processing during uptake into glyoxysomes or by proteolysis during differentiation to limit activity. This may also be the reason for the very low M_r values obtained for ICL from *Helianthus annus* with M_r 28,500 (Zemlyanukhin *et al.*, 1984) and *Zea mays* with M_r 32,000 (Zemlyanukhin *et al.*, 1986).

The native M_r of the *E. coli* ICL is 180,000 (Robertson and Reeves, 1987; MacKintosh and Nimmo, 1988), suggesting that like most other ICLs, the *E. coli* enzyme is a tetramer of subunits with identical M_r. Indeed,the sequence of the amino-terminus has been determined for the *E. coli* ICL indicating that the subunits are probably identical (MacKintosh, 1987). However, ICL from *Glycine max* (Ruchti and Windmer, 1984) and *Candida tropicalis* (Uchida *et al.*, 1986) exist as dimers while the *Cucumis savitus* ICL has been reported to be a pentamer (Lamb *et al.*, 1978) although another group reported that the enzyme from the same source was a tetramer. There is evidence that the *N. crassa* ICL is stable as a dimer (Johanson *et al.*, 1974b).

The amino acid composition has been determined for ICL from several organisms; *E. coli* (Hoyt *et al.*, 1988), *Pseudomonas indigofera* (Shiio *et al.*, 1965a), thermophilic *Bacillus* (Chell *et al.*, 1978), *N. crassa* (Johanson *et al.*, 1974b), *Citrullus vulgaris* (Jameel *et al.*, 1984), *Candida tropicalis* (Uchida *et al.*, 1986), *Pinus pinea* (Pinzauti *et al.*, 1986)

and *Linum usitatissum* (Khan and McFadden, 1982). The data is shown in table 1.2.

By comparing the amino acid composition of pairs of proteins, Marchalonis and Weltman (1971) found that 98% of unrelated proteins have an SΔQ value ($S\Delta Q=\Sigma(x_{ij}-x_{kj})^2$), where i and k indicate the proteins to be compared and x_j is the percent of residues of a given amino acid type j in the monomer) of greater than 100 and no case was the value less than 50. Using this method of comparison, it was shown that for every ICL for which the amino acid composition was known, the SΔQ value was less than 50 for each comparison. This indicates that the amino acid composition of all of ICL from various organisms is conserved.

As well as revealing the similiarity between proteins, the amino acid composition can also give information on the physical properties of an individual enzyme. Vanni *et al.* (1990) calculated the hydrophobicity average (Bigelow, 1967), the "non-polar side chains (Waugh, 1954), the polar : non-polar volume ratio (Fisher, 1964), the polar : non-polar residues ratio (Hatch, 1965), the ratio of (Asx+Glx) : (Lys+His+Arg) and the relative specific volume. When the values for each ICL are compared, they are similar. The only exception is that the (Asx+Glx) : (Lys+His+Arg) ratio for ICL from *E. coli*, *P. indigofera* and *Bacillus*, are similar to each other with values of 1.84, 1.73 and 1.86 but the values for the plant enzymes are all in the range from 1.34 to 1.55, thus the plant ICLs are more basic. This is reflected in the pI values of the plant ICLs compared to the prokaryotic ICLs. The *E. coli* and *Bacillus* ICLs have pI values of 4.6 (Robertson and Reeves, 1987) and 4.5 (Chell *et al.*, 1978) respectively, while the plant ICLs appear to have very little charge at pH7.0-7.7 (Giachetti *et al.*, 1987).

1.4.2 Requirement Of ICL for Magnesium:

ICL from *E. coli* requires Mg^{2+} for activity as does ICL from all other sources with the possible exception of ICL from *Thiobacillus novellus*, which shows greater activity with Co^{2+} (McCarthy and Charles, 1973). Hoyt *et al.* (1988) tested a variety of metal ions for their ability to activate ICL. No activity was observed in the absence of any metal ion and the highest activity was observed with Mg^{2+}. With Mn^{2+}, Co^{2+}, Ni^{2+}, and Sr^{2+}, the activities were 54%, 17%, 7% and 3% of the activity in the presence of Mg^{2+} respectively. The optimum Mg^{2+} concentration was found to be 5mM, concentrations above this had an inhibitory effect on ICL

Table 1.2: Amino Acid Compositions Of ICL From Various Organisms

The amino acid compositions are expressed as the molar percentage obtained as follows:

(1) Matsuoka and McFadden (1988)

(2) and (4) Johanson *et al.* (1974b)

(3) Chell *et al.* (1978)

(5) Jameel *et al.* (1984a)

(6) Uchida *et al.* (1986)

(7) Pinzauti *et al.* (1986)

(8) Khan and McFadden (1982).

The SΔQ values were calculated with respect to the *E. coli* ICL.

Amino Acid	*E. coli* (1)	*P. indigofera* (2)	*Bacillus* (3)	*N. crassa* (4)	*C. vulgaris* (5)	*C. tropicalis* (6)	*P. pinea* (7)	*L. usitatissimum* (8)
Asx	7.27	9.66	9.66	10.09	9.12	9.70	9.58	9.27
Thr	5.90	5.86	6.65	5.25	7.18	5.50	6.29	5.52
Ser	6.53	4.39	5.50	4.94	7.01	4.90	5.57	4.35
Glx	12.39	11.77	15.67	12.30	11.44	11.80	11.01	10.87
Pro	6.35	3.93	2.93	3.83	3.19	3.60	3.51	2.59
Gly	8.15	8.61	7.37	6.45	6.95	7.10	7.57	8.88
Ala	12.84	13.02	13.15	12.13	10.64	11.90	11.20	12.13
Val	5.19	7.63	5.46	5.33	5.14	4.70	6.31	9.10
Met	2.36	2.09	1.86	2.36	1.98	1.90	3.70	2.53
Ile	4.38	4.07	5.59	5.15	4.74	6.00	3.95	4.74
Leu	7.37	6.85	8.00	8.07	7.78	7.00	7.89	6.63
Tyr	4.14	4.10	3.61	3.33	2.90	3.80	2.76	3.73
Phe	3.91	3.46	4.18	2.86	4.22	3.60	4.00	4.20
Lys	4.69	5.59	6.23	7.91	5.77	9.00	6.00	5.95
His	1.90	2.28	2.28	1.82	2.62	2.60	2.56	2.47
Arg	4.08	4.61	5.21	4.65	6.56	4.50	5.92	4.86
Trp	0.87	1.05	1.30	2.16	1.28	1.70	0.87	1.30
Cys	1.76	1.03	0.84	1.36	1.48	0.70	1.00	1.15
SΔQ	-	9	14	21	25	27	17	20

(MacKintosh, 1987). This effect is unlikely to be a result of the increased concentration of the chloride ion, which inhibits ICL (see section 1.4.5) as this occurs at higher Cl^- concentrations.

Giatechetti *et al.* (1988) have shown that for the *Pinus pinea* ICL, the Mg^{2+}-isocitrate complex is the true substrate and that Mg^{2+} also acts as a non-essential activator of ICL. The model for the action of Mg^{2+} in catalysis is shown in figure 1.5. In this model there are two Mg^{2+} binding sites; one high affinity site and one low affinity site, the catalytic site. Free Mg^{2+} binds to the catalytic site of the activated enzyme (Mg-E) but not the non-activated form (E). Neither free isocitrate nor the Mg^{2+}-isocitrate complex can bind to the Mg-E-Mg form. This may be an explanation for the inhibition of ICL at high Mg^{2+} concentrations. Consistent with this was the finding that isocitrate did not protect ICL from *N. crassa* against 3-bromopyruvate inactivation unless Mg^{2+} was present even though Mg^{2+} does not itself protect (Roche *et al.*, 1971).

Mg^{2+} also protected ICL against thermal inactivation (Shiio *et al*, 1965b; Pinzauti *et al.*, 1983; Tsukamoto *et al*, 1986; Galassi *et al*, 1988) and proteolysis (Galassi *et al.*, 1988). It has been suggested that there may be a third Mg^{2+} binding site since ICL from *Pinus pinea* when exhaustively dialysed against EDTA showed the presence of residual Mg^{2+} when submitted to atomic absorption (Vanni *et al.*, 1990). However, the stoichiometery was not calculated and it is still a possibility that the residual Mg^{2+} is bound to the high affinity activation site.

1.4.3 K_m Of ICL From *E. coli* For Isocitrate:

The K_m of ICL from *E. coli* ML308 for isocitrate decreases with pH and was found to be 63μM at pH7.3, 32μM at pH6.8 and 7μM at pH6.3 (MacKintosh and Nimmo, 1988). These values compare with 18μM and 24μM at pH6.8 for *E. coli* strains Bm and W respectively (Ashworth and Kornberg, 1963), 8μM at pH7.5 in *E. coli* $D_5H_3G_7$ (Robertson and Reeves, 1987), 604μM at pH7.5 (LaPorte *et al.*, 1984) and 3mM (conditions not stated) (Bautista *et al.*, 1979).

The values for the K_m for isocitrate given by Robertson and Reeves (1987) and LaPorte *et al.* (1984) could be misleading due to the assay conditions used to obtain these values.

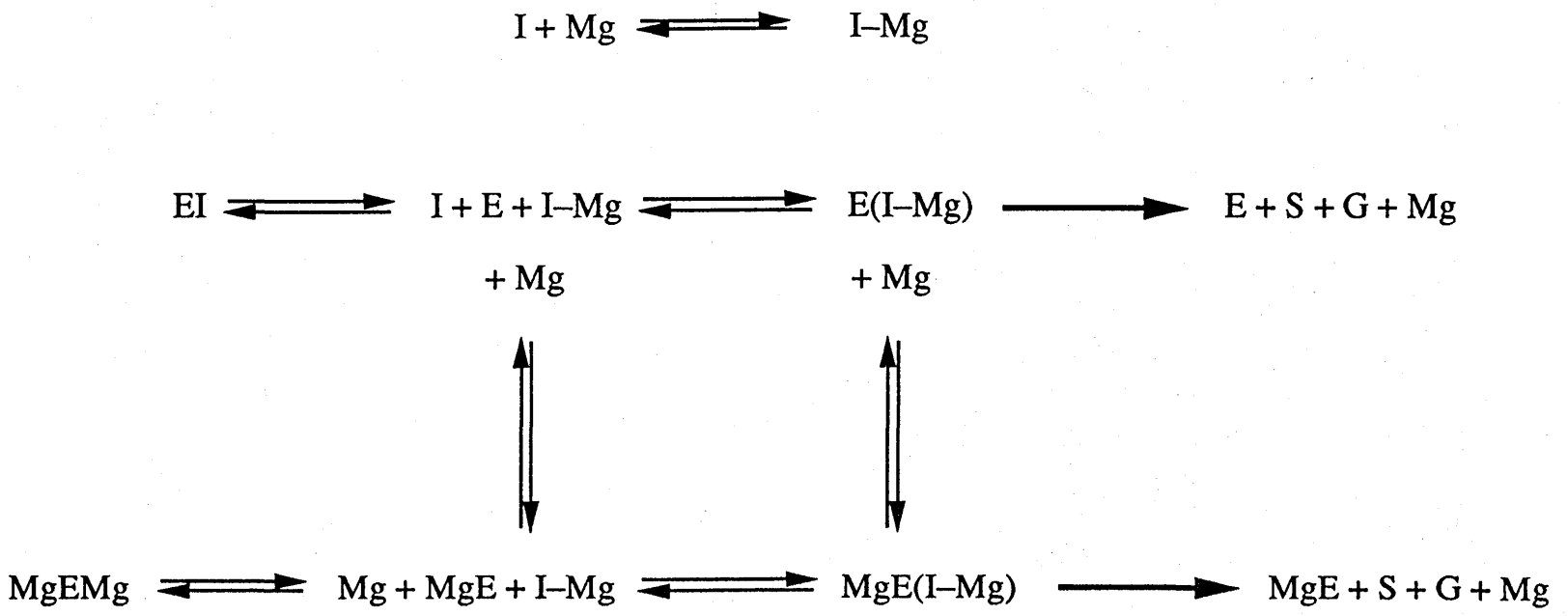


Figure 1.5: Model For Binding Of Magnesium By Isocitrate Lyase

Redrawn from Vanni *et al.*, 1990, where I represents isocitrate, Mg is Mg2+, E is free, non-activated enzyme, S is succinate; G is glyoxylate and MgE is activated enzyme.

LaPorte *et al.* used the stopped phenylhydrazine assay in the presence of 200mM NaCl. The chloride ion was shown to be a competitive inhibitor of ICL (see below; MacKintosh and Nimmo, 1988) and ICL from *E. coli* ML308 assayed under these conditions gave a K_m of 600μM instead of 63μM (MacKintosh, 1987). Robertson and Reeves used a Tris buffer, which has been shown to give lower activities compared to Mops as the buffer for the *E. coli* enzyme (Hoyt *et al.*, 1988). This could be the result of formation of a N-hemiacetal with glyoxylate and Tris (Roche *et al.*, 1971) or the chloride ion in the Tris buffer being more inhibitory than the Mops anion.

Table 1.3 shows the K_m values of ICL from various organisms and the conditions in which they were obtained. The K_m value for the *E. coli* ML308 ICL is in the same order of magnitude as all the other ICLs.

1.4.4 Effect OF pH On ICL:

The optimal pH for ICL from *E. coli* ML308 was found to be pH7.3 (MacKintosh and Nimmo, 1986; MacKintosh, 1987). This value is in agreement with that obtained for ICL from *E. coli* $D_5H_3G_7$ (Robertson and Reeves, 1987). The pH optima for ICL from various organisms are very diverse (Vanni *et al.*, 1990).

Although it can be seen that the K_m for isocitrate decreases with pH for the *E. coli* enzyme over the pH range 6.3-7.3, very little other work has been done on the effect of pH on *E. coli* ICL. More comprehensive studies have been done for the *N. crassa* (Rogers and McFadden, 1977) and *P. indigofera* (Rogers and McFadden, 1976). For the *N. crassa* enzyme, a plot of pK_m against pH results in a biphasic curve with the top linear portion of the curve having a gradient of -1 and the bottom non-linear portion tending to flatten, indicating that the K_m for isocitrate depends on a single dissociating group on the enzyme-substrate complex. The pK_a of this group was determined to be 8.5 (Rogers and McFadden, 1977). This dissociation involves the free enzyme since the plot of pK_m against pH reflects the ionization of either the free substrate or the free enzyme (Fersht, 1977). With *P. indigofera* ICL, a plot of pK_m against pH gives a straight line with gradient -1, indicating that, like the *N. crassa* enzyme, the K_m is dependent on a single dissociating group. Similarly, in the *N. crassa*

Table 1.3: K_m Values Of ICL From Various Organisms

Organism	K_m for Isocitrate (mM)	pH And Buffer	Reference
Thermophilic *Bacillus*	0.020	6.8(Imd)	Chell *et al.*, 1978
P. indigofera	0.056	7.7(Mops)	Rogers and McFadden, 1976
N. crassa	0.050	6.8(Mops)	Johanson *et al.*, 1974a
P. aerugenosa	0.450	7.6(Tris)	Smith and Gunsalus, 1957
Chlorella pyrenoidosa	0.023	6.8(Tra)	John and Syrett, 1967
S. cerevisiae	1.200	6.0(P_i)	Olsen, 1959
S. lypolytica	0.220	7.0(Imd)	Matsuoka *et al.*, 1984
Citrullus vulgaris	0.250	7.5(Mops)	Jameel *et al.*, 1985
Cucumis sativus	0.039	6.8(Mops)	Lamb *et al.*, 1978
Cucumis sativus	0.100	7.4(Hepes)	Frevert and Kindl, 1978
Glycine max	1.800	8.0(P_i)	Ruchti and Widmer, 1986a
Helianthus annuus	0.667	7.5(P_i)	Zemlyanukhin *et al.*, 1984
Linum usitatissimum	0.289	7.5(P_i)	Khan *et al.*, 1977
Lupinus alba	0.035	7.0(Imd)	Vincenzini *et al.*, 1986
Pinus densiflora	0.330	7.6(Tris)	Tsukamoto *et al.*, 1986
Pinus pinea	0.080	7.0(Hepes)	Pinzauti *et al.*, 1982;1986
Ricinus communis	0.290	7.0(P_i)	Malhorta and Srivastava, 1982
C. elegans	0.540	7.4(Mops)	Colonna and McFadden, 1975
Turbatrix aceti	0.660	7.5(Tris)	Reiss and Rothstein, 1974a
Hyalomma dromedarii	2.380	7.4(P_i)	Kamel and Fahmy, 1982

enzyme, the ionizing group is on the free enzyme and not on the enzyme-substrate complex and has a pK_a of 5.8-6.0 (Rogers and McFadden, 1976).

1.4.5 Effect Of Anions On ICL:

MacKintosh and Nimmo (1988) studied the effect of anions on ICL from *E. coli* ML308 in order to try to estimate the K_m for isocitrate *in vivo*. KCl and NaCl were competitive inhibitors with respect to isocitrate, indicating that the chloride ion is probably the inhibitor. The slope replots were non-linear, concave upwards suggesting that the chloride ion interacts in more than one way at the active site. Sulphate and phosphate are linear competitive inhibitors of the cleavage reaction, indicating that they interact with a single site in the active site.

Johanson *et al.* (1974a) investigated the effects of phosphate, sulphate, nitrate, chloride and acetate on ICL from *N. crassa* and *P. indigofera*. Similar to *E. coli* ICL, phosphate and sulphate were linear competitive inhibitors with respect to isocitrate of ICL from both organisms, while chloride and nitrate were non-linear competitive inhibitors with the shape of the slope replots concave upwards.

The data from all three organisms is consistent with the masking of two adjacent sites by a divalent anion and multiple binding of monovalent anions to adjacent sites. Johanson *et al.* (1974a) suggested that the adjacent sites may coordinate Mg^{2+} which is thought to interact with glyoxylate in the active site. Another possibility is that monovalent anions interact independently with any or all of three possible binding sites for the carboxyl groups of isocitrate and that only one divalent anion can bind at any one time (MacKintosh, 1987).

1.4.6 Potential Metabolic Regulators Of ICL:

Since ICL is the first enzyme of the glyoxylate bypass, it might be expected to play an important role controlling the pathway and perhaps to be susceptible to feedback inhibition. Indeed, the simulations of LaPorte *et al.* (1984) show that changes in ICL activity can have significant effects on the partitioning of flux between the glyoxylate bypass and the TCA cycle. ICL has not been shown to be activated as yet, however, it has been shown from many

organisms to be inhibited by various metabolites and metabolic precursors. As well as the products, glyoxylate and succinate, acetate, citrate, 2–oxoglutarate, fumarate, malate and oxaloacetate, pyruvate, PEP, 3-PG, ATP and ADP all inhibit ICL. As mentioned in section 1.2.4., the most important effectors of ICDH kinase/phosphatase *in vivo* are likely to be pyruvate, PEP, NADPH, adenine nucleotides and isocitrate (Nimmo and Nimmo, 1984; El-Mansi *et al.*, 1985). It therefore seems likely that theses metabolites may also exert control of flux through the glyoxylate bypass by interacting with ICL as well as ICDH kinase/phosphatase.

It has been shown that PEP, or a metabolite close to it can control the synthesis of ICL (section 1.3.5). ICL was found to be inhibited by PEP at pH6.8 (Ashworth and Kornberg, 1963) with a K_i of 0.13mM. As a result, the authors suggested that variations in the intracellular concentration of PEP could allosterically control ICL activity. The intracellular pH in *E. coli* is pH7.3 - 7.6 (e.g. Paden *et al.*, 1981). At pH7.3, it was found that PEP was a non-competitive inhibitor and not an allosteric effector, with a K_i of 0.91mM, nearly one order of magnitude higher than at pH6.8 (MacKintosh and Nimmo, 1988). Since the intracellular concentration of PEP is thought to be 0.22mM during growth on acetate (Lowry *et al.*, 1971) and does not rise after addition of glucose, it seems unlikely that PEP has any effect on flux through ICL *in vivo*.

Glutamate, aspartate, citrate, fumarate, malate, fructose 1,6-bisphosphate, fructose-6-phosphate, ATP, ADP, AMP, GTP, $NADP^+$, NAD^+ and NADH inhibited ICL activity less than 10%, each at a concentration of 1.5mM in the presence of 0.15mM D_s-isocitrate (MacKintosh and Nimmo, 1988).

Better inhibitors were 2-oxoglutarate and 3-PG. 3-PG gave linear competitive inhibition in the cleavage reaction with a K_i of 0.8mM at pH7.3 and 2-oxoglutarate gave non-competitive inhibition with a K_i of 1.35mM at pH7.3 (MacKintosh and Nimmo, 1988). The intracellular concentration of 2-oxoglutarate during growth on acetate is 0.01mM (Lowry *et al.*, 1971) which is much lower than the K_i of 135mM for inhibition of ICL by this metabolite at pH7.3. It is therefore unlikely that 2-oxoglutarate inhibition is significant *in vivo*. LaPorte and Koshland (1983) reported the intracellular concentration of 3-PG to be 2.5mM during growth on acetate, which is much higher than the K_i of 0.8mM at pH7.3 for inhibition of ICL. A 3-PG

concentration of 2.5mM is sufficient to raise the K_m of ICL for isocitrate from 63μM to 260μM at pH7.3 and is therefore likely to be physiologically significant (MacKintosh and Nimmo, 1988).

Since 3-PG also activates ICDH phosphatase and inhibits ICDH kinase (LaPorte and Koshland, 1983; Nimmo and Nimmo, 1984), it is also possible that 3-PG is the metabolite which controls expression of the *aceA* gene (section 1.3.5). Therefore, small changes in the intracellular concentration of 3-PG could have a significant effect on the partitioning of flux at the branchpoint of the glyoxylate bypass and TCA cycle by changing the phosphorylation state of ICDH, the expression of *aceA*, and the activity of ICL.

1.4.7 K_m Of ICL For Isocitrate *in vivo*:

The intracellular pH of *E. coli* is thought to be in the range 7.3-7.6 (Paden *et al.*, 1981). Therefore, the K_m of ICL for isocitrate in intact cells is at least 63μM and probably higher due to the effects of metabolic inhibitors and anions. *E. coli* ML308 was grown in the presence of sulphate and phosphate anions. The intracellular anion concentration was not determined (MacKintosh and Nimmo, 1988), but it is known that these anions can be taken up into the cell. As mentioned above, the intracellular concentration of 3-PG is sufficient to increase the K_m to 260μM. The presence of phosphate and sulphate is likely to increase the K_m even more.

It has been shown that the flux through ICDH during growth on acetate is close to the maximal catalytic capacity of the enzyme (Walsh and Koshland, 1984). In *E. coli* ML308, during growth on acetate, ICL and ICDH have similar maximal catalytic capacities (Nimmo *et al.*, 1987). Flux through ICL is about half of that through ICDH (Walsh and Koshland, 1984; Holms, 1986). Therefore, ICL must be almost half saturated *in vivo* and so the K_m of ICL in intact cells must be close to the intracellular concentration of isocitrate during growth on acetate, which was found to be 0.57mM (El-Mansi *et al.*, 1985).

1.4.8 Phosphorylation Of ICL:

As well as regulation by 3-PG, it has recently been shown that ICL may also be regulated by phosphorylation (Hoyt and Reeves, 1988; Robertson *et al.*, 1987;1988).

Robertson *et al.* (1987) found that after incubation of partially purified ICL with [γ^{32}P]-ATP, a phosphorylated protein was found to co-migrate with purified ICL after two dimensional electrophoresis. The phosphorylated amino acid was acid labile, indicating that the amino acid was not serine, threonine or tyrosine. Pre-incubation with DEPC, a histidine modifying agent, prevented incorporation of ^{32}P (Robertson *et al.*, 1988) suggesting that the phosphorylated amino acid is a histidine or an adjacent residue. Alkaline hydrolysis, followed by anion exchange chromatography showed that the phosphorylated residue was a histidine.

The phosphorylated ICL co-migrated with catalytically active ICL after two dimensional electrophoresis (Robertson *et al.*, 1987) suggesting that the phospho-ICL was catalytically active. This does not however indicate that ICL is active only when phosphorylated. Indication that this was the case came from treatment of ICL with potato acid phosphatase (Robertson *et al.*, 1988). Incubation of ICL with acid phosphatase resulted in inactivation corresponding to an increase in the isoelectric point, consistent with a dephosphorylation event. In addition, acid phosphatase treatment of ^{32}P-ICL, labelled *in vitro*, resulted in loss of radioactivity from the band corresponding to ICL. This contradicts an earlier report that phosphorylation of ICL does not affect enzyme activity (MacDonald and Chang, 1985).

ICL was immunoprecipitated from *E. coli* $D_5H_3G_7$ extracts grown in the presence of ^{32}P inorganic phosphate. When analysed on SDS gels and Western blots, the ICL was found to be labelled, indicating that ICL is phosphorylated *in vivo* (Hoyt and Reeves, 1988).

The potato acid phosphatase experiment indicated that the active form of ICL is the phospho-ICL form (Robertson *et al.*, 1988). However, partially purified ICL, as well as being inactivated by acid phosphatase, can be labelled by incubating with ^{32}P-ATP. These results indicate that either, i) partially purified ICL contains a mixed population of ICL enzymes in different phosphorylation states, ii) proteolysis occurs duting the acid phosphatase treatment (which is carried out at pH5.0), which would account for the change in the isoelectric point, or iii) phospho-ICL is the active form but can be further phosphorylated without affecting the enzyme activity. Since only one band is present on isoelectric focusing gels, ICL is only present in only one form. Incubation of ^{32}P-ICL at pH5.0 does not result in any loss of radioactivity from the band corresponding to ICL (Robertson *et al.*, 1988), therefore the first

two possibilities seem unlikely, although the possibility that there is a contaminating protease in the phosphatase still exists. Further investigation into ICL phosphorylation is therefore required. Until it has been demonstrated that a significant proportion of ICL molecules can be phosphorylated and that the extent of phosphorylation correlates with changes in ICL activity, the case that ICL is controlled by phosphorylation remains not proven.

1.4.9 Kinetic Mechanism Of ICL From *E. coli* ML308:

Two possible kinetic mechanisms have been proposed for ICL from various organisms. I will first describe results from our group, on ICL from *E. coli* ML308, then those from other groups which are sometimes contradictory. The kinetic mechanism of ICL from *E. coli* ML308 has been well characterised (MacKintosh, 1987; MacKintosh and Nimmo, 1988). The results are summarised in table 1.4.

For the condensation (reverse) reaction, linear double reciprocal plots were obtained when either glyoxylate or succinate concentrations were varied at a fixed concentration of the other substrate. In both cases, the lines intersected on the abcissa, indicating that neither substrate affected the K_m of the other. PEP and glycolate were chosen as likely analogues of succinate and glyoxylate respectively. Glycolate was a competitive inhibitor with respect to glyoxylate and a non-competitive inhibitor with respect to succinate i.e. the K_m for succinate was not affected by the presence of glycolate. PEP was a non-competitive inhibitor with respect to glyoxylate and a competitive inhibitor with respect to succinate. These results can be explained by a random order equilibrium mechanism, where both substrates bind randomly at equilibrium to the enzyme and binding of one substrate does not affect the binding of the other.

For the cleavage reaction, succinate and PEP gave non-competitive inhibition with respect to isocitrate. This is inconsistent with a random order mechanism for which both succinate and PEP would be expected to give competitive inhibition. It was therefore postulated that a ternary enzyme-isocitrate-succinate complex could exist but not the analogous enzyme-isocitrate-glyoxylate complex (MacKintosh and Nimmo, 1988; MacKintosh, 1987) (figure 1.6). If this mechanism is correct then isocitrate may interact preferentially with ICL at the glyoxylate sub-site so that isocitrate can bind at the glyoxylate sub-site even when the succinate sub-site is occupied but not at the succinate sub-site when the glyoxylate sub-site is occupied.

Table 1.4: Kinetic Properties Of ICL From *E. coli* ML308
Taken from MacKintosh and Nimmo (1988).

Reaction	Parameter	Type Of Inhibition	Value (mM) pH7.3	pH6.8
Cleavage	K_m for isocitrate		0.063	0.032
	K_i for phosphoenolpyruvate	Non-competitive	0.91	0.10
	K_i for 3-phosphoglycerate	Competitive	0.80	0.36
	K_i for 2-oxoglutarare	Non-competitive	1.35	
	K_i for succinate	Non-competitive	1.19	0.30
Condensation	K_m for succinate		0.59	
	K_m for glyoxylate		0.13	
	K_i for phosphoenolpyruvate versus glyoxylate	Non-competitive	0.86	
	K_i for phosphoenolpyruvate versus succinate	Competitive	0.59	
	K_i for glycolate versus glyoxylate	Competitive	0.16	
	K_i for glycolate versus succinate	Non-competitive	0.26	

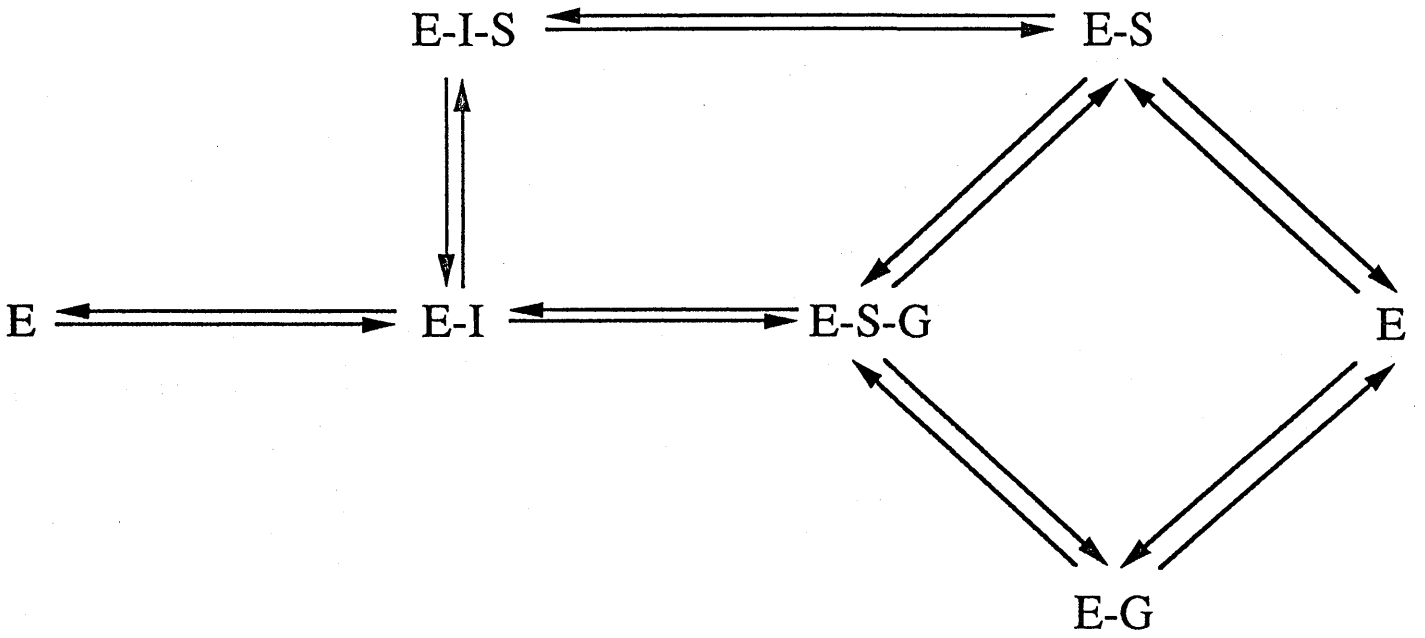


Figure 1.6: Postulated Mechanism Of Isocitrate Lyase From *E. coli* ML308

Redrawn from MacKintosh and Nimmo, 1988. E, enzyme; S, succinate; G, glyoxylate; I, isocitrate. The interconversions between E-S-G and E-I are slow relative to the binding steps.

Therefore, glyoxylate should be a competitive inhibitor with respect to isocitrate. This however was not tested because glyoxylate interferes with both assays used and unfortunately the authors did not test the effect of the glyoxylate analogue, glycolate, which like glyoxylate, should give competitive inhibition with respect to isocitrate. Glyoxylate analogues, oxalate, malonate and *meso*-tartrate all show competitive inhibition with respect to both glyoxylate and isocitrate in ICL from other sources (Jameel *et al.*, 1985; Vincenzini *et al.*, 1986; Hoyt *et al.*, 1988).

The kinetic mechanism of the *E. coli* ML308 ICL is different from that of ICL from most other sources so far, which tend to favour a compulsory order mechanism with glyoxylate binding first in the reverse reaction. In all cases, where studied, variation of one substrate at a fixed concentration of the other in the condensation reaction indicates either a random order rapid equilibrium mechanism or a compulsory order mechanism (e.g. Williams *et al.*, 1971; Johanson *et al.*, 1974a; Kahn and McFadden, 1982; Vincenzini *et al.*, 1986; MacKintosh and Nimmo, 1988; Rua *et al.*, 1990). The succinate analogue, itaconate showed linear uncompetitive inhibition with respect to both isocitrate and glyoxylate for ICL from *E. coli* $D_5H_3G_7$ (Hoyt *et al.*, 1988), *Phycomyces blakeseeanus* (Rua *et al.*, 1990), *Lupinus* (Vincenzini *et al.*, 1986) and *Pseudomonas indigofera* (Williams *et al.*, 1971), indicating that for itaconate to bind to ICL, glyoxylate has to bind first. PEP also showed linear uncompetitive inhibition with respect to isocitrate, while succinate showed either linear non-competitive or linear mixed inhibition with respect to isocitrate (Williams *et al.*, 1971; Johanson *et al.*, 1974a; Kahn and McFadden, 1982; Jameel *et al.*, 1985; Vincenzini *et al.*, 1986; Rua *et al.*, 1990). This inhibition pattern is consistent with a compulsory order mechanism with succinate leaving first in the cleavage reaction.

However, at higher concentrations of succinate, PEP or itaconate, the inhibition patterns changed. At higher itaconate concentrations, inhibition changed from linear non-competitive to uncompetitive with respect to isocitrate in the *Citrus vulgaris* enzyme (Jameel *et al.*, 1985). In the *Phycomyces blakeseeanus* enzyme, itaconate and PEP inhibition became non-competitive or mixed with respect to isocitrate (Rua *et al.*, 1990). At high PEP concentrations, inhibition of the *N. crassa* enzyme became non-competitive (Johanson *et al.*,

1974a). Also, for succinate inhibition, linearity was lost at high concentrations in the *Citrus vulgaris*, *N. crassa* and *P. indigofera* enzymes (Williams *et al.*, 1971; Johanson *et al.*, 1974a; Jameel *et al.*, 1985). Therefore the compulsory order mechanism breaks down at high concentrations. Direct binding of ^{14}C-succinate has been observed for the *P. indigofera* enzyme (Rittenhouse and McFadden, 1974) and there is competition between succinate and the reaction intermediate, 3-nitropropionate, in the absence of glyoxylate (Schloss and Cleland, 1982). Also, succinate alone affords protection of *E. coli* ICL against chemical modification by iodoacetate (Nimmo *et al.*, 1989) These results are not consistent with a compulsory order mechanism. Therefore, a technically random order equilibrium with a highly preferred compulsory order has been proposed (Williams *et al.*, 1971; Johanson *et al.*, 1974a; Rua *et al.*, 1990).

This preferred compulsory order mechanism predicts that PEP should show uncompetitive inhibition with respect to glyoxylate and that succinate should give mixed inhibition with respect to isocitrate where reciprocal plots against succinate should become non-linear in the presence of glycolate. This is clearly not the case for the *E. coli* ML308 enzyme.

The kinetic results for the binding of succinate and glyoxylate with respect to isocitrate by the *Linum usitatissimum* (flax) enzyme (Khan and McFadden, 1982) are very similar to those for the *E. coli* ML308 enzyme. Also itaconate and PEP inhibition is linear non-competitive with respect to isocitrate, implying that both dead-end inhibitors can bind to ICL in the absence of isocitrate which is consistent with a random order mechanism. Despite this evidence, the authors still favoured a compulsory order mechanism.

1.4.10 Active Site Cysteine Residue:

Studies with 3-bromopyruvate (3-BrP) indicated that there was a cysteine residue at the active site of ICL. This glyoxylate analogue has been shown to inactivate ICL from *P. indigofera* (McFadden *et al.*, 1968; Roche and McFadden, 1969; Roche *et al.*, 1971), *N. crassa* (Johanson *et al.*, 1974b), *L. usitatissimum* (Khan and McFadden, 1982) and *Citrullis vulgaris* (Jameel *et al.*, 1985) and follows saturation kinetics. 3-BrP binds reversibly to a specific site prior to alkylation resulting in carboxymethylation of one cysteine residue per

subunit and complete loss of activity. Further indication that the Cys residue modified is in the active site is shown by protection from 3-BrP inactivation by substrates and products. In the presence of Mg^{2+}, isocitrate offers competitive protection with a K_s similar to the K_m for the substrate (Roche *et al.*, 1971; Jameel *et al.*, 1985). For the *P. indigofera* enzyme, glyoxylate plus succinate but neither product on its own offers protection against 3-BrP inactivation (Roche *et al.*, 1971). The authors therefore proposed that the Cys residue is part of the succinate sub-site since the reaction mechanism requires binding of glyoxylate prior to succinate. However, the data do not rule out the possibility that 3-BrP binds at a specific site and alkylates a residue elsewhere. For the *N. crassa* and *L. usitatissum* enzymes, glyoxylate and glyoxylate plus succinate protect against 3-BrP inactivation (Johanson *et al.*, 1974b; Khan and McFadden, 1982) and the authors proposed that the Cys residue was in either the glyoxylate sub-site or both glyoxylate and succinate sub-sites.

The sequence around the active site Cys residue was identified from ICL from *E. coli* ML308 (Nimmo *et al.*, 1989). Iodoacetate inhibited ICL in a pseudo first order process resulting in incorporation of one carboxymethyl group per subunit. Isocitrate, succinate and glyoxylate all protected against inactivation with K_s values close to their respective K_m values, suggesting that the reactive group was in the active site. After trypsin digestion, the carboxymethylated peptide was purified and sequenced. The sequence of this peptide was Cys-Gly-His-Met-Gly-Gly-Lys, which corresponds to amino acids 195-201 in the *E. coli* K12 sequence (Matsuoka and McFadden, 1988; Rieul *et al.*, 1988). The sequence around the active site Cys was later confirmed for the *E. coli* K12 enzyme (Ko and McFadden, 1990).

The rate of inactivation of ICL from *E. coli* ML308 by iodoacetate is greatly reduced below pH7.0. Assuming that the protonated form of the enzyme is unreactive, the data correspond to the titration of a single ionizing group with a pK_a of 7.1 (Nimmo *et al.*, 1989). A similar result was obtained for the *E. coli* K12 enzyme with a group of pK_a 7.4 (Ko and McFadden, 1990). Using circular dichroism, no conformational change in the pH range 6.5-7.5 was detected and the pK_a of 7.1 is lower than the value expected for the sulphydryl of Cys. It therefore seems possible that there is another group in the active site whose protonation affects the accessibility of the Cys to iodoacetate, 3-BrP and isocitrate.

The presence of an essential sulphydryl group would explain why addition of

sulphydryl reagents to buffers enhanced stability of ICL during purification and storage (McFadden and Howes, 1963; Vanni *et al.*, 1979; Giachetti *et al.*, 1982; Robertson and Reeves, 1987; MacKintosh and Nimmo, 1988).With the exception of ICL from *Pinus densiflora* (Tsukamoto *et al.*, 1986), which is inhibited at concentrations of 1-4mM, all ICLs studied so far can be activated by several sulphydryl reagents such as glutathione (GSH), cysteine, β-mercaptoethanol, dithiothreitol (DTT) or 1,4-dithioerythritol (DTE) (Cionni *et al.*, 1981; Giachetti *et al.*, 1982).

1.4.11 Other Active Site Groups:

Diethylpyrocarbonate (DEPC) has been shown to inactivate ICL from *L. usitatissum*, *Citrullis vulgaris* and *Zea mays* with non-saturating kinetics (Khan and McFadden, 1982; Jameel *et al.*, 1985; Igamberdiev and Zemlyanukin, 1987). DEPC treatment resulted in an increase in A_{240}, which was abolished by treatment with hydroxylamine concominant with restoration of ICL activity, indicating that the group with which DEPC reacted with was His. For the *Citrullis vulgaris* enzyme DEPC modified five His residues per subunit, but only four in the presence of isocitrate (Jameel *et al.*, 1985). Similarly, there was only one essential His residue in the *L. usitatissum* enzyme (Khan and McFadden, 1982). Isocitrate, glyoxylate plus succinate and glyoxylate protected against inactivation by DEPC in the *Citrullis vulgaris* enzyme (Jameel *et al.*, 1985), while both products on their own protected for the *L. usitatissum* enzyme (Khan and McFadden, 1982). Removal of the C-terminal His from the *N. crassa* and *P. indigofera* enzymes resulted in complete inactivation (McFadden *et al.*, 1968; Johanson *et al.*, 1974b). It is not known whether or not the terminal residue is the same as that involved in the active site. It has been suggested that the groups with pK_a 6.1 and 6.9 in the *N. crassa* and *P. indigofera* enzymes essential for V_{max} are His residues (Rogers and McFadden, 1977).

A recent report indicates that two His residues, His266 and 306, are modified by DEPC in the *E. coli* ICL, following saturating kinetics, (Ko *et al.*, 1991). These residues are not modified in the presence of isocitrate, indicating that they are in the active site. His 266 has been previously been suggested to be the residue phosphorylated resulting in activation of the

enzyme (Matsuoka and McFadden, 1988).

Work done by our group (Rua, Robertson and Nimmo, submitted) also indicates that His306 can be modified by DEPC and that this modification results in inactivation of ICL from *E. coli* ML308. Unlike the *E. coli* K12 enzyme however, His 266 was not found to be modified by DEPC. Isocitrate, glyoxylate and succinate offer protection against DEPC inactivation, suggesting that His306 is in the active site. However, the K_s values obtained for the protection by isocitrate, glyoxylate and succinate are much higher than the K_m values and the K_s values from the protection of inactivation by iodoacetate (Nimmo *et al.*, 1989). Also His306 in the *E. coli* ICL sequence is not conserved between other ICLs although other residues in this region are highly conserved, and there is a greater effect of pH change on isocitrate binding to the active site than on protection of ICL from DEPC by this substrate. Thus, it appears that His306 is probably not involved in the catalytic mechanism of *E. coli* ICL and it seems more likely that reaction with DEPC may cause some non-specific conformational change. This does not rule out the possibility that another His residue, perhaps one close to the active site Cys195, may be involved in catalysis.

The *Citrullis vulgaris* enzyme has been shown to be inactivated by the analogue of the inhibitor itaconate, itaconate epoxide (Jameel *et al.*, 1985). Itaconate and succinate offered protection against inactivation, suggesting that the reactive group is in the active site. This inhibition is reversed by hydroxylamine, indicating that an ester linkage has been formed between itaconate epoxide and a carboxylate of an Asp or a Glu residue in the enzyme.

1.4.12 Catalysis:

Considering the chemical reaction, the cleavage of isocitrate to succinate and glyoxylate, is the reverse of an aldol cleavage. The first step in the cleavage reaction is the deprotonation of the hydroxyl group on C2 of isocitrate by the enzyme, resulting in a carbanion at C3 which can then extract a proton from the enzyme, cleaving the C2-C3 bond to give succinate and glyoxylate. In the reverse reaction, the enzyme must first extract a proton from the α-C of succinate, resulting in either a succinate carbanion or the conjugate base of succinate. This may then form a C-C bond with the keto group of glyoxylate, concominant with protonation of the keto group by the enzyme (figure 1.2). However, the α-hydrogen of

succinate is very non-acidic, with a pK_a estimated to be in the region of 24 (Walsh, 1979).

In a similar reaction catalysed by citrate lyase, the enzyme cleaves citrate by forming a S-citryl-enzyme intermediate, which stabilises the α-carbanion, reducing the pK_a of the α-hydrogen to around 14 (Walsh, 1979). ICL, however, does not proceed by this mechanism.

It is clear that ICL acts as an acid-base catalyst, although the amino acids involved in either catalysis or substrate binding have not been identified. However, chemical modification of ICL has shown that Cys195 in the *E. coli* enzyme may be involved in some way (Nimmo *et al.*, 1989; Ko and McFadden, 1990). Also, although it has been shown that neither His266 or His306 is involved, the results do not rule out the possibility that another His residue plays some role in the catalytic mechanism (Ko *et al.*, 1991; Rua, Robertson and Nimmo, submitted). Either Asp or Glu could also be involved in the catalytic mechanism (Jameel *et al.*, 1985), as well as an Arg (El-Gul and McFadden, unpublished). Jameel *et al.* (1985) suggested that the residues modified by itaconate epoxide, either a Glu or an Arg, could be involved in coordination of the Mg^{2+} ion of the Mg^{2+}-isocitrate complex due to their negative charge.

Although chemical modification has indicated that the above residues may be involved in the catalytic mechanism, it is possible that modification of the above residues either causes or conformational change in the enzyme which prevents substrate binding or that the reagents used in the modification bind to the enzyme at a specific site and modify an adjacent residue. Therefore, more information is required before the catalytic mechanism of ICL can be determined.

1.5 Aims Of This Thesis:

This introduction has described how in *E. coli* flux through the glyoxylate bypass is controlled by a reversible phosphorylation of ICDH catalysed by the bifunctional ICDH kinase/phosphatase. The K_m of ICL for isocitrate is much higher than that of ICDH and it was proposed that during growth on acetate as the sole carbon source, inactivation of ICDH permits the intracellular isocitrate concentration to rise to a level that can allow flux through the glyoxylate bypass. Much work has been carried out on ICDH, while comparatively little is known about the mechanism of ICL despite the fact that the metabolic effectors of ICDH

kinase/phosphatase are dependent on flux through this enzyme.

When I started my research, the sequence of ICL or the gene encoding this enzyme, *aceA* had not been determined from any source, although the *ace* operon had been cloned from *E. coli* ML308 in Glasgow (El-Mansi *et al.*, 1987) and work was underway to identify an active site Cys residue.

The first objective was to subclone the *aceA* gene from the plasmid containing the *ace* operon, pEM9 to enable overexpression of ICL and determination of the nucleotide sequence of *aceA*. During the course of this work, several groups determined the nucleotide sequence of this gene from several organisms, including *E. coli* K12. This allowed comparison of the derived amino acid sequences which were already thought to be similar from comparison of the amino acid compositions.

Identification of the active site Cys residue as Cys195 by chemical modification by iodoacetate, allowed site directed modification of this residue, replacing it with a Ser or an Ala. By this strategy, in conjunction with X-ray diffraction studies, it was hoped to determine whether Cys195 was an essential residue. Using kinetic studies of the mutant enzymes and comparison with the wild type enzyme, it was hoped to determine the role of Cys195 in catalysis and/or substrate binding.

Chapter 2
Materials And Methods

2.1 Materials:

General chemicals of analytical reagent grade were obtained from BDH Ltd., Poole, U.K.; Boehringer Mannheim, Lewes, U.K.; Formachem Ltd, Strathaven, U.K.; FSA Laboratory Supplies, Loughborough, U.K.; Koch-Light Ltd, Haverhill, U.K.; May and Baker, Dagenham, U.K..

Ampicillin, tetracyclin, dithiothreitol (DTT), ethidium bromide, Ficoll, Mops buffer, ribonuclease A (RNase A), benzamidine hydrochloride, bovine serum albumin (for Bradford's protein determination), Coomassie Brilliant Blue G250, DL-isocitrate (trisodium salt), lysozyme, phenylmethylsulphonyl fluoride (PMSF), Tris buffer and Triton X-100 were obtained from Sigma (London) Chemical Co., Poole, Dorset, U.K..

Acrylamide, alkaline phosphatase (calf intestine), ammonium sulphate (enzyme grade), HCl (Aristar), 2-mercaptoethanol, N,N'-methylenebisacrylamide, N,N,N',N'-tetramethylenediamine, ATP, lactate dehydrogenase (pig heart) and NADH were obtained from Boehringer Mannheim, Lewes, U.K..

Amberlite MB3, bromophenol blue, caesium chloride, iodoacetic acid, polyethylene glycol 6000 and SDS were obtained from BDH Chemicals Ltd., Poole, Dorset, U.K..

Xylene cyanol was obtained from IBI Ltd., Cambridge, U.K..

Agarose, 5-bromo-4-chloro-3-indoyl-β-galactoside (X-gal), exonuclease III (Exo III), isopropyl-β-thiogalactoside (IPTG), the Klenow fragment of *E. coli* DNA polymerase I, nuclease S1, phenol (ultrapure), all restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase and urea (ultrapure) were obtained from Gibco, BRL Ltd., Paisley, U.K..

Bactotryptone, yeast extract and bactoagar (agar) were obtained from Difco, Detroit, USA. Oxoid No.1 agar was obtained from Oxoid Ltd., London, U.K..

Oligonucleotides were synthesised on an Applied Biosystems Model 280A DNA synthesiser using phosphoramidate chemistry by Dr. Veer Math (Department Of Biochemistry).

FPLC ion-exchange (Mono Q) pre-packed columns, Sephacryl S-300 superfine and Phenyl Sepharose were obtained from Pharmacia, Milton Keynes, Bucks., U.K..

[α-^{35}S]-dATP (code SJ.304), the oligonucleotide-directed *in vitro* mutagenesis system version 2 (code No. RPN.1523) and the Amersham M13 sequencing kit (code No. N4502) were obtained from Amersham International plc., Amersham, U.K..

The Sequenase version 2.0 sequencing kit was obtained from United States Biochemical Corporation (distributed by Cambridge BioScience, Cambridge, U.K.).

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

2.2 General Biochemical Methods:

2.2.1 pH Measurements:

All pH measurements were made with a Radiometer pH probe calibrated at room temperature using standards of pH7 and pH4 prepared from tablets obtained from the manufacturer.

2.2.2 Conductivity Measurements:

Conductivity measurements were made on a Radiometer conductivity meter, type CDM2e (Radiometer, Copenhagen, Denmark).

2.2.3 Glassware And Plastics:

Glassware was washed in Haemo-sol solutions (Alfred Cox (Surgical) Ltd., U.K.), rinsed with tap water and then distilled water and dried in an oven. Plastic tips for micropipettes and Eppendorf tubes were taken from a newly opened bag.

2.2.4 Distilled Water:

Glass distilled water stored in polythene containers was used in all experiments.

2.2.5 Protein Estimations:

Protein concentrations were determined by the method of Bradford (1976). This method is based on the binding of Coomassie Brilliant Blue to the protein. The reagent was prepared by dissolving 100mg Coomassie Brilliant Blue G250 in 50ml of 95% (v/v) ethanol and 85% (w/v) orthophosphoric acid. The reagent was made up to 1 litre with distilled water and filtered. For each of the assays, a standard curve in the range 0–25μg protein was obtained using a $1mg.ml^{-1}$. BSA solution which was made up assuming that a $1mg.ml^{-1}$. BSA solution has an absorbance of 0.62 at 280nm. To each sample 2.5ml of Bradford Reagent was added, the samples vortexed and allowed to stand for 2 minutes. The absorbances were recorded at 595nm. The amount of protein present in the unknown samples was estimated from the standard curve. Protein concentrations of the fractions eluted from columns were monitored at 280nm.

2.2.6 Dialysis:

Dialysis membranes (Scientific Instruments Centre Ltd., London), were boiled for 5 min. in 1% (w/v) EDTA, pH7.0, stored in 20% ethanol and rinsed in distilled water prior to use.

2.2.7 Preparation Of Chromatographic Media:

Phenyl-Sepharose and Sephacryl S-300 superfine were obtained pre-swollen, resuspended in starting buffer and poured. Phenyl-Sepharose was regenerated, after use, by washing with 24% (v/v) ethanol and then distilled water.

All columns were poured at room temperature and packed at higher flow rates than would be used during a run, then moved to a 4°C cold room. Columns were stored in 0.02% (w/v) sodium azide and equilibrated with starting buffer before use.

2.2.8 Spectrophotometric Determination Of Nucleic Acid Concentrations:

Nucleic acid concentrations were determined spectrophotometrically at 260nm (Maniatis *et al.*, 1982). In a 1 cm path length quartz cuvette an absorbance of 1.0 corresponds to 50 $\mu g.ml^{-1}$. for double stranded DNA, 40 $\mu g.ml^{-1}$. for single stranded DNA and approximately

20 $\mu g.ml^{-1}$. for single stranded oligonucleotides.

2.2.9 Sterilisation:

Solutions for the preparation of growth media and solutions used in manipulation of nucleic acids were autoclaved in a B&T Autoclave 225E (laboratory Thermal Equipment U.K.) at 15 p.s.i. for 25 min., except carbon sources which were autoclaved at 5 p.s.i. for 50 min..

Because of their probable heat lability, the following compounds were sterilised by filtration through 0.22 μm pore-sized Millex G.V. filters (Millipore Ltd., U.K.) into sterile bottles: ampicillin, tetracyclin, methionine, vitamin B_1 and IPTG.

2.3 Microbiological Techniques:

2.3.1 Bacterial Strains, Plasmids And Phage:

Escherichia coli ML308 (ATCC15224) was originally obtained from the American Type Culture Collection (Rockville, Maryland, USA). It has been maintained in Glasgow for twenty years. It has the genetic configuration $i^-z^+y^+a^+$ for the *lac* operon. In all other respects, the strain is regarded as wild type. Because the defective repressor is produced, synthesis of the operon (β-galactosidase, β-galactosidase permease and thiogalactoside transacetylase) is constitutive and therefore, assays of β-galactosidase represent a convenient spot test indicative of *E. coli* ML308.

Escherichia coli KAT-1 is a mutant derived from *Escherichia coli* LE392 (Murray *et al.*, 1977), itself a K12 derivative, by transposon Tn10 insertion in the *aceA* gene (Maloy and Nunn, 1982). The genotype of *Escherichia coli* KAT-1 is thus F-, *hsd*R514 (Ecok r^-m^-), *sup*E44, *sup*F58, *lac*YI or (*lac*IZY)6, *gal*K2, *gal*T22, *met*BI, *trp*R55, λ-, *ace*B+, *ace*A::Tn10. *Escherichia coli* KAT-1 was obtained from W.D. Nunn (University of California, Irvine, USA).

Escherichia coli TG-1 was obtained from Amersham International plc. (Amersham, U.K.) and has the genotype Δ*lac pro*, *thi*, *sup*E, [F' *tra*D36, *pro*AB, *lac*I^qZ M15], *hsd*Δ5 (Ecok r^-m^-).

Plasmid pGLW11 is a derivative of pKK223-3 carrying ampicillin resistance, the *lac* promoter-operator region, *lacZ'* and the *lacI* gene. This plasmid was constructed in Glasgow and obtained from I.S. Hunter.

Plasmid pEM9 carries a 11kb insert, containing the *ace* operon, ligated into the Cla I site of pAT153 which also carries the genes for ampicillin and tetracycline resistance (El-Mansi *et al.*, 1987).

Plasmid pUC18 (Yanish-Perron *et al.*, 1985) carries the gene encoding ampicillin resistance and was obtained from BRL, Gibco Ltd., Paisley, U.K.

λ*c*857*s*7 DNA was obtained from BRL, Gibco Ltd., Paisley, U.K.. This DNA was digested with appropriate restriction enzymes and used exclusively as molecular weight markers. From this point onward λ*c*857*s*7 DNA is referred to simply as λ DNA.

Phage M13mp8 was obtained from Amersham International plc. Amersham U.K..

2.3.2 Storage Of Bacteria:

Organisms were stored in L-broth, containing 40% (v/v) glycerol at -20°C and at -80°C or as a stab (L-agar) at room temperature for long term storage. Short term storage was on a suitably sealed plate at 4°C or as a 10ml L-broth culture at 4°C. Samples were checked for homogeneity microscopically and periodically plated out on L-agar plates, containing 10μg.ml^{-1}. 5-bromo-4-chloro-indoxy-β-galactoside (X-gal) in the case of *E. coli* ML308, and containing 12.5μg.ml^{-1}. tetracyclin for *E. coli* KAT-1 or containing the appropriate antibiotic(s) for strains carrying plasmids.

2.3.3 Growth Media:

Unless otherwise stated, all media were sterilised by autoclaving at 15 p.s.i. for 25 min..

a) Rich Media:

L-Broth- (per litre)	Bactotryptone	10g
	Yeast extract	5g
	NaCl	10g
	(+5ml 20% (w/v) glucose)	
	pH7.5 with NaOH	
L-Agar-(per litre)	As for L-broth, +15g Difco-agar.	
2xTY-Medium- (per litre)	Bacto-tryptone	16g
	Yeast extract	10g
	NaCl	5g
H-Agar- (per litre)	Bacto-tryptone	10g
	NaCl	8g
	Bacto-agar	12g
H-Top Agar- (per litre)	Bacto-tryptone	10g
	NaCl	8g
	Bacto-agar	8g

20% (w/v) glucose was autoclaved separately at 5 p.s.i. for 50 min. and added where appropriate to prepared media.

b) Minimal Media: The minimal medium used was based on the M9 salts medium and contained per litre of distilled water:

NH_4Cl	1g
$MgSO_4.7H_2O$	0.13g
KH_2PO_4	3g
Na_2HPO_4	6g

After autoclaving, vitamin B_1 (thiamine hydrochloride; final concentration $2\mu g.ml^{-1}$.), $CaCl_2$ (final concentration 0.1mM), and glucose (final concentration $2g.l^{-1}$.) was added. The $CaCl_2$ and glucose were autoclaved together prior to the addition to the M9 salts while the vitamin B_1 was filter sterilised through a 0.22μm filter (Millipore).

For solid media, the salts were autoclaved separately from the agar to avoid precipitation problems. Hence, the M9 salts were made up as a 10 times concentration stock solution and autoclaved separately before being mixed with the agar solution. Oxoid No. 1 agar was used at 12.5 $g.l^{-1}$. final concentration. After mixing the M9 salts and the agar, the various supplements were added after cooling to 55°C.

c) Specific Supplements To Defined Media: *E. coli* KAT-1/pEM9 required addition of vitamin B_1 to a final concentration of $2\mu g.ml^{-1}$., methionine to a final concentration of $50\mu g.ml^{-1}$. and ampicillin to a final concentration of $100\mu g.ml^{-1}$..

2.3.4 Antibiotics:

Ampicillin (Amp) was used at a final concentration of $100\mu g.ml^{-1}$.. A stock solution of $25mg.ml^{-1}$. was filter sterilised and stored at -20°C. Hot L-agar was cooled to 55°C before ampicillin was added. L-Amp plates were stable for at least 4 weeks if stored at 4°C.

Tetracycline (Tet) was used at a final concentration of $12.5\mu g.ml^{-1}$.. A stock solution of $12.5mg.ml^{-1}$. in 50% (v/v) ethanol was stored in the dark at -20°C. Hot agar was cooled to 55°C before tetracycline was added. L-Tet plates were used on the day of preparation.

2.3.5 Growth Of Bacteria:

10ml of media in a 20ml universal bottle was inoculated from a loop of a single colony from a plate or from 10µl of a glycerol stock. 100ml of media in a 250ml conical flask was inoculated with 500µl of an overnight 10ml L-broth culture. For larger cultures, 500ml L-broth in a 2.5 litre conical flask was inoculated with 10ml of an overnight culture. All cultures were grown on an orbital shaker at 37°C.

2.3.6 Measurement Of Growth:

Bacterial cell density was measured as an apparent absorption at either 420nm or 600nm. The measurements were carried out in 1 cm light path cuvettes using a Unicam SP500 spectrophotometer (Pye Unicam Instruments Ltd., U.K.) equipped with a Gilford photoelectric detector and digital read out. If necessary, the culture samples were diluted 1:10 to give an absorbance of less than 0.5.

2.3.7 Harvesting Of Bacteria:

Cells were harvested by centrifugation at 8,000rpm for 15 min at 4°C (MSE 18 with an 8x250ml rotor).

2.3.8 Disruption Of Bacteria:

a) Ultrasonic Disruption: A sample of culture (2.5ml) was added to 2.5ml of chilled BSA ($10mg.ml^{-1}$.) in 0.15M NaCl in a 2-dram vial within a brass holder (Holms and Bennett, 1971) surrounded by an ice/water slurry. The sample was ultra-sonicated (Dawe Soniprobe type 1130A from Dawe Instruments Ltd., London) for three 30 sec. periods alternating with two 30 sec. cooling periods.

b) French Pressure Cell Disruption: Bacterial cell pellets were gradually resuspended in 2 volumes of 10mM potassium phosphate pH6.5, 0.5M KCl, 2mM $MgCl_2$, 1mM benzamidine, 1mM DTT, 1.2mM PMSF at 4°C and disrupted by 2 passages through a French pressure cell (Cat. No. 4-3398A, American Instruments Company, Maryland, U.S.A.) at a pressure of 98

MPa (14,300 p.s.i.). The pressure cell, capacity 40ml, was cooled on ice before use.

2.4 General Recombinant DNA Techniques:

2.4.1 Large Scale Plasmid Purification:

Both methods employed are based on the alkali lysis method of Birnboim and Doly (1979) and were routinely used to purify milligram quantities of plasmid or RF M13 DNA.

a) Using CsCl Gradients: 500 ml of L-broth, containing the appropriate antibiotic, was inoculated with 10 ml of an overnight culture of plasmid transformed cells. This culture was incubated at 37°C on an orbital shaker overnight (16hrs). The bacterial cells were harvested by centrifugation for 10 min at 7,000 rpm, 4°C (MSE 18, 6x250ml rotor). The cell pellet was washed once in GTE (50mM glucose, 25mM Tris. HCl, pH8.0, 10mM EDTA). The cells were then resuspended in 16 ml GTE. Lyzozyme was added to a final concentration of 18 $mg.ml^{-1}$. and the suspension was incubated at room temperature for 30 min. 40 ml of 0.2N NaOH, 1% (w/v) SDS were added, gently mixed by swirling and incubated on ice for 5 min. Lysis was achieved by addition of 20 ml of ice cold 5M potassium acetate, pH4.8, followed by rocking to mix and incubation on ice for 15-60 min. Bacterial cell debris was removed by centrifugation (10,000 rpm, 10 min, 4°C, MSE 18, 8x50ml rotor). The supernatant was filtered through gauze to remove any floating cell debris. The nucleic acids were precipitated by addition of 0.7 volumes of isopropanol. After incubation at room temperature for 30 min, the nucleic acid was recovered by centrifugation (Sorvall HB-4, 12,000 rpm, 30 min, 25°C). The pellet was washed with 70% (v/v) ethanol, dried briefly under vacuum and resuspended in 5 ml 50mM Tris.HCl, pH8.0, 10mM EDTA. 5.5g of CsCl and 0.5 ml ethidium bromide, $10mg.ml^{-1}$., were added and the concentration of CsCl adjusted to exactly $1g.ml^{-1}$., using a refractometer with a sodium lamp to check the concentration. The gradient was then set up by centrifugation at 55,000 rpm for 16 hrs at 25°C in a Beckman Ti70.1 rotor in polycarbonate tubes. The DNA bands were usually visible, however if the bands were not visible, long wavelength U.V. light was used to visualise the plasmid band which migrates in front of the host chromosomal DNA. The plasmid band was removed by inserting a micro capillary tube,

attached to a pump, to just below the plasmid band. Ethidium bromide was removed by extracting several times with TE (10mM Tris. HCl, pH8.0, 1mM EDTA) saturated 1-butanol. The DNA was then diluted to double the volume with distilled water and precipitated twice with ethanol (2.4.10) and resuspended in 0.2 ml TE. The DNA concentration and purity (A_{260}/A_{280}) was then measured spectrophotometrically.

b) Using QIAGEN-Tips: 100 ml of L-broth, containing the appropriate antibiotic, was inoculated with 10 ml of an overnight culture of plasmid transformed cells. This culture was incubated at 37°C on an orbital shaker overnight (16hrs). The bacterial cells were harvested by centrifugation for 10 min at 7,000 rpm, 4°C (MSE 18, 8x50ml rotor). The cell pellet was then resuspended in 4 ml 100$\mu g.ml^{-1}$. RNaseA, 50mM Tris.HCl, pH8.0, 10mM EDTA. 4 ml 200mM NaOH, 1% (w/v) SDS was added, mixed gently by rocking and incubated at room temperature for 5 min. Lysis was achieved by addition of 4 ml 2.55M potassium acetate, pH4.8. The lysate was then centrifuged at 10,000 rpm for 30 min at 4°C (MSE 18, 8x50ml rotor) The supernatant was removed promptly and centrifuged again at 10,000 rpm for 30 min at 4°C (MSE 18) to obtain a particle free lysate. The lysate was then loaded onto a QIAGEN-tip 100 equilibrated with 3 ml 750mM NaCl, 50mM Mops, 15% (v/v) ethanol, pH7.0, 0.15% (v/v) Triton X-100 and allowed to flow through the QIAGEN-tip 100 by gravity. The QIAGEN-tip 100 was then washed in 10 ml 1.0M NaCl, 50mM Mops, 15% (v/v) ethanol, pH7.0 and the plasmid DNA was eluted from the QIAGEN-tip 100 with 5 ml 1.25M NaCl, 50mM Mops, 15% (v/v) ethanol, pH8.2. The QIAGEN-tip 100 was then re-equilibrated with 3 ml 750mM NaCl, 50mM Mops, 15% (v/v) ethanol, pH7.0, 0.15% (v/v) Triton X-100 and the run-through from the first loading step re-applied. The QIAGEN-tip 100 was washed and the plasmid DNA was eluted as above. The two eluates were then combined. The plasmid DNA was precipitated by addition of 0.7 volumes of isopropanol, incubated at room temperature for 15 min and centrifuged at 10,000 rpm at 25°C for 30 min (Beckman J2-21, 8x50ml rotor). The DNA pellet was washed in 2 ml 70% (v/v) ethanol, allowed to dry for 5-15 min and resuspended in 200µl TE (10mM Tris.HCl, pH8.0, 1mM EDTA). The DNA concentration and purity (A_{260}/A_{280}) was then measured spectrophotometrically.

2.4.2 Small Scale Plasmid Purification:

The method employed for small scale plasmid and RF M13 DNA was based on the method of Birnboim and Doly (1979). 1.5 ml. of an overnight culture was harvested in a microfuge and the cell pellet resuspended in 100 µl GTE. 10 µl 10mg.ml^{-1}. Lyzozyme was added and incubated at room temperature for 5 min. 200 µl freshly made 0.2N NaOH, 1% (w/v) SDS was added, mixed gently and incubated at 4°C for 5 min.. Lysis was achieved by addition of 150 µl ice cold 5M potassium acetate, pH4.8. The lysate was incubated in ice for 5-15 min. and the cell debris removed by centrifugation in a microfuge. The supernatant was then removed into a fresh microfuge tube and the plasmid DNA was then precipitated by addition of 0.5 volumes of 7.5M ammonium acetate and 2.5 volumes of cold (-20°C) ethanol. After incubation at -80°C for 30-60 min., the plasmid DNA was recovered by centrifugation in a microfuge at 4°C (cold room) for 30 min. The pellet was washed with 200 µl 70% (v/v) ethanol, air dried for 5-15 min. and resuspended in 30 µl TE. A 5 µl sample was sufficient for digestion by restriction enzymes. During the digestion, 1µl boiled RNase A, 10mg.ml^{-1}, was added to the digest for the last 15 min. of the incubation at 37°C.

2.4.3 Digestion Of DNA With Restriction Enzymes:

The methods used were as described in Sambrook *et al.* (1989). Restriction digests were carried out using the BRL React buffers which were provided with each batch of enzyme. There are ten different React buffers with a range of salt concentrations, each one suitable for a range of enzymes. Analytical digests were carried out in a volume of 10 or 20 µl at 37°C. Preparative digests were carried out in larger volumes. When DNA was digested with two restriction enzymes, the endonuclease requiring the lower salt buffer was used first. After the recommended duration of digestion the salt concentration was adjusted and the second enzyme added.

2.4.5 Filling Recessed Termini Using Klenow Polymerase:

After digestion of the DNA with restriction enzymes, the DNA was ethanol precipitated and redissolved in 10 µl TE. 1 µl of each of the four deoxynucleoside triphosphates at concentrations of 5 mM, 2 µl 100mM Tris.HCl, pH8.0, 50mM $MgCl_2$ and 1 unit of the

Klenow fragment of *E. coli* DNA polymerase I for each µg of DNA were added. Distilled water was added to give a final volume of 20 µl and the reaction was carried out at room temperature for 15-30 min.

2.4.5 Agarose Gel Electrophoresis:

DNA was separated at room temperature on horizontal submerged agarose gels as described by Sambrook *et al.*, (1989). The Tris-borate (TBE) and Tris-acetate (TAE) buffer systems were employed. A 0.8% (w/v) agarose gel was used to size restriction fragments accurately in the range of 0.8 - 10 kb. Samples for agarose gels were prepared by addition of 0.2 volumes of 10 mM-Tris/HCl pH 7.2, 20% (w/v) ficoll, 0.5% (w/v) bromophenol blue and 10 $mg.ml^{-1}$. ethidium bromide. Ethidium bromide (0.5 $\mu g.ml^{-1}$.) was added to both gel and buffer, and stained DNA bands visualised on a long wave U.V. transilluminator (U.V. Products Inc.). Restriction fragment markers of known size were run alongside the unknown fragments, usually λ DNA digested with *Hin*dIII, or with *Eco*RI+*Hin*dIII.

2.4.6 Recovery Of DNA From Agarose Gels:

The method employed was based on the use of the GENECLEAN II kit (Bio 101 inc.). The agarose gel was made up and run in TAE buffer and the DNA band of interest was visualised by ethidum bromide staining and low energy long-wave U.V. transillumination. The desired band was excised, placed in a 1.5 ml microfuge tube and weighed. 3 volumes of saturated NaI was added and then incubated at 50°C, mixing in a vortex occasionally until the gel slice had melted. 5µl of glassmilk (Bio 101 Inc.) solution was added and incubated on ice for 5 min. The glassmilk/DNA complex was then pelleted by centrifugation in a microfuge for 10 sec. and the supernatant was removed. The pellet was then washed three times in 500 µl NEW WASH (Bio 101 Inc.), resuspending the glassmilk/DNA complex and centrifuging in a microfuge each time. The DNA was then eluted from the glassmilk by adding 10-50 µl TE and incubating at 50°C for 5 min. and removing the glassmilk by centrifugation in a microfuge for 30 sec.. For the highest recovery, the elution step was repeated and the eluates pooled.

2.4.7 Phenol/Chloroform Extraction of Nucleic Acids:

Restriction digests and other solutions containing DNA were deproteinised by extracting with phenol and chloroform. The volume of the sample was measured and adjusted to 100µl (usually) and an equal volume of TE-saturated phenol:chloroform:isoamyl alcohol (25:24:1) was added. This mixture was vortexed, allowed to stand until the two layers separated and then vortexed again. The two layers were then separated by centrifugation in a microfuge for 5 min.. The top layer was removed into a fresh tube and 100 µl TE added to the bottom, phenol layer. The mixture was vortexed and the two layers separated as in the first extraction. The two aqueous , top, layers were pooled and extracted in an identical manner with an equal volume of TE-saturated chloroform/isoamyl alcohol (24:1) which removes any residual traces of phenol. The DNA was recovered from the aqueous layer by ethanol precipitation (section 2.4.9).

2.4.8 Extraction Of DNA With Butanol:

When the volume of a restriction digest or other DNA solution was too large for ethanol precipitation in a single microfuge tube, this procedure was used to reduce the volume of the sample (Maniatis *et al.*, 1982). The volume of the DNA sample was measured and an equal volume of butan-2-ol was added and the solution vortexed. The layers were separated by centrifugation in a microfuge for 30-60 secs. The top, butan-2-ol, layer was removed and discarded and the extraction repeated until the desired volume was achieved. Traces of butan-2-ol were removed by extraction with TE-saturated ether. The top layer, containing the ether was discarded and the tube allowed to stand to remove any remaining ether by evaporation. The DNA was recovered from the aqueous layer by ethanol precipitation (section 2.4.9).

2.4.9 Ethanol Precipitation:

Sodium acetate was added to the DNA solution to a final concentration of 0.3M by adding 0.1 volumes of a 3.0M solution. 2.5 volumes of cold ethanol (-20°C) were added and mixed. The mixture was then incubated at -20°C overnight or alternatively, at -80°C for 1 hour and the precipitated DNA recovered by centrifugation in a microfuge at 4°C for 30 min. or for

larger volumes, centrifugation at 10,000 rpm for 30 min. at 4°C (MSE 18, 8x50ml rotor). The DNA pellet was washed in 70% (v/v) ethanol and resuspended in a suitable volume of TE.

An alternative to the addition of sodium acetate was to add 0.5 volumes of 7.5M ammonium acetate, however, when the DNA was to be treated with T4 polynucleotide kinase, sodium acetate was used since ammonium ions inhibit this enzyme.

2.4.10 Ligations:

The insert DNA was digested with the appropriate restriction enzymes, purified on an agarose gel and recovered using GENECLEAN. Plasmid vector DNA was linearised by digestion with the appropriate restriction enzymes and treated with bacterial alkaline phosphatase to yield DNA ready to use. This treatment prevented self-ligation of the vector and facilitated a low-background of false positive clones.

For ligation reactions where the insert size was similar to the vector size an equal concentration of foreign DNA to vector DNA was employed. For those reactions where the insert size was much greater than the vector size, at least a four-fold excess in molar concentration of foreign DNA to vector DNA was typical. Depending on the size of insert, ligation mixtures contained 25 ng vector and between 25 to 400 ng insert in a final volume between 5 to 10 µl.

Vector plus insert were pre-incubated at 45°C for 5 minutes to melt any cohesive termini that had reannealed. The ligations were performed overnight in 66 mM-Tris/HCl pH 7.6, 6.6 mM-$MgCl_2$, 0.5 mM-ATP, 10 mM-DTT at 16°C using 0.5 Units (Weiss *et al.*, 1968) of bacteriophage T4 DNA ligase.

2.4.11 Transformation Of *E. coli* With DNA:

a) Preparation Of Competent Cells: A single colony of *E.coli* KAT-1 from an L-agar plate containing tetracycline was used to inoculate 10 ml of L-broth containing tetracycline which was incubated overnight at 37°C on an orbital shaker. 1.0 ml of overnight culture was used to inoculate 100 ml of L-broth containing tetracycline which was grown at 37°C for around 2 hours to an A_{600} of 0.4 - 0.6. The cells were cooled on ice for 10 minutes and gently harvested (Beckman J2-21, 4000 rpm, 10 minutes, 8x50ml rotor, 4°C). When being prepared by the

$CaCl_2$ method (Cohen *et al.*, 1972) the cells were resuspended in one half of the original culture volume of ice-cold, sterile 50 mM-$CaCl_2$, 10 mM-Tris/HCl pH 8.0 and placed in an ice bath for 15 minutes. Following centrifugation (Beckman, J2-21, 4000 rpm, 10 minutes, 8x50ml rotor, 4°C) the cells were resuspended in 1/15th original culture volume 50 mM-$CaCl_2$, 10 mM-Tris/HCl pH 8.0 and mixed in the ratio 3:1 culture to glycerol and cells dispensed into chilled microfuge tubes in 200-400 µl aliquots. These cells were stored at -80°C.

For *E. coli* TG-1, the competent cells were not viable if stored at -80°C and the cells were prepared the day before transformation, since the greatest efficiency of transformation was achieved 24 hours after preparation (Dagert and Ehrlich, 1979).

b) Transformation Of Competent Cells:

Transformations were carried out in sterile 1.5 ml microfuge tubes. Ligation mix also refers to any DNA being used in the transforming process but was usually a ligation mixture. An aliquot of ligation mix containing up to 25 ng plasmid DNA was added to 300 µl aliquots of competent cells and the mixture was incubated on ice for at least 30 minutes. The DNA/cell mix was then heat shocked at 42°C for 2 minutes. 1 ml of L-broth was added to the tubes and they were incubated without shaking at 37°C for at least 1 hour. The cells were then plated onto an appropriately prepared antibiotic-containing L-agar plate inverted and incubated overnight at 37°C.

When transforming with M13 and M13 recombinants, 40 µl IPTG (100mM), 40µl X-gal (2% (w/v) in dimethylformamide), 200µl of fresh *E. coli* cells and the DNA/cell mix were added to 3 ml. of molten H-top agar and poured onto H-agar plates. The plates were then inverted and incubated at 37°C overnight.

2.5 Cloning In M13:

2.5.1 Maintenance Of Host Cells:

E. coli TG-1 was grown in 10 ml. 2xTY, glycerol was added to a final concentration of 50% (v/v) and cells were stored at -80°C. Before transformation with M13 DNA, the culture was plated out on a glucose/minimal medium plate and grown overnight at 37°C. This ensured

that the host cells maintained the F episome essential for infection by M13. The plate was then used for short term storage of the host cells.

2.5.2 Preparation Of Phage Stock:

A single TG-1 colony from the glucose/minimal plate was used to inoculate 10 ml. 2xTY medium which was then grown on an orbital shaker at 37°C overnight. One drop of the overnight culture was then used to inoculate 20 ml. of fresh 2xTY medium which was then grown at 37°C on an orbital shaker for 3 hours. 1 ml of 2xTY medium was inoculated with 100 µl of the 3 hour culture and with a plaque from a plate of *E. coli* TG-1 transformed with M13 or a recombinant. This culture was then grown at 37°C on an orbital shaker for 4 hours. The cells were removed by centrifugation in a microfuge. The supernatant was removed into a fresh microfuge tube and centrifuged as before. The supernatant contained the phage and was stored at -20°C. Before plating out the stock phage was diluted 1 in 10^8 and 10 µl was used for infection.

2.5.3 Preparation Of RF M13 DNA:

a) Small Scale: A host cell lawn was prepared by inoculating 10 ml. 2xTY with a single colony of TG-1 from a glucose/minimal medium plate which was then grown overnight at 37°C on an orbital shaker. One drop of the overnight culture was used to inoculate 20 ml. of 2xTY which was grown for 3-4 hours.

200 µl of this culture was infected by either 10 µl of phage stock, diluted 1 in 10^8, or 10 µl of a single plaque in 1 ml. 2xTY medium, diluted 1 in 10^4. 40 µl IPTG (100mM) and 40µl X-gal (2% (v/v) in dimethylformamide) were added and mixed with 3 ml. molten H-top agar and then poured onto H-agar plates. The plates were inverted and grown overnight at 37°C. 5 ml. 2xTY medium was inoculated with 50 µl of a fresh overnight culture of *E. coli* TG-1. A plaque was picked up, using a Pasteur pipette to ensure that the whole plaque was obtained, and transferred into the 5 ml. culture. This culture was then grown at 37°C on an orbital shaker for 5 hours. The culture was then transferred into 5 microfuge tubes (5x1ml) and centrifuged for

5 min. The RF M13 DNA was then isolated by the alkali lysis method (section 2.4.2).

b)Large Scale: 100 ml 2xTY medium were inoculated with 1 ml. of an overnight culture of *E. coli* TG-1 and grown on an orbital shaker at 37°C until the culture had an A_{550} of 0.3. 1 ml. of phage stock was added to the culture and growth was continued for a further 4 hours. The cells were harvested (MSE 18, 8,000 rpm, 10 min., 4°C, 8x50ml rotor) and the RF M13 DNA isolated using a QIAGEN-Tip 100 (section 2.4.1 b).

2.5.4 Preparation Of Single Stranded DNA:

The host cells were prepared, infected, plated out, the plaques grown and centrifuged as in section 2.5.3 a. The supernatants were transferred into fresh microfuge tubes and the centrifugation repeated. At this stage, one of the supernatants was stored at -20°C and used as a phage stock. To the other four supernatants, 200 µl PEG/NaCl (polyethyleneglycol 6000, 20% (w/v), 2.5M NaCl) was added, mixed by shaking and left to stand at room temperature for 15 min.. The phage particles were then pelleted by centrifugation in a microfuge at room temperature for 15 min.. The supernatant was discarded and the phage pellet resuspended in 100 µl TE. The single stranded DNA was then isolated by extracting with an equal volume of TE-saturated phenol:chloroform;isoamyl alcohol (25:24:1) and an equal volume of TE-saturated chloroform:isoamyl alcohol (24:1) (section 2.4.7) and recovered by ethanol precipitation (section 2.4.9).

2.6 Generation Of Nested Deletions Using Exonuclease III:

The method employed to generate a series of nested deletions was based on that devised by Henikoff (1984) and is outlined in figure 3.6.

2.6.1 Digestion With PstI And BamHI:

20 µg of pAR23 and pAR24 were digested to completion with 10 units each of *Pst*I and *Bam*HI in a total volume of 40 µl according to section 2.4.3. The DNA was then ethanol precipitated and the DNA pellet resuspended in 80 µl distilled water.

2.6.2 Digestion With Exonuclease III:

20 µl of 330mM Tris.HCl, pH8.0, 3.3mM $MgCl_2$ was added to the digested DNA and equilibrated at 37°C for 10 min..5 µl Exonuclease III (65 units. μl^{-1}.) was added and 5µl samples were removed at 30 sec. intervals into 15 µl 0.2N NaCl, 5mM EDTA, pH8.0 and incubated at 70°C for 10 min.. The DNA in each sample was then precipitated with ethanol and resuspended in 38.5 µl distilled water.

2.6.3 Removal Of Single Stranded Overhangs:

10 µl of 1.25M NaCl, 150mM potassium acetate, 25% (v/v) glycerol, 5mM $ZnSO_4$ and S1 nuclease to a final concentration of 67 Vogt units. ml^{-1}. were added to each sample and then incubated at room temperature for 30 min. (One Vogt unit will cause 1 µg of nucleic acid to become perchloric acid soluble per min. at 37°C). The reaction was stopped by addition of 6 µl 0.5M Tris.HCl, pH8.0, 0.125M EDTA and the DNA extracted with phenol and with chloroform. The DNA was recovered by ethanol precipitation and resuspended in 10 µl distilled water.

2.6.4 Ligation And Transformation:

Any remaining overhangs were removed by treatment of the DNA with the Klenow fragment of *E. coli* DNA polymerase I (section 2.4.5). The DNA was precipitated with ethanol, resuspended in 5 µl distilled water and ligated overnight at 25°C (section 2.4.10). Half the sample was used to transform competent *E. coli* TG-1, plated on H-agar and incubated at 37°C overnight.

2.6.5 Analysis Of Deletion Clones:

Each plaque was used to infect 5 ml TG-1 cultures and grown for 5 hours at 37°C. Both RF DNA and single stranded DNA were prepared (sections 2.5.3 a and 2.5.4). Clones suitable for sequencing were selected by digesting the RF DNA with *Eco*RI and *Hin*dIII and running on agarose gels.

2.7 Sequencing By The Chain Termination Method:

Sequencing described in chapter 3 was carried out using Amersham M13 sequencing kit (cat. No. N4502) and protocols described in the 'M13 Cloning and Sequencing Handbook' (Amersham International) were strictly adhered to. Sequencing described in both chapters 3 and 4 was carried out using the United States Biochemical Corporation (USB) sequencing kit and protocols described in 'Step-By-Step Protocols For Sequencing With Sequenase Version 2.0' were strictly adhered to.

The major difference between the two sequencing kits was the enzyme used to synthesise DNA complementary to the M13 template. The Amersham kit contained the Klenow fragment of *E. coli* DNA polymerase I whereas the USB kit contained Sequenase Version 2.0. Sequenase Version 2.0 is a site directed mutant of bacteriophage T7 DNA polymerase which carries no 3'-5' exonuclease activity and gives a high rate of polymerisation of nucleotides (Tabor and Richardson, 1989). The USB sequencing system was found to give clearer gels than the Amersham system that were easier to read and as a result, more sequence could be read from each reaction.

2.7.1 Sequencing With The Klenow Fragment Of *E. coli* DNA Polymerase I:

a) Annealing Primer To Template: 5 µl of single stranded template prepared as described as in section 2.5.4, was annealed for 2 hours at 55°C in a mixture also containing:

1 µl M13 sequencing primer (see below)
1.5 µl 10xKlenow Buffer (100mM Tris.HCl, pH8.5, 100mM $MgCl_2$)
2.5 µl distilled water

The M13 sequencing primer employed in this work was a 17-mer with the sequence 5'-GTA AAA CGA CGA CCA GT-3'. Annealings could be stored at -20°C for up to 1 month.

b) Sequencing Reactions: The annealed primer/template mix was thawed and 1.5 µl (15 µCi) [α^{35}S]-dATPαS at >600Ci. mmol^{-1}. (Amersham SJ.304) and 1 unit of Klenow fragment added. After thorough mixing, 2.5 µl aliquots were spotted just inside the rim of four microfuge tubes, labelled A, C, G and T. 2 µl of the relevent ddNTP/dNTP mix (see below) were added to the individual tubes and the contents mixed by a brief spin in a microfuge. The sequencing reactions were performed at room temperature and after 20 min., 2 µl of a 0.5mM uniform chase solution of all four dNTPs added in the same manner. Following the 15 min. chase, the sequencing reaction was stopped by addition of 4 µl of formamide dye mix (96% (v/v) deionised formamide, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanol, 20mM EDTA).

c) Composition Of Reaction Mixes:

i) deoxyNTP mixes (A°, C°, G° and T°) for sequencing with [α^{35}S]-dATPαS (volumes in µl):

	A°	C°	G°	T°
0.5mM dCTP	20	1	20	20
0.5mM dGTP	20	20	1	20
0.5mM dTTP	20	20	20	1
TE buffer	20	20	20	20

ii) dideoxyNTP solutions:

0.1mM ddATP	0.3mM ddGTP
0.1mM ddCTP	0.5mM ddTTP

iii) dNTP/ddNTP mixes: To each dNTP mix (N°), an equal volume of the corresponding ddNTP mix was added.

For sequencing areas which were particularly G/C rich, 2mM dITP was used instead of 0.5mM dGTP in making up the deoxyNTP mixes and 0.03mM ddGTP instead of 0.3mM when making up the dideoxyNTP mixes.

2.7.2 Sequencing With Sequenase Version 2.0:

a) Annealing Primer To Template: 7 µl of single stranded template prepared as described as in section 2.5.4, was annealed for 2 hours at 55°C in a mixture also containing:

1 µl M13 sequencing primer, -40 primer (3ng. μl^{-1}.) (see below)

2 µl 200mM Tris.HCl, pH7.5, 100mM $MgCl_2$, 250mM NaCl (5xSequenase Buffer)

The primer used was the M13 -40 primer, which has the sequence 5'-GTT TTC CCA GTC ACG AC-3'.

b) Sequencing Reactions: To the annealed primer/template mix, the following were added; 1 µl DTT (100mM), 2 µl labelling mix (1.5 µM dGTP, 1.5µM dCTP, 1.5µM dTTP), 0.5 µl (5µCi) [α^{35}S]-dATPαS at >600Ci. $mmol^{-1}$. (Amersham SJ.304) and 1.5 units of Sequenase Version 2.0. After mixing the labelling reaction was incubated at room temperature for 5 min.. During this incubation, 3.5 µl of the labelling reaction was spotted just inside the rim of four microfuge tubes labelled A, C, G, and T, in which 2.5 µl of the appropriate termination mix (80 µM dATP, 80µM dCTP, 80µM dGTP, 80µM dTTP, 8µM ddNTP (appropriate analogue), 50mM NaCl) had been spotted at the bottom of the tube. After the 5 min. incubation, the tubes were incubated at 37°C for 1 min. and the reagents mixed by a brief spin in a microfuge. After a 5 min. incubation at 37°C, 4 µl of stop mix (95% (v/v) formamide, 20mM EDTA, 0.05% (w/v) Bromophenol Blue, 0.05% (w/v) Xylene Cyanol FF) was added to the lip of each tube and mixed with the reaction by a brief spin in a microfuge. The sequencing reactions were then stored at -20°C.

For sequencing areas which were particularly G/C rich, the laballing mix contained (3.0 µM dITP, 1.5µM dCTP, 1.5µM dTTP) and the termination mixes were (80 µM dATP, 80µM

dCTP, 80μM dITP, 80μM dTTP, 8μM ddNTP (appropriate analogue), 50mM NaCl) except the ddGTP termination mix (80 μM dATP, 80μM dCTP, 160μM dITP, 80μM dTTP, 8μM ddNTP (appropriate analogue), 50mM NaCl).

2.7.3 Polyacrylamide Gel Electrophoresis:

The nested set of primer extended fragments produced by either of the M13/dideoxy DNA sequencing protocols were resolved by electrophoresis on thin polyacrylamide gels using the BRL S2 sequencing gel apparatus (BRL-Gibco).

Gels were composed of the following constituents:

40% Acrylamide	15ml.
(38% acrylamide, 2% bisacrylamide)	
10xTBE	10ml
Urea	50g
Distilled Water	35ml
10% (v/v) ammonium persulphate	350μl
TEMED	50μl
(N, N, N', N', tetramethylethylenediamine)	

Before addition of TEMED and ammonium persulphate, the gel solution was degassed. Gels were pre-run for at least 15 min. at 50mA constant current before electrophoresis in TBE. Samples were prepared by heating to 75-80°C for 2 min. before loading, during which the top of the gel was thoroughly cleared of any unpolymerised acrylamide or urea and the lanes formed using a sharkstooth comb. The heat-denatured samples were loaded immediately onto the gel and electrophoresis carried out at 60W (constant power), 50mA, 2,000V. The duration of electrophoresis depended on the length of sequence required. DNA close to the primer extending 250bp away from the primer was obtained from gels run for 2-5 hours, or until the dye front reached the anode, while sequence from 200-400bp from the primer was obtained from gels run for 8 hours. Routinely, 5 μl of each sample was run until the dye front reached

the anode and 5 µl run for 8 hours in order to obtain the maximum sequence fron each reaction.

All gels were fixed in 10% (v/v) acetic acid, 10% (v/v) methanol for 30 min. following electrophoresis and were then dried onto a sheet of Whatman 3MM paper using a Bio-Rad gel drier, model 1125. The dried gel was autoradiographed using Fuji RX film at room temperature.

2.8 Computer Programs Used In The Analysis Of DNA And Amino Acid Sequences:

A number of programs for the manipulation and analysis of DNA sequences of the UWGCG (University Of Wisconsin Genetics Computer Group) package (Devereux *et al.*, 1984) were run on a VAX computer. This package contains programs for the comparison of DNA sequences with those in GenBank and EMBL (European Molecular Biology laboratory, Heidelberg, Germany) databases. The following UWGCG programs were used:

SeqEd is an interactive editor which allows entering and modification of sequences.

Map displays both strands of a DNA sequence with a restriction map shown above the sequence and possible protein translations shown below.

BestFit makes an optimal alignment of the best segment of similarity between two sequences and inserts gaps if necessary (Smith and Waterman, 1981).

Gap finds an optimal alignment between two sequences by inserting gaps in either sequence. It considers all possible alignments, and creates the alignment with the largest number of matched bases and fewest gaps (Needleman and Wunsch, 1970).

Wordsearch searches for sequences similar to the query sequence in any group of sequences, using a Wilbur and Lipman search (Wilbur and Lipman, 1983).

Segments aligns and displays the segments of similarity found by Wordsearch.

FastA searches for sequences similar to the query sequence in any group of sequences, using a Pearson and Lipman search (Pearson and Lipman, 1988).

TFastA Does a Pearson and Lipman search as for FastA, but first translates the query sequence in all six reading frames.

Strings finds sequences by searching sequence documentation for character patterns eg citrate.

LineUp is a screen editor for editing multiple sequence alignments.

Profile calculates a table (profile) which quantitatively represents a group of aligned sequences.

ProfileSearch uses a profile, created by Profile to search databases for new sequences with similarity to the group.

ProfileSegments makes optimal alignments showing the segments of similarity found by ProfileSearch.

CodonPreference is a frame specific gene finder which tries to recognise protein coding sequences by comparison to a codon frequency table, or by their composition in the third position of each codon.

TestCode identifies protein coding sequences by plotting a measure of the non-randomness of the composition at every third base.

Translate translates nucleotide sequence into peptide sequence.

PeptideStructure makes secondary structure predictions for a peptide sequence using the methods of Chou and Fasman (1978) and Garnier *et al.* (1978).

PlotStructure plots the output of PeptideStructure.

2.9 Site Directed Mutagenesis:

The site directed mutagenesis was carried out using the Amersham Oligonuleotide-directed *in vitro* mutagenesis system version 2 (code No.RPN.1523) and the protocols described in the manual 'Oligonuleotide-directed *in vitro* mutagenesis system version 2' were strictly adhered to.

2.9.1 Preparation Of Oligonucleotides:

The oligonucleotides used for mutagenesis of aceA had the sequence GTG AAG AAA GCC GGT CAC ATG GGC G (Cys195-Ala), GTG AAG AAA TCC GGT CAC ATG GGC G (Cys195-Ser) and were supplied in a solution of 35% (w/w) NH_4OH. The oligonucleotides were precipitated with ethanol (section 2.4.9) and resuspended in TE. The oligonucleotide concentration was measured and adjusted to a final concentration of 20nmol.ml^{-1}..

2.9.2 Phosphorylation Of Oligonucleotides:

The oligonucleotides were phosphorylated with 2 units of T4 polynucleotide kinase at 37°C for 15 min. in 100mM Tris.HCl. pH8.0, 10mM $MgCl_2$, 7mM DTT and 1mM ATP. The T4 polynucleotide kinase was inactivated by heating at 70°C for 10 min.

2.9.3 Annealing Oligonucleotides To Single Stranded Templates:

5µg of single stranded template, pAR24, were mixed with 2.5µl of phosphorylated oligonucleotide, 3.5µl buffer 1 and 17µl of distilled water in a microfuge tube. The tube was heated at 70°C for 3 min, then placed in a 37°C water bath for 30 min and placed on ice.

2.9.4 Synthesis And Ligation Of Mutant DNA Strand:

To the annealed template/oligonucleotide were added, 5µl $MgCl_2$ solution, 19µl Nucleotide mix 1, 6µl distilled water, 6 units of Klenow fragment and 6 units of T4 ligase. The contents of the tube were mixed by pipetting up and down and then incubated at 16°C overnight.

170µl of distilled water and 30µl 5M NaCl were added to the reaction mixture and the excess single stranded (non-mutant) DNA was removed by centrifugation through a nitrocellulose membrane at 1500 rpm for 10 min in a swing out rotor (Sorval HB-4). 100µl of 0.5M NaCl was then spun through the membrane in order to recover any remaining RF DNA. The RF DNA was then recovered by precipitation with ethanol and resuspended in 25µl buffer 2.

2.9.5 Digestion Of The Non-Mutant Strand:

10µl of the RF DNA were incubated at 37°C with 65µl buffer 3 and 5 units *Nci*I for 90 min. 12µl 500mM NaCl, 10µl buffer 4 and 50 units of exonuclease III were added to the nicked RF DNA and the reaction mixture incubated at 37°C for 30 min. The exonuclease III was inactivated by incubation at 70°C for 15 min.

2.9.6 Repolymerisation And Ligation Of The Gapped DNA:

To the RF DNA with the exonuclease III extended NciI nicks, 13µl Nucleotide mix 2, 5µl $MgCl_2$ solution, 3 units of DNA polymerase I and T4 DNA ligase were added. The reaction mixture was incubated at 16°C for 3 hours. The RF DNA was then recovered by precipitation with ethanol, using 1 volume of 4M ammonium acetate, pH5.4 and 4 volumes of ethanol. In this way, unincorporated nucleotides were removed prior to transformation of competent *E. coli* TG-1.

2.10 Polyacrylamide Gel Electrophoresis Of Proteins:

2.10.1 Denaturing Electrophoresis:

Protein samples were analysed by discontinuous slab gel electrophoresis in the presence of SDS according to Laemmli (1970). Glass plates were separated by 1.5mm spacers and sealed with 1.0% (w/v) agarose. Gels were prepared from the volumes of stock solutions shown in table 2.1. The acrylamide, Tris buffer and distilled water were mixed and degassed; SDS, ammonium persuphate and TEMED were added and the gel was poured into the cast immediately. Isopropanol was layered onto the separating gel and rinsed off upon polymerisation. The stacking gel was then made up and poured around a well-forming template. The wells were rinsed with distilled water after polymerisation. The electrophoresis buffer consisted of 25mM Tris/192mM glycine, pH8.8, 0.1% (w/v) SDS. Electrophoresis was carried out at a constant current of 50mA.

Protein samples were mixed with at least an equal volume of sample buffer, consisting of 50mM Tris-HCl pH6.8, 1% (w/v) SDS, 10% (v/v) glycerol, 0.01% (w/v) bromophenol blue and 1% (v/v) 2-mercaptoethanol.

2.10.2 Non-denaturing Electrophoresis:

This was carried out on 7% (w/v) polyacrylamide gels at 4°C at pH8.0 (Davis, 1964). The following stock solutions were stored at 4°C:

A	36.3g	Tris
	0.23ml	TEMED
	48ml	1M HCl
	Distilled Water to 100ml	
B	28g	acrylamide
	0.74g	N,N'-methylenebisacrylamide (bis)
	Distilled Water to 100ml	

Table 2.1: Preparation Of Polyacrylamide Gels

Quantities and stock solutions used.

Stock Solution	Stacking Gel (ml)	Separating Gel (ml)		
		8%	10%	15%
30% (w/v) acrylamide/ 0.8% (w/v) bisacrylamide	1.5	8	10	15
1.5M Tris HCl pH8.8	-	7.5	7.5	7.5
0.5M Tris HCl pH6.8	3.75	-	-	-
H_2O	9.45	14.05	12.05	7.05
10% (w/v) SDS	0.15	0.3	0.3	0.3
10% ammonium persulphate (freshly prepared)	0.15	0.15	0.15	0.15
TEMED	0.01	0.01	0.01	0.01

C	30g	Tris
	144g	glycine
	Distilled Water to 1000ml	

The 'gel monomer' was prepared by mixing 1 volume A, 2 volumes B and 1 volume distilled water. An equal volume of freshly prepared ammonium persuphate (1.4mg.ml^{-1}.) was added. Solutions were degassed separately and then mixed and poured into gel tubes (dimensions 0.5x10cm). 9cm gels were poured, carefully overlayed with distilled water and allowed to set. The electrophoresis buffer was solution C which had been diluted 25-fold with distilled water at 4°C containing 0.1% (v/v) 2-mercaptoethanol. The gels were pre-electrophoresed for 30 min at 3mA per tube at 4°C. 5µl of tracking dye (0.01% (w/v) bromophenol blue in 20% (v/v) glycerol) was layered onto the top of each gel and electrophoresis was carried out until the dye had penetrated the gel. The samples (less than 50µl in 20% (v/v) glycerol) were loaded onto the gels and electrophoresis was performed at 3mA per tube at 4°C until the tracking dye approached the bottom of the gel. The dye front was marked on each gel by inserting a piece of wire through the gel.

2.10.3 Staining Gels For Protein:

Routinely, the gels were stained in 0.1% (w/v) Coomassie Brilliant Blue G250, 50% (v/v) methanol, 10% (v/v) acetic acid for 1 hour at 45°C and destained in several changes of 10% (v/v) methanol, 10% (v/v) acetic acid at 45°C.

2.10.4 M_r Determination By SDS-PAGE:

M_r values were obtained from SDS gels, after Coomassie Blue staining by comparing the electrophoretic mobility (R_f) of the unknown with a standard curve produced by running a mixture of proteins of known M_r on the same gel. The mixture of proteins used to produce the standard curve was obtained from Sigma Chemical Co. and consisted of bovine albumin (66,000), egg albumin (45,000), rabbit muscle glyceraldehyde-3-phosphate dehydrogenase (36,000), carbonic anhydrase from bovine erythrocytes (29,000), trypsinogen from bovine

pancreas (24,000) trypsin inhibitor from soybean (20,100) and α-lactalbumin from bovine milk (14,200).

2.11 Purification Of ICL From *E. coli* KAT-1/pAR9023, KAT-1/pAR654 And KAT-1/pAR655:

Two methods were employed to purify ICL from KAT-1/pAR9023. The first method was based on that developed for the purification of ICL from *E. coli* ML308 (MacKintosh and Nimmo, 1988), while the second method which was also used to purify ICLSer195 and ICLAla195 from *E. coli* KAT-1/pAR654 And KAT-1/pAR655 respectively was based on that of Conder *et al.* (1988). All steps were carried out at 4°C except the FPLC ion-exchange on Mono Q, which was carried out at room temperature.

2.11.1 Purification Of Wild Type ICL From *E. coli* KAT-1/pAR9023:

a) Purification Buffers:

Buffer A:	10mM potassium phosphate pH6.5, 0.5M KCl, 2mM $MgCl_2$, 1mM benzamidine, 1mM DTT, 1.2mM PMSF.
Buffer B_1:	50mM Mops-NaOH pH7.3, 1mM EDTA, 1mM benzamidine, 1mM DTT, 1.2mM PMSF
Buffer B_2:	Buffer B_1 without PMSF.
Buffer B_3:	Buffer B_1 with 40% (v/v) glycerol and 5mM $MgCl_2$.

b) Preparation Of Crude Extract: 2l (4 x 2.5l flasks, each containing 500ml medium) of L-broth containing ampicillin and tetracycline were inoculated with 10ml of an overnight culture of *E. coli* KAT-1/pAR9023 and grown on a shaker at 37°C for 2-3 hours. IPTG was added to a final concentration of 1mM and growth continued overnight. After harvesting, the cells (10-15g) were resuspended in 2 volumes of Buffer A and passed through the French Pressure cell three times. The resulting material was centrifuged at 15,000 rpm for 20 min. at 4°C (MSE 18 with an 8x50ml rotor) and the supernatant was decanted through glass wool to give the crude extract.

c) Ammonium Sulphate Fractionation: Powdered $(NH_4)_2SO_4$ was added gradually to give 30% saturation ($175g.l^{-1}$.), while the pH of the solution was maintained in the range pH 6.5-7.3 by the addition of 5M NH_3. The solution was stirred for 30 min and centrifuged at 18,000 rpm for 10 min at 4°C (MSE 18, 8x50ml rotor) and the pellet was discarded. $(NH_4)_2SO_4$ was added to the supernatant to give 45% saturation (final concentration of $278g.l^{-1}$.) and the solution was stirred and centrifuged as before. The pellet was dissolved in a small volume of Buffer B_1.

d) Gel Filtration On Sephacryl S-300 Superfine: The resuspended pellet from the 30-45% saturated $(NH_4)_2SO_4$ fraction was loaded onto a 2.2cm x 71cm column of Sephacryl S-300 superfine, equilibrated in Buffer B_1. The flow rate was $10ml.hour^{-1}$ and 3ml fractions were collected. Fractions containing ICL activity, up to and including the fraction after the peak, were pooled. Later ICL-containing fractions were discarded in order to eliminate a contaminating band of M_r 80,000 (MacKintosh and Nimmo, 1988).

e) Phenyl-Sepharose Chromotography: Powdered $(NH_4)_2SO_4$ was added to the pooled fractions from the gel filtration on Sephacryl S-300 superfine to bring the concentration to 0.6M. The solution was then loaded onto a 20ml column of phenyl-Sepharose equilibrated in 0.6M $(NH_4)_2SO_4$ in Buffer B_1. The column was washed in this buffer, then a gradient (total volume 150ml) of 0.6M to 0M $(NH_4)_2SO_4$ in Buffer B_1 was applied. The flow rate was $30ml.hour^{-1}$ and 2.5ml fractions were collected. The enzyme pool from this step was dialysed for 4 hours against 2 x 2l Buffer B_2.

f) FPLC Ion-exchange Chromotography On Mono Q: This was carried out at room temperature. One quarter of the enzyme pool from the phenyl-Sepharose chromotography was applied to a Mono Q column equilibrated with Buffer B_2. The enzyme was washed in this buffer and eluted with a gradient of 0M to 0.5M NaCl in Buffer B_2. The flow rate was $3.0ml.min^{-1}$ and 0.5ml fractions were collected. After 3 more runs, the purity was assessed with SDS-PAGE and homogeneous material was pooled.

g) Storage: Pooled enzyme was dialysed overnight into Buffer B_3 and stored at -20°C.

2.11.2 Purification Of ICL, ICLSer195 And ICLAla195 From *E. coli* KAT-1/pAR9023, KAT-1/pAR654 And KAT-1/pAR655 Respectively:

a) Purification Buffers:

Buffer C:	50mM triethanolamine pH7.3, 5mM $MgCl_2$, 1mM EDTA.
Buffer D:	10mM triethanolamine pH7.3, 5mM $MgCl_2$, 1mM EDTA, 50mM NaCl, 1mM 2-mercaptoethanol.
Buffer E:	25mM triethanolamine pH7.3, 5mM $MgCl_2$, 1mM EDTA, 450mM NaCl, 1mM 2-mercaptoethanol.

b) Preparation Of Crude Extract: Crude extracts of *E. coli* KAT-1/pAR9023, KAT-1/pAR654 and KAT-1/pAR655 were prepared as in section 2.11.1b) except that the cells were resuspended in 2 volumes of Buffer C.

c) Ammonium Sulphate Fractionation: This was carried out as described in section 2.11.1c) except that the 30-45% pellet was resuspended in a small volume of Buffer D and dialysed against this buffer until the conductivity was equal to that of Buffer D.

d) FPLC Ion-exchange Chromotography On Mono Q: This was carried out at room temperature. One quarter of the dialysed enzyme pool from step c) was applied to a Mono Q column equilibrated with Buffer D. The enzyme was washed in this buffer and eluted with a gradient of 100-0% Buffer D and 0-100% Buffer E. The flow rate was 3.0ml.min^{-1}. and 0.5ml fractions were collected. After 3 more runs, the purity was assessed with SDS-PAGE and homogeneous material was pooled.

e) Storage: Pooled enzyme was dialysed overnight into Buffer B_3 (section 2.10.1a) and stored at -20°C.

2.12 Isocitrate Lyase Assays:

a) Method 1: In steady-state kinetic experiments ICL was assayed, at 340nm and 37°C, by coupling the formation of glyoxylate to the oxidation of NADH with lactate dehydrogenase (El-Mansi *et al.*, 1987). Each cuvette contained in a final volume of 1ml, 50mM Mops-NaOH pH7.3, 5mM $MgCl_2$, 1mM EDTA, 5mM DL-isocitrate, 0.2mM NADH, 0.1mg.ml^{-1} pig heart lactate dehydrogenase and ICL extract. The reaction was initiated by the addition of enzyme.

b) Method 2: In crude extracts and sonicated extracts, ICL was assayed using a modification of the procedure of Dixon and Kornberg (1959). Each cuvette contained in a final volume of 1ml, 50mM Mops-NaOH pH7.3, 5mM $MgCl_2$, 1mM EDTA, 5mM DL-isocitrate, 4mM phenylhydrazine-HCl. The reaction was started by the addition of enzyme and the change in A_{334} was monitored. The phenylhydrazine-HCl was stored at -20°C in the dark.

Each of the above assay methods are standard methods, the linearity of rate dependent on enzyme concentration was checked and the maximum amount of enzyme used in the assays was set so that the change in absorbance was not greater than 0.2/min for a 1ml assay. In some steady state kinetic experiments, the DL-isocitrate concentration was varied (in the range 0.025-10mM) and also the pH of the assay buffer 50mM Mops-NaOH (pH6.8-7.8), 5mM $MgCl_2$, 1mM EDTA or 50mM Taps-NaOH (pH7.8-9.0), 5mM $MgCl_2$, 1mM EDTA.

2.13 Reaction Of ICL, ICLSer195 and ICLAla195 With Iodoacetate:

In the standard conditions, wild type ICL, ICLSer195 or ICLAla195 at a concentration of 1mg.ml^{-1}. was incubated in 50mM Mops-NaOH pH7.3, 1mM EDTA at 25°C with 2.0mM iodoacetate and other additions as indicated in individual experiments. Samples were assayed for ICL activity at various times after addition of iodoacetate.

In substrate protection experiments, 0.5mM DL-isocitrate, 5mM $MgCl_2$, or both were included in the incubation mixture.

2.14 Fluorimetry:

The analysis was carried out at the Department of Biological and Molecular Sciences, University Of Stirling by N. Price and H.G. Nimmo. ICL at a concentration of 0.3mg.ml^{-1}. in 50mM Mops-NaOH pH7.3, 5mM $MgCl_2$, 1mM EDTA was excited at 290nm and the intrinsic emission spectra measured at 20°C on a Perkin-Elmer fluorescence spectrophotometer, model L5 50 with slit widths of 2.5nm.

2.15 Circular Dichroism:

Circular dichroism analysis of wild type ICL, ICLSer195 and ICLAla195 was carried out in 10mM Mops-NaOH pH7.3 at a protein concentration of 0.3mg.ml^{-1}. with slit width 0.02cm and 0.5cm for the far U.V. (190-150nm) and near U.V. (260-320nm) respectively. The analysis was carried out at the Department of Biological and Molecular Sciences, University Of Stirling by N. Price, S. Kelly and H.G. Nimmo on a Jobin-Yvon dichrograph III equipped with a temperature controlled cell holder. Ellipticity values (θ) (mean residue molar ellipticity in deg.cm^2.decimole^{-1}) were calculated using the following equation:

$$\theta = \frac{\text{m.r.w. x sensitivity x displacement x 3300}}{\text{concentration x pathlength}}$$

where m.r.w. is the mean residue weight.

displacement is the difference between sample and reference at any given wavelength.

concentration = 0.3mg.ml^{-1}.

pathlength = 1cm.

Chapter 3
Cloning And
Sequencing Of *aceA*

3.1 Introduction:

Recombinant DNA technology is playing an increasing role in the study of enzyme structure, function and regulation of metabolic pathways. For example, using an adjustable expression vector, it is possible to evaluate the contribution of an individual enzyme to the control of flux through a metabolic pathway. This was done for citrate synthase in *E. coli* growing on different carbon sources (Walsh and Koshland, 1985). Cloning and overexpression of the appropriate genes also facilitates the purification of a desired protein in large amounts. It is now much quicker and easier to obtain sequence information at the nucleotide level than at the protein level. Recombinant DNA technology has also been of value in the identification of multi-gene families. Site-directed mutagenesis and protein engineering has been used in the identification of specific residues and domains important for the activity or regulation of an enzyme.

At the outset of this project, the kinetic properties of ICL from *E. coli* had been studied in great detail (e.g. MacKintosh and Nimmo, 1988; Hoyt *et al.*, 1988), see section 1.4. The *ace*-operon had been cloned from *E. coli* ML308 (El-Mansi *et al.*, 1987) and ICL had been overexpressed. During the early stages of the project work on determination of the three-dimensional structure of this enzyme was intitiated (Abeysinghe *et al.*, 1991) and chemical modification experiments, using iodoacetate, allowed identification of a cysteine residue at the active site (Nimmo *et al.*, 1989). Another group identified the same residue using 3-BrP to modify ICL (Ko and McFadden, 1990). This created the opportunity for site directed mutagenesis of ICL. Our objectives were therefore to sequence *aceA* in order to facilitate the resolution of the three dimensional structure and to use site directed mutagenesis to investigate the role of this cysteine residue in enzyme activity.

During the course of this work, both Cozzone's group (Rieul *et al.*, 1988) and MacFadden's group (Matsuoka and MacFadden, 1988) determined the sequence of the *aceA* gene from *E. coli* K12. There were however, differences in the two sequences and it was hoped that by sequencing the *aceA* gene from *E. coli* ML308, these differences could be resolved.

3.2 Cloning Of The *aceA* Gene From *E. coli* ML308:

3.2.1 Subcloning *aceA* From pEM9:

In order to facilitate the isolation of large amounts of the enzymes of the glyoxylate bypass and ICDH kinase/phosphatase and to study the control of expression of the *ace* operon, this operon was cloned from *E. coli* ML308 (El-Mansi *et al.*, 1987). Plasmid pCTS, which contained the *ace* operon of *E. coli* K12, was used as a probe in conjunction with complementation, to clone the *ace* operon into pAT153. The resulting clone, pEM9 (figure 3.1), was shown to contain all three genes of the *ace* operon. *E. coli* KAT-1/pEM901, a *Cla*I-*Hin*dIII subclone of pEM9 (figure 3.1), could not grow on acetate. When grown on glycerol and challenged with acetate, KAT-1/pEM901 expressed ICL and MSA activities but not ICDH kinase/phosphatase. Unlike pEM9, pEM901 did not show any proteins which co-migrated with ICDH kinase/phosphatase on SDS gels after *in vitro* transcription/translation, but it did direct the synthesis of proteins which co-migrated with ICL and MSA. These results were taken to suggest that pEM9 encodes all three structural genes of the *ace* operon and that the *Hin*dIII site in pEM901 lies either within or upstream of *aceK*, within the *aceA-aceK* intercistronic region.

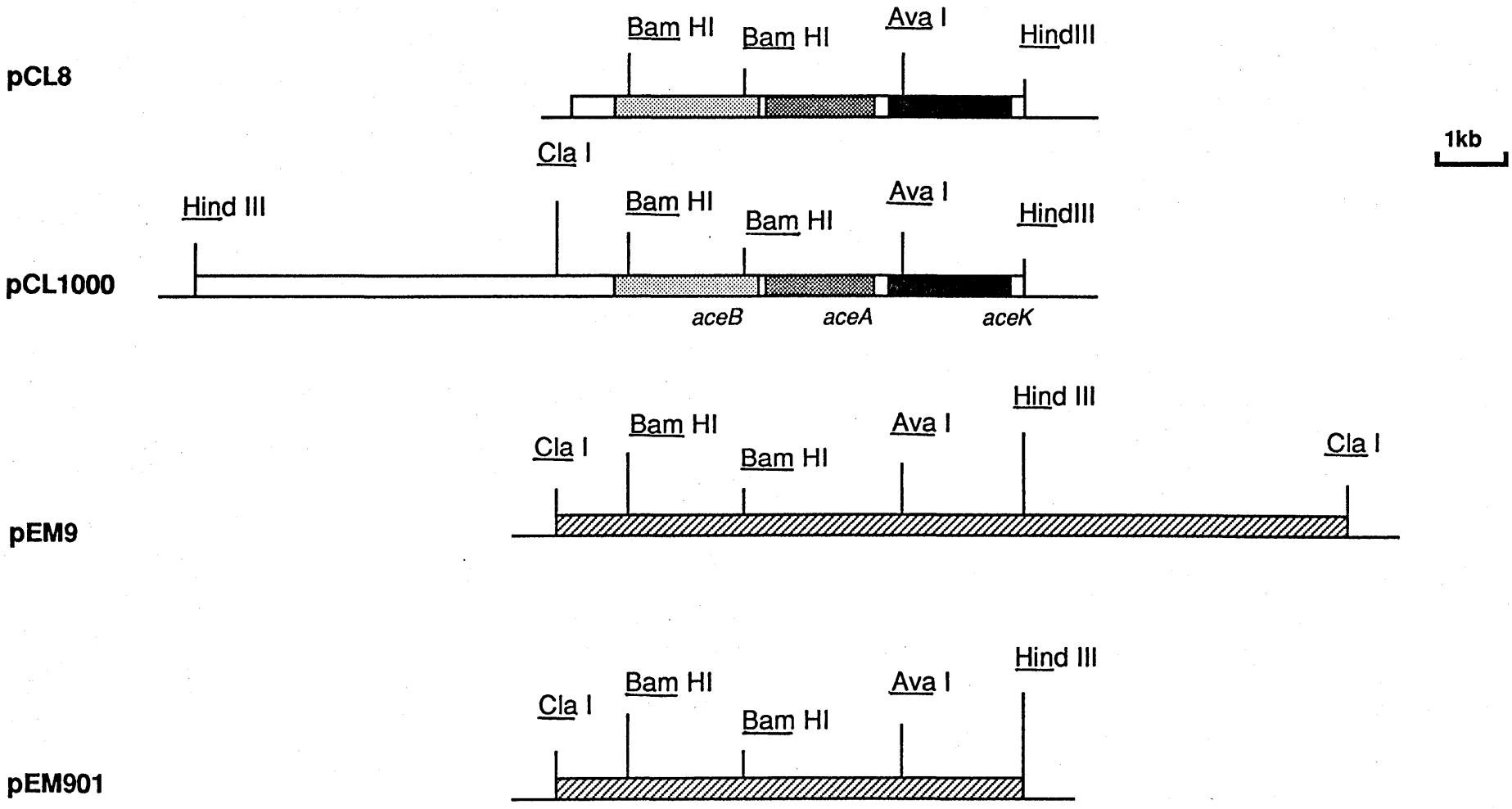
Shortly after the start of this project however, LaPorte's group reported both the sequence of *aceK* (Klumpp *et al.*, 1988) and cloning and analysis of the *ace* operon from *E. coli* K12. Using the derived restriction map, they determined the position of *aceK* on their plasmid, pCL1000. They determined the functional map by deleting pCL1000 from one end in conjunction with enzyme assays, SDS gel electrophoresis of products of maxicell labelling and complementation. Their conclusions are summarised in figure 3.1. By comparing the restriction map of pEM9 with that of pCL1000, it is clear that contrary to the assumption that the *Hin*dIII site is upstream or within *aceK*, this gene is within the *Cla*I-*Hin*dIII fragment.

Comparison of the restriction maps of pEM9 and pCL1000 revealed a high degree of similarity over the region encoding the operon in pCL1000 (figure 3.1). Since both plasmids were isolated from *E. coli* and the restriction maps show this high degree of similarity, it seemed probable that the functional maps may also be very similar.

The *aceA* gene lies within a 2.3kb *Bam*HI-*Ava*I region of pCL1000. Due to the high

Figure 3.1: Comparison of restriction maps of clones encoding the complete *ace* operon

The restriction and functional maps of pCL1000 and pCL8 were redrawn from Chung *et al.* (1988) and the restriction maps of pEM9 and pEM901 were redrawn from El-Mansi *et al.* (1987).


pCL8
Bam HI
Bam HI
Ava I
HindIII
1kb
Cla I
pCL1000
Hind III
Bam HI
Bam HI
Ava I
HindIII
aceB
aceA
aceK
pEM9
Hind III
Cla I
Bam HI
Bam HI
Ava I
Cla I
pEM901
Hind III
Cla I
Bam HI
Bam HI
Ava I

degree of similarity between the restriction maps of pEM9 and pCL1000, it seemed likely that *aceA* from pEM9 is encoded in the corresponding *Bam*HI-*Ava*I fragment. This fragment was therefore subcloned from pEM9.

In order to express ICL from the cloned *aceA* gene, an expression vector was essential since it is the second gene in the operon and as such is transcribed from the promoter upstream of *aceB*. The expression vector chosen was pGLW11, a derivative of pKK223-3. pGLW11 contains the hybrid *tac* promoter (*ptac*) (DeBoer *et al.*, 1983), derived from the *trp* and *lac* promoters and regulated by the *lac* repressor, upstream of a multiple cloning site. Also encoded on pGLW11 is the gene for ampicillin resistance and the *lacI* gene (Calos, 1978). Expression of genes under the control of *ptac* is achieved by addition of isopropyl β-thiogalactopyranoside (IPTG) to the culture medium.

The sequencing strategy which was to be adopted required the generation of a series of nested deletions using exonuclease III (Exo III). In order to determine the sequence of both strands, these deletions had to be generated on both strands.There are two ways by which the sequence of both strands of an insert can be determined, i) cloning the insert into both M13mp8 and its "sister" vector, M13mp9, which contains the same multiple cloning site but in the opposite orientation with respect to the primer or ii) by cloning the insert in both orientations in the same vector, M13mp8. Exo III cannot initiate digestion from DNA which has a 3' overhang of four or more bases. The only restriction enzyme site in the multiple cloning site of the sequencing vector, M13mp8, which generates this type of overhang is *Pst*I. The insert therefore had to be cloned in opposite orientations with respect to the *Pst*I site. This could not be achieved by the first method, hence *aceA* was cloned in both orientations.

pEM9 was digested with *Bam*HI and *Ava*I. The fragments were then blunt-ended by filling in the overhangs using the Klenow fragment of DNA polymerase and a mixture of all four dNTPs. The 2.3kb fragment was then purified from an agarose gel and ligated with pGLW11 which had been digested with *Sma*I. The ligation mix was then used to transform competent *E. coli* KAT-1 and the cells plated out onto L-agar containing ampicillin. Transformants were screened by assaying for ICL activity in sonicated extracts after addition of IPTG and glucose to overnight cultures. The restriction map of the plasmid from one of the

positive transformants, designated, pAR9023 is shown in figure 3.2. Other transformants were shown to contain plasmids with 2.3kb inserts but could not express any ICL activity. The restriction map of one of these plasmids is shown in figure 3.2. When compared with the restriction map of pAR9023, it is clear that this plasmid contains the same insert in the opposite direction. Comparison of the restriction map of pAR9023 with a more detailed map of pEM9 indicates that *aceA* is transcribed in the direction from *Bam*HI to *Ava*I. This is consistent with the functional map of pCL1000. Thus *aceA* is found in the same fragment and transcribed in the same direction in both pEM9 and pCL1000, and this is a further indication that the functional maps of these plasmids may be similar.

3.2.2 Expression Of ICL From *E. coli* KAT-1/pAR9023:

One of the project aims was to carry out site directed mutagenesis of ICL and characterize the mutant enzymes. In order to do this, the enzymes have to be expressed and purified. A long-term aim was to express the wild type and mutant enzymes at different levels and study the effect of controlled expression on the partitioning of flux at the branchpoint between the TCA cycle and the glyoxylate bypass.

In order to optimize the conditions for expression, the effects of [IPTG] and time after addition of IPTG were studied. IPTG was added at different concentrations to cultures of *E. coli* KAT-1/pAR9023 growing on L-broth containing ampicillin in log phase. Samples were taken at appropriate time intervals, the absorbance was measured and the cells were harvested. Sonicated extracts were assayed for ICL activity. Figure 3.3 shows the effect of time after addition of IPTG on ICL activity and A_{600}. There is a linear increase in the ICL activity with time for a given IPTG concentration. As the IPTG concentration increases, the rate of increase of ICL activity also increases. This is due to an increase in the rate of ICL synthesis rather than an increase in the growth rate of cultures with higher IPTG concentrations since the growth rate decreases with increased IPTG concentration (figure 3.3B). From these results, it appears that maximal expression of ICL occurs at an IPTG concentration of 1mM.

The rate of increase in specific activity increases with IPTG concentration (figure 3.4), therefore the level of ICL activity cannot be regulated precisely by adjusting the IPTG concentration since for any given IPTG concentration, the specific activity increases with time.

Figure 3.2: Restriction maps of pAR9023 and pAR9024

The shaded boxes represent the 2.3kb insert with the arrow indicating the direction of transcription. The distances between the restriction sites are indicated. There are 4 PvuII sites. The exact position of the middle 2 sites are unknown and the possibilities are indicated by the thin lines. The direction of transcription from the *tac* promoter is indicated by a thin arrow.

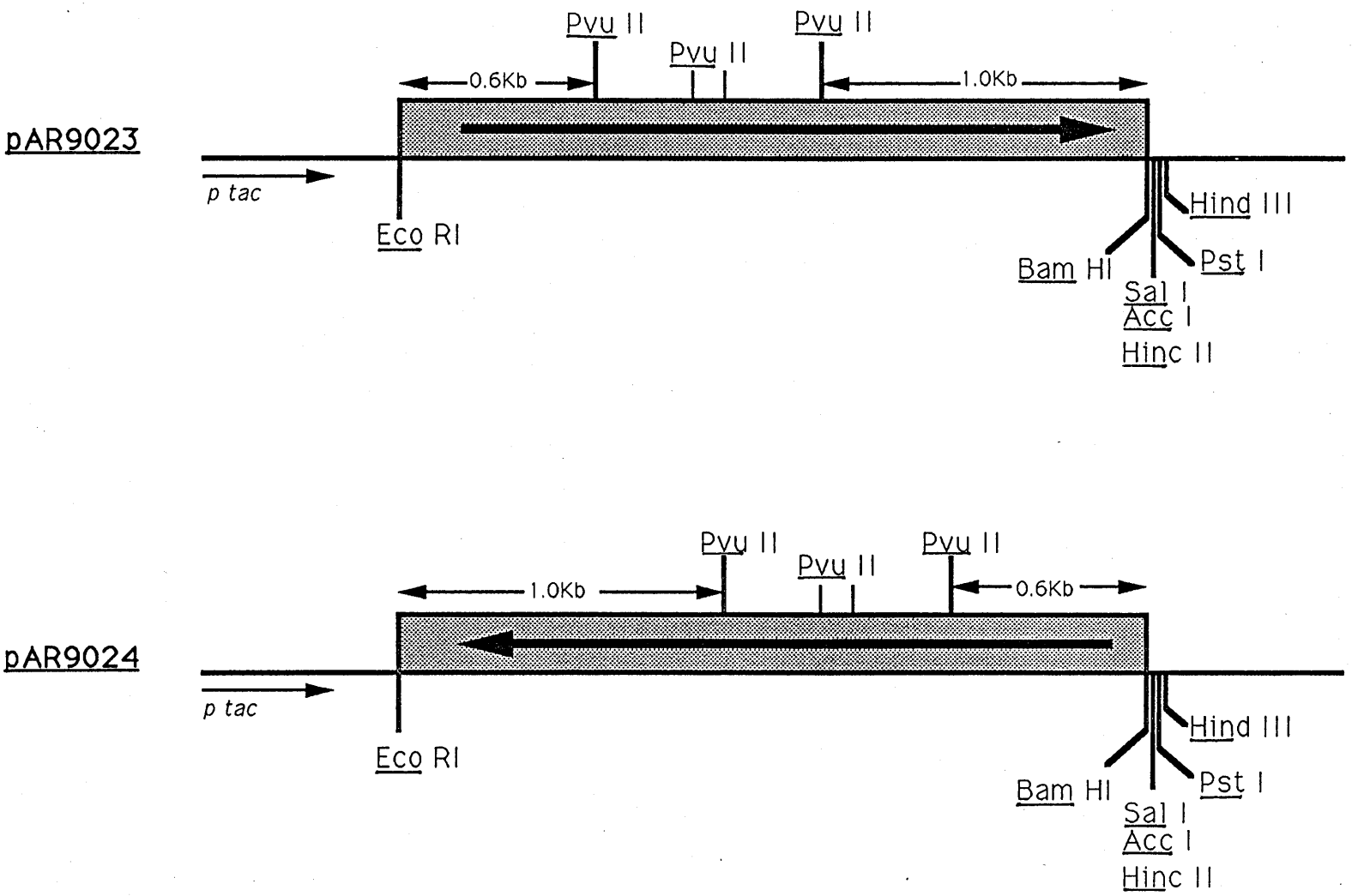

pAR9023
Pvu II
Pvu II
Pvu II
0.6Kb
1.0Kb
p tac
Eco RI
Hind III
Bam HI
Pst I
Sal I
Acc I
Hinc II
pAR9024
Pvu II
Pvu II
Pvu II
1.0Kb
0.6Kb
p tac
Eco RI
Hind III
Bam HI
Pst I
Sal I
Acc I
Hinc II

Figure 3.3: Effect of IPTG concentration on growth and expression of ICL activity from *E.coli* KAT-1/pAR9023.

E.coli KAT-1/pAR9023 was grown in 100ml of L-broth containing 100μg.ml^{-1}. ampicillin. At an A_{600} of approximately 1.0, IPTG was added to various concentrations as indicated. Samples were taken at time intervals and the A_{600} was measured (B). The samples were then sonicated and assayed for ICL activity (A).

- ○ 0 mM IPTG
- △ 0.25 mM IPTG
- ▲ 0.50 mM IPTG
- ◇ 0.75mM IPTG
- ● 1.00 mM IPTG

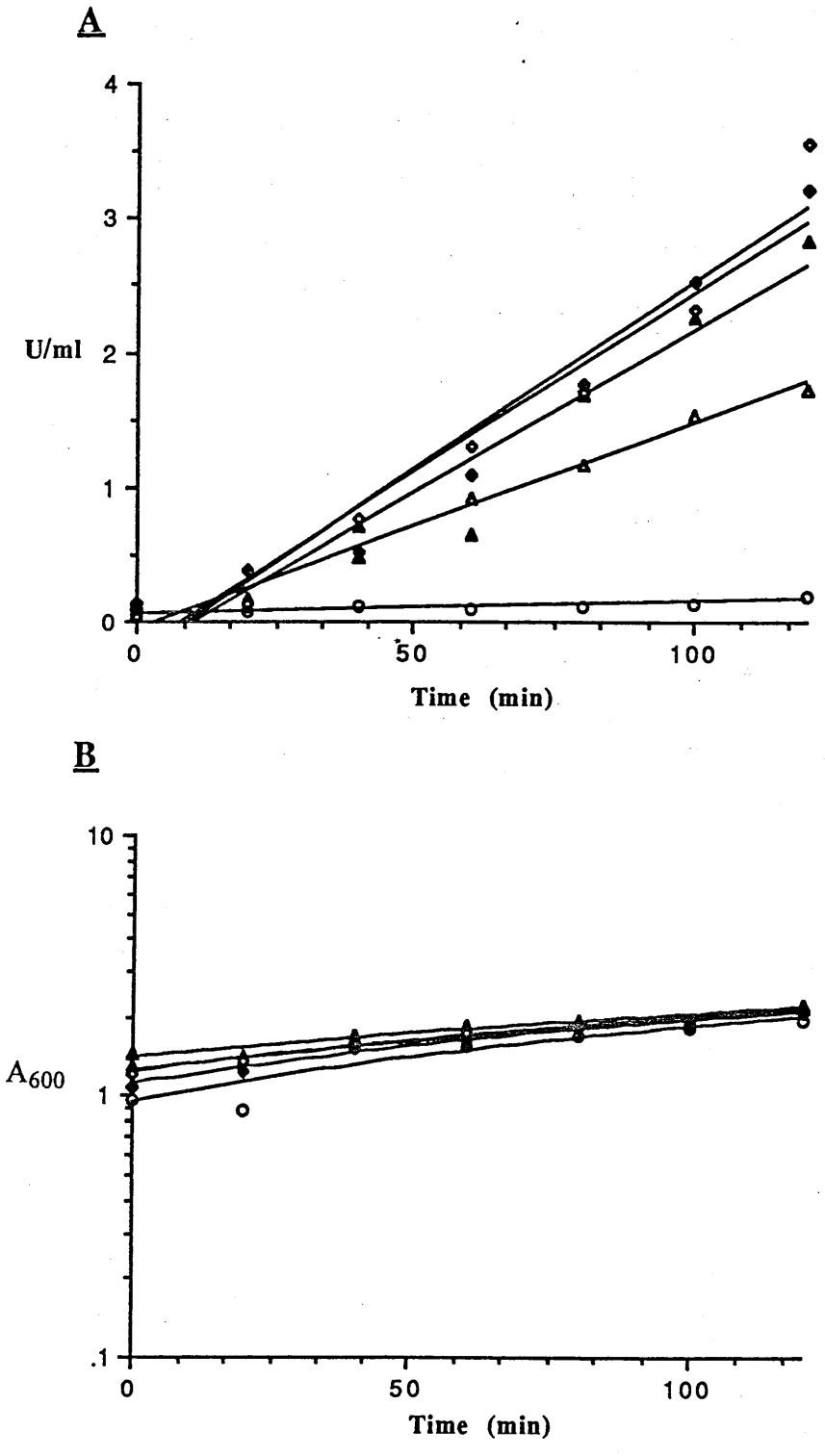

A
4
3
U/ml 2
1
0
0
50
100
Time (min)
B
10
A600
1
.1
0
50
100
Time (min)

Figure 3.4: Effect of IPTG concentration on the rate of increase in specific activity of ICL in *E.coli* KAT-1/pAR9023.

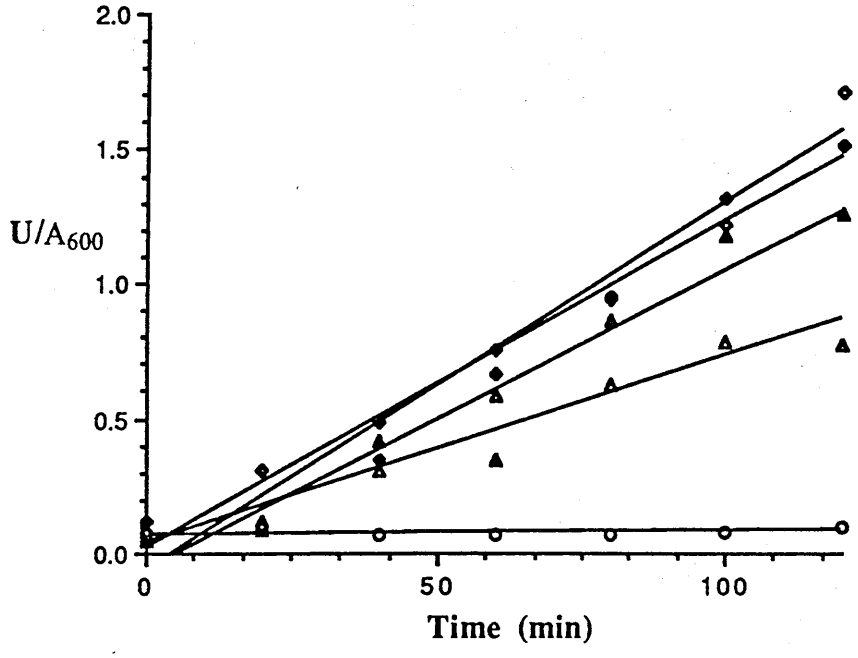
E.coli KAT-1/pAR9023 was grown in 100ml of L-broth containing 100μg.ml^{-1}. ampicillin. At an A_{600} of approximately 1.0, IPTG was added to various concentrations. Samples were taken at time intervals and the A_{600} was measured. The samples were then sonicated and assayed for ICL activity and the specific activity calculated.

(A) shows a plot of increase in specific activity with respect to time at the IPTG concentrations indicated:

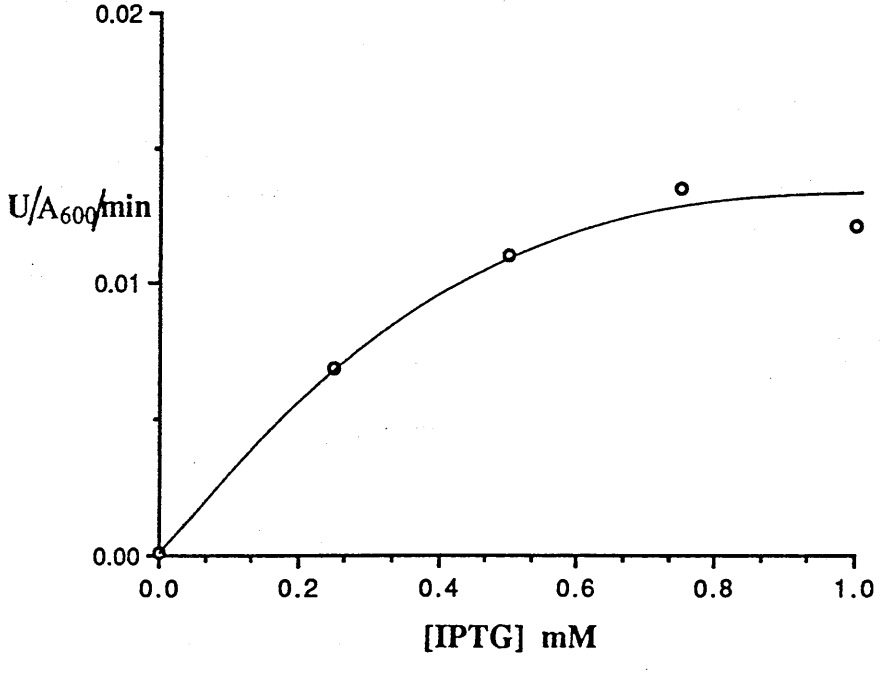
- ○ 0 mM IPTG
- ▲ 0.25 mM IPTG
- ▲ 0.50 mM IPTG
- ◆ 0.75mM IPTG
- ● 1.00 mM IPTG

(B) shows a plot of the rate of increase in specific activity against IPTG concentration.

A


2.0
1.5
1.0
0.5
0.0
U/A600
0
50
100
Time (min)

B


0.02
0.01
0.00
U/A600/min
0.0
0.2
0.4
0.6
0.8
1.0
[IPTG] mM

KAT-1/pAR9023 was grown on L-broth containing ampicillin and IPTG added to a final concentration of 1mM. At various time intervals, 1ml samples were taken, the cells were harvested and extracts were prepared by boiling the cell pellets in sample buffer, containing SDS. Samples were then run on an SDS gel (figure 3.5). There is a relative increase in the intensity of a band which co-migrates with pure ICL and accounts for a substantial fraction of the total cell protein in tracks loaded with KAT-1/pAR9023. Tracks loaded with KAT-1/ pGLW11, KAT-1/ pAR9024 and KAT-1 do not have this band, indicating that this band is a result of expression of a gene in the insert under the control of *ptac*. Since the band co-migrates with pure ICL, the insert of pAR9023 probably contains the entire coding region of *aceA*. Because a large proportion of the total cell protein is ICL in KAT-1/pAR9023, it is not surprising that there is a decrease in growth rate in the presence of increasing IPTG concentrations as there must be a great burden placed on the cell by expressing ICL to this level.

With the tracks loaded with KAT-1/pAR9023, as the intensity of the ICL band increases, the intensity of a band with a slighter higher M_r decreases relative to other bands. This higher M_r band appears in all tracks except that for ML308. Therefore, it is likely that the gene encoding this protein is on the KAT-1 chromosome, rather than on pGLW11. The identity of this protein is not known, although there are several possibilities. *E. coli* KAT-1 contains a transposon insertion in the *aceA* gene, therefore this protein could be encoded by the transposon.

3.3 Sequence Of *aceA*:

3.3.1 Sequencing Strategy:

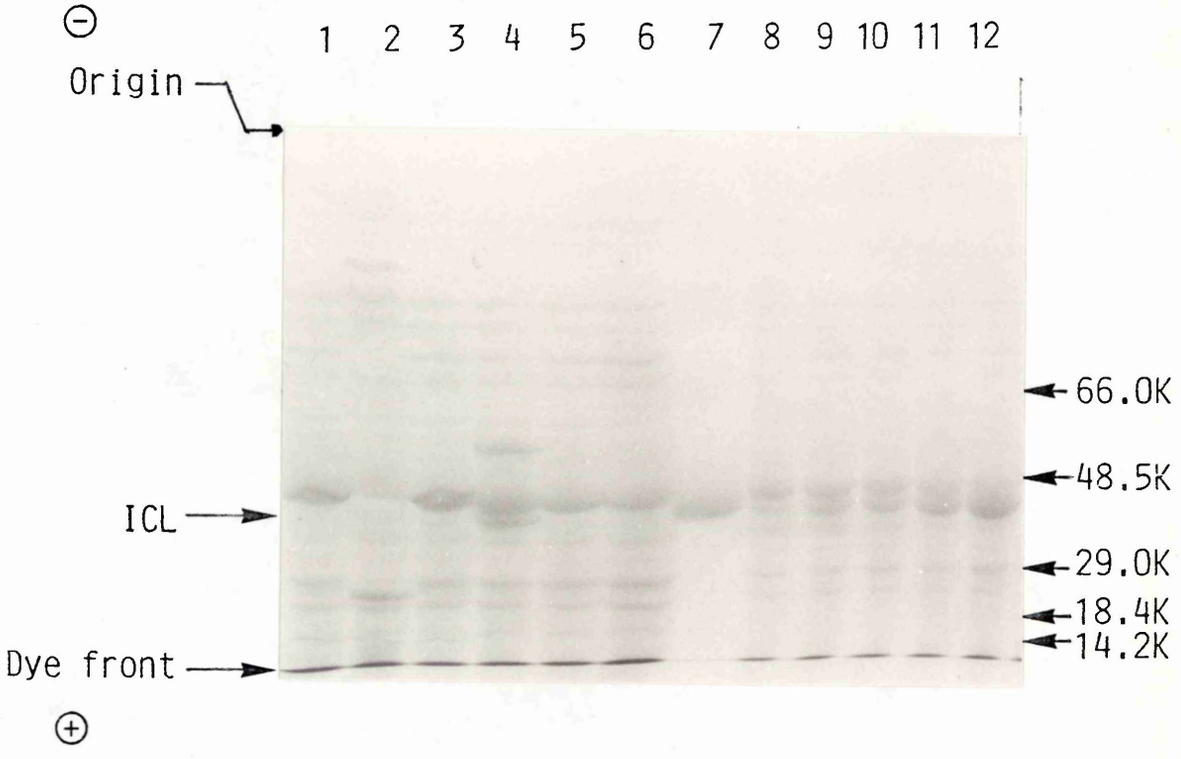
Only approximately 200-400 nucleotides can be read from each sequencing reaction by the Sanger dideoxynucleotide method (Sanger *et al.*, 1977). Therefore in general, large DNA fragments are sequenced by subcloning smaller fragments, each of which are small enough that their complete sequence can be read from a single sequencing reaction. By determining the sequence of a number of overlapping subclones, the complete sequence of the large DNA

Figure 3.5: Overexpression of ICL from *E.coli* KAT-1/pAR9023

IPTG and glucose were added to 10ml overnight cultures to final concentrations of 1mM and 0.1% (w/v) respectively. 1ml sampes were taken at time intervals and harvested in a microfuge. The cell pellets were resuspended in 200µl SDS sample buffer and boiled for 5min in the presence of 1% (v/v) 2-mercaptoethanol and loaded onto a 10% SDS-polyacrylamide gel.

The mobility of the molecular weight markers are indicated and the lanes are as follows:

1) *E.coli* KAT-1 (120min after addition of IPTG)
2) *E.coli* ML308 (120min after addition of IPTG)
3) *E.coli* KAT-1/pGLW11 (120min after addition of IPTG)
4) *E.coli* KAT-1/pEM901 (120min after addition of IPTG)
5) *E.coli* KAT-1/pAR9024 (no IPTG added)
6) *E.coli* KAT-1/pAR9024 (120min after addition of IPTG)
7) Purified ICL
8) *E.coli* KAT-1/pAR9023 (0min after addition of IPTG)
9) *E.coli* KAT-1/pAR9023 (30min after addition of IPTG)
10) *E.coli* KAT-1/pAR9023 (60min after addition of IPTG)
11) *E.coli* KAT-1/pAR9023 (90min after addition of IPTG)
12) *E.coli* KAT-1/pAR9023 (120min after addition of IPTG)


⊖
1 2 3 4 5 6 7 8 9 10 11 12
Origin
66.0K
48.5K
ICL
29.0K
18.4K
14.2K
Dye front
⊕

fragment can be determined. It is a prerequisite that the sequence of both strands is determined since comparison of the complementary strands can reveal sequencing errors. Many strategies have been devised to generate the subclones for sequencing.

A number of strategies involve digesting the large DNA fragment at random. This is referred to as shotgun sequencing. Digestion of the DNA with restriction enzymes which recognize four base sites followed by ligation into the single-stranded bacteriophage M13 has been used with great success. The development of M13 vectors with a variety of unique restriction sites in the multiple cloning site has greatly aided this method (Messing, 1981). Even if the restriction enzymes used to digest the DNA do not generate ends suitable for ligation into the sites of the multiple cloning site, oligonucleotide linkers containing the appropriate restriction site can be ligated and cleaved prior to ligation with M13 (Scheller *et al.*, 1977; Sanger *et al.*, 1980; Heidecker *et al.*, 1980). One approach is to cleave the large DNA fragment by sonication (Messing, 1983) or by digestion with DNase I in the presence of Mn^{2+} (Anderson, 1981), which causes the DNase I to cut double stranded DNA at approximately the same position on both strands (Melgar and Goldthwaite, 1968). In both of these methods the resulting DNA fragments have to be treated with either nuclease *Bal*31 or DNA polymerase I to ensure that the ends are blunt prior to ligation.

If the restriction map of the large fragment is known, then suitable subclones can be derived from double digests so that the orientation of the clone is known. Suitable pairs of vectors have been engineered which contain the same multiple cloning sites in the opposite orientations relative to the primer. This greatly decreases the number of clones that have to be sequenced. However, the determination of the restriction map can be very time consuming and may well prove to be very difficult as in some cases no convenient restriction sites can be found, possibly due to unusual features of the DNA sequence.

For these reasons, several methods have been devised which are less random and involve the generation of sets of overlapping deletions. In one method, random cuts are introduced into the insert by DNase I in the presence of Mn^{2+} followed by treatment with DNA polymerase I to blunt the ends. Oligonucleotide linkers are then ligated which can be cleaved by a restriction enzyme recognizing a unique site in the multiple cloning site, proximal to the

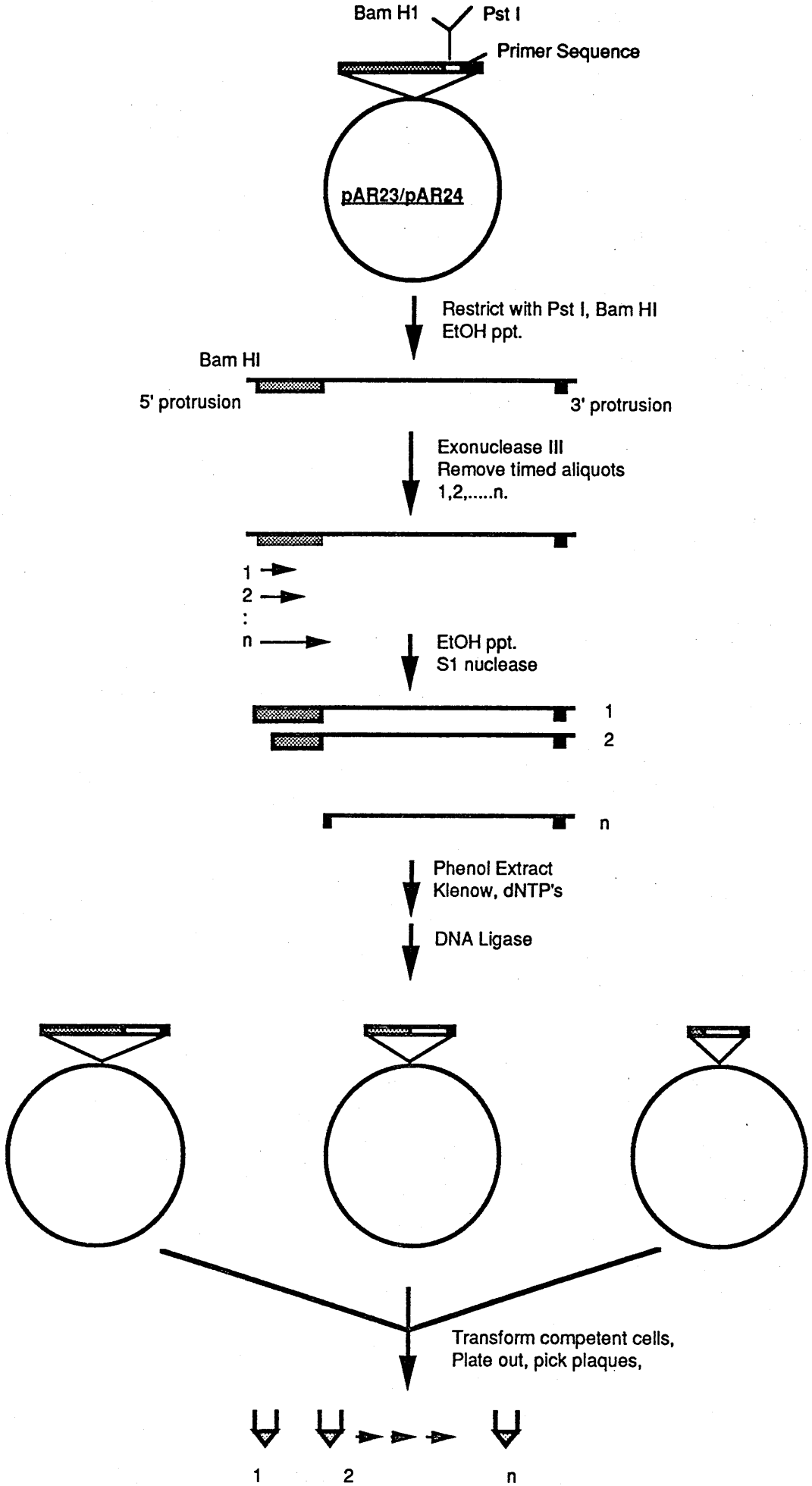
primer. The recombinant phage DNA is then digested by this enzyme and size fractionated on agarose gels. Fragments of an appropriate size can then be recircularised after excision from gel slices and used to transform *E. coli* (Frischauf *et al.*, 1980). A similar method was devised by Barnes *et al.* (1983) and Barnes (1983) except that the DNA was nicked by DNase I in the presence of ethidium bromide, which relieves supercoils, followed by extension of the gap by digestion with Exo III. The single-stranded region left by Exo III is then removed by *Bal*31 to leave flush ends. Oligonucleotide linkers are then added and cleaved as above. The DNA is then recircularised and then used to transform *E. coli*. The clones are fractionated by size by running intact phage on agarose gels. More progressive deletions can be generated using nuclease *Bal*31, which shortens the insert by digesting both strands. After digestion with an enzyme which cuts near the insert, the DNA is then incubated with *Bal*31 and cut with another enzyme at the opposite end of the insert. The fragment is then isolated from an agarose gel and religated into M13 (Poncz *et al.*, 1982).

The strategy chosen to sequence the *aceA* gene is based on deletion by Exo III (figure 3.6) and involves two useful properties of this enzyme; i) Exo III digests DNA in a 3'-5' direction from a 3'-hydroxyl at a very uniform rate, leaving single stranded DNA and ii) Exo III cannot however initiate deletion from ends with an 3'-overhang of four or more nucleotides (Wu *et al.*, 1976; Henikoff, 1984).

Double stranded, replicative form (RF) DNA is digested with two restriction enzymes at unique sites between the insert and the primer. The enzyme which cuts nearest the primer leaves a four base 5' overhang, thus ensuring unidirectional deletion from the primer. Exo III then digests the DNA from the 3'-hydroxyl left by the other enzyme resulting in a length of single-stranded DNA. Aliquots are removed at time intervals and the reaction stopped by incubation at 70°C in the presence of EDTA. S1 nuclease is then added to each aliquot to digest the single-stranded DNA left by the exonuclease reaction. After the ends have been filled in using Klenow polymerase and a mixture of all four dNTPs, the fragment is recircularised using T4 DNA ligase and used to transform *E. coli*. Clones are picked for sequencing by cutting out the inserts and determining their size by running the fragments on agarose gels. This method is very rapid as it does not involve a gel purification step and the deletion can be carried out in a single reaction mixture.

Figure 3.6: Strategy used to create a series of nested deletions, redrawn from Henikoff (1984).

The insert is shown by the shaded box, while the primer sequence is shown by the black box.


Bam H1
Pst I
Primer Sequence
pAR23/pAR24
Restrict with Pst I, Bam HI
EtOH ppt.
Bam HI
5' protrusion
3' protrusion
Exonuclease III
Remove timed aliquots
1,2,.....n.
1
2
:
n
EtOH ppt.
S1 nuclease
1
2
n
Phenol Extract
Klenow, dNTP's
DNA Ligase
Transform competent cells,
Plate out, pick plaques,
1
2
n

3.3.2 Determination Of Rate Of Digestion By Exonuclease III:

The number of nucleotides which can be read from a sequencing reaction is between 200-400. Therefore, it is necessary to generate a set of overlapping deletions from which the entire sequence of the insert can be determined. For this reason, the determination of the rate of digestion of insert DNA by Exo III is essential.

A pilot experiment was set up in which M13mp8 RF DNA was digested with *Bam*HI and *Pst*I thus generating a 5'-hydroxyl at the overhang left by *Bam*HI and a four base 3'-overhang generated by *Pst*I. The DNA was then digested with Exo III. Aliquots were removed at various time intervals and the reaction stopped by incubation at 70°C in the presence of 3.75mM EDTA. After digestion with S1 nuclease, samples were run on agarose gels and the sizes of the resulting fragments calculated in order to examine the extent of digestion by Exo III (figure 3.7).

From the gel, the rate of digestion by Exo III was 360 nucleotides per minute. Therefore, in order to generate a set of overlapping deletions covering the insert encoding *aceA*, it was necessary to take aliquots at 30sec. intervals.

3.3.3 Generation Of A Set Of Nested Deletions In Both Strands To Sequence *aceA*:

The first step was the recloning of the inserts of pAR9023 and pAR9024 in M13mp8. M13mp8, pAR9023 and pAR9024 were digested with *Bam*HI and *Eco*RI. The 2.3kb inserts of pAR9023 and pAR9024, as well as the large fragment of M13mp8 were purified from agarose gels and ligated yielding pAR23 and pAR24, from pAR9023 and pAR9024 respectively. RF DNA was isolated from transformants containing recombinant plasmids and the orientation of the inserts determined from the restriction map (figure 3.8). The position of the primer in pAR23 and pAR24 is at the opposite side of the insert from the *tac* promoter in pAR9023 and pAR9024, therefore the sequence derived from pAR24 should be the sense strand of the *aceA* gene.

In order to generate the 5'-overhang for initiation of digestion by Exo III and the four base 3'-overhang to block digestion into the insert, pAR23 and pAR24 were digested with

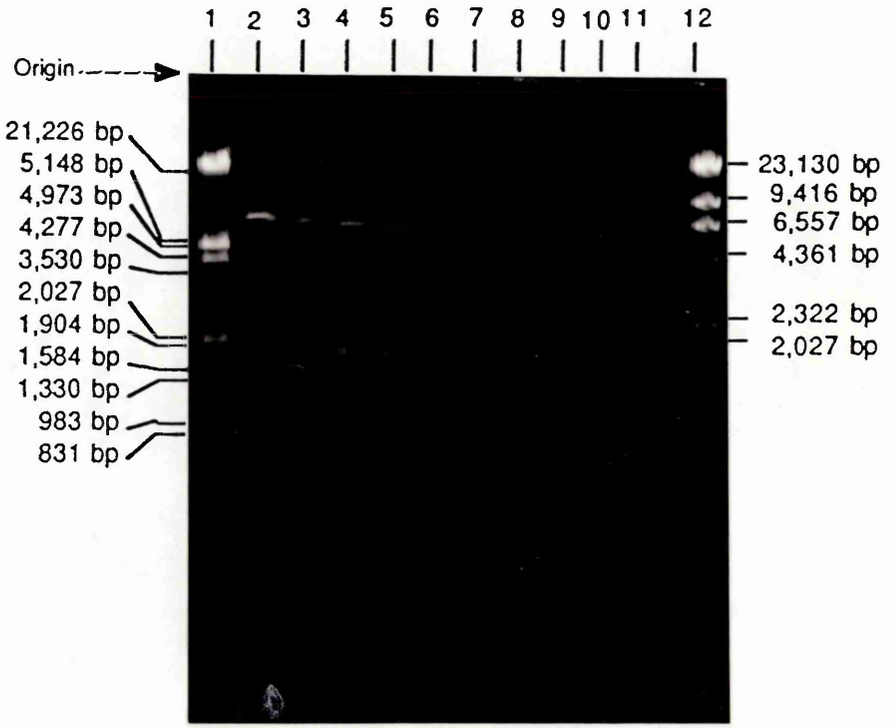
Figure 3.7: Determination of the rate of digestion by Exonuclease III.

(A) RF M13mp8 was digested with BamHI and PstI according to methods 2.6.1, deleted using Exo III, according to methods 2.6.2 and aliquots taken at time intervals. The aliquots were then incubated with S1 nuclease as in methods 2.6.3, and the DNA was then run on an 0.8% (w/v) agarose gel. The lanes are as indicated:

1) Lambda / HindIII, EcoRI
2) RF M13mp8 / ExoIII, S1 (0min)
3) " " (1min)
4) " " (2min)
5) " " (3min)
6) " " (4min)
7) " " (6min)
8) " " (8min)
9) " " (10min)
10) " " (12min)
11) " " (14min)
12) Lambda / HindIII

(B) The sizes of the fragments were calculated and plotted against time to calculate the rate of digestion by ExoIII.

A


1 2 3 4 5 6 7 8 9 10 11 12
Origin
21,226 bp
5,148 bp
4,973 bp
4,277 bp
3,530 bp
2,027 bp
1,904 bp
1,584 bp
1,330 bp
983 bp
831 bp
23,130 bp
9,416 bp
6,557 bp
4,361 bp
2,322 bp
2,027 bp

B

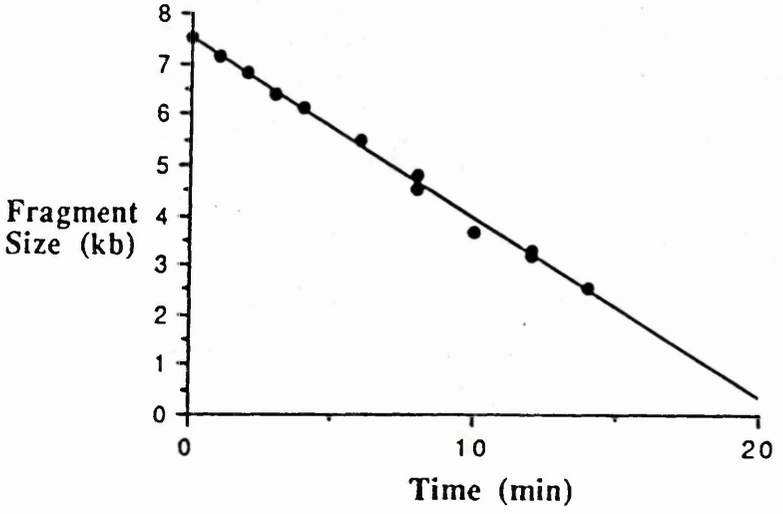

Fragment Size (kb)
0
1
2
3
4
5
6
7
8
0
10
20
Time (min)

Figure 3.8: Restriction maps of pAR23 and pAR24.

(A) pAR23 and pAR24 were digested with various restriction enzymes as follows:

1) Lambda / HindIII
2) pAR23 / BamHI, PvuII
3) " / PvuII
4) " / EcoRI, PvuII
5) " / BamHI, EcoRI, PvuII
6) Lambda / HindIII, EcoRI
7) Lambda / HindIII
8) pAR24 / BamHI, PvuII
9) " / PvuII
10) " / EcoRI, PvuII
11) " / BamHI, EcoRI, PvuII
12) Lambda / HindIII, EcoRI

(B) Restriction maps of pAR23 and pAR24. The shaded boxes represent the 2.3kb insert with the arrow indicating the direction of transcription. The distances between the restriction sites are indicated. The exact position of the PvuII site is unkown and the posible sites are indicated by the thin lines. The direction of DNA synthesis from the universal primer is shown.

Figure 3.9: Exonuclease III deletion of pAR23 and pAR24.

The RF DNA from individual clones derived from the digestion of pAR23 and pAR24 was isolated and digested by BamHI and EcoRI to liberate the insert. The sizes of the inserts were estimated from agarose gels.

(A) Distribution of deletion breakpoints for pAR23.

(B) Distribution of deletion breakpoints for pAR24.

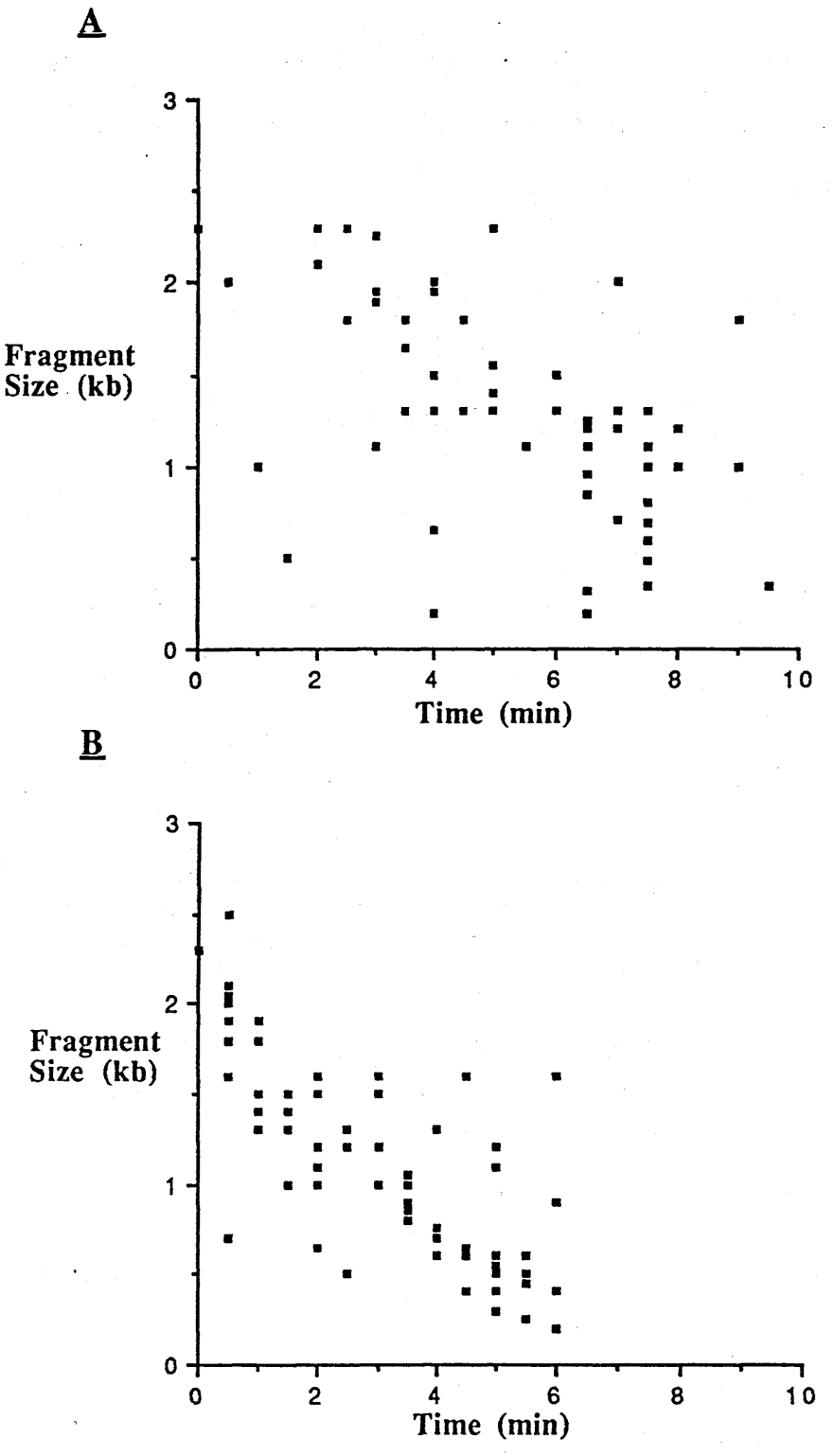
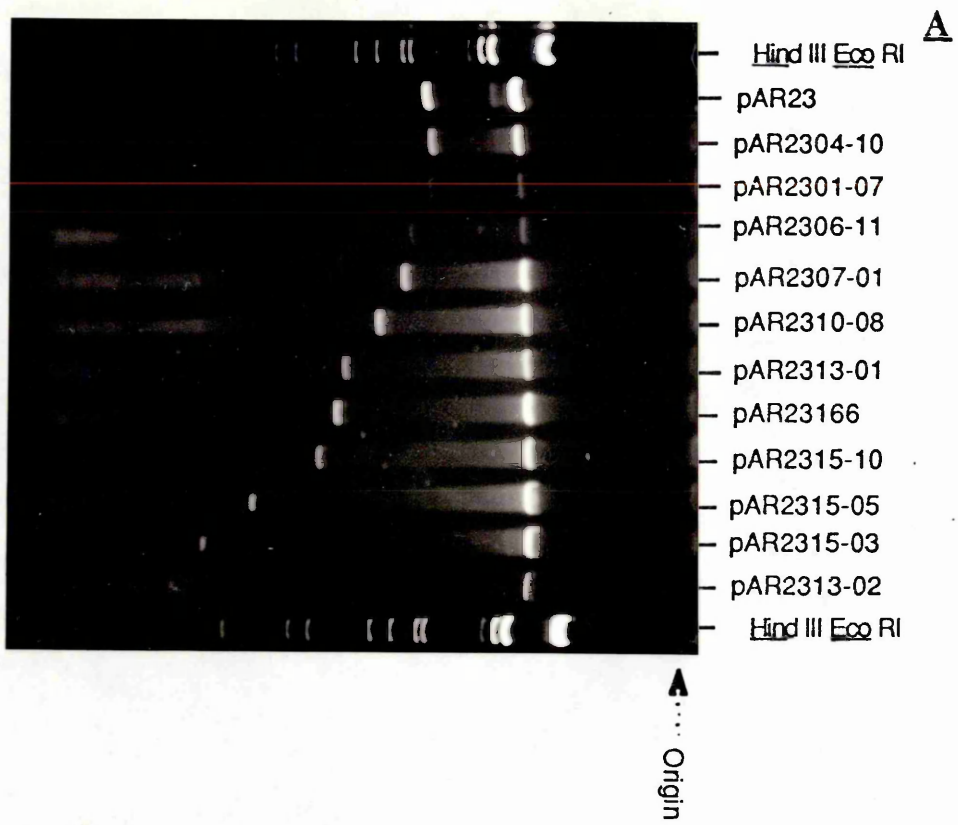
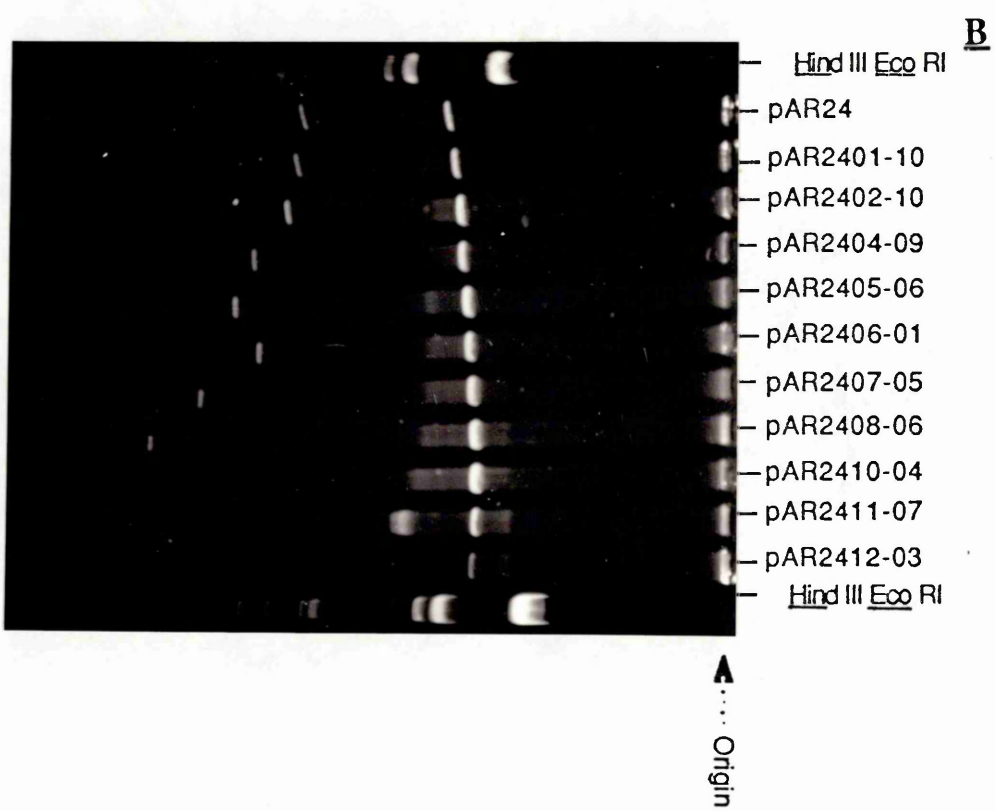

A
Fragment Size (kb)
3
2
1
0
0
2
4
6
8
10
Time (min)
B
Fragment Size (kb)
3
2
1
0
0
2
4
6
8
10
Time (min)

Figure 3.10: Examples Of ExoIII Deletion Of pAR23 and pAR24.

Rf DNA was isolated from individual plaques (methods 2.5.3a), digested with HindIII and EcoRI according to methods 2.4.3 and separated on 0.8% (w/v) agarose gels. Gel A) shows deletion clones from pAR23; gel B) shows deletions from pAR24.


A
Hind III Eco RI
pAR23
pAR2304-10
pAR2301-07
pAR2306-11
pAR2307-01
pAR2310-08
pAR2313-01
pAR23166
pAR2315-10
pAR2315-05
pAR2315-03
pAR2313-02
Hind III Eco RI
Origin


B
Hind III Eco RI
pAR24
pAR2401-10
pAR2402-10
pAR2404-09
pAR2405-06
pAR2406-01
pAR2407-05
pAR2408-06
pAR2410-04
pAR2411-07
pAR2412-03
Hind III Eco RI
Origin

should have covered the other gap, of approximately 100 nucleotides, it was not filled (figure 3.11). These subclones seemed to have deletions other than from the primer since many of them showed identity with sequence 5' of their predicted position at their 5' end and with 3' regions at their 3' ends. The possibility that this region contained direct repeats of 5' or 3' regions could be ruled out since each subclone showed different regions of identity and the sequence was not the same on both strands. One possibility is that this region has the potential to form stem-loop structures. If these structures are present during incubation with S1 nuclease, the single stranded region of the loop will be digested by this enzyme, thus deleting it on religation. Consistent with this, two groups (Cortay *et al.*, 1988; Klumpp *et al.*, 1988) found that the *aceA-aceK* intercistronic region had the potential to form several different stem-loop structures (section 1.3.6, figure 1.4). When the nucleotide sequence of the insert was analysed, the gap in the sequence corresponds to this region. Another possibility is that there is an advantage for M13 to delete these regions as they may interfere with packaging of the single-stranded DNA. This may explain the 10-fold difference in the number of colonies obtained for each time point compared to that obtained by Henikoff (1984). However the position of the deletion endpoint was, in general, similar to that predicted from restriction digests of each clone. The nucleotide sequence of the coding strand of the insert is shown in figure 3.12.

During the course of this work, the sequence of *aceA* was determined for *E. coli* K12 by two independent groups (Rieul *et al.*, 1988; Matsuoka and MacFadden, 1988). When these sequences were compared to that of pAR24, there is a very high degree of similarity over the full length of the *aceA* gene from both K12 sequences, although there are differences (section 3.4.2). This is more evidence that pAR9023 contains the entire coding sequence of *aceA*.

The nucleotide sequence of *aceB* (Byrne *et al.*, 1988) and *aceK* (Cortay *et al.*, 1988; Klumpp *et al.*, 1988) were determined for *E. coli* K12. Since the *Bam*HI-*Ava*I fragment of pCL1000 containing the *aceA* gene also contains part of the *aceB* and *aceK* genes, the sequence of the insert of pAR24 was compared to *aceB* and *aceK*. Position 17-333 of pAR24 showed 100% identity to the 3' end of the *aceB* sequence (position 17-79 are sequenced on only one strand, positions 1-16 are part of M13). The complete sequence of the DNA after the

Figure 3.11: The sequencing scheme of *aceA*.

The 2.3kb insert of pAR9023, containing the gene encoding ICL (dark shaded box), MSA (light shaded box) and ICDH kinase/phosphatase (black box) are shown, with the untranslated regions shown in the unshaded boxes. The arrows indicate the direction of second strand synthesis and length of readable sequence from each clone. The hatched box indicates the region which is not covered by any of the deletion clones. The clones above the insert were all generated from digestion of pAR24 by Exo III and the clones below the insert were generated from pAR23.

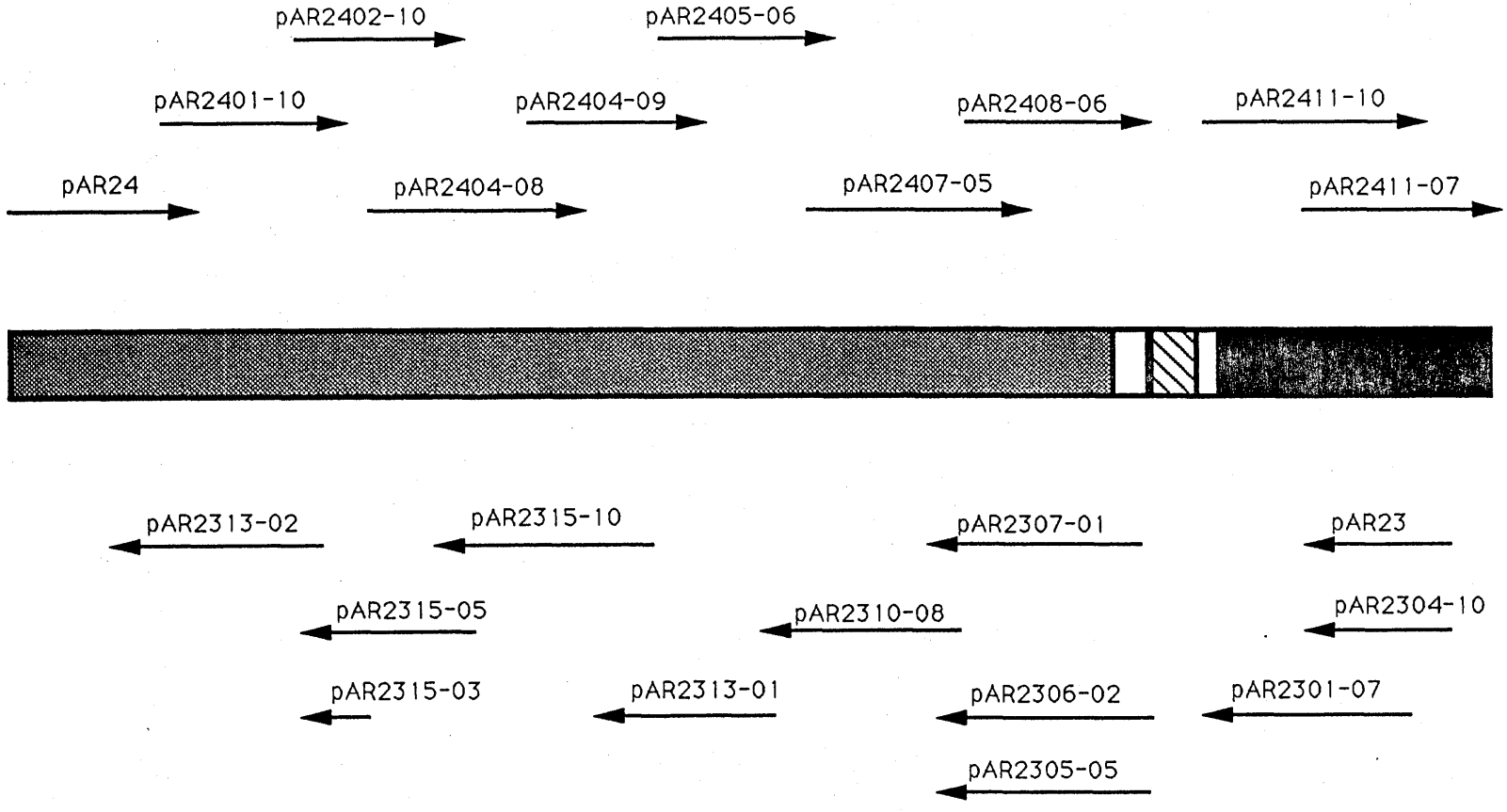

pAR2402-10
pAR2405-06
pAR2401-10
pAR2404-09
pAR2408-06
pAR2411-10
pAR24
pAR2404-08
pAR2407-05
pAR2411-07
pAR2313-02
pAR2315-10
pAR2307-01
pAR23
pAR2315-05
pAR2310-08
pAR2304-10
pAR2315-03
pAR2313-01
pAR2306-02
pAR2301-07
pAR2305-05

Figure 3.12: Nucleotide Sequence Of The Insert Of pAR24.

Nucleotides 1 to 1642 include the 3' end of *aceB*, the *aceB-aceA* intercistronic region, *aceA* and part of the *aceA-aceK* intercistronic region (pAR24-pAR2408-06). Nucleotides 1'-393' include part of the *aceA-aceK* intercistronic region and the 5' end of the *aceK* gene (pAR2411-10-pAR2411-07).

```
  1  GGTCGNCGGA TCCCCGATCC ATCATCAAAA AACGTTGAGC AATGGCAAAC CGGTGACCAA AGCCTTGTTC CGCCAGATGC TGGGCGAAGA GATGAAAGTC   100
101  ATTGCCAGCG AACTGGGCGA AGAACGTTTC TCCCAGGGGC GTTTTGACGA TGCCGCACGC TTGATGGAAC AGATCACCAC TTCCGATGAG TTAATTGATT   200
201  TCCTGACCCT GCCAGGCTAC CGCCTGTTAG CGTAAACCAC CACATAACTA TGGAGCATCT GCACATGAAA ACCCGTACAC AACAAATTGA AGAATTACAG   300
301  AAAGAGTGGA CTCAACCGCG TTGGGAAGGC ATTACTCGCC CATACAGTGC GGAAGATGTG GTGAAATTAC GCGGTTCAGT CAATCCTGAA TGCACGCTGG   400
401  CGCAACTGGG CGCAGCGAAA ATGTGGCGTC TGCTGCACGG TGAGTCGAAA AAAGGCTACA TCAACAGCCT CGGCGCACTG ACTGGCGGTC AGGCGCTGCA   500
501  ACAGGCGAAA GCGGGTATTG AAGCAGTCTA TCTGTCGGGA TGGCAGGTAG CGGCGGACGC TAACCTGGCG GCCAGCATGT ATCCGGATCA GTCGCTCTAT   600
601  CCGGCAAACT CGGTGCCAGC TGTGGTGGAG CGGATCAACA ACACCTTCCG TCGTGCCGAT CAGATCCAAT GGTCCGCGGG CATTGAGCCG GGCGATCCGC   700
701  GCTATGTCGA TTACTTCCTG CCGATCGTTG CCGATGCGGA AGCCGGTTTT GGCGGTGTCC TGAATGCCTT TGAACTGATG AAAGCGATGA TTGAAGCCGG   800
801  TGCAGCGGCA GTTCACTTCG AAGATCAGCT GGCGTCAGTG AAGAAATGCG GTCACATGGG CGGCAAAGTT TTAGTGCCAA CTCAGGAAGC TATTCAGAAA   900
901  CTGGTCGCGG CGCGTCTGGC AGCTGACGTG ACGGGCGTTC CAACCCTGCT GGTTGCCCGT ACCGATGCTG ATGCGGCGGA TCTGATCACC TCCGATTGCG  1000
1001  ACCCGTATGA CAGCGAATTT ATTACCGGCG AGCGTACCAG TGAAGGCTTC TTCCGTACTC ATGCGGGCAT TGAGCAAGCG ATCAGCCGTG GCCTGGCGTA  1100
1101  TGCGCCATAT GCTGACCTGG TCTGGTGTGA AACCTCCACG CCGGATCTGG AACTGGCGCG TCGCTTTGCA CAAGCTATCC ACGCGAAATA TCCGGGCAAA  1200
1201  CTGCTGGCTT ATAACTGCTC GCCGTCGTTC AACTGGCAGA AAAACCTCGA CGACAAAACT ATTGCCAGCT TCCAGCAGCA GCTGTCGGAT ATGGGCTACA  1300
1301  AGTTCCAGTT CATCACCCTG GCAGGTATCC ACAGCATGTG GTTCAACATG TTTGACCTGG CAAACGCCTA TGCCCAGGGC GAGGGTATGA AGCACTACGT  1400
1401  TGAGAAAGTG CAGCAGCCGG AATTTGCCGC CGCGAAAGAT GGCTATACCT TCGTATCTCA CCAGCAGGAA GTGGGTACAG GTTACTTCGA TAAAGTGACG  1500
1501  ACTATTATTC AGGGCGGCAC GTCTTCAGTC ACCGCGCTGA CCGGCTCCAC TGAAGAATCG CAGTTCTAAG CAACAACAAC CGTTGCTGAC TGTAGGCCGG  1600
1601  ATAAGGCGTT CACGCCGCAT CCGGCAATCG GTGCACGATG CC                                                              1642

  1' CCGCATCCGG CAATTCTCTG CTCCTGATGA GGGCGCTAAA TGCCGCGTGG CCTGGAATTA TTGATTGCTC AAACCATTTT GCAAGGCTTC GATGCTCAGT   100'
101' ATGGTCGATT CCTCGAAGTG ACCTCCGGTG CGCAGCAGCG TTTCGAACAG GCCGACTGGC ATGCTGTCCA GCAGGCGATG AAAAACCGTA TCCATCTTTA   200'
201' CGATCATCAC GTTGGTCTGG TCGTGGAGCA ACTGCGCTGC ATTACTAACG GCCAAAGTAC GGACGCGGCA TTTTTACTAC GTGTTAAAGA GCATTACACC   300'
301' CGGCTGTTGC CGGATTACCC GCGCTTCGAG ATTGCGGAGA GCTTTTTTAA CTCCGTGTAC TGTCGGTTAT TTGACCACCG CTCGCTTACT CCC         393'
```

gap shows 100% identity to both sequences for *aceK*. Thus it seems that the *Bam*HI-*Ava*I fragment of pEM9 is very similar to the corresponding fragment in pCL1000, indicating that the functional maps of the two clones are very similar.

3.4 Sequence Analysis:

3.4.1 Translation:

The sequence shown in figure 3.12 was translated from position 80-1642 in the three forward reading frames using the CODONPREFERENCE program in the GCG package (Gribskov *et al.*, 1984). This program calculates the GC bias in the third position of a codon and identifies rare codons by comparing them with codon frequency tables for frequently used *E. coli* codons. This program identified three open reading frames extending a short distance into the insert from position 1 (figure 3.13). Two of these reading frames contained many rare codons, the other contained very few. When this reading frame was translated using the TRANSLATE program of the GCG package, the amino acid sequence showed 100% identity to the deduced amino acid sequence at the carboxyl-terminus of *aceB* (Byrne *et al.*, 1988). This open reading frame extends to position 232 after which there is a translation termination codon, TAA. As well as the stop codon at position 233-235, there is another stop codon in the same reading frame at 245-247 (TAA) further downstream and several other stop codons upstream of position 233 in the other two reading frames, 187-189 and 196-198 (TGA) in one reading frame and at 204-206 (TGA) and 228-230 (TAG) in the other reading frame.

The CODONPREFERENCE program also revealed a large open reading frame extending from approximately 250-1550 with very few rare codons (figure 3.13). Downstream of the translation termination codon after the *aceB* gene, there are two ATG translation initiation codons in-frame with each other, starting at positions 250 and 265. When translated, they reveal an open reading frame which extends to position 1566. However, the only recognisable ribosome binding site, Shine-Dalgarno sequence (Shine and Dalgarno, 1974), downstream of *aceB* is at position 252-255, 10-13 bases upstream of the second ATG. Therefore, the open reading frame which will be translated extends from 265-1566, after which there is a stop

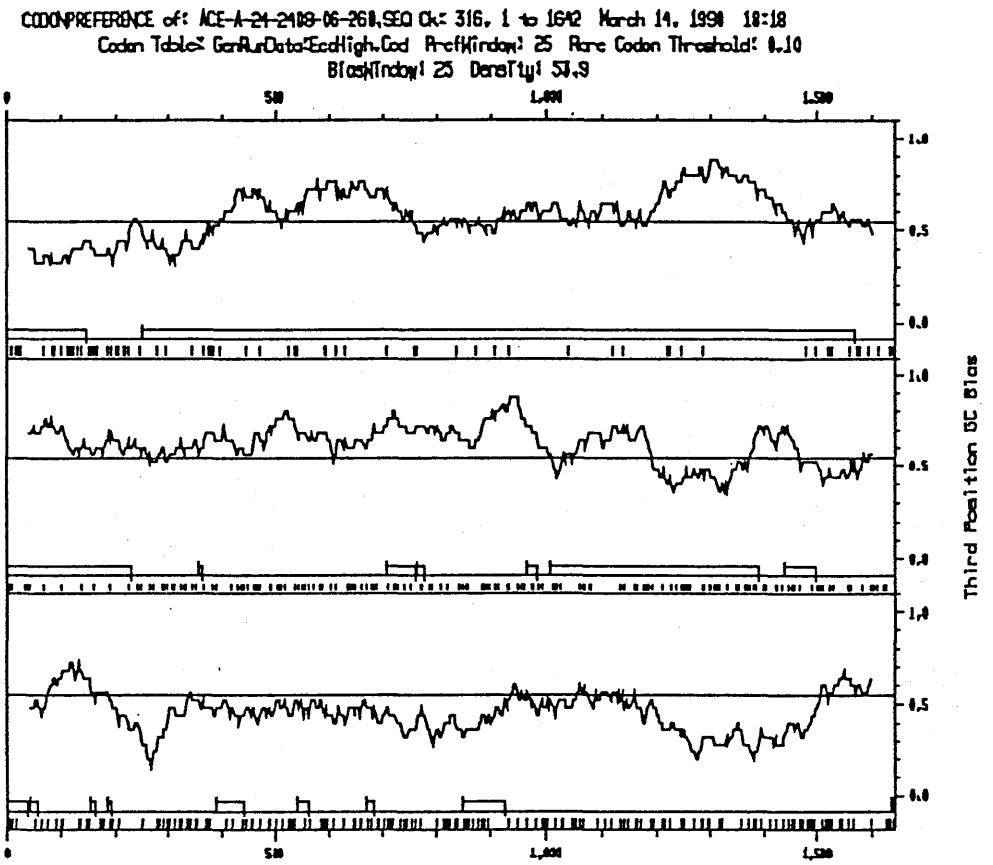


Figure 3.13: CODONPREFERENCE Table For Nucleotides 1-1642 Of pAR24.

Each panel represents one reading frame showing the third base GC bias of the reading frame. Open reading frames are represented by the open boxes at the bottom of each panel and rare codons are represented by vertical bars below the reading frame panels.

codon, TAA (1567-1569). There are also stop codons in the other reading frames, at positions 1551-1553 (TGA) and 1538-1540 (TGA). This open reading frame encoded 434 amino acids corresponding to a 47,491 dalton protein. This is similar to the value obtained for the subunit molecular weight of ICL from *E. coli* of between 45,000-48,000 (Robertson and Reeves, 1987; MacKintosh and Nimmo, 1988). The N-terminal amino acid sequence of the protein translated from this open reading frame is similar to the N-terminal amino acid sequence derived from pure ICL (37 residues sequenced) (MacKintosh, 1987) and also contains the sequence around the active site cysteine residue reported by Nimmo *et al.* (1989). Therefore, this open reading frame corresponds to the *aceA* gene.

There are 29 nucleotides between the translation termination codon of *aceB* and the translation initiation codon of *aceA* and this region contains no sequence with the potential to form stable stem-loop structures, which is consistent with the very similar levels of expression of ICL and MSA (Cortay *et al.*, 1989).

After the gap in the sequence of the insert, there is a third open reading frame, which has its startsite at position 40' and extends to the end of the insert. There is a recognisable Shine-Dalgarno sequence at position 29'-32', 8 bases upstream of the ATG. The derived amino acid sequence of this open reading frame shows 100% identity to the derived amino acid sequence of ICDH kinase/phosphatase (Cortay *et al.*, 1988; Klumpp *et al.*, 1988). The nucleotide sequence and derived amino acid sequence of the above genes are shown in figure 3.14.

Downstream of the stop codon at the 3' end of *aceA*, is a region with 100% identity to the first REP (repetitive, extragenic, palindromic) element found within the *aceA-aceK* intercistronic region (Cortay *et al.*, 1988; Klumpp *et al.*, 1988). When compared to the *aceA-aceK* intercistronic region of the K12 sequence, only the first 3 bases at the 5' end of REP-2 and the last 12 bases at the 3' end of REP-3 and the first partial element are present. Since the second and third REP elements have opposite orientations relative to each other in the K12 sequences, they could have formed a stable stem-loop structure, which either could not be sequenced, or which was deleted during the Exo III/S1 or cloning in M13 steps. The second partial REP element (Cortay *et al.*, 1988), which includes the Shine-Dalgarno sequence and

Figure 3.14: Nucleotide Sequence And Derived Amino Acid Sequence Of The Insert Of pAR24.

The nucleotide sequence is shown with the derived amino acid sequence below. The putitive ribosome binding sites are indicated by the bold print and underlined. The REP and the partial REP elements are underlined.

1160 GTCGCTTTGCACAAGCTATCCACGCGAAATATCCGGGCAAACTGCTGGCTTATAACTGCTCGCCGTCGTTCAACTGGCAGAAAAACCTCGACGACAAAACTATTGCCAGCTTCCAGCAGC 1279

rgArgPheAlaGlnAlaIleHisAlaLysTyrProGlyLysLeuLeuAlaTyrAsnCysSerProSerPheAsnTrpGlnLysAsnLeuAspAspLysThrIleAlaSerPheGlnGlnG

1280 AGCTGTCGGATATGGGCTACAAGTTCCAGTTCATCACCCTGGCAGGTATCCACAGCATGTGGTTCAACATGTTTGACCTGGCAAACGCCTATGCCCAGGGCGAGGGTATGAAGCACTACG 1399

lnLeuSerAspMetGlyTyrLysPheGlnPheIleThrLeuAlaGlyIleHisSerMetTrpPheAsnMetPheAspLeuAlaAsnAlaTyrAlaGlnGlyGluGlyMetLysHisTyrV

1400 TTGAGAAAGTGCAGCAGCCGGAATTTGCCGCCGCGAAAGATGGCTATACCTTCGTATCTCACCAGCAGGAAGTGGGTACAGGTTACTTCGATAAAGTGACGACTATTATTCAGGGCGGCA 1519

alGluLysValGlnGlnProGluPheAlaAlaAlaLysAspGlyTyrThrPheValSerHisGlnGlnGluValGlyThrGlyTyrPheAspLysValThrThrIleIleGlnGlyGlyT

1520 CGTCTTCAGTCACCGCGCTGACCGGCTCCACTGAAGAATCGCAGTTCTAAGCAACAACAACCGTTGCTGACTGTAGGCCGGATAAGGCGTTCACGCCGCATCCGGCAATCGGTGCACGAT 1639

hrSerSerValThrAlaLeuThrGlySerThrGluGluSerGlnPheEnd **REP-1**

1640 GCC 1642

1' CCGCATCCGGCAATTCTCTGCTCCTGATGAGGCGCTAAATGCCGCGTGGCCTGGAATTATTGATTGCTCAAACCATTTTGCAAGGCTTCGATGCTCAGTATGGTCGATTCCTCGAAGTG 120'

MetProArgGlyLeuGluLeuLeuIleAlaGlnThrIleLeuGlnGlyPheAspAlaGlnTyrGlyArgPheLeuGluVal

121' ACCTCCGGTGCGCAGCAGCGTTTCGAACAGGCCGACTGGCATGCTGTCCAGCAGGCGATGAAAAACCGTATCCATCTTTACGATCATCACGTTGGTCTGGTCGTGGAGCAACTGCGCTGC 240'

ThrSerGlyAlaGlnGlnArgPheGluGlnAlaAspTrpHisAlaValGlnGlnAlaMetLysAsnArgIleHisLeuTyrAspHisHisValGlyLeuValValGluGlnLeuArgCys

241' ATTACTAACGGCCAAAGTACGGACGCGGCATTTTTACTACGTGTTAAAGAGCATTACACCCGGCTGTTGCCGGATTACCCGCGCTTCGAGATTGCGGAGAGCTTTTTTAACTCCGTGTAC 360'

IleThrAsnGlyGlnSerThrAspAlaAlaPheLeuLeuArgValLysGluHisTyrThrArgLeuLeuProAspTyrProArgPheGluIleAlaGluSerPhePheAsnSerValTyr

361' TGTCGGTTATTTGACCACCGCTCGCTTACTCCC 393'

CysArgLeuPheAspHisArgSerLeuThrPro

```
80  CTGGGCGAAGAGATGAAAGTCATTGCCAGCGAACTGGGCGAAGAACGTTTCTCCCAGGGGCGTTTTGACGATGCCGCACGCTTGATGGAACAGATCACCACTTCCGATGAGTTAATTGAT  199
    LeuGlyGluGluMetLysValIleAlaSerGluLeuGlyGluGluArgPheSerGlnGlyArgPheAspAspAlaAlaArgLeuMetGluGlnIleThrThrSerAspGluLeuIleAsp

200 TTCCTGACCCTGCCAGGCTACCGCCTGTTAGCGTAAACCACCACATAACTATGGAGCATCTGCACATGAAAACCCGTACACAACAAATTGAAGAATTACAGAAAGAGTGGACTCAACCGC  319
    PheLeuThrLeuProGlyTyrArgLeuLeuAlaEnd                                    MetLysThrArgThrGlnGlnIleGluGluLeuGlnLysGluTrpThrGlnProA

320 GTTGGGAAGGCATTACTCGCCCATACAGTGCGGAAGATGTGGTGAAATTACGCGGTTCAGTCAATCCTGAATGCACGCTGGCGCAACTGGGCGCAGCGAAAATGTGGCGTCTGCTGCACG  439
    rgTrpGluGlyIleThrArgProTyrSerAlaGluAspValValLysLeuArgGlySerValAsnProGluCysThrLeuAlaGlnLeuGlyAlaAlaLysMetTrpArgLeuLeuHisG

440 GTGAGTCGAAAAAAGGCTACATCAACAGCCTCGGCGCACTGACTGGCGGTCAGGCGCTGCAACAGGCGAAAGCGGGTATTGAAGCAGTCTATCTGTCGGGATGGCAGGTAGCGGCGGACG  559
    lyGluSerLysLysGlyTyrIleAsnSerLeuGlyAlaLeuThrGlyGlyGlnAlaLeuGlnGlnAlaLysAlaGlyIleGluAlaValTyrLeuSerGlyTrpGlnValAlaAlaAspA

560 CTAACCTGGCGGCCAGCATGTATCCGGATCAGTCGCTCTATCCGGCAAACTCGGTGCCAGCTGTGGTGGAGCGGATCAACAACACCTTCCGTCGTGCCGATCAGATCCAATGGTCCGCGG  679
    laAsnLeuAlaAlaSerMetTyrProAspGlnSerLeuTyrProAlaAsnSerValProAlaValValGluArgIleAsnAsnThrPheArgArgAlaAspGlnIleGlnTrpSerAlaG

680 GCATTGAGCCGGGCGATCCGCGCTATGTCGATTACTTCCTGCCGATCGTTGCCGATGCGGAAGCCGGTTTTGGCGGTGTCCTGAATGCCTTTGAACTGATGAAAGCGATGATTGAAGCCG  799
    lyIleGluProGlyAspProArgTyrValAspTyrPheLeuProIleValAlaAspAlaGluAlaGlyPheGlyGlyValLeuAsnAlaPheGluLeuMetLysAlaMetIleGluAlaG

800 GTGCAGCGGCAGTTCACTTCGAAGATCAGCTGGCGTCAGTGAAGAAATGCGGTCACATGGGCGGCAAAGTTTTAGTGCCAACTCAGGAAGCTATTCAGAAACTGGTCGCGGCGCGTCTGG  919
    lyAlaAlaAlaValHisPheGluAspGlnLeuAlaSerValLysLysCysGlyHisMetGlyGlyLysValLeuValProThrGlnGluAlaIleGlnLysLeuValAlaAlaArgLeuA

920 CAGCTGACGTGACGGGCGTTCCAACCCTGCTGGTTGCCCGTACCGATGCTGATGCGGCGGATCTGATCACCTCCGATTGCGACCCGTATGACAGCGAATTTATTACCGGCGAGCGTACCA  1039
    laAlaAspValThrGlyValProThrLeuLeuValAlaArgThrAspAlaAspAlaAlaAspLeuIleThrSerAspCysAspProTyrAspSerGluPheIleThrGlyGluArgThrS

1040 GTGAAGGCTTCTTCCGTACTCATGCGGGCATTGAGCAAGCGATCAGCCGTGGCCTGGCGTATGCGCCATATGCTGACCTGGTCTGGTGTGAAACCTCCACGCCGGATCTGGAACTGGCGC  1159
     erGluGlyPhePheArgThrHisAlaGlyIleGluGlnAlaIleSerArgGlyLeuAlaTyrAlaProTyrAlaAspLeuValTrpCysGluThrSerThrProAspLeuGluLeuAlaA
```

the ATG startsite is present.

3.4.2 Comparison Of The *E. coli* ICL Sequences With Each Other:

The *aceA* gene from *E. coli* K12 was sequenced by two other groups during the course of this work (Matsuoka and MacFadden, 1988; Rieul *et al.*, 1988) and by another group (Byrne, EMBL library, accession number Pir:S05692) after the ML308 gene was sequenced. These sequences will be referred to as the Cozzone, McFadden and Byrne sequences respectively. Figure 3.15 shows a LINE UP of the amino acid sequences derived from the different *aceA* sequences.

The derived amino acid sequences from MacFadden's and Cozzone's groups differ in two regions, between amino acid 101 and 117 and between amino acid 419 and the carboxy-terminus. The ML308 sequence is the same as the MacFadden sequence between 101 and 117 and the same as the Cozzone sequence at the carboxy-terminus.The Cozzone sequence also differs from the ML308 and MacFadden's sequences at amino acids 215 with a Pro instead of an Ala and 338, with a Glu instead of a Gln. The Byrne sequence is identical to that from ML308. The nucleotide sequences of the *aceA* genes are compared in Figure 3.16. The Cozzone sequence has no C residue at position 300 in the nucleotide sequence and an extra T at position 352. This causes a shift in the reading frame between amino acids 101 and 117. The Cozzone nucleotide sequence also differs from the others at bases 643 and 644, with GC instead of CG and at 1013, with a G instead of a C. These differences account for the amino acid changes at 215 and 338. The MacFadden sequence has an extra G between bases 1255 and 1256 which results in a change in the reading frame causing early translation termination resulting in a protein with a predicted of M_r 47,200.

An important difference between the ICL sequences from *E. coli* is that the MacFadden amino acid sequence places a His at the carboxy-terminus. ICL from several organisms has been shown to have a His at the carboxy-terminus (McFadden *et al.*, 1968; Johanson *et al.*, 1974b; Khan and McFadden, 1982) and removal of this carboxy-terminal His resulted in loss of activity in the *P. indigofera* enzyme (McFadden *et al.*, 1968). MacFadden used five independent deletion clones from both strands to determine the nucleotide sequence of the 3'-

Figure 3.15: Comparison of The *E. coli* ICL Derived Amino Acid Sequences.

The *aceA* ML308 sequence is the sequence of the *aceA* gene from pAR24. The Cozzone sequence was obtained from Rieul *et al.* (1988), the McFadden sequence from Matsuoka and McFadden (1988) and the Byrne sequence taken from the EMBL database, accession number Pir:S05692. The sequences were aligned using the LINEUP programme of the UWGCG package. Differences are indicated by an asterisk.

```
            1                                                                                                          100
   Cozzone  MKTRTQQIEE LQKEWTQPRW EGITRPYSAE DVVKLRGSVN PECTLAQLGA AKMWRLLHGE SKKGYINSLG ALTGGQALQQ AKAGIEAVYL SGWQVAADAN
 Macfadden  MKTRTQQIEE LQKEWTQPRW EGITRPYSAE DVVKLRGSVN PECTLAQLGA AKMWRLLHGE SKKGYINSLG ALTGGQALQQ AKAGIEAVYL SGWQVAADAN
     Byrne  MKTRTQQIEE LQKEWTQPRW EGITRPYSAE DVVKLRGSVN PECTLAQLGA AKMWRLLHGE SKKGYINSLG ALTGGQALQQ AKAGIEAVYL SGWQVAADAN
aceA ML308  MKTRTQQIEE LQKEWTQPRW EGITRPYSAE DVVKLRGSVN PECTLAQLGA AKMWRLLHGE SKKGYINSLG ALTGGQALQQ AKAGIEAVYL SGWQVAADAN

          101********** *******                                                                                     200
   Cozzone  WRPACIRISR SIRQTRCPAV VERINNTFRR ADQIQWSAGI EPGDPRYVDY FLPIVADAEA GFGGVLNAFE LMKAMIEAGA AAVHFEDQLA SVKKCGHMGG
 Macfadden  LAASMYPDQS LYPANSVPAV VERINNTFRR ADQIQWSAGI EPGDPRYVDY FLPIVADAEA GFGGVLNAFE LMKAMIEAGA AAVHFEDQLA SVKKCGHMGG
     Byrne  LAASMYPDQS LYPANSVPAV VERINNTFRR ADQIQWSAGI EPGDPRYVDY FLPIVADAEA GFGGVLNAFE LMKAMIEAGA AAVHFEDQLA SVKKCGHMGG
aceA ML308  LAASMYPDQS LYPANSVPAV VERINNTFRR ADQIQWSAGI EPGDPRYVDY FLPIVADAEA GFGGVLNAFE LMKAMIEAGA AAVHFEDQLA SVKKCGHMGG

          201              *                                                                                        300
   Cozzone  KVLVPTQEAI QKLVPARLAA DVTGVPTLLV ARTDADAADL ITSDCDPYDS EFITGERTSE GFFRTHAGIE QAISRGLAYA PYADLVWCET STPDLELARR
 Macfadden  KVLVPTQEAI QKLVAARLAA DVTGVPTLLV ARTDADAADL ITSDCDPYDS EFITGERTSE GFFRTHAGIE QAISRGLAYA PYADLVWCET STPDLELARR
     Byrne  KVLVPTQEAI QKLVAARLAA DVTGVPTLLV ARTDADAADL ITSDCDPYDS EFITGERTSE GFFRTHAGIE QAISRGLAYA PYADLVWCET STPDLELARR
aceA ML308  KVLVPTQEAI QKLVAARLAA DVTGVPTLLV ARTDADAADL ITSDCDPYDS EFITGERTSE GFFRTHAGIE QAISRGLAYA PYADLVWCET STPDLELARR

          301                                       *                                                               400
   Cozzone  FAQAIHAKYP GKLLAYNCSP SFNWQKNLDD KTIASFQEQL SDMGYKFQFI TLAGIHSMWF NMFDLANAYA QGEGMKHYVE KVQQPEFAAA KDGYTFVSHQ
 Macfadden  FAQAIHAKYP GKLLAYNCSP SFNWQKNLDD KTIASFQQQL SDMGYKFQFI TLAGIHSMWF NMFDLANAYA QGEGMKHYVE KVQQPEFAAA KDGYTFVSHQ
     Byrne  FAQAIHAKYP GKLLAYNCSP SFNWQKNLDD KTIASFQQQL SDMGYKFQFI TLAGIHSMWF NMFDLANAYA QGEGMKHYVE KVQQPEFAAA KDGYTFVSHQ
aceA ML308  FAQAIHAKYP GKLLAYNCSP SFNWQKNLDD KTIASFQQQL SDMGYKFQFI TLAGIHSMWF NMFDLANAYA QGEGMKHYVE KVQQPEFAAA KDGYTFVSHQ

          401                  ******************435
   Cozzone  QEVGTGYFDK VTTIIQGGTS SVTALTGSTE ESQF
 Macfadden  QEVGTGYFDK VTTIIQGGDV FSHRADRLH .....
     Byrne  QEVGTGYFDK VTTIIQGGTS SVTALTGSTE ESQF
aceA ML308  QEVGTGYFDK VTTIIQGGTS SVTALTGSTE ESQF
```

Figure 3.16: Comparison Of The Nucleotide Sequences Of The *aceA* Genes From *E. coli*.

The *aceA* ML308 sequence is the sequence of the *aceA* gene from pAR24. The Cozzone sequence was obtained from Rieul *et al.* (1988), the McFadden sequence from Matsuoka and McFadden (1988) and the Byrne sequence taken from the EMBL database, accession number Pir:S05692. The sequences were aligned using the LINEUP programme of the UWGCG package. Differences are indicated by an asterisk.

```
Macfadden  661 TGACGTGACG GGCGTTCCAA CCCTGCTGGT TGCCCGTACC GATGCTGATG CGGCGGATCT GATCACCTCC GATTGCGACC CGTATGACAG CGAATTTATT ACCGGCGAGC  770
  Cozzone      TGACGTGACG GGCGTTCCAA CCCTGCTGGT TGCCCGTACC GATGCTGATG CGGCGGATCT GATCACCTCC GATTGCGACC CGTATGACAG CGAATTTATT ACCGGCGAGC
    Byrne      TGACGTGACG GGCGTTCCAA CCCTGCTGGT TGCCCGTACC GATGCTGATG CGGCGGATCT GATCACCTCC GATTGCGACC CGTATGACAG CGAATTTATT ACCGGCGAGC
Aceaml308      TGACGTGACG GGCGTTCCAA CCCTGCTGGT TGCCCGTACC GATGCTGATG CGGCGGATCT GATCACCTCC GATTGCGACC CGTATGACAG CGAATTTATT ACCGGCGAGC

Macfadden  771 GTACCAGTGA AGGCTTCTTC CGTACTCATG CGGGCATTGA GCAAGCGATC AGCCGTGGCC TGGCGTATGC GCCATATGCT GACCTGGTCT GGTGTGAAAC CTCCACGCCG  880
  Cozzone      GTACCAGTGA AGGCTTCTTC CGTACTCATG CGGGCATTGA GCAAGCGATC AGCCGTGGCC TGGCGTATGC GCCATATGCT GACCTGGTCT GGTGTGAAAC CTCCACGCCG
    Byrne      GTACCAGTGA AGGCTTCTTC CGTACTCATG CGGGCATTGA GCAAGCGATC AGCCGTGGCC TGGCGTATGC GCCATATGCT GACCTGGTCT GGTGTGAAAC CTCCACGCCG
Aceaml308      GTACCAGTGA AGGCTTCTTC CGTACTCATG CGGGCATTGA GCAAGCGATC AGCCGTGGCC TGGCGTATGC GCCATATGCT GACCTGGTCT GGTGTGAAAC CTCCACGCCG

Macfadden  881 GATCTGGAAC TGGCGCGTCG CTTTGCACAA GCTATCCACG CGAAATATCC GGGCAAACTG CTGGCTTATA ACTGCTCGCC GTCGTTCAAC TGGCAGAAAA ACCTCGACGA  990
  Cozzone      GATCTGGAAC TGGCGCGTCG CTTTGCACAA GCTATCCACG CGAAATATCC GGGCAAACTG CTGGCTTATA ACTGCTCGCC GTCGTTCAAC TGGCAGAAAA ACCTCGACGA
    Byrne      GATCTGGAAC TGGCGCGTCG CTTTGCACAA GCTATCCACG CGAAATATCC GGGCAAACTG CTGGCTTATA ACTGCTCGCC GTCGTTCAAC TGGCAGAAAA ACCTCGACGA
Aceaml308      GATCTGGAAC TGGCGCGTCG CTTTGCACAA GCTATCCACG CGAAATATCC GGGCAAACTG CTGGCTTATA ACTGCTCGCC GTCGTTCAAC TGGCAGAAAA ACCTCGACGA

Macfadden  991 CAAAACTATT GCCAGCTTCC AGCAGCAGCT GTCGGATATG GGCTACAAGT TCCAGTTCAT CACCCTGGCA GGTATCCACA GCATGTGGTT CAACATGTTT GACCTGGCAA 1100
  Cozzone      CAAAACTATT GCCAGCTTCC AGGAGCAGCT GTCGGATATG GGCTACAAGT TCCAGTTCAT CACCCTGGCA GGTATCCACA GCATGTGGTT CAACATGTTT GACCTGGCAA
    Byrne      CAAAACTATT GCCAGCTTCC AGCAGCAGCT GTCGGATATG GGCTACAAGT TCCAGTTCAT CACCCTGGCA GGTATCCACA GCATGTGGTT CAACATGTTT GACCTGGCAA
Aceaml308      CAAAACTATT GCCAGCTTCC AGCAGCAGCT GTCGGATATG GGCTACAAGT TCCAGTTCAT CACCCTGGCA GGTATCCACA GCATGTGGTT CAACATGTTT GACCTGGCAA

Macfadden 1101 ACGCCTATGC CCAGGGCGAG GGTATGAAGC ACTACGTTGA GAAAGTGCAG CAGCCGGAAT TTGCCGCCGC GAAAGATGGC TATACCTTCG TATCTCACCA GCAGGAAGTG 1210
  Cozzone      ACGCCTATGC CCAGGGCGAG GGTATGAAGC ACTACGTTGA GAAAGTGCAG CAGCCGGAAT TTGCCGCCGC GAAAGATGGC TATACCTTCG TATCTCACCA GCAGGAAGTG
    Byrne      ACGCCTATGC CCAGGGCGAG GGTATGAAGC ACTACGTTGA GAAAGTGCAG CAGCCGGAAT TTGCCGCCGC GAAAGATGGC TATACCTTCG TATCTCACCA GCAGGAAGTG
Aceaml308      ACGCCTATGC CCAGGGCGAG GGTATGAAGC ACTACGTTGA GAAAGTGCAG CAGCCGGAAT TTGCCGCCGC GAAAGATGGC TATACCTTCG TATCTCACCA GCAGGAAGTG

Macfadden 1211 GGTACAGGTT ACTTCGATAA AGTGACGACT ATTATTCAGG GCGGCGACGT CTTCAGTCAC CGCGCTGACC GGCTCCACTG AAGAATCGCA GTTCTAA 1307
  Cozzone      GGTACAGGTT ACTTCGATAA AGTGACGACT ATTATTCAGG GCGGC.ACGT CTTCAGTCAC CGCGCTGACC GGCTCCACTG AAGAATCGCA GTTCTAA
    Byrne      GGTACAGGTT ACTTCGATAA AGTGACGACT ATTATTCAGG GCGGC.ACGT CTTCAGTCAC CGCGCTGACC GGCTCCACTG AAGAATCGCA GTTCTAA
Aceaml308      GGTACAGGTT ACTTCGATAA AGTGACGACT ATTATTCAGG GCGGC.ACGT CTTCAGTCAC CGCGCTGACC GGCTCCACTG AAGAATCGCA GTTCTAA
```

```
Macfadden    1 ATGAAAACCC GTACACAACA AATTGAAGAA TTACAGAAAG AGTGGACTCA ACCGCGTTGG GAAGGCATTA CTCGCCCATA CAGTGCGGAA GATGTGGTGA AATTACGCGG  110
  Cozzone      ATGAAAACCC GTACACAACA AATTGAAGAA TTACAGAAAG AGTGGACTCA ACCGCGTTGG GAAGGCATTA CTCGCCCATA CAGTGCGGAA GATGTGGTGA AATTACGCGG
    Byrne      ATGAAAACCC GTACACAACA AATTGAAGAA TTACAGAAAG AGTGGACTCA ACCGCGTTGG GAAGGCATTA CTCGCCCATA CAGTGCGGAA GATGTGGTGA AATTACGCGG
Aceaml308      ATGAAAACCC GTACACAACA AATTGAAGAA TTACAGAAAG AGTGGACTCA ACCGCGTTGG GAAGGCATTA CTCGCCCATA CAGTGCGGAA GATGTGGTGA AATTACGCGG

Macfadden  111 TTCAGTCAAT CCTGAATGCA CGCTGGCGCA ACTGGGCGCA GCGAAAATGT GGCGTCTGCT GCACGGTGAG TCGAAAAAAG GCTACATCAA CAGCCTCGGC GCACTGACTG  220
  Cozzone      TTCAGTCAAT CCTGAATGCA CGCTGGCGCA ACTGGGCGCA GCGAAAATGT GGCGTCTGCT GCACGGTGAG TCGAAAAAAG GCTACATCAA CAGCCTCGGC GCACTGACTG
    Byrne      TTCAGTCAAT CCTGAATGCA CGCTGGCGCA ACTGGGCGCA GCGAAAATGT GGCGTCTGCT GCACGGTGAG TCGAAAAAAG GCTACATCAA CAGCCTCGGC GCACTGACTG
Aceaml308      TTCAGTCAAT CCTGAATGCA CGCTGGCGCA ACTGGGCGCA GCGAAAATGT GGCGTCTGCT GCACGGTGAG TCGAAAAAAG GCTACATCAA CAGCCTCGGC GCACTGACTG

Macfadden  221 GCGGTCAGGC GCTGCAACAG GCGAAAGCGG GTATTGAAGC AGTCTATCTG TCGGGATGGC AGGTAGCGGC GGACGCTAAC CTGGCGGCCA GCATGTATCC GGATCAGTCG  330
  Cozzone      GCGGTCAGGC GCTGCAACAG GCGAAAGCGG GTATTGAAGC AGTCTATCTG TCGGGATGGC AGGTAGCGGC GGACGCTAA. CTGGCGGCCA GCATGTATCC GGATCAGTCG
    Byrne      GCGGTCAGGC GCTGCAACAG GCGAAAGCGG GTATTGAAGC AGTCTATCTG TCGGGATGGC AGGTAGCGGC GGACGCTAAC CTGGCGGCCA GCATGTATCC GGATCAGTCG
Aceaml308      GCGGTCAGGC GCTGCAACAG GCGAAAGCGG GTATTGAAGC AGTCTATCTG TCGGGATGGC AGGTAGCGGC GGACGCTAAC CTGGCGGCCA GCATGTATCC GGATCAGTCG

Macfadden  331 CTCTATCCGG CAAACTCGGT G.CCAGCTGT GGTGGAGCGG ATCAACAACA CCTTCCGTCG TGCCGATCAG ATCCAATGGT CCGCGGGCAT TGAGCCGGGC GATCCGCGCT  440
  Cozzone      CTCTATCCGG CAAACTCGGT GTCCAGCTGT GGTGGAGCGG ATCAACAACA CCTTCCGTCG TGCCGATCAG ATCCAATGGT CCGCGGGCAT TGAGCCGGGC GATCCGCGCT
    Byrne      CTCTATCCGG CAAACTCGGT G.CCAGCTGT GGTGGAGCGG ATCAACAACA CCTTCCGTCG TGCCGATCAG ATCCAATGGT CCGCGGGCAT TGAGCCGGGC GATCCGCGCT
Aceaml308      CTCTATCCGG CAAACTCGGT G.CCAGCTGT GGTGGAGCGG ATCAACAACA CCTTCCGTCG TGCCGATCAG ATCCAATGGT CCGCGGGCAT TGAGCCGGGC GATCCGCGCT

Macfadden  441 ATGTCGATTA CTTCCTGCCG ATCGTTGCCG ATGCGGAAGC CGGTTTTGGC GGTGTCCTGA ATGCCTTTGA ACTGATGAAA GCGATGATTG AAGCCGGTGC AGCGGCAGTT  550
  Cozzone      ATGTCGATTA CTTCCTGCCG ATCGTTGCCG ATGCGGAAGC CGGTTTTGGC GGTGTCCTGA ATGCCTTTGA ACTGATGAAA GCGATGATTG AAGCCGGTGC AGCGGCAGTT
    Byrne      ATGTCGATTA CTTCCTGCCG ATCGTTGCCG ATGCGGAAGC CGGTTTTGGC GGTGTCCTGA ATGCCTTTGA ACTGATGAAA GCGATGATTG AAGCCGGTGC AGCGGCAGTT
Aceaml308      ATGTCGATTA CTTCCTGCCG ATCGTTGCCG ATGCGGAAGC CGGTTTTGGC GGTGTCCTGA ATGCCTTTGA ACTGATGAAA GCGATGATTG AAGCCGGTGC AGCGGCAGTT

Macfadden  551 CACTTCGAAG ATCAGCTGGC GTCAGTGAAG AAATGCGGTC ACATGGGCGG CAAAGTTTTA GTGCCAACTC AGGAAGCTAT TCAGAAACTG GTCGCGGCGC GTCTGGCAGC  660
  Cozzone      CACTTCGAAG ATCAGCTGGC GTCAGTGAAG AAATGCGGTC ACATGGGCGG CAAAGTTTTA GTGCCAACTC AGGAAGCTAT TCAGAAACTG GTGCCGGCGC GTCTGGCAGC
    Byrne      CACTTCGAAG ATCAGCTGGC GTCAGTGAAG AAATGCGGTC ACATGGGCGG CAAAGTTTTA GTGCCAACTC AGGAAGCTAT TCAGAAACTG GTCGCGGCGC GTCTGGCAGC
Aceaml308      CACTTCGAAG ATCAGCTGGC GTCAGTGAAG AAATGCGGTC ACATGGGCGG CAAAGTTTTA GTGCCAACTC AGGAAGCTAT TCAGAAACTG GTCGCGGCGC GTCTGGCAGC
```

end of *aceA*. In support of the ML308 sequence, four independent deletion clones, one from the sense strand and three from the antisense strand, were used to identify the sequence of this region, as well as several clones on the sense strand which had undergone rearrangement elsewhere in the sequence. The other *E. coli aceA* sequences are identical to the ML308 sequence over this region (Klumpp *et al.*, 1988; Rieul *et al.*, 1988; Byrne, EMBL library, accession number Pir:S05692). Comparing the predicted amino acid sequence of the ML308 ICL to ICL from *B. napus*, *R. communis*, *G. hirsutum* and *C. tropicalis*, indicates that the ML308 sequence is probably correct (figure 3.17; section 3.4.3). The plant and yeast ICLs do not have a His at the carboxy-terminus, suggesting that a His at the carboxy-terminus is not essential for activity.

Each of the differences occurs in only one of the sequences and all the other sequences are in agreement over the region where there is a difference, indicating that the differences are probably due to sequence errors, strain variation or mutations occurring during cloning. Since both the Cozzone and the MacFadden sequences are from *E. coli* K12, it seems unlikely that the differences are due to strain variation. Unfortunately, the strain from which Byrne obtained the sequence was not given. No information is given about the sequencing strategy used by either Byrne's or Cozzone's groups, however, MacFadden's group sequenced templates from several different Exo III deletion breakpoints covering the region at the carboxy-terminus, where there is a difference, on both strands. Therefore, it seems unlikely that MacFadden's group made errors in reading the sequence. If the difference is due to spontaneous mutation, then it occurred during the cloning since the same sequence was obtained on several templates on both strands.

3.4.3 Comparison Of The Amino Acid Sequences Of *E. coli* ICL With ICL From Other Organisms:

The genes encoding ICL from *Brassica napus* (Comai *et al.*, 1989), the n-alkane assimilating yeast *Candida tropicalis* (Atomi *et al.*, 1990) castor bean (*Riccinus communis*) (Beeching and Northcote, 1987) and cotton (*Gossypium hirsutum*) (Turley *et al.*, 1990) have all been sequenced. The *E. coli* enzyme shows 41.7% identity with the *R. communis* ICL,

```
          278* *  **  *  *   *        **        *   ****  ******                    *    ***** * *          *        ** 376
      B.napus  AFAQIADLIW METASPDLNE CTQFAEGVKS KTPEVMLAYN LSPSFNWDAS GMTDQQMMEF IPRIARLGYC WQFITLAGFH ADALVVDTFA KDYAR.RGML
C.tropicalis  AFAPYADLIW MESALPDYNQ AKEFADGVKA AVPDQWLAYN LSPSFNWN.K AMPADEQETY IKRLGQLGYV WQFITLAGLH TTALAVDDFA NQYSQ.IGMR
  G.hirsutum  AFAPHADLIW METSSPDMVE CTRFAEGVKS MHPEIMLAYN LSPSFNWDAS GMTDEHMRDF IPRIAKLGFC WQFITLAGFH ADALVTDTFA RDFAR.RGML
   R. communis  AFAPHADIIW METAKPDFAE CTAFAEGVKS MHPEIMLAYN LSPSFNWDAS GMTDEQMRDF IPRIARLGFC WQFITLGGFH ADALVIDTFA KDYAR.RGML
E.coli ML308  AYAPYADLVW CETSTPDLEL ARRFAQAIHA KYPGKLLAYN CSPSFNWQ.K NLDDKTIASF QQQLSDMGYK FQFITLAGIH SMWFNMFDLA NAYAQGEGMK

          377 *    *  *       *      **   *     *        **  *  *       * **                                          434
      B.napus  AYVERIQREE ...RSNGVDT LAHQKWSGAN YYDRYLKTVQ GGISSTAAMG KGVTEEQFKE TWTRPGAAGM GEGTSLVVAK SRM
C.tropicalis  AYGQTVQQPE ...IEKGVEV VKHQKWSGAN YIDGLLRMVS GGVTSTAAMG AGVTEDQFKE TKAKV
  G.hirsutum  AYVEKIQREE ...RNNGVDT LAHQKWSGAN FYDRYLKTVQ GGISSTAAMG KGVTEEQFKE TWTRPGAGNI GSEGNLVVAK ARM
   R. communis  AYVERIQREE ...RKNGVDT LAHQKWSGAN YYDRYLKTVQ GGISSTAAMG KGVTEEQFKE TWTRPGAMEM GSAGSEVVAK ARM
E.coli ML308  HYVEKVQQPE FAAAKDGYTF VSHQQEVGTG YFDKVTTIIQ GGTSSVTALT GSTEESQF
```

Figure 3.17: LINEUP Of ICL Sequences:
Sequences of ICL from *B. napus*, *G. hirsutum*, *R. communis*, *C. tropicalis* and *E. coli* ML308 were aligned using the LINEUP programme of the UWGCG package. The sequences are numbered according to the *E. coli* ML308 sequence and asterisks indicate conserved residues.

```
                          1         *    *    *     *  * * *    * *              *     *             ***         **   85
     B.napus  MAASFSVPSM IMEEEGRFEA EVAEVQTWWS SERFKLTRRP YTARDVVALR GHLKQGYASN EM.AKKLWRT LKSHQANGTA SRTFGALDPV QVTMMAKH.L
C.tropicalis  MAYTKID....INQEEADFQK EVAEIKKWWS EPRWRKTKRI YSAEDIAKKR GTLKIAYPSS QQ.SDKLFKL LEKHDAEKSV SFTFGALDPI HVAQMAKY.L
   G.hirsutum  MAASFSVPSM IMEEEGRFET EVAEVQAWWN SERFKLTRRP YSARDVVALR GSLKQSYGSN EM.AKKLWTT LKTHQANGTA SRTFGALDPV QVTMMAKH.L
   R. communis  MAASFSGPSM IMEEEGRFEA EVAEVQAWWN SERFKLTRRP YTARDVVALR GNLKQSYASN EL.AKKLWRT LKTHQANGTA SRTFGALDPV QVTMMAKH.L
  E.coli ML308  .......... ....MKTRTQ QIEELQKEWT QPRWEGITRP YSAEDVVKLR GSVNPECTLA QLGAAKMWRL LHGESKKGYI N.SLGALTGG QALQQAKAGI

          86  * ****             **   **    **  **         * *           *       *    ** **    * **       * *   * **180
     B.napus  DTIYVSGWQC SSTHTSTNEP GPDLADYPYD TVPNKVEHLF FAQQYHDRKQ REARMSMSRE ERAKTPFVDY LKPIIADGDT GFGGTTATVK LCKLFVERGA
C.tropicalis  DSIYVSGWQC SSTASTSNEP SPDLADYPMD TVPNKVEHLW FAQLFHDRKQ REERLNMTKE ERANTPYIDF LRPIIADADT GHGGITAIIK LTKLFIERGA
   G.hirsutum  DSIYVSGWQC SSTHTTTNEP GPDLADYPYD TVPNKVEHLF FAQQYHDRKQ REARMSMSRE ERARTPYVDY LKPIIADGDT GFGGTTATVK LCKLFVERGA
   R. communis  DSIYVSGWQC SSTHTTTNEP GPDLADYPYD TVPNKVEHLF FAQQYHDRKQ REARMSMSRE ERARTPYVDY LKPIIADGDT GFGGTTATVK LCKLFVERGA
  E.coli ML308  EAVYLSGWQV AADANLAASM YPDQSLYPAN SVPAVVERIN NTFRRADQIQ WSAGI..... EPGDPRYVDY FLPIVADAEA GFGGVLNAFE LMKAMIEAGA

        181*  ****      ******** ****   * *   *** *    *  *     * ****  ** * *    *  *   ** *                        255
     B.napus  AGVHIEDQSS VTKKCGHMAG KVLVAVSEHI NRLVAARLQF DVMGTETVLV ARTDAVAATL IQSNIDSRDH QFILGVTNPS LRGKSLSSLL AEGMAVGNNG
C.tropicalis  AGIHIEDQAP GTKKCGHMAG KVLVPVQEHI NRLVAIRASA DIFGSNLLAV ARTDSEAATL ITSTIDHRDH YFIIGATNPE ..SGDLAALM AEAEAKGIYG
   G.hirsutum  AGVHIEDQSS VTKKCGHMAG KVLVAVSEHI NRLVAARLQF DVMGVETVLV ARTDAVAATL IQTNVDTRDH QFILGATNPN LRGKSLANML AEGMAAGKNG
   R. communis  AGVHIEDQSS VTKKCGHMAG KVLVAISEHI NRLVAARLQF DVMGVETLLV ARTDAEAANL IQSNVDTRDH QFILGVTNPN LRGKSLATLL ATGMANGKTG
  E.coli ML308  AAVHFEDQLA SVKKCGHMGG KVLVPTQEAI QKLVAARLAA DVTGVPTLLV ARTDADAADL ITSDCDPYDS EFITG..... ..........  ..........

        255                                                                                  *   **  *       *  ** 277
     B.napus  PALQAIEDQW LSSARLMTFS DAVVEALKAM NLSENEK.SR RVNEWLNHAR YENCLSNEQG RELAAKLGLK NLFWDWDLPR TREGFYRFQG SVTAAVVRGW
C.tropicalis  DELARIETEW TKKAGLKLFH EAVIDEIKAG NYSNKEALIK KFTDKVNPLS HT...SHKEA KKLAKELTGK DIYFNWDVAR AREGYYRYQG GTQCAVMRGR
   G.hirsutum  PQLQAIEDNW LAIAQLKTFS ECVMDAIKSM NITEDEK.RR RMNEWMNHSS YDKCLSNEQA REIAERLGLQ NLFWDWDLPR TREGFYRFRG SVMAAIVRGW
   R. communis  AELQATEDNW LAMAQLKTFP ECVMDAIKNM NAGEDEK.RR RMNEWMNHTS YDKCLSYEQG REIADRMGLK NLFWDWDLPR TREGFYRFKG SVMAAVVRGR
  E.coli ML308  .......... .......... .......... .......... .......... .......... .......... ........ER TSEGFFRTHA GIEQAISRGL
```

42.2% identity with the *G. hirsutum* enzyme, 42.0% identity with the *B.napus* enzyme and 41.6% identity with ICL from *C.tropicalis* (figure 3.17). This value is above the 25% level which is considered to indicate homology between proteins (Doolittle, 1986). There is extensive identity between the three plant enzymes, all of which show 86-90% identity to one another with the highest identity (90%) between the cotton and castor bean enzymes. The yeast ICL shows between 54-55% identity to the plant enzymes.

As well as indicating residues which may be of importance in an enzyme, comparison of sequences with a high degree of identity can also indicate possible sequencing errors. The C-terminal sequence of the *E. coli* ML308 enzyme is **TEESQF**. This sequence is very similar to the other sequences, indicating that this sequence is likely to be the correct sequence and not the **HRRDRLH** predicted by MacFadden's group. There is no sequence similar to that predicted by MacFadden's group in any of the other C-terminal regions.

Considering the differences between the *E. coli* ML308 sequence and the Cozzone sequence, the similar analysis reveals that there may be sequencing errors in the Cozzone sequence. When comparing the *E. coli* ML308 sequences to all three plant ICLs and the yeast ICL, there are several conserved residues; Pro107, Asp108, Tyr112, Pro113 andVal117 are all conserved. The same region in the Cozzone sequence corresponds to the region where this sequence differs from the other *E. coli* sequences and does not show conservation of any of the above residues. Alanine 215 in the ML308 sequence is also conserved in all the other ICL sequences whereas at the same position in Cozzone's sequence there is a proline. This is further indication that the sequence of the *E. coli* ML308 *aceA* gene is correct.

There are several regions of considerable identity between the *E. coli* enzyme and the other ICLs. There is a large block of identity between amino acids 141 and 241 in the *E. coli* ML308 sequence and where there are differences, many of them are conservative changes. There are also several smaller blocks of identity. There are conserved sequences, **YxSGWQ**, (position 89-94 in the *E. coli* ML308 sequence), **LAYNxSPSFNW** (position 314-324) and **QFITLxGxH** (position 349-356) where the gaps are conservative changes, except for a Cys instead of a Leu at position 318. An important conserved region is the region around the active site cysteine residue, Cys195 in the *E. coli* enzyme;

GAAxxHxEDQxxxxKKCGHMxGKVLVxxxExIxxLVAxR. It is clear from the high degree of identity between all of the ICL sequences that ICL has been highly conserved through evolution and that the active site is very highly conserved, even though the eukaryotic ICLs isolated so far have a much higher subunit M_r than the *E. coli* enzyme.

The major difference between the *E. coli* ICL and the other three is the presence of an insert, of 103 amino acids in the plant and 98 in the yeast, after Gly 255. There are also extensions at the N-terminus and the C-terminus.

Between the three plant enzymes, the region of greatest variation is the insert, although there are differences scattered throughout the length of the proteins. This is also reflected in the secondary structure predictions using the PEPTIDE STRUCTURE and PLOT STRUCTURE programmes of the UWGCG package (data not shown). The N-terminus of all three plant ICL enzymes contains the sequence **MAASFSxPSMIxEE**. The C-terminal extension is also very similar. Although the yeast enzyme has a shorter N-terminal extension and also a shorter C-terminal extension, although they are similar in sequence to the plant extensions.

Since ICL is found in glyoxysomes in eukaryotic organisms, but there is no compartmentation in *E. coli*, it seems likely that the signal for sequestration of ICL into glyoxysomes may be present in the amino-terminus, the carboxy-terminus or the insert sequence. Also there seems to be no post-translational processing of ICL during import in to the glyoxysome, so the signal for uptake of ICL into these organelles must be encoded in the mature peptide (Riezman *et al.*, 1980; Roberts and Lord, 1981; Lazarow and Fujiki, 1985; Borst, 1986). The peptide signal for uptake of proteins into peroxisomes was identified to be **Ser-Lys/His-Leu** at the carboxy-terminus (Gould *et al.*, 1988). The carboxy-terminal tripeptide of the *B.napus* ICL is **Ser-Arg-Met**, which is similar to that for peroxisomal proteins. However, the other plant ICLs have **Ala-Arg-Met** at the carboxy-terminus. It is therefore unlikely that this is the signal for uptake of ICL into glyoxysomes. Within the insert, the sequence **Ala-Thr-Leu** in the *R. communis* amino acid sequence (position 285-287) is the same as a region within the uricase, soybean nodulin 35 polypeptide, which has been indicated to be part of the signal for uptake into peroxisomes (Volokita and Somerville, 1987). This sequence in the soybean enzyme corresponds to **Phe-Leu-Thr-Leu** in cucumber malate synthase and spinach glycolate oxidase. However, this sequence is similar to **Phe-Ile-Thr-Leu** which is

conserved between all the ICL sequences, including *E. coli*, which has no glyoxysomes. It has been suggested that two clusters of basic residues, each cluster within a five residue stretch, could be the signal for uptake of proteins into glyoxysomes (Wierenga *et al.*, 1987). Within the insert, there is a cluster of basic residues in the *R. communis* and *G. hirsutum* ICL of **Lys-Arg-Arg-Arg**, and in the *B. napus* ICL, **Lys-Ser-Arg-Arg**. This or any of the other sequences may be the signal for uptake into glycosomes. More information on enzymes present in glyoxysomes is required in order to identify the sequence responsible for uptake into these organelles.

3.5 A Consensus Isocitrate Binding Site Between ICL And ICDH:

3.5.1 Comparison Of ICL And ICDH From *E. coli*:

While the reactions catalysed by ICL and ICDH are quite different, and ICDH contains an unusual NADP-binding domain, both enzymes bind isocitrate and it seemed interesting to investigate whether these competing enzymes could conceivably possess similar isocitrate binding sites, structure or sequence motifs although the residues in contact with isocitrate in ICDH are dispersed.

The gene encoding ICDH, *icd*, has now been cloned and sequenced (Thorsness and Koshland, 1987). When the amino acid sequences of ICL and ICDH are aligned using the BESTFIT program of the UWGCG package, there is 14% identity over the full length of the enzymes. Between positions 159-237 in ICL and 158-233 in ICDH there are 14 identities and 20 conservative changes with only two gaps over the 81 residues. Several of the conserved residues are also conserved in all the other ICL enzymes. At position 192-194 in the ICL sequence, the tri-peptide **Val-Lys-Lys** is conserved between the ICL and ICDH from *E. coli*. The plant enzymes all have the sequence **Val-Thr-Lys-Lys** and the *C.tropicalis* ICL has the sequence **Gly-Thr-Lys-Lys** at the same position, suggesting that the double Lys could be important for isocitrate binding or the structure of the active site since the next residue in ICL is the active site Cys. At position 213-214 in the ICL sequence, the di-peptide **Leu-Val** is conserved between all the other ICL sequences and the ICDH. Other residues which are conserved between ICDH and all the ICL sequences are Gly164, Lys173, His197 and Ala216

in the *E. coli* ICL sequence. These residues are close to the active site Cys195 in *E. coli* ICL, which was identified by two independent groups by chemical modification studies using iodoacetate (Nimmo *et al.*, 1989) and 3-bromopyruvate (Ko and McFadden, 1990).

3.5.2 Secondary Structure Analysis Of ICL And ICDH:

The secondary structure of all the ICL enzymes and ICDH were predicted using the PEPTIDE STRUCTURE and PLOT STRUCTURE programs of the GCG package.This program makes use of the two methods to predict the secondary structure devised by Garnier *et al.* (1978) and Chou and Fasman (1978).

Between positions 159-237 in the ICL sequence and 158-233 in the ICDH sequence, the secondary structures of both enzymes, predicted by both the Chou and Fasman and the Garnier methods are similar. The predicted secondary structure for ICDH over this region by the Garnier method contains a long α-helix, followed by a short turn and another α-helix. At the corresponding region in the ICL sequence, there are two α-helices, a short region of random coil, followed by a β-strand, another α-helix and a β-strand. The position and length of the α-helices in both enzymes are very similar and the β-strands in the ICL correspond to turns in the ICDH. Similarly, by the Chou and Fasman method the ICDH structure over this region is a long α-helix, a short β-strand, an α-helix of similar length, two turns a short β-strand and an α-helix, while for ICL, this method predicts two α-helices, one turn, and a long α-helix. Again, the length of the secondary structures are similar. The main difference between the two methods of predicting the structure is the the prediction of turns; where one method predicts a turn, the other predicts a β-strand. Since in both methods, the positions of the α-helices are similar, the secondary structures could be similar over this region. The secondary structure predictions from the PEPTIDE STRUCTURE and PLOTSTRUCTURE programs of the UWGCG package are shown in figure 3.18.

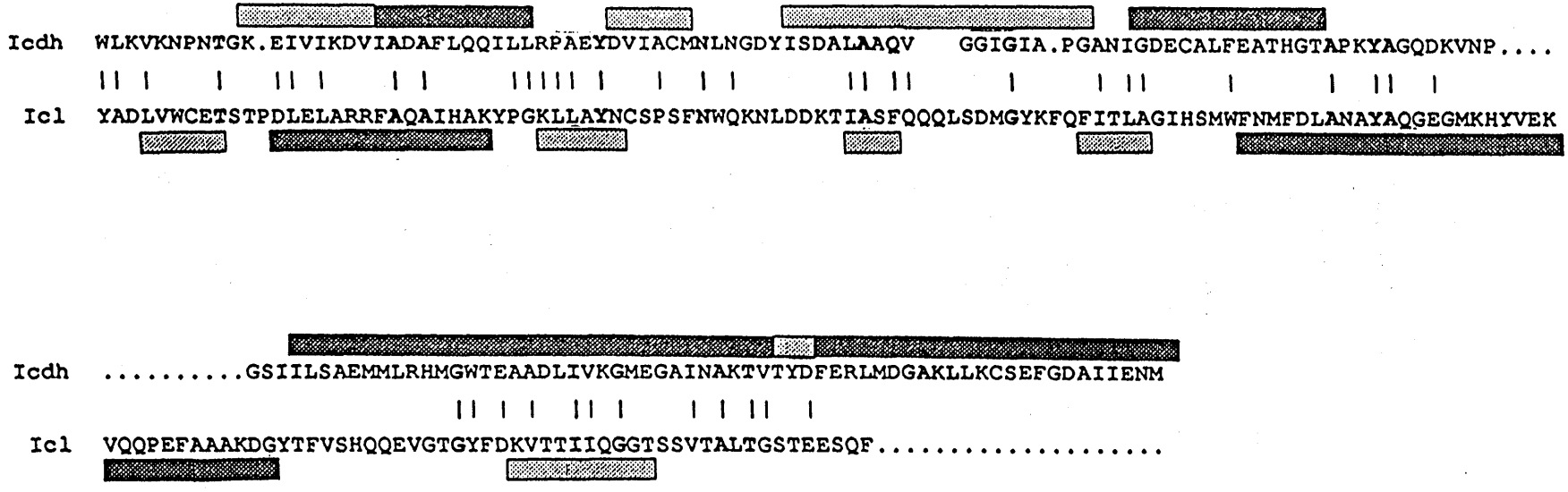
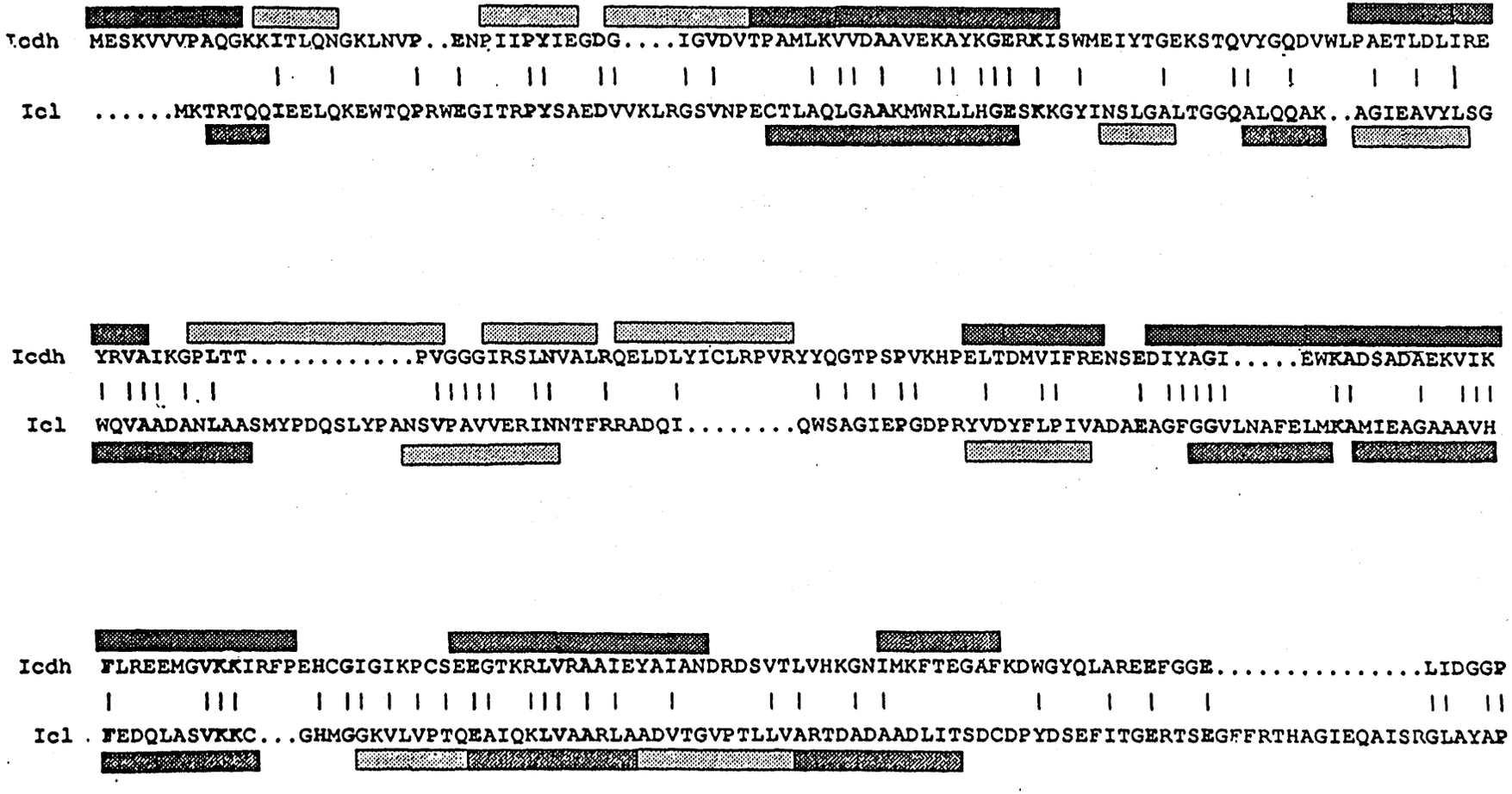


Figure 3.18: Comparison Of The Derived Amino Acid Sequence And Secondary Structure Prediction Of ICL With ICDH

Alignment of amino acid sequences of *E. coli* ICL and *E. coli* ICDH. Residues which are identical are indicated by bold print with lines connecting also in bold print. Lines connect positions where there are conservative changes. β-strands and α-helices are indicated by light and dark shaded boxes respectively. The secondary structures of were predicted by the Garnier method of the PEPTIDESTRUCTURE programme of the UWGCG package.


Icdh MESKVVVPAQGKKITLQNGKLNVP..ENPIIPYIEGDG....IGVDVTPAMLKVVDAAVEKAYKGERKISWMEIYTGEKSTQVYGQDVWLPAETLDLIRE
Icl ......MKTRTQQIEELQKEWTQPRWEGITRPYSAEDVVKLRGSVNPECTLAQLGAAKMWRLLHGESKKGYINSLGALTGGQALQQAK..AGIEAVYLSG
Icdh YRVAIKGPLTT.....PVGGGIRSLNVALRQELDLYICLRPVRYYQGTPSPVKHPELTDMVIFRENSEDIYAGI.....EWKADSADAEKVIK
Icl WQVAADANLAASMYPDQSLYPANSVPAVVERINNTFRRADQI.....QWSAGIEPGDPRYVDYFLPIVADAEAGFGGVLNAFELMKAMIEAGAAAVH
Icdh FLREEMGVKKIRFPEHCGIGIKPCSEEGTKRLVRAAIEYAIANDRDSVTLVHKGNIMKFTEGAFKDWGYQLAREEFGGE.....LIDGGP
Icl FEDQLASVKKC...GHMGGKVLVPTQEAIQKLVAARLAADVTGVPTLLVARTDADAADLITSDCDPYDSEFITGERTSEGFFRTHAGIEQAISRGLAYAP

3.5.3 Comparison Of The Predicted ICL Secondary Structure With The Actual Structure Of ICDH:

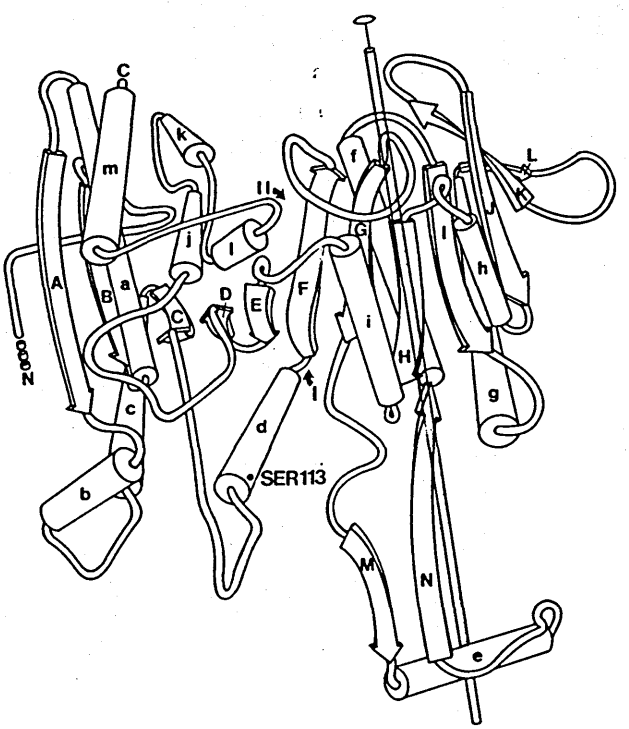
The three dimensional structure of ICDH has now been solved (Hurley *et al.*, 1989) and is shown in figure 3.19a. ICDH contains three domains: a large α+β domain, residues 1-124 and 318-416 (where the α-helices and β-strands do not alternate), a smaller α/β domain, residues 125-157 (where the α-helices and β-strands alternate), both of which are built around a 12-stranded β-sheet and an α/β clasp-like domain which involves both subunits.

The clasp domain consists of residues 158-202 from both subunits (figure 3.19b). The region of greatest sequence similarity between ICDH and ICL covers these residues 158-202 and also helix-f (figure 3.19a). When the predicted ICL structure is aligned with the actual structure of ICDH (figure 3.20), it can be seen that they are similar over the region 159-237 in the ICL sequence and 158-233 in the ICDH sequence. There is an α-helix in the ICL corresponding to helix-e in ICDH, a β-strand corresponding to strand-N and another α-helix corresponding to helix-f. At strand-M, however, the corresponding predicted structure for ICL is an α-helix.

Assuming that the two structures are similar, then the active site Cys195 of ICL, which is at the C-terminus of the helix corresponding to helix-e, will point into the front pocket on the other subunit, thus allowing it to interact with the isocitrate molecule which was shown to bind in this pocket (Hurley *et al.*, 1990a).

The residues which are involved in binding isocitrate and coordinating the Mg^{2+} in ICDH are Ser113, which is the residue phosphorylated, Asn115, Arg119, Arg129, Tyr160, Lys230, Arg153, which all coordinate isocitrate and Asp307, Asp283, Asp311, which all coordinate the Mg^{2+}. Ser113 lies at the N-terminus of helix-d and Asn115 and Arg119 also form part of this helix. In ICL the predicted structure corresponding to helix-d is a β-strand, however, the predicted structure for ICDH is also a β-strand. Of course, where the secondary predictions for ICDH are incorrect but similar to the predicted structure for ICL, the secondary structure of ICL may in fact be similar to that of ICDH. Ser113 which forms a hydrogen bond with the carboxyl group at C5 of isocitrate is not conserved in ICL, although the corresponding residue is an Arg in the *E. coli* enzyme and a His in all the other ICLs.

A)



B)

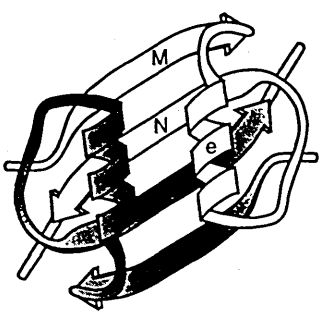


Figure 3.19: Three Dimensional Structure Of ICDH

A) Diagram of the ICDH monomer, taken from Hurley *et al.* (1989). Helices are labelled in alphabetical order as they occur in the chain while the β-strands are labelled as they occur in the β-sheet. The front and rear pockets are indicated by I and II respectively. The rod represents the crystallographic twofold axis of symmetry relating the two subunits.

B) Daigram of the clasp domain, looking down the crystallographic twofold axis of symmetry. The shaded part represents the second subunit.

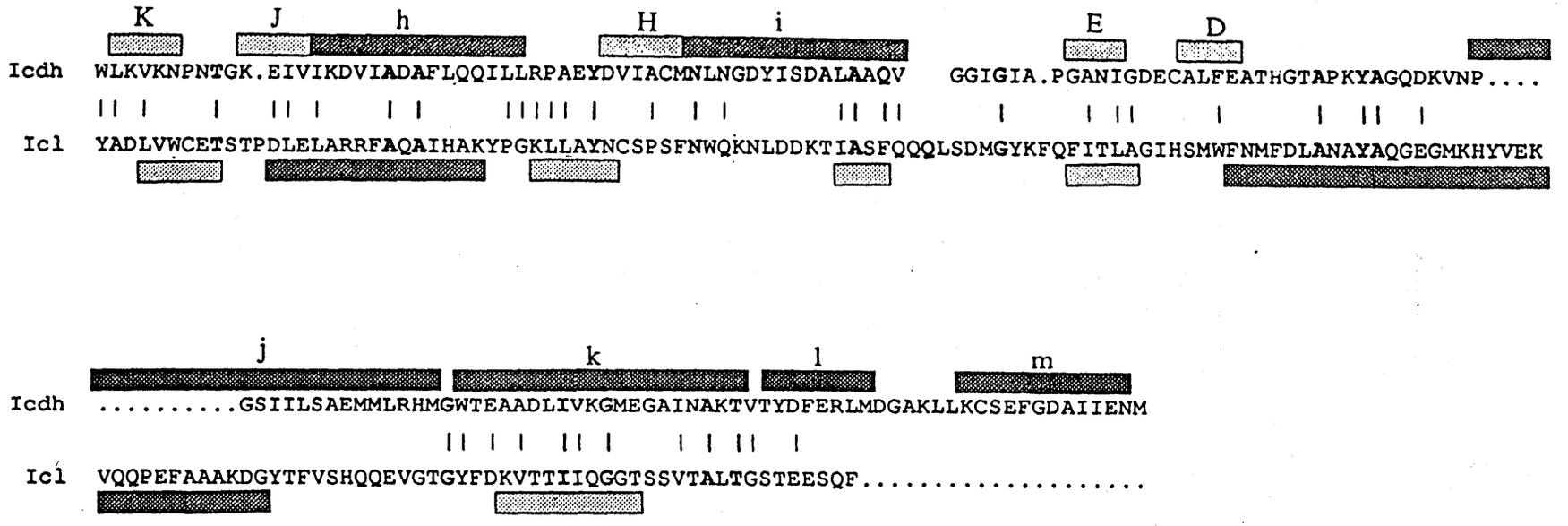
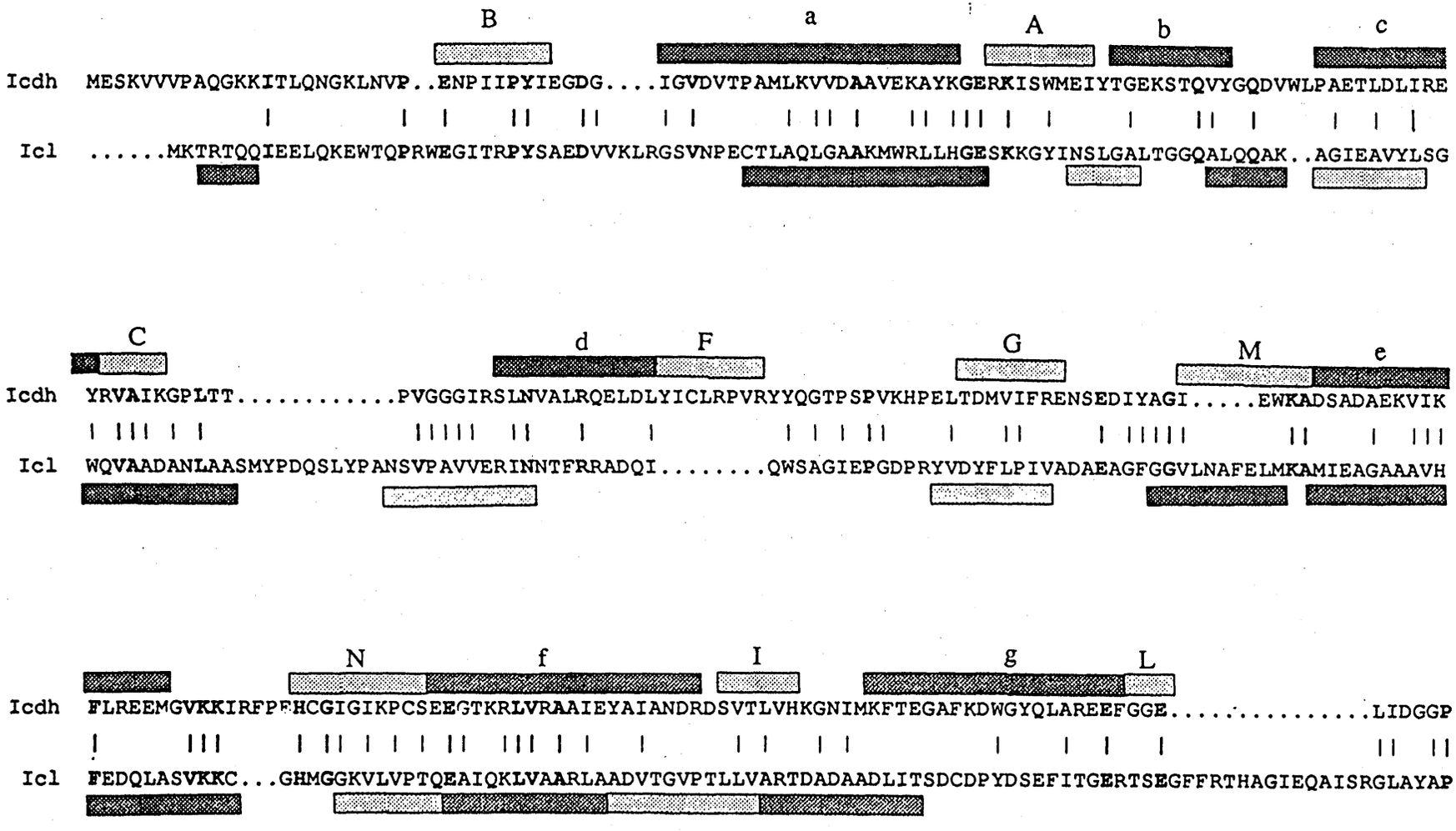


Figure 3.20: Comparison Of The Derived Amino Acid Sequence And Secondary Structure Prediction Of ICL With The Actual Secondary Structure Of ICDH

Alignment of amino acid sequences of *E. coli* ICL and *E. coli* ICDH. Residues which are identical are indicated by bold print with lines connecting also in bold print. Lines connect positions where there are conservative changes. β-strands and α-helices are indicated by light and dark shaded boxes and labelled with capital letters and lowercase letters respectively. The actual secondary structure of ICDH was taken from Hurley *et al.* (1989) while the secondary structure of ICL was predicted by the Garnier method of the PEPTIDESTRUCTURE programme of the UWGCG package.


B
a
A
b
c
Icdh MESKVVVPAQGKKITLQNGKLNVP..ENPIIPYIEGDG....IGVDVTPAMLKVVDAAVEKAYKGERKISWMEIYTGEKSTQVYGQDVWLPAETLDLIRE
Icl ......MKTRTQQIEELQKEWTQPRWEGITRPYSAEDVVKLRGSVNPECTLAQLGAAKMWRLLHGESKKGYINSLGALTGGQALQQAK..AGIEAVYLSG
C
d
F
G
M
e
Icdh YRVAIKGPLTT.....PVGGGIRSLNVALRQELDLYICLRPVRYYQGTPSPVKHPELTDMVIFRENSEDIYAGI.....EWKADSADAEKVIK
Icl WQVAADANLAASMYPDQSLYPANSVPAVVERINNTFRRADQI.....QWSAGIEPGDPRYVDYFLPIVADAEAGFGGVLNAFELMKAMIEAGAAAVH
N
f
I
g
L
Icdh FLREEMGVKKIRFPFHCGIGIKPCSEEGTKRLVRAAIEYAIANDRDSVTLVHKGNIMKFTEGAFKDWGYQLAREEFGGE.....LIDGGP
Icl FEDQLASVKKC...GHMGGKVLVPTQEAIQKLVAARLAADVTGVPTLLVARTDADAADLITSDCDPYDSEFITGERTSEGFFRTHAGIEQAISRGLAYAP

Therefore there are conservative changes at this position. Asn115 and Arg119 are conserved in the *E. coli* ICL but not in the other ICLs, Asp corresponding to Asn115 and His corresponding to Arg119. Although an Asp at position 115 would result in charge repulsion, a His at position 119 could interact with isocitrate in a similar manner to Arg. Other substitutions of isocitrate binding residues are Asp283 in ICDH to Gln in the *E. coli* ICL and either Glu or Asp in the other ICLs and Asp311 to a Lys in *E. coli* ICL, changing the charge, and either a Glu, Asp or Gln in the other ICLs. Tyr160 in ICDH corresponds to a Phe in all ICLs. Therefore, the above residues in the ICLs which are either conserved or are conservative changes when compared to ICDH may be important for isocitrate binding and where there are differences, these may account for the difference in the K_m of ICL and ICDH for isocitrate. Site-directed mutagenesis of these residues could confirm whether or not these residues are important or not.

Although the secondary structure predictions for ICL from *E. coli* are similar to the actual structure of ICDH, this can only be speculation and whether or not a consensus isocitrate binding site exists between ICDH and ICL will not be answered until the three dimensional structure of ICL is solved.

3.6 Discussion:

The evidence presented in this chapter shows that the *aceA* gene of *E. coli* ML308 is indeed in the BamHI-AvaI fragment of pEM9. When induced by the addition of IPTG, KAT-1/ pAR9023 expresses a protein of the same molecular weight as ICL, concomitant with an increase in ICL activity. The nucleotide sequence of the 2.3kb insert has the capacity to encode a 434 amino acid protein of molecular weight 47,491, similar to the reported molecular weight of ICL, 45,000 - 48,000 (1988; Robertson and Reeves, 1987; MacKintosh and Nimmo). The nucleotide sequence and derived amino acid sequence of this reading frame is very similar to the other *E. coli* ICLs and the amino acid sequence is also similar to the amino acid sequences of ICL from other organisms. The amino acid sequence at the amino-terminus is very similar to the sequence obtained by direct peptide sequencing of the amino-terminus of ICL from *E. coli* ML308 (MacKintosh, 1987). The difference between these sequences is an Asn residue at

position 8 in the peptide sequence while there is an Ile residue at the same postion in the derived amino acid sequence. The reason for this difference is not clear. However, for all the *E. coli* ICLs sequences, there is an Ile residue at this postition. Also, several templates and several clones were used to determine the nucleotide sequence of this region on both strands. The derived amino acid sequence also contains the sequence around the active site cysteine (Nimmo *et al.*, 1989; Ko and McFadden, 1990) and identify this cysteine residue as Cys195.

Since ICL activity is expressed from pAR9023 with the only open reading frame long enough, extending over 1301bp, encoding a peptide of molecular weight 47,491, these results confirm that ICL contains only one type of subunit.

By placing the *aceA* gene under the control of the *tac* promoter of pGLW11, it was hoped that the contribution of ICL to the control of flux through the central metabolic pathways, during growth on different carbon sources could be measured. Similar systems have been used for this purpose in the past. Walsh and Koshland (1985) placed the gene encoding citrate synthase from *E. coli* under the control of the *tac* promoter in a vector containing the *lacI*^{q} gene. By growing the host cells containing this plasmid in the presence of different concentrations of IPTG, they found that citrate synthase was expressed at different specific activities. Using this system, they reported that citrate synthase was rate limiting for flux through the TCA cycle and the glyoxylate bypass as well as for growth, during growth on acetate but not glucose. Using a similar system, LaPorte *et al.* (1985) showed that the ICDH kinase/phosphatase could compensate for high levels of ICDH expressed from the *tac* promoter by phosphorylating ICDH to a higher degree, keeping the level of active ICDH constant.

In both of the above examples, the level of the enzyme enzyme expressed by the *tac* promoter was dependent on the concentration of IPTG; in the case of citrate synthase, the specific activity could be altered by altering the IPTG concentration. However, expression of ICL from *E. coli* KAT-1/pAR9023 did not give similar results. Different IPTG concentrations did not result in different levels of specific activity of ICL. Instead, at any given IPTG concentration, the ICL activity increased linearly with time (figure 3.3), hence the rate of increase in specific activity of ICL and not the specific activity itself was dependent on the concentration of IPTG (figure 3.4).

As the culture grows, the total number of plasmid molecules increases, increasing the number of the *aceA* genes. This results in increased expression in the presence of IPTG, which binds to the *lac* repressor, preventing it from binding to the *lac* operator and repressing transcription from the *tac* promoter. However since the constitutive *lacI* gene is also present on the vector, the total number of repressor molecules also increases. The vector pGLW11 was reported to contain the up-mutation in the promoter of the *lacI* gene, $lacI^q$ (Calos, 1978), thus expressing higher levels of the repressor (I.S. Hunter, personal communication). Therefore, as the culture grows, the IPTG, which is present at a constant concentration, should eventually be bound to *lac* repressor so that there is excess *lac* repressor. This should result in repression of transcription from the *tac* promoter and control the specific activity of ICL. However, it appears that the *lac* repressor is not present at sufficiently high levels in *E. coli* KAT-1/ pAR9023 for controlled expression of ICL. The nucleotide sequence of the promoter upstream of the *lacI* gene in the vector, pGLW11, was examined (C. Boyd, personal comunication) for the presence of the mutation in $lacI^q$ which causes increased expression of the repressor. It was found that pGLW11 did not contain the $lacI^q$ gene, as reported, but the wild type *lacI* gene. This presumably explains why the expression of ICL could not be controlled in a similar manner to that reported for citrate synthase (Walsh and Koshland, 1985). However, this system allows very high expression of ICL (figure 3.5) and thus aids purification of large quantities of the enzyme (section 4.4).

The amino-terminus of ICL from several organisms, *Ps. indigofera* (McFadden *et al.*, 1968), *L. usitattissimum* (Kahn and McFadden, 1982) and *E. coli* (Matsuoka and McFadden, 1988; Rieul *et al.*, 1988) has been identified as methionine. The carboxy-terminus from *N. crassa*(Johanson *et al.*, 1974b) *Ps. indigofera* (McFadden *et al.*, 1968) and flax (*L. usitattissimum*) (Kahn and McFadden, 1982) has been shown to be histidine, while from a thermophillic *Bacillus* the carboxy-terminus is a threonine (Chell *et al.*, 1978) and DNA sequencing showed the carboxy-terminus to be methionine in *R. communis* (Beeching and Northcote, 1987). The amino acid composition has been calculated for ICL from *Ps. indigofera* (Shiio *et al.*, 1965), thermophillic *Bacillus* (Chell *et al.*, 1978), *N. crassa* (Johanson *et al.*, 1974b), *C. tropicalis* (Uchida *et al.*, 1986), *P. pinea* (Pinzauti *et al.*, 1986), *L. usitattissimum* (Khan and McFadden, 1982) and watermelon (*Citrullus vulgaris*)

(Jameel *et al.*, 1984). The relative amino acid compositions of the above enzymes and from the derived amino acid sequence of the *R. communis* ICL were used to calculate their relatedness to each other using the Marchalonis and Weltman (1971) method (Vanni *et al.*, 1990). The SΔQ values between ICL from all of the above organisms were all less than 50, which indicates that the enzymes are related. The inhibition patterns of several of the ICL enzymes are similar. Also cysteine has been reported to be at or near the active site in ICL from *N. crassa*, *Ps. indigofera*, *E. coli*, *L. usitattissimum* and *Citrullus vulgaris*. The amino acid sequence around the active site cysteine has been determined (Nimmo *et al.*, 1989) for the *E. coli* ICL and confirmed by Ko and McFadden (1990) to be **CGHMGGK** and this sequence is very highly conserved in ICLs sequenced to date. Because of the similarities between the ICL from different sources, it seems likely that they are evolutionarily related, even though the prokaryotic ICL has a subunit M_r of 45,000-48,000 and the eukaryotic ICL has a subunit M_r of 65,000-68,000.

Comparison of the ML308 derived amino acid sequence of ICL with those from *B.napus*, *R. communis*, *G.hirsutum* and *C.tropicalis* indicates a high degree of identity, confirming the suggestion that all ICL enzymes share a common active site structure despite the difference in the molecular weights of the prokaryotic and eukaryotic enzymes. There are two regions of high identity, with the *E. coli* ICL sequence split by a large insert in the plant and yeast enzymes (figure 3.17). There are also extensions at the carboxy- and amino-termini. Within the regions of high identity, there are several regions which show very high identity (figure 3.17), suggesting that these regions are important and could form part of the active site. Since the kinetics of both the forward and reverse reactions are similar between the prokaryotic and eukaryotic ICLs, as are the kinetics of inhibition and residues at the active site, it appears that during evolution, the structural gene encoding ICL has aquired an insert without altering the structure of the active site.

The results presented in this chapter show that ICL from a wide range of sources, plant, fungi and bacteria, is very highly conserved. Because of the very high degree of identity between the various ICLs the sequence data alone does not provide any likely candidates for important residues. However in conjunction with chemical modification experiments it will

provide information on which residues to change using site-directed mutagenesis.

As mentioned above, the sequence around the active site cysteine has been determined and comparing this to the *E. coli* sequence, this residue was identified as Cys195. This residue is conserved between all ICLs sequenced so far (figure 3.17) as is the sequence around this residue. Using site directed mutagenesis, the role of Cys195 in either catalysis or substrate binding can be investigated.

His has been implicated as an active site residue in ICL from a number of organisms, *L. usitattissimum* (Khan and McFadden, 1982), *Citrullus vulgaris* (Jameel *et al.*, 1985), and *E. coli* (Ko *et al.*, 1991). The sequence around the modified His is not known as yet for ICL from *L. usitattissimum* and *Citrullus vulgaris*. In *E. coli* it has recently been reported that there are two His residues which react with DEPC in the *E. coli* enzyme, His266 and His306 (Ko *et al.*, 1991). However work done here in Glasgow indicates that only His306 reacts with DEPC and that this residue is unlikely to be at or near the active site (Rua, Robertson and Nimmo, submitted). His306 is not conserved in any of the other ICL sequences, all of which place a Lys at this position, the K_s values for protection against inactivation by DEPC. This does not however prove that there are no His residues in the active site, only that DEPC does not react with any residues in the active site. There are eight His residues in the *E. coli* ICL, but since only four are conserved, at positions 184, 197, 356 and 399, it would not be a difficult task to use site directed mutagenesis to change all of these residues and identify which is involved in binding of substrate or catalysis.

Itaconate epoxide, which is an analogue of succinate, has been used as a modifying agent for ICL from *Citrullus vulgaris* (Jameel *et al.*, 1985). The inactivation caused by itaconate epoxide is protected against by itaconate, a competitive inhibitor with respect to succinate and an uncompetitive inhibitor with respect to glyoxylate, succinate and isocitrate. Using hydroxylamine to reactivate the modified ICL, it was shown that an ester was formed between ICL and itaconate epoxide, indicating that the group modified was a carboxylate of either Glu or Asp. Since succinate and itaconate protect against modification by itaconate epoxide, the authors suggest that the residue modified is at the succinate sub-site. They also suggest that the residue modified could be involved in coordinating the Mg^{2+}. There are nine conserved Glu residues, at positions 10, 122, 141, 186, 208, 260, 289, 386 and 431, and

thirteen conserved Asp residues, at positions 31, 108, 132, 149, 157, 187, 221, 234, 246, 249, 284, 294 and 409. Since there are so many Asp and Glu residues, it would be easier to identify the modified residue by isolating and sequencing a peptide, labelled by reacting ICL with radioactive itaconate epoxide and digested with a protease. Site directed mutagenesis could then be used to clarify the role of this residue in catalysis or substrate binding. Since the active site Cys residue has been identified as Cys195, this residue is an obvious candidate for site directed mutagenesis and this is the topic of the next chapter.

Chapter 4
Site Directed
Mutagenesis Of ICL

4.1 Introduction:

There have been several studies which have identified possible active site residues in ICL which may be important in catalysis and/or substrate binding. Inhibition of ICL from *L. usitatissimum* (Khan and McFadden, 1982), *Citrullis vulgaris* (Jameel *et al.*, 1985) and *Zea mays* (Igamberdiev and Zemlyanukin, 1987) with diethylpyrocarbonate (DEPC) implicated an active site histidine, since isocitrate and the products offered protection against inhibition. Itaconate epoxide, an analogue of the competitive inhibitor of succinate binding, itaconate, implicated the carboxylate group of either aspartate or glutamate at the active site of the *Citrullis vulgaris* enzyme (Jameel *et al.*, 1985). Ko and McFadden (1990) identified two His residues in the *E. coli* enzyme which reacted with DEPC, His266 and His306. It seems unlikely that either of these residues is at the active site, since work done here in Glasgow indicated that only His306 was modified by DEPC and the K_s values for protection against DEPC inactivation by the substrates and products are much higher than the respective K_m values (Rua, Robertson and Nimmo, submitted).

Through studies with sulphydryl reagents and sulphydryl-modifying reagents, ICL from many sources has been shown to have an essential cysteine residue (section 1.3.10). The glyoxylate analogue, 3-bromopyruvate (3-BrP) was shown to inhibit ICL from *P. indigofera* (McFadden *et al.*, 1968; Roche and McFadden, 1969; Roche *et al.*, 1971), *N. crassa* (Johanson *et al.*, 1974b), *L. usitatissimum* (Khan and McFadden, 1982) and *Citrullis vulgaris* (Jameel *et al.*, 1985) following saturation kinetics with 3-BrP reversibly binding to a specific site prior to alkylation of the enzyme. Isocitrate and the products, glyoxylate and succinate, offered protection against this inactivation, indicating that the reactive group was at or near the active site.

A more detailed study was carried out on the *E. coli* ML308 enzyme by our group (Nimmo *et al.*, 1989) using iodoacetate. Iodoacetate inhibited ICL from this organism in a pseudo-first-order process resulting in carboxymethylation of one cysteine residue per subunit. Isocitrate, glyoxylate and succinate protected against inhibition by iodoacetate with K_s values similar to their respective K_m values. Also, the decrease in reactivity of the enzyme with iodoacetate over the pH range 8.1-6.1 was very similar to the decrease in the K_m for isocitrate

over the same pH range. This strongly suggests that the reactive group is in the active site. The carboxymethylated peptide was isolated and sequenced after tryptic digestion. The amino acid sequence of this peptide was CMCys-Gly-His-Gly-Gly-Lys. Comparison of this sequence with the derived amino acid sequence of ICL from *E. coli* ML308 (figure 3.13) indicated that the modified residue was Cys195. This residue and many of the surrounding residues are conserved between *E. coli* ML308, *E. coli* K12 (Matsuoka and McFadden, 1988; Rieul *et al.*, 1988), *Riccinus communis* (Beeching and Northcote, 1987), *Brassica napus* (Comai *et al.*, 1989), *Candida tropicalis* (Atomi *et al.*, 1990) and *Gossypium hirsutum* (Turley *et al.*, 1990) (figure 3.17). This suggests that Cys195 may be an active site residue and as such may play an important role in either catalysis, substrate binding or both in all of the above ICLs.

Based on the chemical modification of ICL from *P. indigofera* by 3-BrP and protection by isocitrate and products, glyoxylate and succinate, Roche *et al.* (1971) tentatively proposed a catalytic mechanism indicating a role for an active site Cys in protonating the C3 of isocitrate. However, an alternative mechanism has been proposed by Vanni *et al.* (1990) in which no role for an active site Cys has been indicated, instead, these authors prefer a mechanism in which a histidine protonates the C3 of isocitrate.

The catalytic mechanism proposed by Roche *et al.* (1971) indicates that the S-H group of the active site cysteine residue in the *P. indigofera* ICL donates a proton to the C3 of isocitrate during the cleavage reaction and accepts a proton from succinate in the condensation reaction. Therefore, the ionization of the active site Cys residue is critical for catalysis, if this catalytic mechanism is correct. It was for this reason that it was decided to change Cys195 in the *E. coli* ML308 ICL to a serine, thus replacing the S-H group with a O-H group which should have a higher pK_a and an alanine, thus replacing the O-H with -H which results in an enzyme with no ionizable group at position 195.

4.2 Site Directed Mutagenesis Of ICL:

4.2.1 Strategy:

The strategy used was based on the method of Eckstein and co-workers (Taylor *et al.*, 1985a; Taylor *et al.*, 1985b; Nakamaye and Eckstein, 1986; Sayers *et al.*, 1988) and involves annealing an oligonucleotide carrying the specific mutation to a single stranded template. Klenow polymerase is then used to synthesise double stranded DNA which is then ligated to form a closed, circular duplex. Any excess single stranded template DNA is then removed by binding to a nitrocellulose filter in conditions which allow free passage of double stranded DNA. During the extension step, dCTP is replaced with dCTPαS so that all the newly synthesised DNA contains dCMPαS instead of dCMP. The restriction enzyme, *Nci* I , which cannot cut a strand which contains dCMPαS in its recognition site, is then used to digest the duplex. Since Nci I cannot cleave the strand with dCMPαS, this results in the duplex being nicked in the template strand while the newly synthesised strand, containing the mutation remains intact. The nicks in the template strand are then extended using Exo III which digests in a 3'-5' direction. The Nci I restriction site is positioned 3' of the end of the M13 polylinker on the template strand so that the wild type template of the insert in the polylinker is quickly deleted. The Exo III digestion is not carried through to completion so that a double stranded region remains, most likely in the M13 sequence, and can be used as a primer for repolymerisation. DNA polymerase I is then used to extend the double stranded region using the intact strand, carrying the mutation, as the template. This results in the newly synthesised double stranded DNA carrying the specific mutation. The double stranded DNA is then ligated and used to transform competent host cells. This strategy is outlined in figure 4.1 and described in detail in section 2.9.

4.2.2 Design Of Oligonucleotides For The Mutagenesis Reaction:

For the most efficient annealing of the oligonucleotide carrying the mutation to the template, the mutation should be towards the middle, at least 3 nucleotides from the 3' end and several bases from the 5' end. This is to avoid the 3'-5' exonuclease activity of the Klenow polymerase or any contaminating exonuclease activity from deleting the mismatch within the

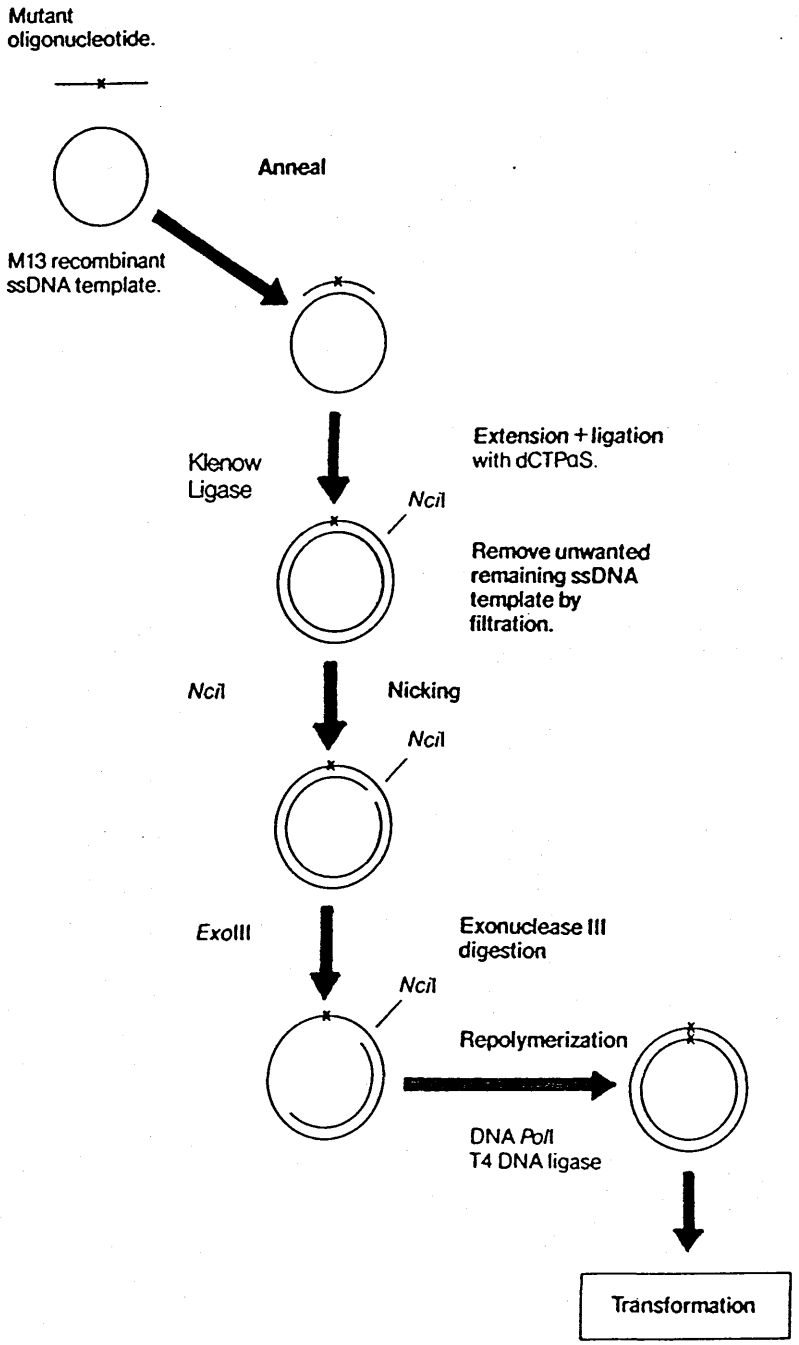


Figure 4.1: Site Directed Mutagenesis Scheme

Single stranded template is represented by a single circle while double stranded DNA is represented by a double circle. The asterisk indicates the position of the point mutation.

oligonucleotide during the extension step. Figure 4.2 shows the nucleotide sequence and the corresponding amino acid sequence around Cys195. Also shown are the sequences of the two oligonucleotides, each 25 nucleotides long with the codon for Cys195, TGC, changed to TCC, Ser, and GCC, Ala in oligonucleotides 654 and 655 respectively. These oligonucleotides should anneal to the template, pAR24 at positions 838-862 in the nucleotide sequence of the insert.

4.2.3 The Mutagenesis Reaction:

Oligonucleotides 654 and 655 were phosphorylated at their 5' ends using T4 polynucleotide kinase in order to provide a 5'-phosphate for ligation to form the closed duplex. The oligonucleotides were then annealed to the single stranded template, pAR24, and then treated as described above (section 4.2.1) and in figure 4.1.

In conjunction with the mutagenesis of *aceA* by oligonucleotides 654 and 655 on pAR24, a control mutagenesis reaction was set up. The control reaction uses a mutated M13mp8 which has a stop codon at position 6298, within the *lacZ'* gene resulting in no β-galactosidase activity, and an oligonucleotide which changes the base at position 6298 back to the wild type, restoring β-galactosidase activity. Thus, the reaction can be evaluated by the reversion of white plaques to blue in the control.

Samples were taken at each step of the mutagenesis reaction, after extension and ligation, filtration and nicking, Exo III deletion and after repolymerisation. These samples were then run on an agarose gel to asess each step of the reaction (figure 4.3).

For all three reactions, sample 1, taken after extension and ligation, shows the presence of single stranded, nicked and covalently closed circular DNA. For the 654 and 655 reactions, there appears to be a large proportion of nicked and single stranded DNA. The excess single stranded DNA may be due to the annealing conditions, while the large proportion of nicked DNA indicates that much of the oligonucleotide is not phosphorylated, however, the control oligonucleotide is phosphorylated but also shows a large proportion of nicked DNA at this stage. The results of sample 2, taken after filtration through nitrocellulose and nicking with NciI, indicates that in all cases, the filtration step removed all of the single stranded template.

```
      Leu Ala Ser Val Lys Lys Cys Gly His Met Gly Gly Lys Val Leu Val
      CTG GCG TCA GTG AAG AAA TGC GGT CAC ATG GGC GGC AAA GTT TTA GTG

Oligo-Ala:        GTG AAG AAA GCC GGT CAC ATG GGC G
                              Ala

Oligo-Ser:        GTG AAG AAA TCC GGT CAC ATG GGC G
                              Ser
```

Figure 4.2: Oligonucleotide Sequences:

The derived amino acid sequence of ICL is shown between Leu189 and Val204 with the nucleotide sequence underneath. The sequence of the oligonucleotides used in the mutagenesis reactions are shown. Where there is a difference in the derived amino acid sequence, the amino acid change is indicated.

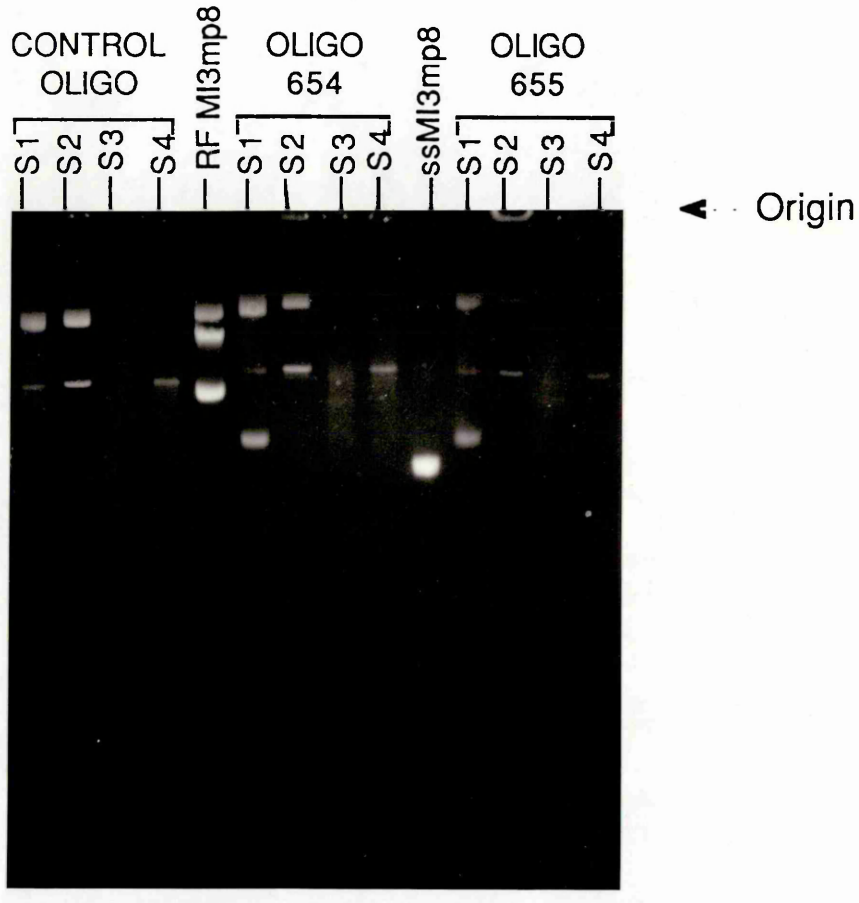


Figure 4.3: Analysis Of Mutagenesis

1.0% agarose gel showing samples from each step of the mutagenesis reactions with the control reaction, and reactions with mutant oligonucleotides 654 and 655. S1-S4 indicate samples taken after extension and ligation; after filtration and nicking with Nci I; after extension of the nick with exonuclease III and after repolymerisation and ligation respectively.

In all reactions, sample 3 shows a smear, indicating efficient digestion by Exo III. Sample 4, taken after repolymerisation shows a single band corresponding to covalently closed circular DNA in all three reactions.

The repolymerised covalently closed circular DNA was then used to transform competent TG-1 cells. The results of the transformation are shown in table 4.1. The reason for the low number of plaques is not known. A control transformation with 1ng of RFM13mp8 resulted in 283 plaques, indicating that the competence of the host cells was high. As can be seen from table 4.1, 33 of the 39 plaques were blue in the control mutagenesis reaction, indicating 84.5% efficiency of mutagenesis.

4.2.4 Nucleotide Sequence Of Mutants:

Single stranded template DNA was prepared from 6 plaques of each of the two mutagenesis reactions, 654 and 655. These were then sequenced using an oligonucleotide, oligo 700, as the primer which anneals to the insert of pAR24 at positions 701-717 (table 4.2). Five of the six templates of the 654 reaction had the Cys195 codon, TGC, changed to TCC, the codon for Ser, at positions 847-849, while all six of the 655 reaction templates had GCC at the same positions, changing the Cys195 to Ala. An example of each sequence is shown in figure 4.4.

In order to rule out any possibility that the oligonucleotides had annealed to pAR24 anywhere other than at the target site in the *aceA* gene, one template was chosen from each mutagenesis reaction, pAR24-654 and pAR24-655, corresponding to reactions with oligonucleotides 654 and 655 respectively, and sequenced again. This was done using oligonucleotides which allowed the complete sequence of the mutated *aceA* genes to be determined when used as primers in the sequencing reactions (table 4.2, figure 4.5).

Apart from the changes at positions 847-849, the target site for site directed mutagenesis, there were no other sequences within the *aceA* gene to which the mutagenic oligonucleotides bound. Also, since when either oligos 654 or 655 were used as primers for sequencing, there were no overlapping sequences on the gel, it is clear that these oligonucleotides annealed at only the predicted site on pAR24.

Table 4.1: Results Of Site Directed Mutagenesis Reactions

Plate	Number Of Colonies	
	White	Blue
No DNA	0	0
Site Directed Mutagenesis With Oligo-654	10	0
Site Directed Mutagenesis With Oligo-655	8	0
RF M13mp8	0	283
RF pAR24	316	0
Control Site Directed Mutagenesis	6	33

Table 4.2: Oligonucleotide Primers For Sequencing *aceA* Mutants

Oligonucleotide	Sequence	Position To Which Oligonucleotide Anneals On pAR24
Oligo-230	CGTAAACCACCACATAA	231- 247
Oligo-480	ACTGGCGGTCAGGCGCT	481- 497
Oligo-700	GCTATGTCGATTACTTC	701- 717
Oligo-1090	GCCTGGCGTATGCGCCA	1091-1107
Oligo-1330	ACAGCATGTGGTTCAAC	1331-1347
Oligo-1610	CACGCCGCATCCGGCAA	1611-1627

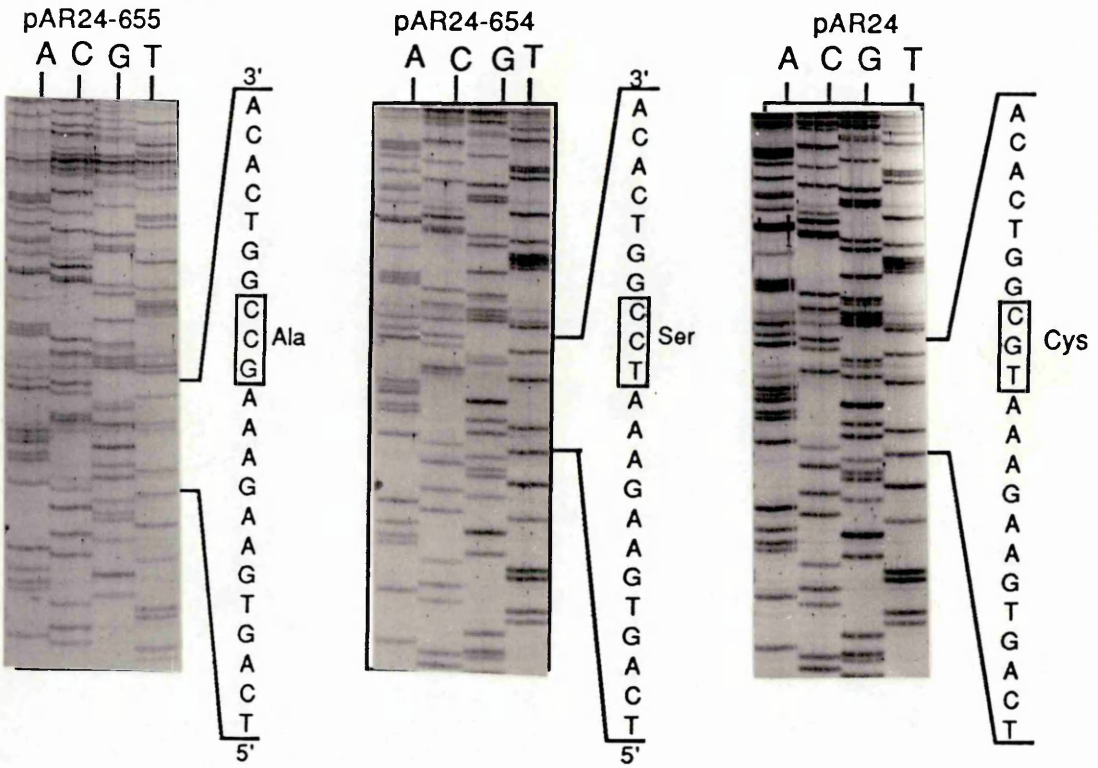


Figure 4.4: Sequencing Of pAR654 And pAR655

Sequencing reactions were carried out according to section 2.7.2, using oligo700 as a primer, and run on a polyacrylamide gel, section 2.7.3. A portion of the sequencing gel is shown for pAR24, pAR654 and pAR655 with the corresponding sequence alongside. The mutated codon is boxed and the resulting amino acid is indicated.

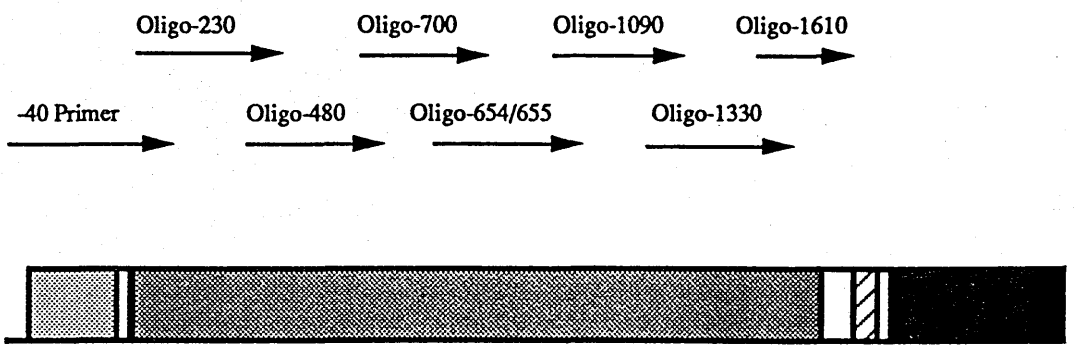


Figure 4.5: Sequencing Strategy For Mutant *aceA* Genes

The 2.3kb insert of pAR24, containing the genes encoding ICL (dark shaded box), MSA (light shaded box) and ICDH kinase/phosphatase (black box) are shown, with the untranslated regions shown in the unshaded boxes. The arrows indicate the direction of second strand synthesis and the length of readable sequence from each oligonucleotide primer. The hatched box indicates the region which was not sequenced by either this strategy or by Exo III deletion.

4.3 Expression Of Mutant *aceA* Gene Produts, ICLSer195 And ICLAla195:

4.3.1 Cloning Of The Inserts Of pAR24-654 And pAR24-655 Into pGLW11:

In order to characterise the effects of replacing Cys195 with Ser and Ala, the mutant enzymes first had to be expressed and purified. The mutant *aceA* genes are in the wrong orientation with respect to the *lac* promoter in pAR24-654 and pAR24-655. Also, M13 inhibits host cell growth and this makes M13 an inefficient expression vector. Therefore, it was decided to use pGLW11 as the expresson vector. This also had the advantage that the wild type *aceA* gene had been cloned into this vector, pAR9023, and so the activities of ICLSer195 and ICLAla195 could be compared directly to that of the wild type enzyme.

If the inserts of pAR24-654 and pAR24-655 are digested with *Bam*HI and *Eco*RI and ligated into pGLW11 digested with the same enzymes, the *aceA* gene will be in the wrong orientation with respect to the *tac* promoter for expression. An alternative strategy was to digest pAR24-654 and pAR24-655 with *Bam*HI and *Eco*RI, fill in the sticky ends using Klenow polymerase and ligate into pGLW11 digested with SmaI in the same way that pAR9023 was subcloned (section 3.2.1). The correct orientation could then be checked by mapping the *Pvu*II sites. When this was tried, the only colonies obtained did not contain any insert and corresponded to religated pGLW11.

To get round this problem of relatively inefficient ligation of blunt ended insert into blunt ended vector compared with sticky end ligation, pAR24-654 and pAR24-655 were digested with *Bam*HI and *Eco*RI, the ends filled in using Klenow polymerase and the 2.3kb fragment ligated into pUC19 digested with Sma I after agarose gel purification (section 2.4.6). This strategy had the advantage that recombinant clones could be selected by their white colour after growth of transformed TG-1 cells on L-plates containing ampicillin, IPTG and X-gal. The *Pvu*II sites were then mapped and two clones, pAR19-654 and pAR19-655 were selected which resulted from the ligation of the inserts of pAR24-654 and pAR24-655 into pUC19 respectively. These clones had 1.0kb *Bam*HI-*Pvu*II and 0.6kb *Eco*RI-*Pvu*II fragments in the inserts, so that the *aceA* genes are transcribed in the *Eco*RI-*Bam*HI direction (figure 3.2, section 3.2.1). pAR19-654 and pAR19-655 were then digested with *Bam*HI and *Eco*RI and

ligated into pGLW11 digested with the same enzymes after agarose gel purification of the appropriate fragments (2.3kb for pAR19-654 and pAR19-655 and 4.85kb for pGLW11). This strategy is outlined in figure 4.6. The PvuII sites were then mapped in both resulting plasmids, pAR654 and pAR655, encoding ICLSer195 and ICLAla195 respectively, to check the orientation with respect to the *tac* promoter. These plasmids were then used to transform competent *E. coli* KAT-1 cells.

4.3.2 Expression Of ICLSer195 And ICLAla195 From pAR654 And pAR655 Respectively:

In order to determine whether or not pAR654 or pAR655 expressed an active ICL enzyme, an expression experiment similar to that for KAT-1/pAR9023 was set up (see section 3.2.2).

IPTG was added to a final concentration of 1mM to 100ml cultures of *E. coli* KAT-1/pAR9023, *E. coli* KAT-1/pAR654 and *E. coli* KAT-1/pAR655 growing on L-broth containing ampicillin when the A_{420} was approximately 0.5. Two 1ml samples were taken from each culture at various time intervals. One of the samples was harvested by centrifugation and resuspended in SDS sample buffer containing β-mercaptoethanol, boiled and run on an SDS gel. Figure 4.7 shows a gel of samples from *E. coli* KAT-1/pGLW11, *E. coli* KAT-1/pAR654 and *E. coli* KAT-1/pAR655 at various times after addition of IPTG. The other sample was used for mearurement of A_{420}, sonication and assay of ICL activity. Figure 4.8 shows the effect of time after addition of IPTG on ICL activity relative to A_{420} for the various cultures.

Figure 4.7 shows a band which co-migrates with purified ICL and increases in intensity with time after addition of IPTG for *E. coli* KAT-1/pAR654 and *E. coli* KAT-1/pAR655, but not *E. coli* KAT-1/pGLW11.This suggests that the bands are the result of expression of ICLSer195 and ICLAla195 respectively. Similarly, a band which co-migrated with ICL and increased in intensity with time was observed with *E. coli* KAT-1/pAR9023 (see section 3.2.2 and figure 3.5).

Figure 4.8 shows a graph of ICL activity versus A_{420}. *E. coli* KAT-1/pGLW11 shows little or no ICL activity which remains constant over the time course, while *E. coli*

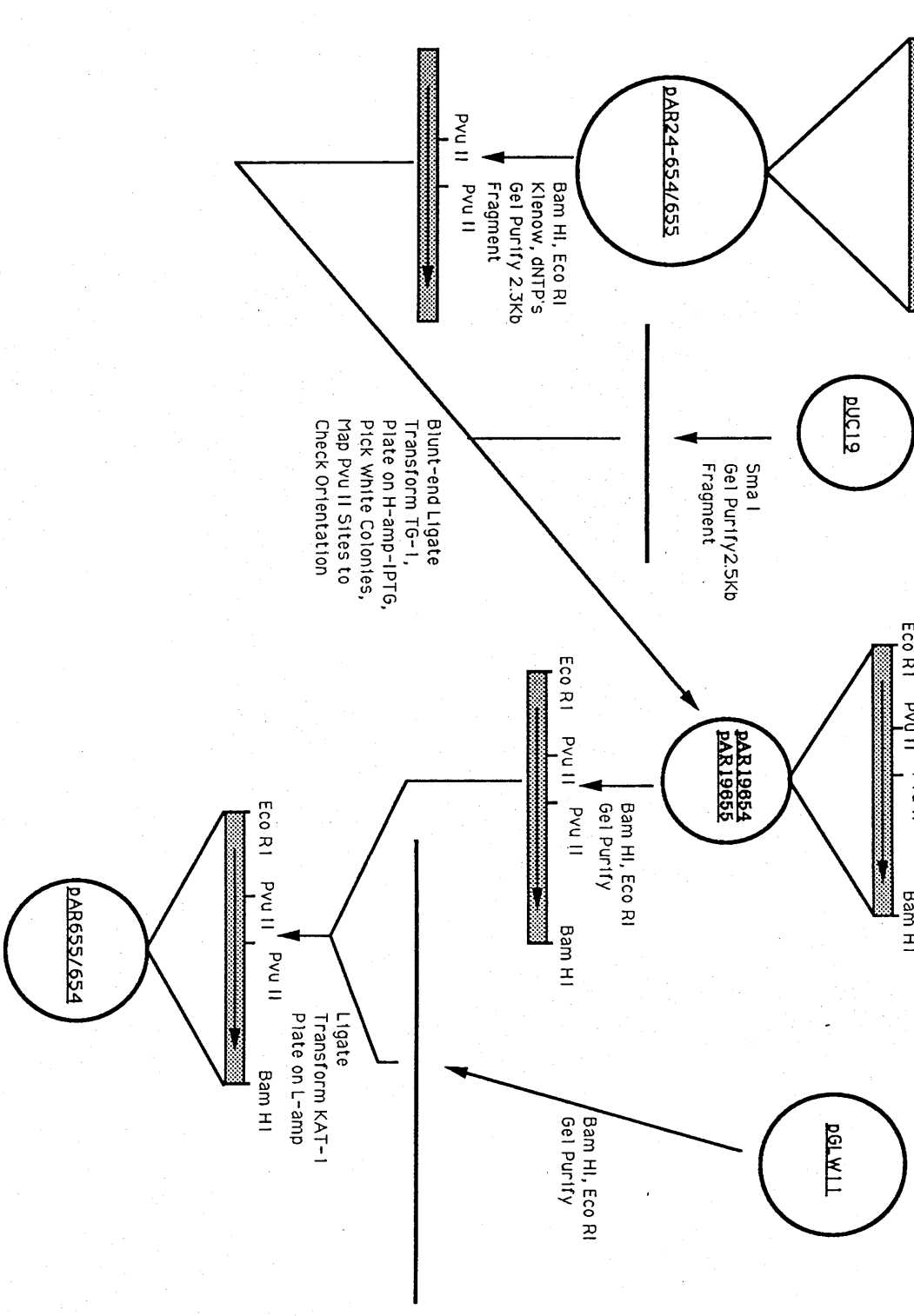


Figure 4.6: Strategy For Constructing Expression Cassettes For Mutant *aceA* Genes

The mutated inserts of pAR24 are represented by the shaded boxes with the arrows indicating the direction of transcription.

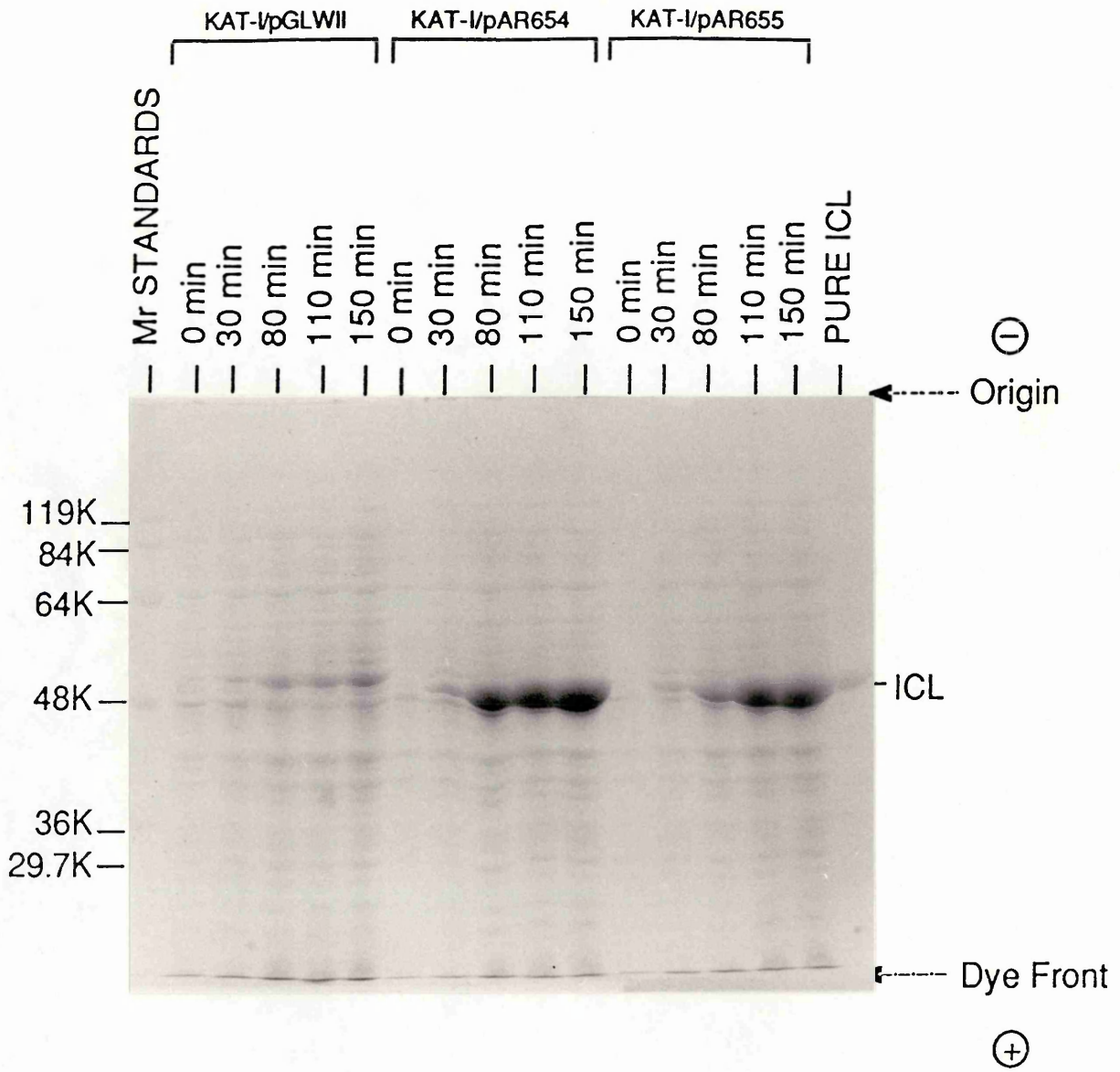


Figure 4.7: Expression Of ICLSer195 And ICLAla195 From *E. coli* KAT-1/ pAR654 And KAT-1/pAR655 Respectively

IPTG and glucose were added to cultures to final concentrations of 1mM and 0.1% (w/v) respectively when the A_{420} was approximately 0.5. 1ml sampes were taken at the time intervals indicated and harvested in a microfuge. The cell pellets were resuspended in 200µl SDS sample buffer and boiled for 5min in the presence of 2-mercaptoethanol and loaded onto a 10% SDS-polyacrylimide gel.

Figure 4.8: Expression Of Mutant ICL Enzymes

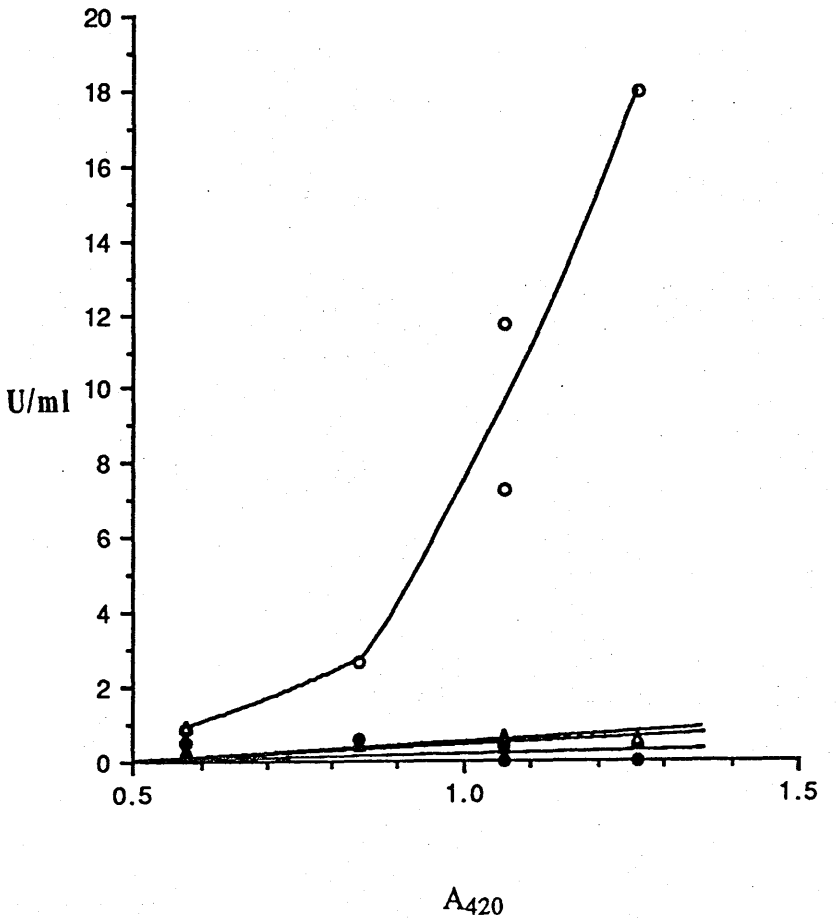
100ml cultures of *E. coli* KAT-1/pGLW11, KAT-1/pAR654 and KAT-1/pAR655 were grown in the presence of Amp (section 2.3.4) until the A_{420} was approximately 0.5. IPTG was then added to a final concentration of 1mM and 1ml samples were taken at 0min., 30min., 60min., 90min. and 120min.. The A_{420} was measured and the samples sonicated (section 2.3.8) and assayed for ICL activity (section 2.12b).

●) *E. coli* KAT-1/pGLW11

○) *E. coli* KAT-1/pAR9023

▲) *E. coli* KAT-1/pAR654

△) *E. coli* KAT-1/pAR655


20
18
16
14
12
10
8
6
4
2
0
U/ml
0.5
1.0
1.5
A420

KAT-1/pAR9023 shows an increase in ICL activity with A_{420} suggesting that this activity is due to expression from the insert of pAR9023. Although *E. coli* KAT-1/pAR654 and *E. coli* KAT-1/pAR655 express a protein which co-migrates with wild type ICL, there is very little ICL activity. The ICL activity, although very low, does increase with A_{420}. It therefore appears that the mutations Cys195 to Ser and Ala, have resulted in a dramatic decrease in ICL activity in the mutant enzymes.

4.4 Purification Of Mutant Enzymes, ICLSer195 And ICLAla195:

4.4.1 Introduction:

In order to characterise ICLSer195 and ICLAla195 in terms of catalytic mechanism and thus determine the role of Cys195 in either catalysis, substrate binding or both, the proteins must first be purified. The kinetic properties of ICLSer195 and ICLAla195 can then be determined and compared to those of the wild type enzyme.

Since ICLSer195 and ICLAla195 have very little activity associated with them (figures 4.7 and 4.8), it would be very difficult to monitor their purification by enzymic assays. It was therefore decided to establish a method of purification for the wild type ICL and to use this procedure to purify the two mutant enzymes and to monitor their purification on SDS gels. Since the mutant enzymes constitute a large proportion of the total cell protein (figure 4.7), it was thought that monitoring the purification of ICLSer195 and ICLAla195 by the presence of an intense band which co-migrates with wild type ICL should be easier than by assaying enzyme activity where the activity might be too little to determine.

4.4.2 Purification Of w.t. ICL From *E. coli* KAT-1/pAR9023:

Initially, the purification procedure developed in this laboratory was employed to purify ICL from *E. coli* KAT-1/pAR9023 after induction with IPTG (Section 2.11.1) (MacKintosh, 1987; MacKintosh and Nimmo, 1988). This procedure involved treatment of the crude extract with protamine sulphate, $(NH_4)_2SO_4$ fractionation, acid treatment, gel filtration through Sephacryl S-300, ion exchange on Phenyl Sepharose and finally Mono Q using FPLC. Many

of the above steps did not give a significant increase in purity, probably due to the fact that a significant proportion of the total cellular protein was ICL (figure 4.7). Indeed, much of the ICL activity had to be discarded at the gel filtration and ion exchange on Phenyl Sepharose, in order to avoid contaminating proteins. For these reasons, it was decided to use a modification of the method used by Conder *et al.* (1988) to purify ICL from an overexpressing strain of *E. coli* K12. This procedure is described in section 2.11.2. Because of the large amount of protein after the $(NH_4)_2SO_4$ fractionation, several runs on the Mono Q column had to be done. Careful pooling of fractions after ion exchange chromatography on Mono Q using the FPLC was essential to avoid contaminants with Mr 80,000 and 41,000, both of which are eluted from the column after ICL and correspond to shoulders on the elution profile (figure 4.9). This purification procedure produced a high yield and specific activity (table 4.3) and could be carried out in less than 24 hours. When dialysed into the 50mM Mops-NaOH pH7.3, 1mM EDTA, 1mM benzamidine, 1mM DTT, 1.2mM PMSF, 40% (v/v) glycerol, the enzyme was stable for at least 6 months when stored at -20°C.

4.4.3 Purification Of ICLSer195 and ICLAla195:

The purification procedure above was used to purify both mutant proteins, ICLSer195 and ICLAla195. In both cases, new Mono Q columns were used in order to rule out completely the possibility of cross contamination. Typical elution profiles of ICLSer195 and ICLAla195 from the Mono Q chromatography are shown in figures 4.10 and 4.11 respectively. Again, careful pooling of samples was essential to avoid the contaminants, however, this was greatly facilitated by the high resolution of the new columns.

Figure 4.12 shows an SDS gel of samples after each stage of the purification of ICLSer195, ICLAla195 and wild type ICL. This gel reveals the presence of 3 contaminating bands with M_r 24,000, 16,000 and 11,500, although the intensity of these bands with respect to ICL is very low. These bands co-purify with ICL and could not be separated from ICL by gel filtration through Sephacryl S-300 superfine,or by chromatography on Phenyl Sepharose or Mono Q. The contaminating bands co-eluted with ICL. Non-denaturing gel electrophoresis revealed only one band in all three enzymes (figure 4.13). This suggests that there could be a sub-population of ICL where one or two peptide bonds in one or more subunits have been

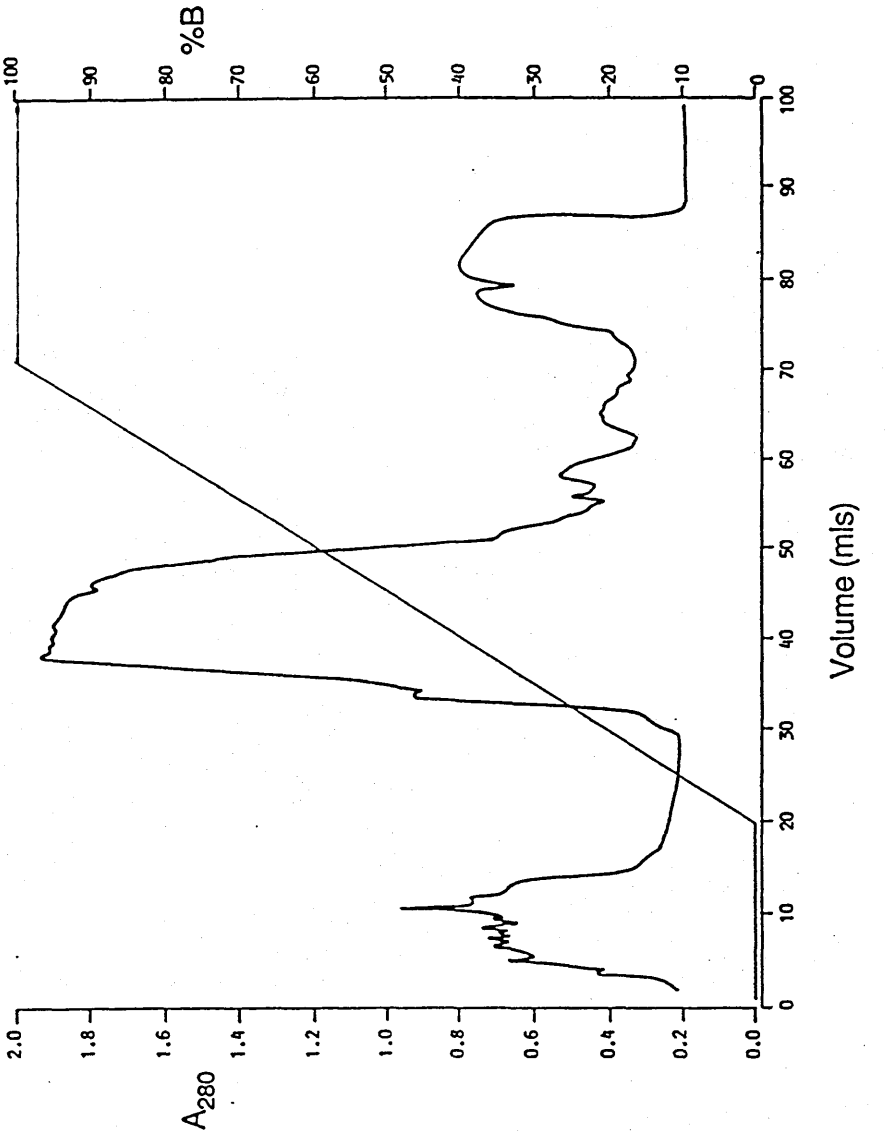


Figure 4.9: Elution Profile Of Wild Type ICL From Mono Q

Chromotography of ICL was carried out as described in section 2.11.2e.

Table 4.3 Purification Of ICL, ICLSer195 And ICLAla195:

Step	Protein (mg)	Enzyme Activity ($\mu mol.min^{-1}$.)	Specific Activity ($\mu mol.min^{-1}.mg^{-1}$.)	Yield (%)	Purification Factor
w.t.ICL:					
Crude Extract	7332.0	22489.5	3.07	100	1.0
$(NH_4)_2SO_4$ 30-45%	605.0	10195.4	16.85	45	5.5
Mono Q	312.5	9645.0	30.86	43	10.1
ICLSer195:					
Crude Extract	1840.0	365.70	0.199	100	1.0
$(NH_4)_2SO_4$ 30-45%	335.0	114.15	0.341	31	1.7
Mono Q	112.0	45.01	0.402	12	2.0
ICLAla195:					
Crude Extract	465.0	144.75	0.311	100	1.0
$(NH_4)_2SO_4$ 30-45%	313.5	117.36	0.374	81	1.2
Mono Q	160.8	103.18	0.642	71	2.1

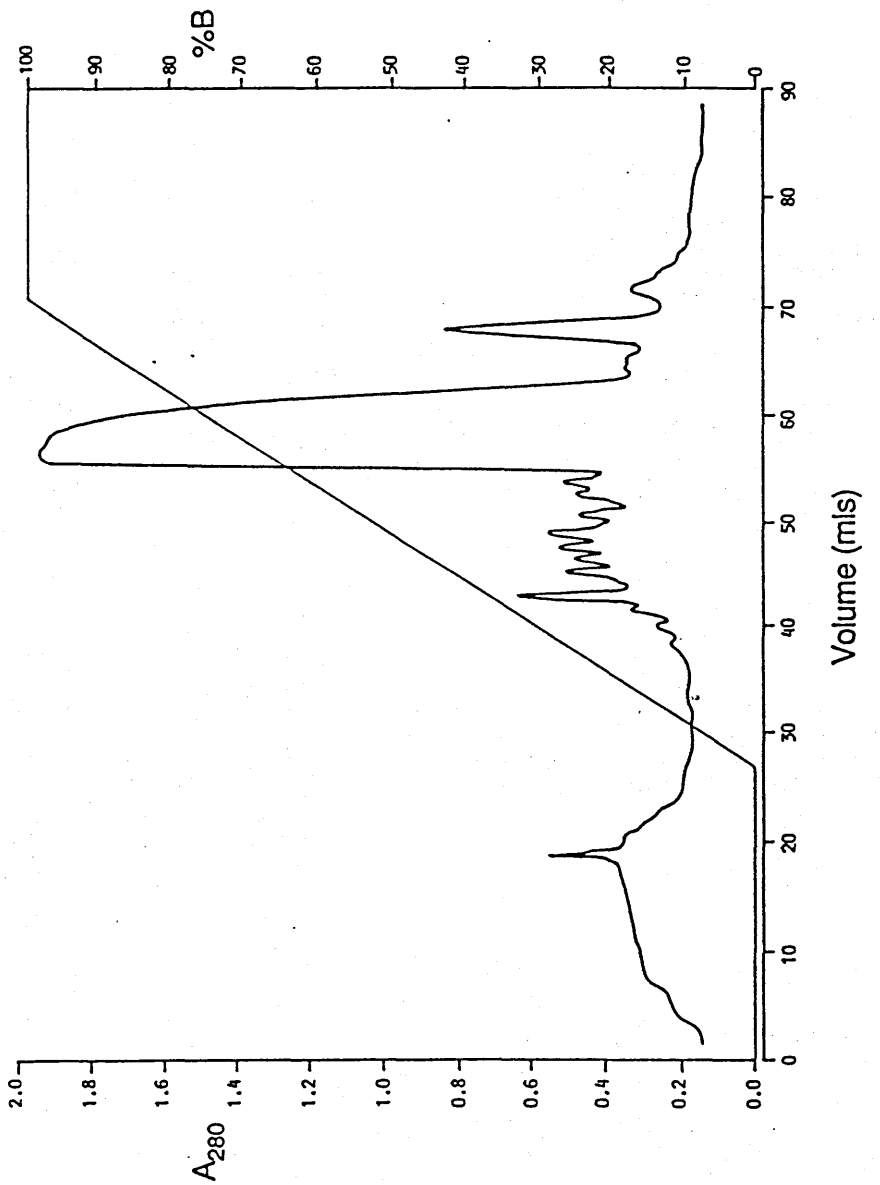


Figure 4.10: Elution Profile Of ICLSer195 From Mono Q

Chromotography of ICLSer195 was carried out as described in section 2.11.2e.

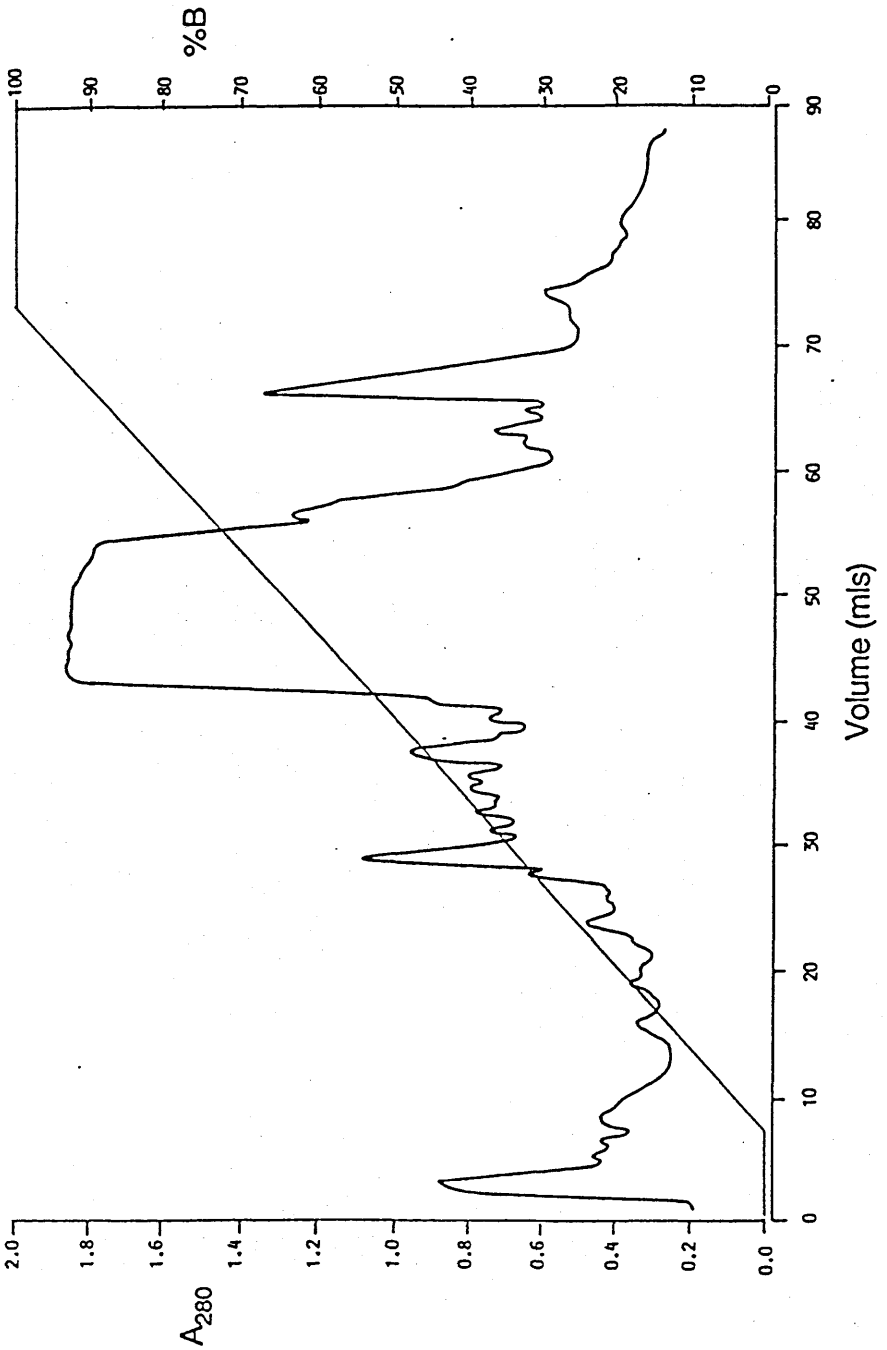


Figure 4.11: Elution Profile Of ICLAla195 From Mono Q

Chromotography of ICLAla195 was carried out as described in section 2.11.2e.

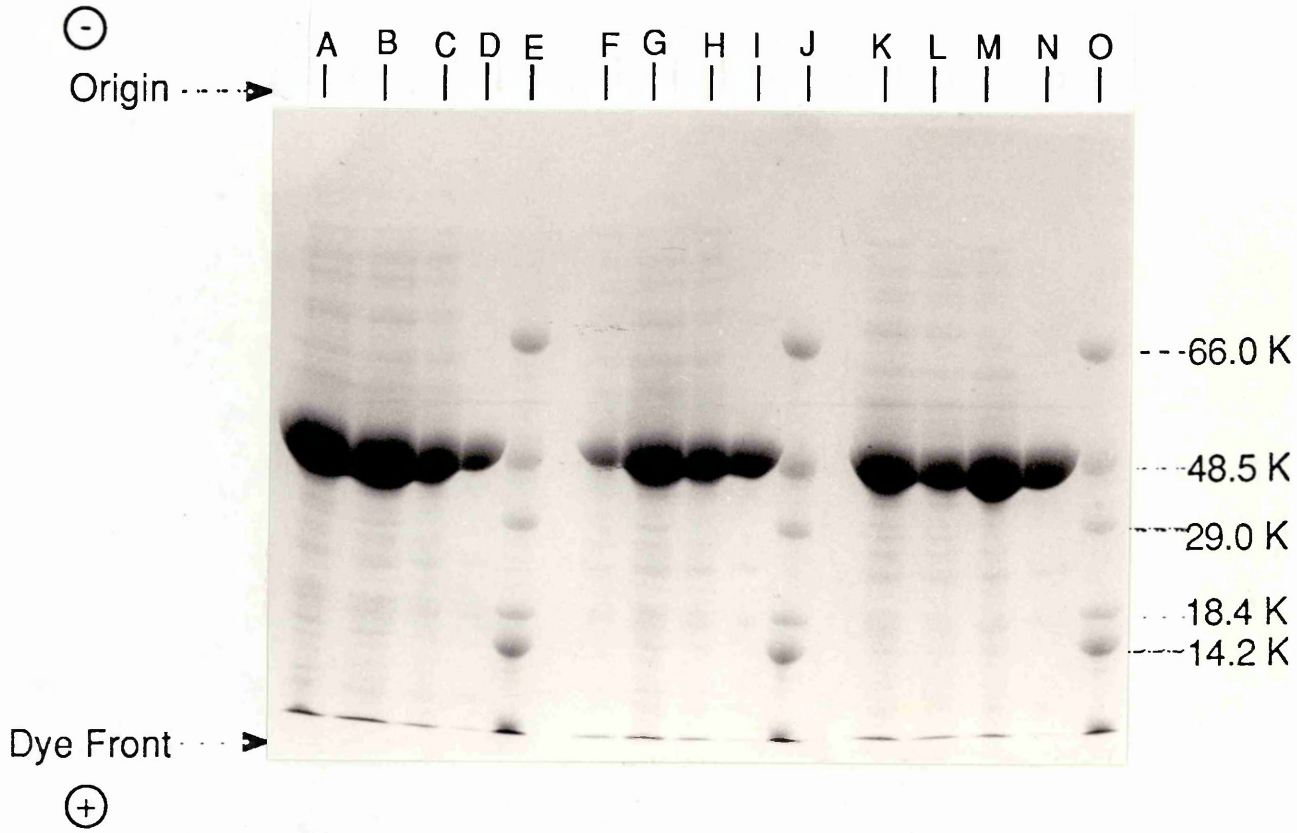


Figure 4.12: Purification Gel Of ICL, ICLSer195 and ICLAla195

10% polyacrylamide gel run in the presence of SDS monitors the purification of ICL, ICLSer195 and ICLAla195 (section 2.11.2).

Lane A) wt ICL, Crude extract; B) wt ICL, Ammonium sulphate 0-30% supernatant; C) wt ICL, Ammonium sulphate 30-45% pellet; D) wt ICL, Mono Q pool; E) M_r standard proteins; F) ICLSer195, Crude extract; G) ICLSer195, Ammonium sulphate 0-30% supernatant; H) ICLSer195, Ammonium sulphate 30-45% pellet; I) ICLSer195, Mono Q pool; J) M_r standard proteins; K) ICLAla195, Crude extract; L) ICLAla195, Ammonium sulphate 0-30% supernatant; M) ICLAla195, Ammonium sulphate 30-45% pellet; N) ICLAla195, Mono Q pool; O) M_r standard proteins.

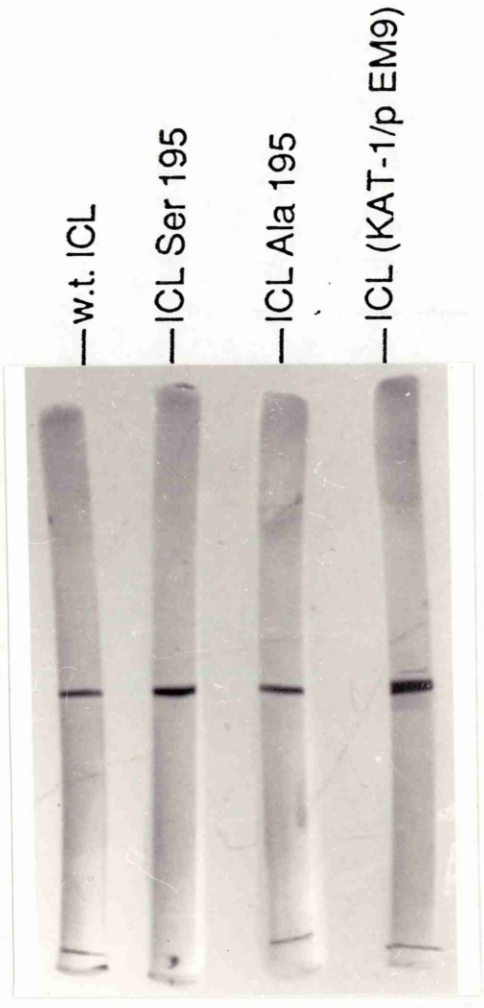


Figure 4.13: Non-Denaturing Polyacrylamide Gel Electrophoresis Of ICL, ICLSer195 and ICLAla195

Purified ICL, ICLSer195 and ICLAla195 run on 7% polyacrylamide tube gels (section 2.10.2).

Lane 1) ICL; 2) ICLSer195; 3) ICLAla195; 4) ICL purifued from *E. coli* KAT-1/pEM9

cleaved without the resulting peptides dissociating from the enzyme.

The purification results are shown in Table 4.3. The specific activities of the purified enzymes, ICLSer195 and ICLAla195 were much lower than that of the wild type ICL, only 1.3% and 2.4%.of the wild type enzyme respectively. This suggested that Cys195, although not essential, does indeed play an important role in the activity of ICL from *E. coli* ML308. This did not however indicate whether Cys195 is involved in catalysis, substrate binding or both. Further characterization of ICLSer195 and ICLAla195 was required in order to define the role of Cys195.

4.5 Characterisation Of ICLSer195 And ICLAla195:

4.5.1 Conformation Of ICLSer195, ICLAla195 And Wild Type ICL:

It is possible that replacing Cys195 with either Ala or Ser may reduce ICL activity by altering the conformation of the enzyme significantly. It was therefore decided to investigate the conformations of each of the mutant ICL enzymes and compare them to that of the wild type enzyme. The fact that ICLSer195, ICLAla195 and the wild type ICL were purified using the same procedure suggested that the conformation and physical properties of all three ICL enzymes might be very similar. However, a more detailed study of the conformation of these enzymes was required.

Using the secondary structure prediction programs of the UWGCG package, PEPTIDESTRUCTURE and PLOTSTRUCTURE, the secondary structures of ICLSer195 and ICLAla195 were predicted and compared to that of the wild type enzyme. There were no differences between either ICLSer195 and ICLAla195 or between the mutant enzymes and the wild type ICL using both the Garnier and Chou-Fasman methods of prediction (results not shown).

Examination of the secondary structures of ICLSer195, ICLAla195 and the wild type using circular dichroism (C.D.) was carried out in collaboration with Dr. N.C. Price at the University of Stirling. While C.D. spectra in the far U.V. range (190-250nm) reflect transitions of peptide bonds and therefore the secondary structure, C.D. spectra in the near

U.V. range (250-330nm) reflect transitions of the amino acid side chains, since there is no overlap with peptide bond transitions above 240-250nm. C.D. spectra in the near U.V. range therefore reflect the environment and conformation of amino acid side chains, hence both the secondary and tertiary structures contribute to the C.D. spectra in this range.

All C.D. analysis was carried out in 10mM Mops pH7.3, at a protein concentration of 0.3mg.ml^{-1} with a slit width of 0.02cm for the far U.V. spectra and 0.5cm for the near U.V. spectra (Section 2.15). The results were plotted as ellipticity, θ (deg.cm^2.d.mol^{-1}) against wavelength (nm). Figure 4.14 shows the C.D. spectra of ICLSer195, ICLAla195 and wild type ICL in the far U.V. range (190-250nm), while figure 4.15 shows the C.D. spectra of all three ICL enzymes in the near U.V. range (260-320nm). For both the near and far U.V. ranges, there is very little difference in the C.D. spectra of the different enzymes, indicating that both the secondary and tertiary structures of all three ICL enzymes are very similar.

Using the CONTIN procedure for determining the amount of each type of secondary structure from the C.D. spectra, the wild type ICL was shown to have 32 ± 2.6% α-helix, 37 ± 2.7% β-sheet and 31 ± 4.6% the remainder. These values compare with 32.7% α-helix using the Garnier method of secondary structure prediction and 29.5% α-helix by the Chou-Fasman method. However, the amount of β-sheet obtained by either method of secondary structure prediction is much different from the experimentally determined value, being 19.1% by the Garnier method and 48.4% by the Chou-Fasman method.

In addition to the C.D. spectra, ICLSer195, ICLAla195 and wild type ICL were analysed by fluorimetry at 20°C. (Section 2.14). This technique principally reflects the environment of tryptophan, tyrosine and phenylalanine residues (mainly Trp because of energy transfer). The emission spectra in the wavelength range 300-400nm after excitation at 290nm are shown in figure 4.16. As with the C.D. spectra, the fluorescence spectra of ICLSer195, ICLAla195 and the wild type enzyme are very similar, indicating that the tertiary structures of all three ICL enzymes are very similar.

The evidence presented above indicates that replacing Cys195 with either Ala or Ser does not alter the overall conformation of ICL. In addition, the fact that both ICLSer195 and ICLAla195 comigrate with the wild type enzyme on non-denaturing gel electrophoresis

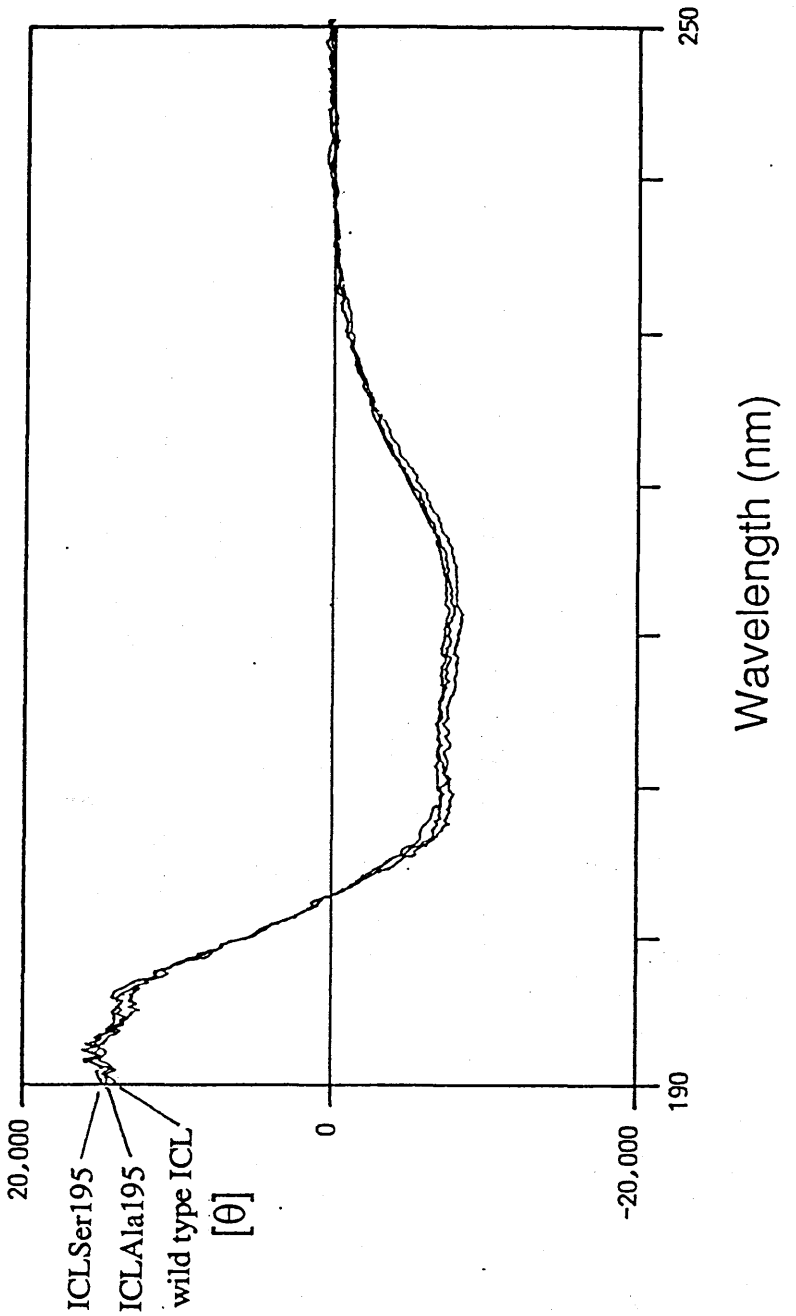


Figure 4.14: Circular Dichriosm Of w.t. ICL, ICLSer195 And ICLAla195, Far U.V. Spectrum.

Results are plotted as elipticity (units per mean molar ellipticity) against wavelength. Data obtained as described in section 2.15.

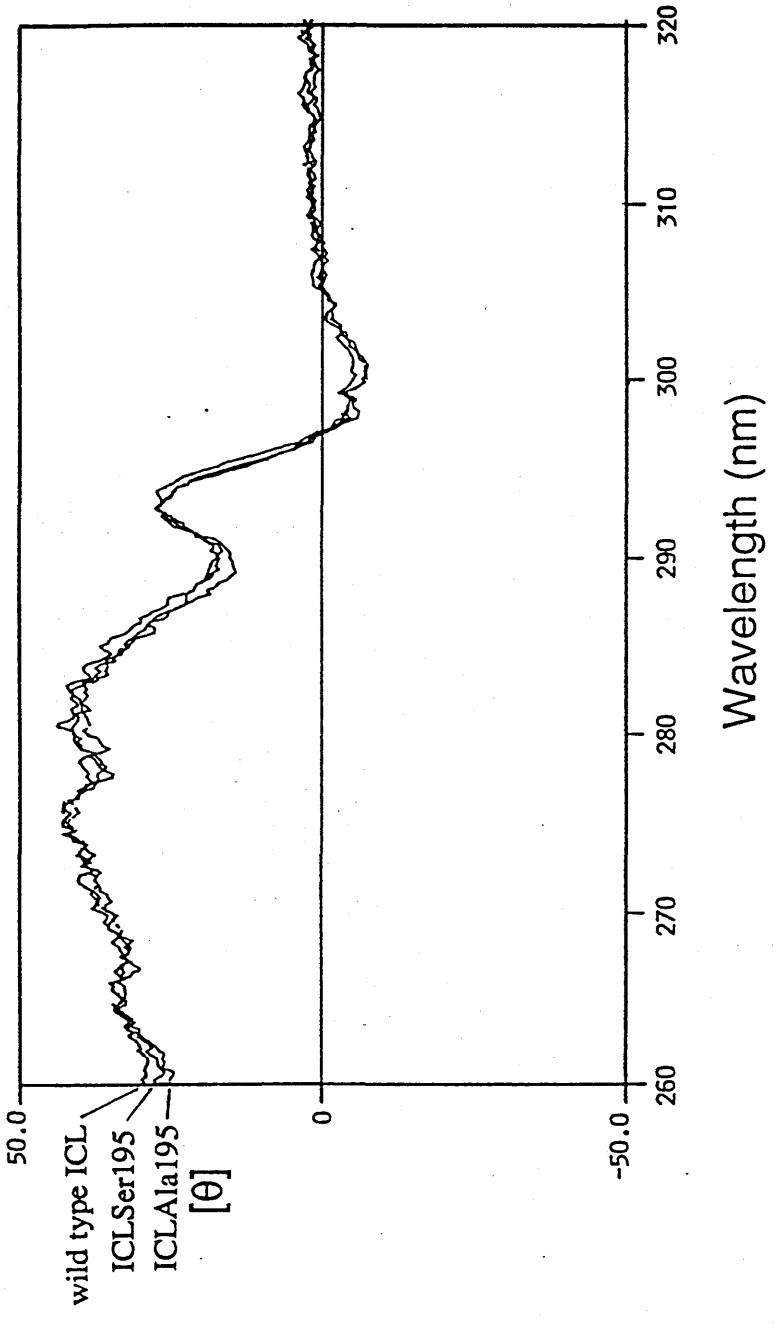


Figure 4.15: Circular Dichriosm Of w.t. ICL, ICLSer195 And ICLAla195, Near U.V. Spectrum.

Results are plotted as for figure 4.14.

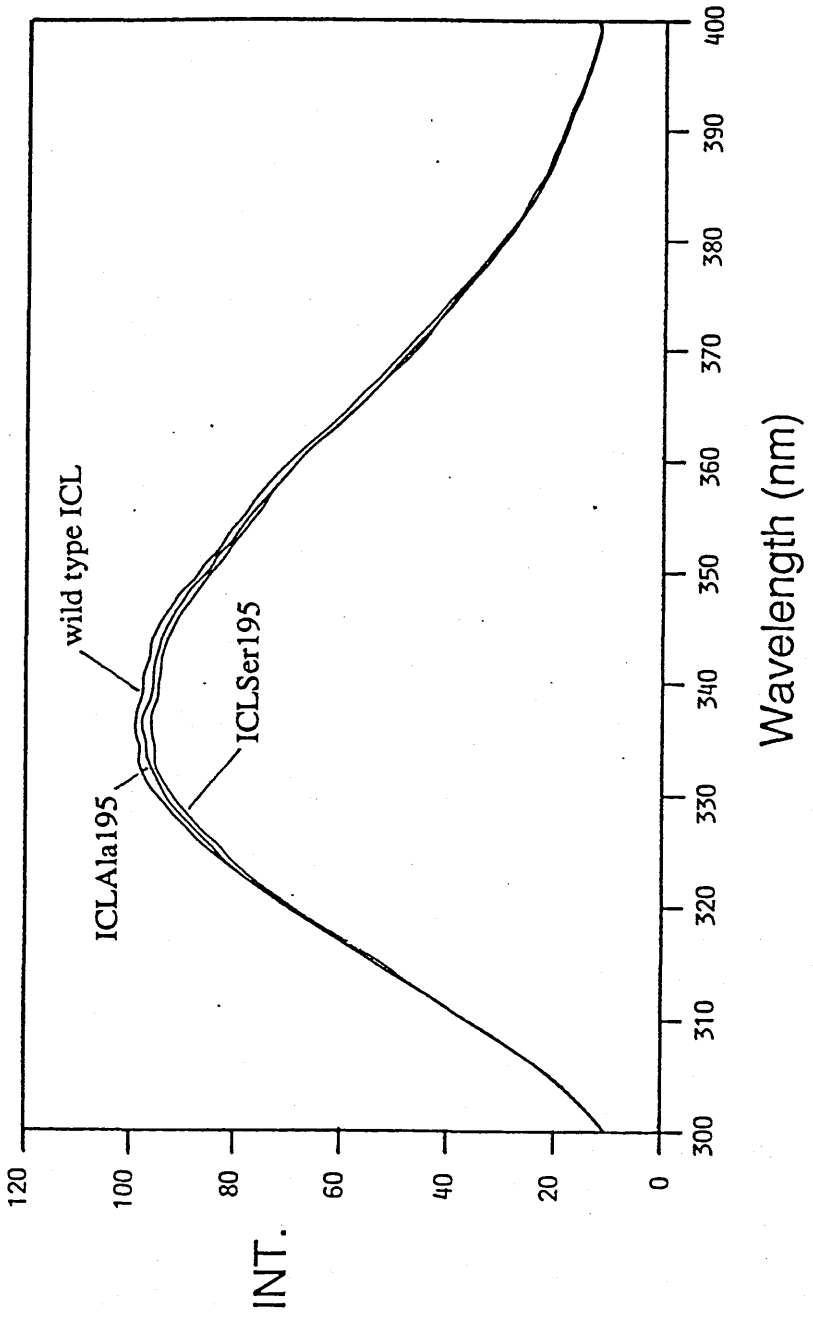


Figure 4.16: Fluorimetric Analysis Of w.t. ICL, ICLSer195 and ICLAla195

Results are plotted as intrinsic fluoresence against wavelength (nm). Data obtained as described in section 2.14.

indicates that replacing Cys195 with either Ser or Ala does not interfere with subunit-subunit interactions. This conclusion was supported a preliminary attempt to purify ICLSer195, in which it was found that this enzyme was eluted from a Sephacryl S-300 gel filtration column at the same position as wild type ICL.

4.5.2 Sensitivity To Inhibition By Iodoacetate :

Since pAR24-654 and pAR24-655 were maintained in the host strain TG-1, which contains the wild type *aceA* gene and pAR654 and pAR655 were maintained in the host, KAT-1, which is not an *aceA* deletion mutant but a Tn10 :: *aceA* insertion and therefore also contains the complete *aceA* gene, the possibility of homologous recombination exists between the wild type chromosomal *aceA* gene and the mutant *aceA* genes on the plasmids in either pAR654, pAR655, pAR24-654 and pAR24-655. If this were the case, then any ICL activity in the mutant ICLs might be due to the presence of a sub-population of plasmids with the wild type *aceA* gene and not due to the mutated gene.

In order to rule out the possibility that the activity of the mutant ICL enzymes arose from a sub-population of wild type enzyme contaminating the purified ICLSer195 and ICLAla195 mutants, the sulphydryl modifying reagent, iodoacetate, was used to inactivate ICLSer195, ICLAla195 and the wild type ICL. Iodoacetate reacts with free sulphydryl groups and was used to identify the active site Cys residue (Nimmo *et al.*, 1989). If as predicted, Cys195 has been replaced with Ala and Ser in ICLAla195 and ICLSer195 respectively, then the rate of inactivation of these enzymes by iodoacetate should be different to that of the wild type enzyme. Indeed the mutants might be almost completely resistant to iodoacetate since in the wild type ICL only Cys195 was modified by iodoacetate (Nimmo *et al.*, 1989).

Figure 4.17 shows the effect of iodoacetate on ICLSer195, ICLAla195 and wild type ICL. After 40min, the activity of the wild type ICL is only 4.4% of the activity before addition of iodoacetate, while 54.8% and 64.8% of the initial activity remains after incubation of ICLAla195 and ICLSer195 with iodoacetate for the same time. The half-life of ICLSer195 and ICLAla195 is therefore very much longer than that of the wild type enzyme, which is 8min, and so the activities of ICLSer195 and ICLAla195 are not due to a sub-population of the wild

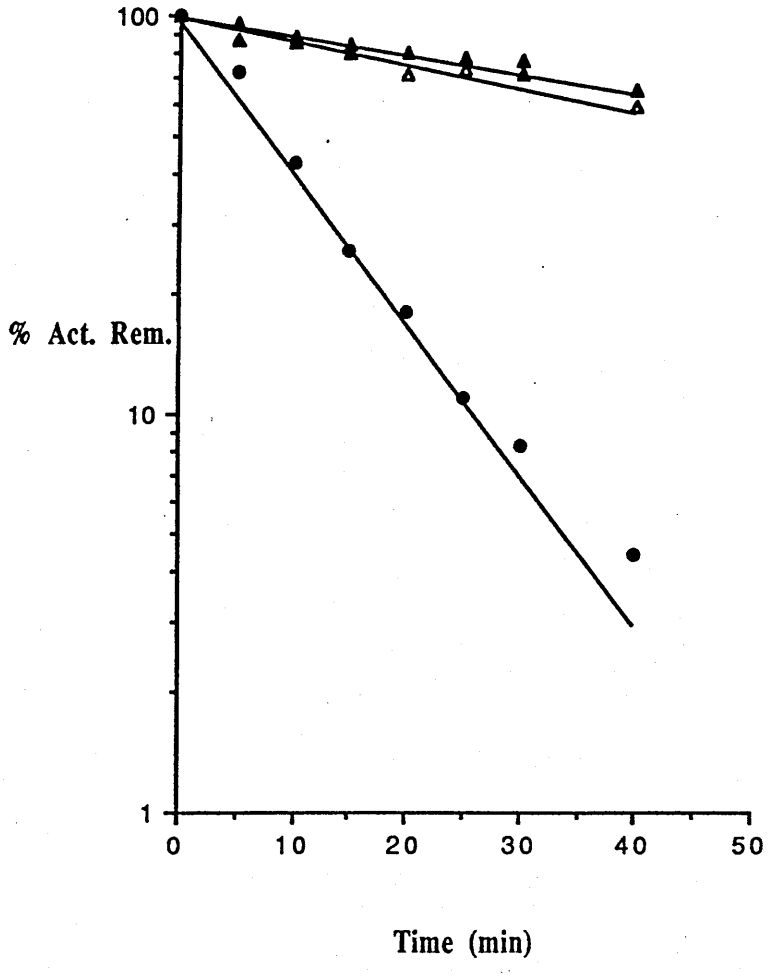


Figure 4.17: Inhibition Of w.t. ICL, ICLSer195 and ICLAla195 With Iodoacetate

The enzyme (1mg. ml^{-1}) was incubated with iodoacetate (2.0mM) as described i section 2.13:

●) wt ICL

▲) ICLAla195

▲) ICLSer195

type ICL present in the purified mutant enzymes either by homologous recombination or contamination during purification of the enzymes. Although the half life of the mutant ICL enzymes is longer than the wild type ICL, there is a decrease in the activity of the mutant enzymes over the time of incubation with iodoacetate. This would suggest that iodoacetate reacts slowly with another sulphydryl group or groups on the enzyme.

In order to establish whether or not the reactive group on the enzyme is at or near the active site, ICLSer195, ICLAla195 and wild type ICL were incubated with 2mM iodoacetate in the presence of 5mM Mg^{2+}, 5mM isocitrate and 5mM Mg^{2+} plus 5mM isocitrate. In the latter case the reaction will reach equilibrium and so glyoxylate and succinate will also be present. Isocitrate protected against iodoacetate inactivation for the wild type enzyme, increasing the half life from 5min to 11min (figure 4.18), while Mg^{2+} increased the half life to 12min (figure 4.19). With 5mM Mg^{2+} plus 5mM isocitrate, the half life was greater than 60 min (figure 4.20). With the mutant enzymes, neither Mg^{2+} (figure 4.19), nor isocitrate (figure 4.18) nor Mg^{2+} plus isocitrate (figure 4.20), had a significant effect on the sensitivity of either enzyme to iodoacetate. Therefore, the group on the mutant enzymes which reacts with iodoacetate is probably not at or near the active site, since the substrates do not protect against inactivation by this reagent.

4.5.3 Kinetic Properties Of ICLSer195, ICLAla195 And Wild Type ICL:

It has already been established that the reduction in the specific activity caused by replacing Cys195 of ICL with Ser or Ala does not result from a large change in the conformation of the enzyme (section 4.5.1). Therefore, the reduced specific activity of ICLSer195 and ICLAla195 compared with that of the wild type enzyme must be due to differences in either substrate binding, catalysis, or both. In order to discriminate between the above possibilities, the K_m and k_{cat} of ICLSer195 and ICLAla195 were compared to those of wild type ICL (table 4.4).

As shown in table 4.4., the k_{cat} values of ICLSer195 and ICLAla195 are approximately 100-fold and 30-fold lower than that of the wild type enzyme respectively. This implies that Cys195 must play some role in the catalytic mechanism, although this residue is obviously not

Figure 4.18: Isocitrate Protection Of w.t. ICL, ICLSer195 And ICLAla195 Against Inactivation By Iodoacetate

Enzyme (1mg. ml-1) was incubated with iodoacetate (2.0mM) in the presence and absence of isocitrate as described in section 2.13:

A)

○) wt ICL.

●) wt ICL, preincubated with 5mM isocitrate.

△) ICLSer195.

▲) ICLSer195, preincubated with 5mM isocitrate.

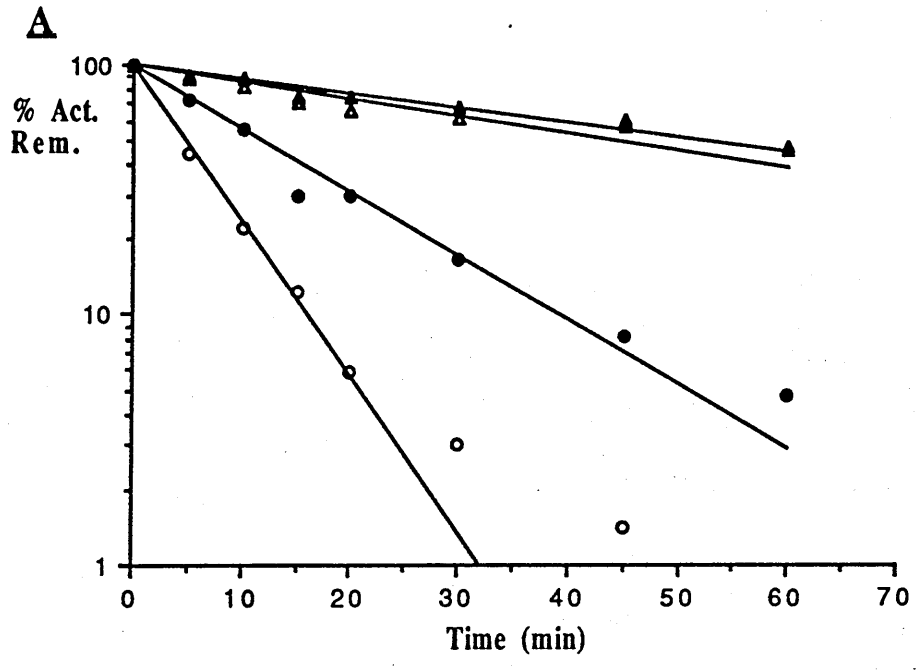
B)

○) wt ICL.

●) wt ICL, preincubated with 5mM isocitrate.

□) ICLAla195.

■) ICLAla195, preincubated with 5mM isocitrate.


A
100
% Act.
Rem.
10
1
0
10
20
30
40
50
60
70
Time (min)

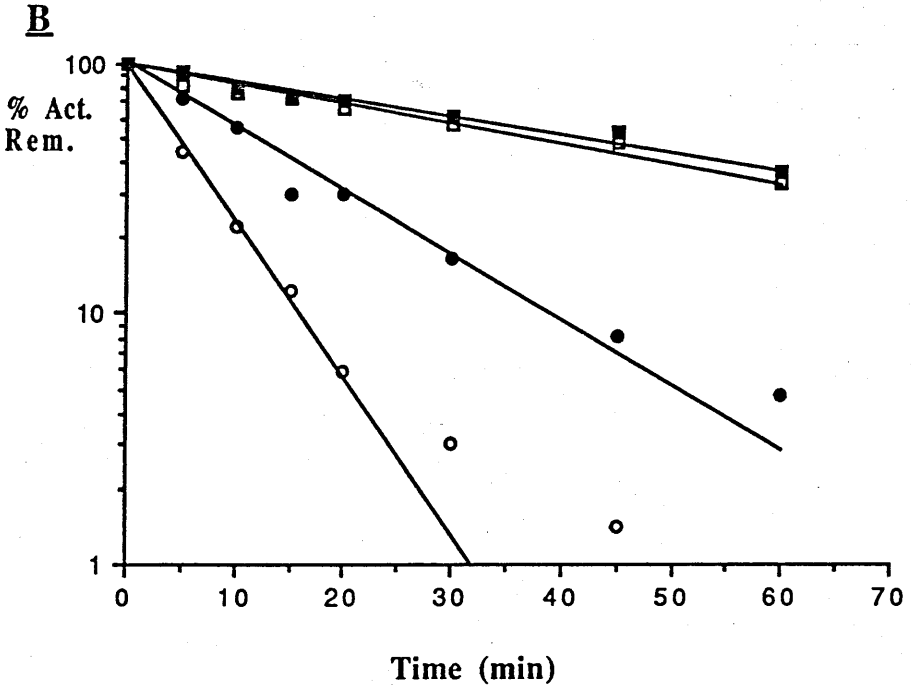

B
100
% Act.
Rem.
10
1
0
10
20
30
40
50
60
70
Time (min)

Figure 4.19: Mg^{2+} Protection Of w.t. ICL, ICLSer195 And ICLAla195 Against Inactivation By Iodoacetate

Enzyme (1mg. ml-1) was incubated with iodoacetate (2.0mM) in the presence and absence of isocitrate as described in section 2.13:

A)

○) wt ICL.

●) wt ICL, preincubated with 5mM $MgCl_2$.

△) ICLSer195.

▲) ICLSer195, preincubated with 5mM $MgCl_2$.

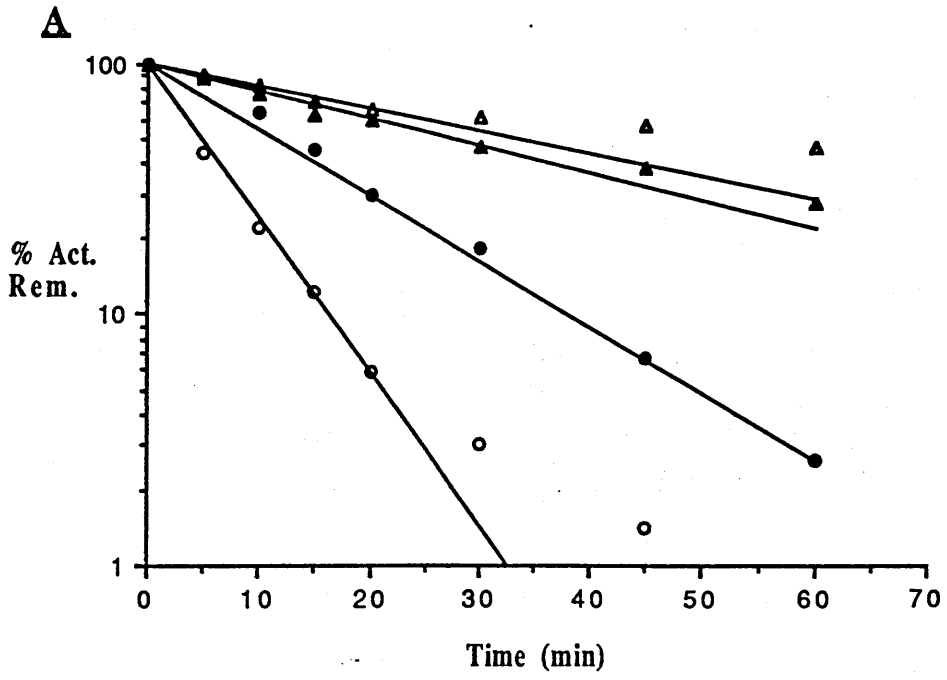
B)

○) wt ICL.

●) wt ICL, preincubated with 5mM $MgCl_2$.

□) ICLAla195.

■) ICLAla195, preincubated with 5mM $MgCl_2$.


A
100
10
1
% Act.
Rem.
0
10
20
30
40
50
60
70
Time (min)

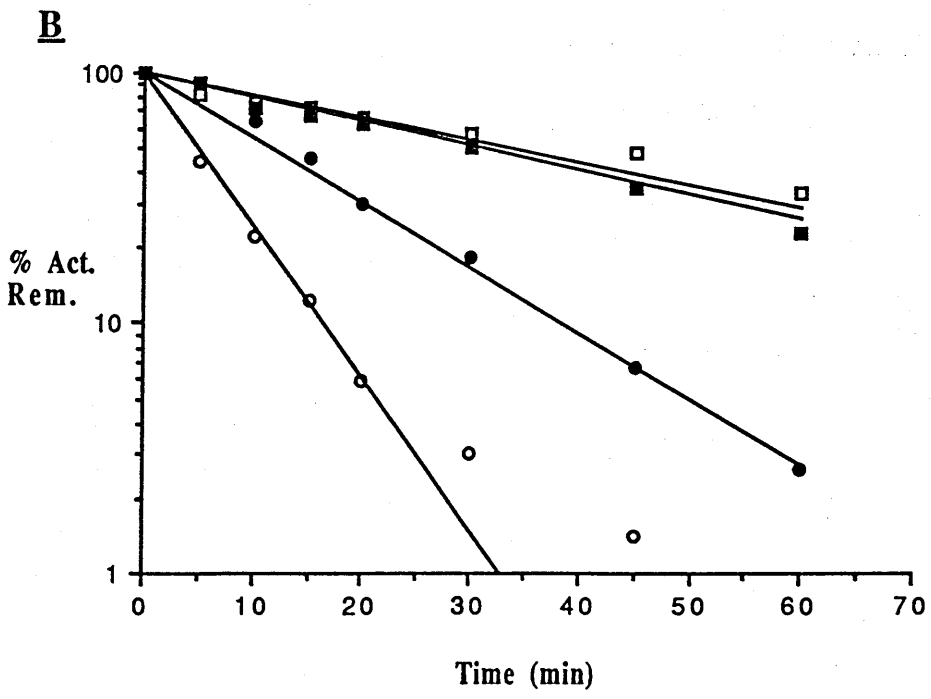

B
100
10
1
% Act.
Rem.
0
10
20
30
40
50
60
70
Time (min)

Figure 4.20: Mg^{2+}-Isocitrate Protection Of w.t. ICL, ICLSer195 And ICLAla195 Against Inactivation By Iodoacetate

Enzyme (1mg. ml-1) was incubated with iodoacetate (2.0mM) in the presence and absence of isocitrate as described in section 2.13:

A)

○) wt ICL.

●) wt ICL, preincubated with 5mM $MgCl_2$ plus 5mM isocitrate.

△) ICLSer195.

▲) ICLSer195, preincubated with 5mM $MgCl_2$ plus 5mM isocitrate.

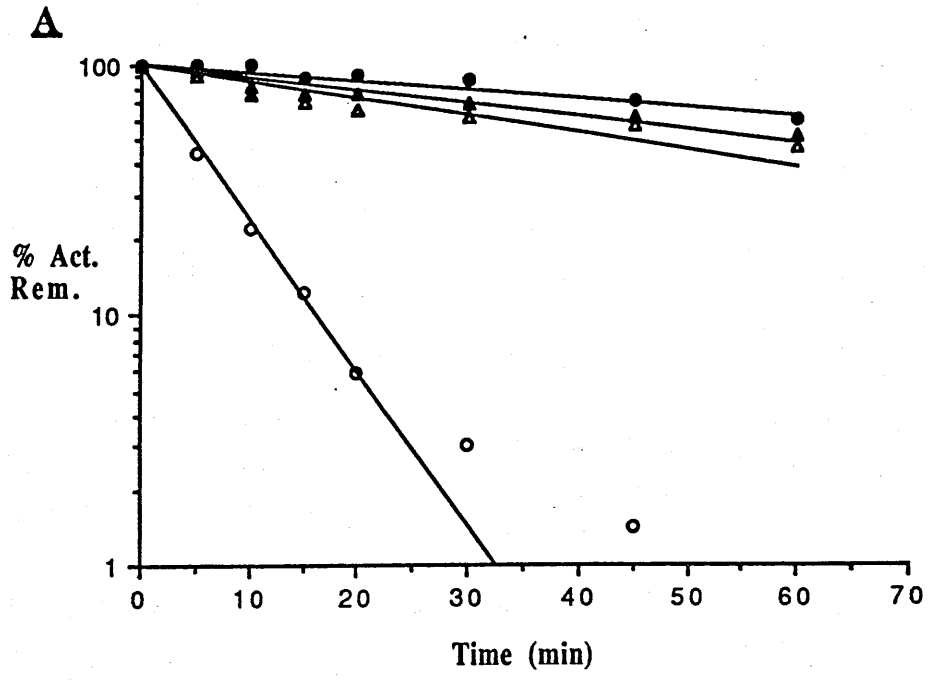
B)

○) wt ICL.

●) wt ICL, preincubated with 5mM $MgCl_2$ plus 5mM isocitrate.

□) ICLAla195.

■) ICLAla195, preincubated with 5mM $MgCl_2$ plus 5mM isocitrate.


A
100
10
1
% Act.
Rem.
0
10
20
30
40
50
60
70
Time (min)

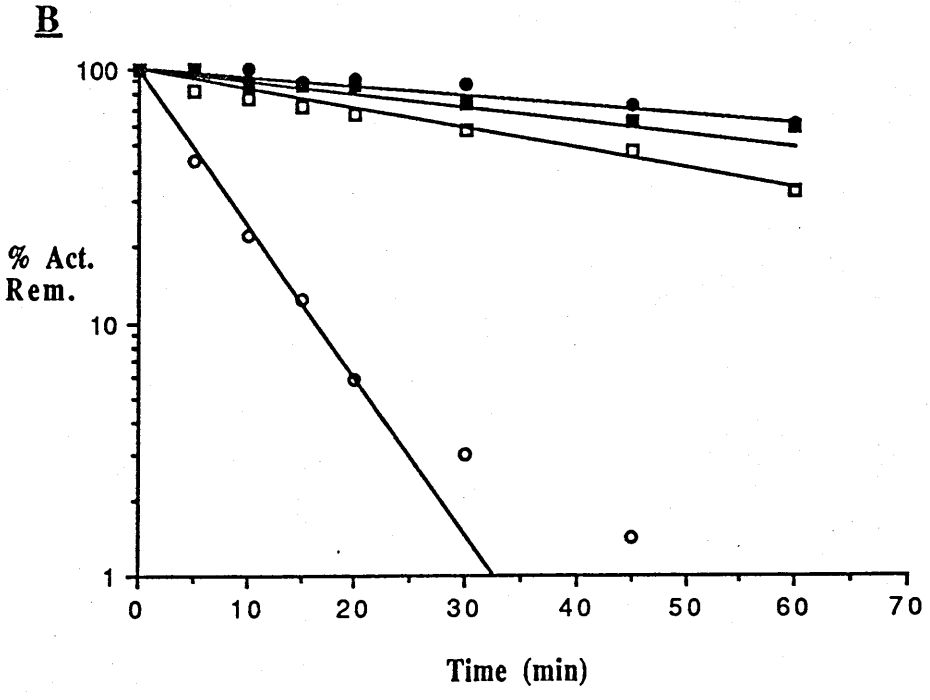

B
100
10
1
% Act.
Rem.
0
10
20
30
40
50
60
70
Time (min)

Table 4.4: Kinetic Properties Of ICLSer195, ICLAla195 And Wild Type ICL

	K_m (μM)	k_{cat} (sec^{-1})	k_{cat}/K_m ($sec^{-1}.M^{-1}.$)
w.t. ICL	76±6 (n=4)	28.5±2.1 (n=4)	$3.75x10^5$
ICL Ser195	272±24 (n=3)	0.296±0.025 (n=3)	$1.089x10^3$
ICL Ala195	641±37 (n=3)	0.937±0.072 (n=3)	$1.462x10^3$

essential since there is still some activity associated with the mutant enzymes.

ICLSer195 and ICLAla195 also have higher K_ms for the substrate, isocitrate, when compared to the K_m of the wild type enzyme, by factors of 3.6-fold and 8.4-fold respectively. Assuming that the mutants obey an equilibrium mechanism, as does the wild type (MacKintosh and Nimmo, 1988), it can be concluded that the side chain of Cys 195, CH_2-S-H, affects the binding of isocitrate. As expected, replacing the CH_2-S-H of Cys with the CH_2-O-H group of Ser did not reduce the affinity of the enzyme as much as elimination of the CH_2-X-H group by replacing it with -CH_3 of Ala. The fact that the values for the K_ms of both mutant ICLs are different from the wild type enzyme supports the work with iodoacetate which indicated that the activity of the mutant enzymes was not due to the presence of a sub-population of wild type ICL. Thus it appears that the side-chain of Cys195 is important for both substrate binding and catalysis, although this residue is not essential since both mutants, ICLSer195 and ICLAla195, have activity. Interestingly, ICLAla195 has a higher k_{cat} than ICLSer195. This was not expected since it was thought that the S–H group of Cys195 played some part in the reaction mechanism. One possibility is that a water molecule is occupying the space created by replacing Cys195 with Ala so providing an O–H group which is likely to be more reactive than that of –CH_2-O–H in ICLSer195 (section 4.6).

4.5.4 Effect Of pH On ICLSer195, ICLAla195 And Wild Type ICL:

There has been very little work done on the effect of pH on catalysis or substrate binding for ICL from *E. coli*. The most comprehensive studies have been done on ICL from *P. indigofera* and *N. crassa* (Rogers and McFadden, 1976; 1977). These workers found that catalysis was dependent on two groups with pK_as of 6.1 and 8.6 in the *N. crassa* enzyme and 6.9 and 8.6 in the *P. indigofera* enzyme. Also the K_m was dependent on a single ionizing group on the enzyme-substrate complex with pK_a of 8.5 in the *N. crassa* enzyme, and in the *P. indigofera* enzyme a single group on the free enzyme or free substrate with pK_a of 5.8. It seems likely, however, that the pK_a is that of a group on the free enzyme since the pK_as of the substrate are 3.29, 4.71 and 6.40 (Dawson *et al.*, 1986). The authors suggested that since ICL from both these organisms is susceptible to sulphydryl reagents and that catalysis of both

enzymes is dependent on a group with pK_a of 8.6, this group could be the active site cysteine since the pK_a is within the range for the S-H of cysteine residues (8.3-8.6), although the environment of the Cys in the enzyme could alter its pK_a quite substantially.

For the *E. coli* enzyme, there is not such a detailed study of the effect of pH on K_m for isocitrate or k_{cat}, although it has been shown that the K_m for isocitrate decreases with pH over the pH range 6.0-8.5 (MacKintosh and Nimmo, 1988; Nimmo *et al.*, 1989). Inactivation of ICL from *E. coli* ML308 by iodoacetate was shown to be dependent on a single ionizing group with pK_a of 7.1 (Nimmo *et al.*, 1989). A similar result was obtained for the reaction of ICL with 3-BrP, which was affected by a single group with a pK_a of 7.4 (Ko and McFadden, 1990). It seems likely that this group is not Cys195, but another group in the active site whose protonation affects the accessibility of Cys195 to iodoacetate, 3-BrP and isocitrate (section 1.3.10).

Therefore it was decided to study the effects of pH on K_m and k_{cat} for wild type ICL and compare them to the effects on ICLSer195 and ICLAla195. Above pH8.5 for wild type ICL and ICLAla195, and above pH9.0 for ICLSer195, the kinetics of substrate binding became sigmoidal preventing meaning comparisons of K_m and k_{cat} values.

The K_m of the wild type ICL for isocitrate declines with pH over the range 6.8 to 8.5. A plot of log K_m against pH is shown in figure 4.21. If the best fit straight line is drawn, this line has a gradient of approximately 0.5, indicating that the K_m of wild type ICL for isocitrate is dependent on more than one ionizing group on the enzyme-substrate complex (Dixon, 1953; Fersht, 1977). Similarly, the plot of log $^1/_{K_m}$ against pH (figure 4.22) also indicates that the K_m is dependent on more than one ionizing group on either the free enzyme or free substrate. Thus the ionizing groups affecting K_m are likely to be the same in the enzyme-substrate complex as those in the free enzyme or the free substrate (Fersht, 1977). A plot of log k_{cat} vs pH (figure 4.23) shows that there is very little change in this pH range, while the plot of log k_{cat}/K_m vs pH (figure 4.24) shows a curve with a straight line portion, gradient -0.5 and flattening towards the top. These data are difficult to interpret in terms of simple models. It is unfortunate that reliable data for k_{cat} and K_m values at pH9.0 and above could not be obtained.

For ICLSer195 and ICLAla195, a plot log K_m vs pH (figure 4.21) give straight lines with slopes of approximately 1, which tend to flatten at the top. This indicates that K_m is

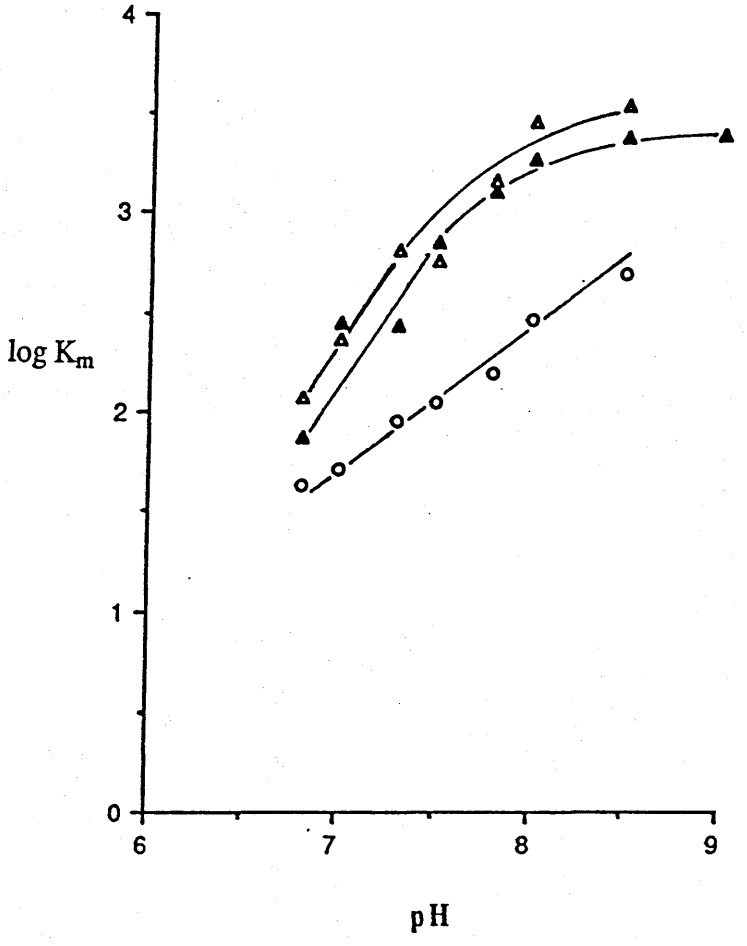


Figure 4.21: Effect Of pH On log K_m Of w.t. ICL, ICLSer195 And ICLAla195

The K_m was measured at each point by assaying by the method described in section 2.12a with 50mM Mops-NaOH, 5mM $MgCl_2$, 1mM EDTA (pH6.8-7.8) or 50mM Taps-NaOH, 5mM $MgCl_2$, 1mM EDTA (pH7.8-9.0) equilibrated to various pH units.

●) wt ICL

▲) ICLAla195

▲) ICLSer195

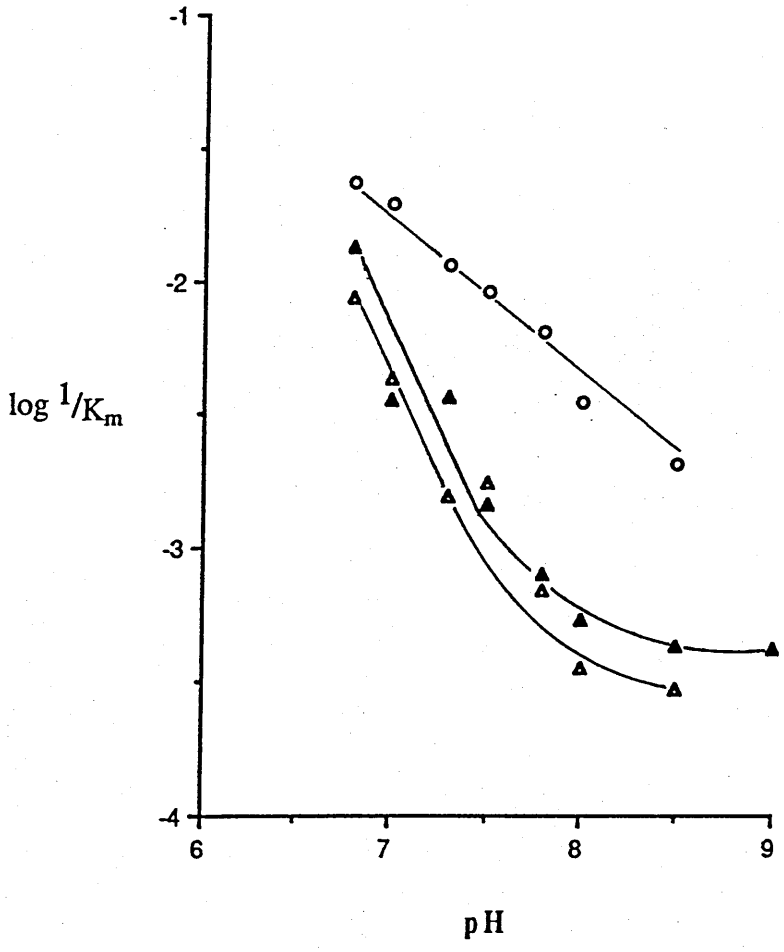


Figure 4.22: Effect Of pH On log $^{1}/_{K_m}$ Of w.t. ICL, ICLSer195 And ICLAla195

Assays were carried out as described for figure 4.21.

●) wt ICL

▲) ICLAla195

▲) ICLSer195

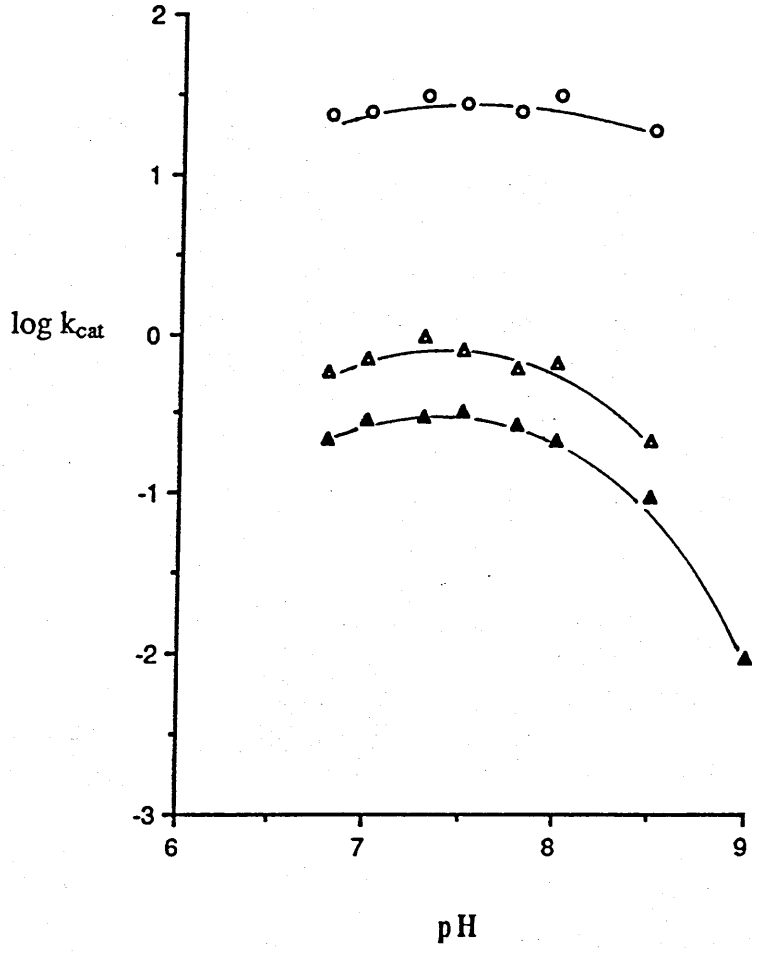


Figure 4.23: Effect Of pH On log k_{cat} Of w.t. ICL, ICLSer195 And ICLAla195

Assays were carried out as described for figure 4.21.

●) wt ICL

▲) ICLAla195

▲) ICLSer195

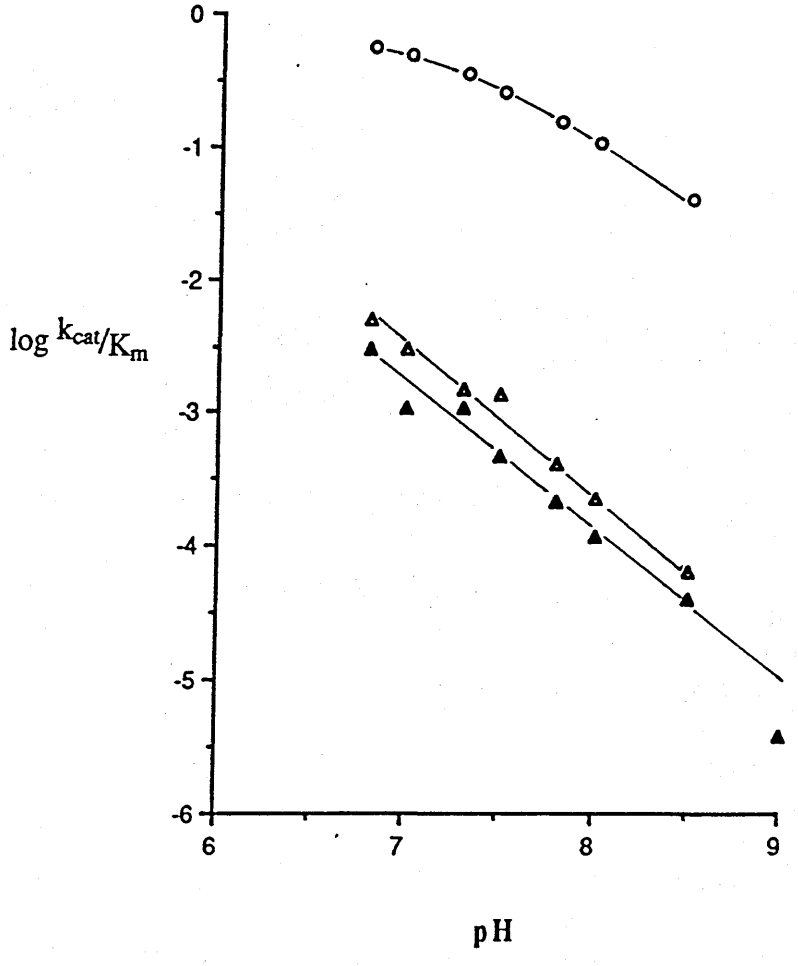


Figure 4.24: Effect Of pH On log k_{cat}/K_m Of w.t. ICL, ICLSer195 And ICLAla195

Assays were carried out as described for figure 4.21.

●) wt ICL

▲) ICLAla195

▲) ICLSer195

dependent on a single dissociating group on the enzyme-substrate complex. Similarly, a plot of log k_{cat} vs pH for ICLSer195 gives a curve with a straight line portion which has gradient -1 and flattens towards the top (figure 4.23). This indicates that k_{cat} is dependent on a single dissociating group with a pK_a of 7.8 on the enzyme-substrate complex. Unfortunately, the pK_a of the group(s) affecting k_{cat} of ICLAla195 could not be determined since no data were obtained above pH8.5. However, since the curve for ICLAla195 is very similar to that of ICLSer195, it seems probable that the same group with pK_a 7.8 also affects k_{cat} of ICLAla195. Thus the pK_a of the group affecting k_{cat} and K_m is the same, 7.8. Similarly, plots of log $^1/_{K_m}$ vs pH for both ICLSer195 and ICLAla195 give straight lines with slopes of 1 and tend to flatten towards the bottom (figure 4.22), indicating that the K_m of both mutant enzymes is dependent on a single dissociating group on the free enzyme or free substrate with pK_a 7.8. A plot of log $^{k_{cat}}/_{K_m}$ vs pH for both ICLSer195 and ICLAla195 (figure 4.24) shows a straight line with a slopes of approximately -1, indicating that a single group with a pK_a outwith the pH range studied affects the catalytic efficiency.

Since the pK_a obtained for both ICLSer195 and ICLAla195 affecting K_m is 7.8, it seems probable that the dissociating group is the same for both enzymes. Also, since the mutant enzymes have one dissociating group less than wild type ICL which affects K_m, it seems likely that one dissociating group on the wild type enzyme has a pK_a of 7.8

The main feature that emerges from this work is that the response of the mutant ICLs to pH is appreciably simpler than that of the wild type enzyme. Thus, it is clear that replacing Cys195 with either Ser or Ala, results in the loss of one dissociating group from the active site and allows the determination of the pK_a of another dissociating group on which K_m and $^{k_{cat}}/_{K_m}$ is dependent. The data suggest that the pK_a of the Cys195 –S–H group is within the pH range studied, 6.8-8.5. It is therefore concievable that the pK_a of 7.1 determined for the reaction between iodoacetate and Cys195 of ICL from *E. coli* ML308 (Nimmo *et al.*, 1989) and 7.4 determined for the reaction of the same residue with 3-BrP from *E. coli* K12 (Ko and McFadden, 1990) is the pK_a of this residue and not that of another residue whose protonation affects the accessibility of these reagents and the substrates to the active site.

Since the pK_a of the group on the enzyme-substrate complex is the same as that of the

group on the free enzyme or free substrate in both mutant enzymes, it seems likely that this group is the same. The pK_a of the ionizing groups on the substrate, isocitrate, are 3.29, 4.71 and 6.40 (Dawson *et al.*, 1986). Therefore, the group with pK_a 7.8 is likely to be on the free enzyme and not the free substrate.

4.6 Discussion:

Site directed mutagenesis can be used to define the role, if any, of proposed active site Cys residues and has in some cases indicated whether the residue is essential for substrate binding or the reaction mechanism. For 3-hydroxy-3-methylglutaryl-CoA reductase from *Pseudomonas mevalonii* (Jordan-Starck and Rodwell, 1989) either Cys156, Cys296 or both were thought to be in the active site based on evidence from chemical modification studies and substrate protection, as were Cys41 of the *Pertussis* toxin S1 subunit (Locht *et al.*, 1990), a cysteine residue in human cyclophilin (Liu *et al.*, 1990), the N-terminal cysteine residue of human asparagine synthetase (Van Heeke and Schuster, 1989) and cysteine residues in receptor-linked protein tyrosine phosphatases (PTPases) in humans and *Drosophila* (Streuli *et al.*, 1989). Site directed mutagenesis has shown that although these residues are at or very near to the active sites of these enzymes, only for the PTPases and asparagine synthetase were the Cys residues essential in that replacement of the Cys residue completely abolished activity (Van Heeke and Schuster, 1989; Streuli *et al.*, 1989). For the asparagine synthetase, site directed mutagenesis also showed that the N-terminal Cys was essential only for the glutamine dependent and not the ammonia dependent activity.

Although Cys195 is at or near the active site of ICL from *E. coli* ML308, and is conserved between all the ICLs sequenced so far, it was not known whether or not this residue was involved in catalysis. It was therefore decided to use site directed mutagenesis to replace Cys195 with othe amino acids, Ala and Ser, in an attempt to discover what role, if any, Cys195 played in the reaction mechanism of ICL from *E. coli* ML308.

It is clear from the evidence in this chapter that Cys195 does indeed play some role in the enzyme's activity, since replacing this residue with either Ser or Ala results in a decrease in k_{cat} and an increase in K_m. Both mutants probably have similar conformations to that of the wild type enzyme based on evidence from circular dichrioism and fluorimetry. Therefore, the

changes in kinetic properties are directly due to replacement of CH_2–S–H of Cys195 with CH_2–O–H and CH_3 in ICLSer195 and ICLAla195 respectively. Interestingly the results of the site directed mutagenesis were not those predicted. It was thought that if Cys195 was an active site residue, then replacing this residue with a Ser would result in an enzyme which had reduced activity and that replacement with Ala would result in an enzyme which had little or no activity since there would be no ionization of this group and no potential nucleophile. Contrary to this prediction was the finding that although ICLSer195 had a lower K_m than ICLAla195, ICLAla195 had a higher k_{cat} than ICLSer195. This can be explained by a water molecule occupying the space createdby replacing CH_2–S–H by CH_3. The O–H group of the water molecule may interact with the substrate in a less deleterious way than that of Ser195.

One possible analogy is with chloramphenicol acyltransferase (Lewendon *et al.*, 1990). It was thought that Ser148 of chloramphenicol acyltransferase was involved in transition state stabilisation via a hydrogen bond to the oxyanion of the putative tetrahedral intermediate. Substitution of Ser148 with Ala resulted in an enzyme with k_{cat} reduced 53-fold and only minor changes in K_m, while substitution with Gly at position 148 resulted in an enzyme with only a 10-fold reduction in k_{cat}. The authors proposed that a water molecule may have partially replaced the hydrogen bonding potential of Ser148. The three dimensional structure of the Ala148 enzyme was isosteric with the wild type enzyme and although the three dimensional structure of the Gly148 enzyme was not determined, it seems likely that the structure of this enzyme is not much different from that of the wild type enzyme.

The pH studies of the wild type ICL and both mutants throw little light on the mechanism of catalysis by ICL because k_{cat} values were fairly constant over the pH range studied. However, it can be suggested that the pK_a of the S–H group of Cys195 is 7.1 in the *E. coli* ML308 enzyme. This value is approximately one pH unit lower than that expected for this residue. A likely explanation is that there is a basic group near to the S–H and thus causes a reduction in the pK_a of Cys195. The identity of the basic group interacting with Cys195 is not known although there are several possibilities. Chemical modification using DEPC has indicated a His residue at the active site of ICL from several organisms (section 1.4.11). From *E. coli*, it has been shown that neither His266 nor His306, which react with DEPC, are at the active site, although the results do not rule out the possibility that another His residue is at or

near the active site. Indeed there is a His residue at position 197, only two residues from Cys195. In the Cys proteases, e.g. papain, the pK_a of the S–H group of the active site Cys is lowered by a His residue in close proximity to the reactive Cys (reviewed in Fersht, 1977). Also, chemical modification of ICL from *Citrullis vulgaris* using itaconate epoxide identified a carboxyl group of either Asp or Glu at the active site (Jameel *et al.*, 1985). Although the authors suggest a role for the carboxyl group in coordinating the Mg^{2+} it is possible that this group acts in the same manner as the His in papain.

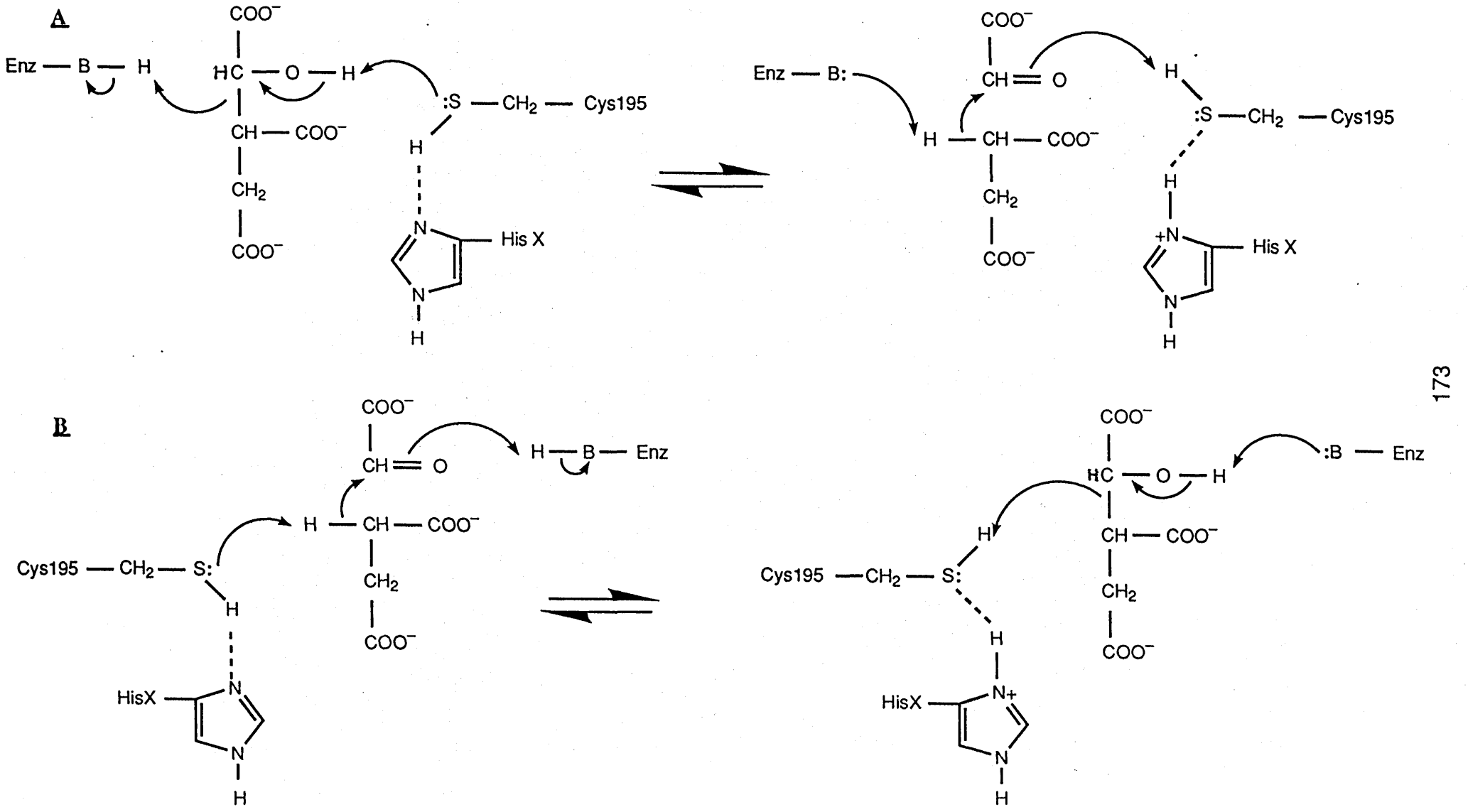
This has interesting consequences for the mechanism of ICL. It is clear in the condensation reaction that the enzyme deprotonates succinate, creating a succinate carbanion or its conjugate base which can then attack the keto group of glyoxylate. The enzyme then acts as a proton donor during the condensation reaction (section 1.4.12). Assuming that the S–H of Cys195 is the basic group which deprotonates succinate, this would explain the fact that ICLSer195 has a much lower k_{cat} than the wild type enzyme, since the pK_a of the O–H of Ser is much higher than that of the S–H of Cys. The data for ICLAla195 are consistent with this proposal, assuming that the presence of a water molecule in the active site is responsible for the activity of this enzyme. Thus the mechanism I propose, based on the differences in k_{cat} between the wild type ICL and ICLSer195 and ICLAla195 and the probable pK_a of 7.1 for Cys195 is that the S atom of Cys195 is made more nucleophilic by the formation of a hydrogen bond between the proton on the sulphydryl group and an adjacent base. The S atom then extracts a proton from the C2 of succinate, creating a succinate carbanion, which then attacks the keto carbon of glyoxylate, concomitant with protonation of the glyoxylate moiety to form isocitrate. This mechanism is outlined in figure 4.25.

There is a second possible mechanism in which the sulphydryl group is involved in deprotonating the substrate. In this mechanism, Cys195 and the adjacent base act in the manner indicated above but instead of deprotonating succinate in the condensation reaction, the S atom deprotonates the hydroxyl group of isocitrate in the cleavage reaction. This results in a carbanion at C3 which can then deprotonate the enzyme to give glyoxylate and succinate (figure 4.25).

Because of the pK_as of the groups affecting the K_m of ICL for isocitrate are so close,

Figure 4.25: Postulated Roles Of Cys195 In The Catalytic Mechanism Of ICL

Two mechanisms are postulated for the role of Cys195 in catalysis. A) the S atom of Cys195 attacks the hydroxyl proton in the cleavage reaction, B) the S atom attacks a proton from the C2 carbon of succinate in the condensation reaction. Solid lines represent covalent bonds, while dashed lines represent hydrogen bonds. Both mechanisms are disscussed in section 4.6, and HisX may be replaced by a carboxyl group of either a Glu or an Asp.


A
Enz — B — H
HC — O — H
COO⁻
CH — COO⁻
CH₂
COO⁻
:S — CH₂ — Cys195
H
N
His X
N
H
Enz — B:
CH = O
H — CH — COO⁻
CH₂
COO⁻
H
:S — CH₂ — Cys195
H
+N
His X
N
H
B
CH = O
H — B — Enz
H — CH — COO⁻
CH₂
COO⁻
Cys195 — CH₂ — S:
H
HisX
N
N
H
HC — O — H
:B — Enz
CH — COO⁻
CH₂
COO⁻
H
Cys195 — CH₂ — S:
H
HisX
N+
N
H

7.8 and 7.1, their values were not previously determined. Replacement of Cys195, one of the two dissociating groups, with either Ala or Ser allowed determination of the pK_a of the second dissociating group, 7.8. Both the identity and role of this residue are unknown. It seems unlikely that this group is the same as that which forms a hydrogen bond with the S–H group of Cys195, since it has the same pK_a in ICLSer195 as in ICLAla195. Even if there is an extra water molecule in the active site of ICLAla195, one might expect the ionisation of this water to influence the dissociation of this residue to a greater degree than the presence of a hydroxyl group at position 195 in ICLSer195. There are several possibilities for the identity of the group with pK_a 7.8. The pK_a is within the range usually found for the imidazole of His in proteins (pK_a 5-8) and for the α-NH_2 of an N-terminal amino acid. It therefore of note that in ICL from *P. indigofera*, *L. usitatissimum*, *Ricinus communis*, *B. napus*, *C. tropicalis*, *G. hirsutum* and *E. coli*, the N-terminal residue is Met (McFadden *et al.*, 1968; Khan and McFadden, 1982; Beeching and Northcote, 1987; Comai *et al.*, 1989; Atomi *et al.*, 1990; Turley *et al.*, 1990). The only exception found so far is that the N-terimnal residue in the thermophilic *Bacillus* is Thr (Chell *et al.*, 1978).

Chapter 5
General Discussion

5.1 Introduction:

It is now well established that the control of flux at the branchpoint between the glyoxylate bypass and the TCA cycle is resolved at least in part by reversible phosphorylation and inactivation of ICDH. The phosphorylation and dephosphorylation of ICDH is catalysed by a bifunctional ICDH kinase/phosphatase (LaPorte and Koshland, 1982; Nimmo *et al.*, 1984; LaPorte and Chung, 1985). Holms and Bennett (1971) proposed that ICDH was inactivated in order to allow flux through the glyoxylate bypass. Work carried out by Koshland's group measuring fluxes through the system and by the Glasgow group studying the individual components of the system, both confirmed that the function of phosphorylation of ICDH was to render this enzyme rate limiting in the TCA cycle during growth on acetate, thus allowing the intracellular concentration of the common substrate of the competing pathways, isocitrate, to rise to a level which facilitates flux through ICL and the glyoxylate bypass.

The mechanism of inactivation of ICDH has now been determined, the enzyme has been crystallized and its three dimensional structure has been determined. The residue which is phosphorylated, Ser113, lies at the entrance of the isocitrate binding pocket and phosphorylation of this residue prevents binding of the substrate by electrostatic repulsion (Thorsness and Koshland, 1987; Dean *et al.*, 1989; Hurley *et al.*, 1989; 1990a; b; Dean and Koshland, 1990).

The system responds to a number of effectors, making it very sensitive to the needs of the cell. An essential component of the system is the difference in the kinetic properties of the two enzymes which compete for the common substrate, isocitrate. The K_m of ICDH for isocitrate is at least one order of magnitude lower than that of ICL for the same substrate. This results in flux through ICDH at concentrations of isocitrate present during growth on glucose or other preferred carbon sources. However, during growth on acetate or fatty acids as the sole carbon source, the V_{max} of ICDH is decreased four-fold, causing the intracellular isocitrate concentration to rise to a level at which ICL can compete effectively for the common substrate, but still allowing flux through ICDH in order maintain a balance between synthesis of biosynthetic precursors and generation of energy and reducing power.

Although a lot of work has been carried out on the function and role of ICDH phosphorylation, there has been comparatively little done on the competing enzyme, ICL. In order to understand more about this branchpoint, it was decided to investigate this enzyme further. In conjunction with D. Rice at the University of Sheffield, work was underway to determine the three dimensional structure of ICL. In order to facilitate this work, and to provide a clone with which to carry out site directed mutagenesis of ICL, the nucleotide sequence of the gene encoding ICL, *aceA*, had to be determined. This provided the start point of the work reported in this thesis.

5.2 Comparison Of ICL From Different Sources:

It has been shown that ICL from various organisms are similar in many ways. For example, all ICLs studied so far have a requirement for Mg^{2+} and most are tetramers of identical subunits. There appear to be two groups according to subunit M_r, with the bacterial enzymes having a subunit Mr of around 45,000 and the eukaryotic enzymes having a subunit M_r of around 65,000. The amino acid composition of ICL from a number of organisms have been determined and it is clear that they are very similar. The relative amino acid composition of these ICL enzymes has been used to determine the relatedness among the various enzymes by the method of Marchalonis and Weltman (1971). The results show that the enzymes are closely related, with SΔQ values of less than 50 in all comparisons (section 1.3.1). In addition the amino acid compositions have been used to measure various physico-chemical properties, such as the hydrophobicity average (Bigelow, 1967), "non-polar side chains" according to Waugh (1954), the polar volume/non-polar volume ratio (Fisher, 1964), the polar/apolar residues ratio (Hatch, 1965) the acid/base ratio and the relative specific volume according to Cohn and Edsall (1943) (reviewed in Vanni *et al.*, 1990).

At the outset of this project, there were no sequence data for ICL from any source. Therefore, one objective of the work was to determine the nucleotide sequence of the gene encoding ICL, *aceA*, from *E. coli* ML308. During the course of the work, the nucleotide sequence of *aceA* from *E. coli* K12 was determined. There were, however, two differences in the nucleotide sequence reported by different groups. It was therefore necessary to determine the nucleotide sequence of the *aceA* gene from *E. coli* ML308 before any site directed

mutagenesis was carried out, since the sequence of the same gene from *E. coli* K12 produced differences. Comparison of the nucleotide sequence of the *aceA* gene from *E. coli* ML308 with that of the two previously sequenced *E. coli* K12 sequences indicated that these differences may be due to sequencing errors and not to strain differences. The most important difference is at the C-terminus of ICL. McFadden's group found a His residue at the C-terminus from the derived amino acid sequence, while for the ML308 enzyme, the C-terminal residue was a Phe, in agreement with Cozzone's and LaPorte's groups. This difference is due to a difference in the nucleotide sequence at position 1255, with McFadden's group having an extra G resulting in a frame shift and early termination. McFadden's group also carried out carboxypeptidase A digestion of ICL and found the C-terminal residue to be His. In support of the sequence of the ML308 sequence, the region encoding the C-terminus was sequenced on both strands using several templates and both the Amersham sequencing system and Sequenase version 2.0. Also, a fourth *aceA* sequence was determined from *E. coli*, although the strain from which it was obtained was not given, which showed 100% identity with the ML308 sequence. Attempts by our group to determine the C-terminal residue by carboxypeptidase digestion of ICL did not give a clear result. The reason for this is not clear although it is thought that the method of detection of amino acids is not suitable for this type of analysis. Another possible method for identifying the C-terminal residue was found as a result of the determination of the derived amino acid sequence of ICL. Digestion of the enzyme with endoproteinase Glu-C and endoproteinase Arg-C should resolve this controversy, assuming that the ML308 sequence is correct. If ICL is first digested to completion with endoproteinase Arg-C, this will result in 17 peptides, the largest of which is 134 amino acids long, the next longest being 71 amino acids long. The largest peptide, which contains the C-terminus, should then be easily purified by chromatography on HPLC. If this peptide is then digested to completion with endoproteinase Glu-C, there should be seven peptides of length 73, 7, 6, 16, 28, 1 and 3 amino acids long. The peptide which is 3 amino acids long and contains the C-terminus should then be easily purified by chromatography on HPLC and sequenced by conventional methods. If the McFadden sequence is correct, then digestion of ICL with endoproteinase Arg-C should result in 19 peptides, with one peptide two amino acids long which contains the C-terminus.

During the course of this work, the genes encoding ICL from *B. napus*, *G. hirsutum*, *R. communnis* and *C. tropicalis* were also sequenced (Beeching and Northcote, 1987; Comai *et al.*, 1989; Turley *et al.*, 1990; Atomi *et al.*, 1990). Comparison of the derived amino acid sequences of ICLs revealed a high degree of identity. There was a very high degree of identity between the plant enzymes, each showing at least 86% identity to another, and also between the plant enzymes and the yeast ICL, with the yeast ICL showing 54-55% identity with the plant enzymes. Thus, the eukaryotic enzymes are highly conserved. More surprisingly, the *E. coli* ML308 sequence also shows a high degree of identity with both the yeast enzyme and the plant enzymes despite the fact that the *E. coli* ICL is a member of the lower M_r group containing the prokaryotic enzymes while the plant and yeast enzymes are all members of the higher M_r group. The areas of identity are interspersed throughout the length of the *E. coli* sequence although there are areas where the amino acid sequences are very highly conserved that may form part of the active site (figure 3.17, see below). The major differences between the *E. coli* ICL sequence and the eukaryotic sequences were the extensions at the N-terminus, the C-terminus and a large insert within the *E. coli* sequence, between residues Gly255 and Glu256. These extensions probably account for the difference in M_r between the prokaryotic and eukaryotic ICLs, although the *E. coli* sequence is, as yet, the only prokaryotic sequence available and may not be typical of the other prokaryotic ICL enzymes. Since there is no compartmentation in *E. coli*, and ICL is found in glyoxysomes in eukaryotes, it is possible that any or all of the additional sequences in the plant enzymes are responsible for sequestration of ICL into these organelles. This possibility is discussed in section 3.4.3.

The origin of the insert is not known but it is thought that there was an ancestral gene encoding an ICL of M_r 16,000. The various ICLs are thought to have arisen by duplication and fusion of either three or four copies of this ancestral gene (Matsuoka and McFadden, 1988). These authors compared the insert region of the *R. communnis* ICL with other regions of the same enzyme and the bacterial enzyme and found that this region showed weak homology to an adjacent region in the *R. communnis* enzyme between amino acids 380 to 470.

5.3 The Active Site Of ICL:

Although both ICDH and ICL bind isocitrate, the K_m of ICL for this substrate is at least one order of magnitude higher than that of ICDH. It is therefore of great interest to find out why this is so. The sequence of the gene encoding ICDH has been determined and the three dimensional structure is now available for ICDH in the presence of either Mg^{2+} plus isocitrate, $NADP^+$ or in the absence of any substrates. It is a long-term aim to modify ICL in such a way that it competes for the common substrate more efficiently so that a mutant organism could direct a greater proportion of isocitrate through the glyoxylate bypass in order to produce more biomass.

It was hoped that the three dimensional structure of this enzyme would be determined by the time I was about to start to generate mutants of the *E. coli* ML308 ICL. However, this was not the case. Crystals of ICL purified from the overexpressing strain KAT-1/pEM9 had been made but they were not suitable for X-ray diffraction, apparently because they were not sufficiently ordered.

Chemical modification studies of ICL from a number of organisms had shown that there may be a Cys, a His and either a Glu or an Asp at the active site. Cys195 was identified as being at the active site of the *E. coli* enzyme independently by both our group using iodoacetate and McFadden's group using 3-BrP to modify the active site residue. This residue is located within a very highly conserved region of the derived amino acid sequence and is conserved in all, the ICL sequences to date. The site directed mutagenesis work reported in Chapter 4 confirmed that Cys195 plays an important role in catalysis and allowed me to make some speculations as to its precise role that are discussed in section 4.6.

The site directed mutagenesis of Cys195 was made feasible because the amino acid sequence around the modified residue was determined. For the other putative active site residues, this sort of data were not available. Without the three-dimensional structure, two other approaches may identify possible active site residues. Chemical modification can identify putative active site residues, which if conserved in other ICLs it can be established whether or not these residues are likely to be important. A second approach is comparison of the ICL sequences with that of ICDH given that the residues which bind isocitrate have been identified for this enzyme from the three-dimensional structure in the presence of Mg^{2+} and isocitrate.

Since ICDH also binds $NADP^+$ but ICL does not, it is very unlikely that the two enzymes are topologically highly related. Nevertheless ICDH and ICL both bind isocitrate. Comparison of the secondary structure predictions of ICL with ICDH indicates that there may be some similarity in the structures of the isocitrate binding pockets, and comparison of the ICL sequence with that of ICDH may indicate which residues may be at the active site (e.g. the His and Asp or Glu residues for which there is chemical modification evidence).

Recently, McFadden's group reported that His266 and His306 could be at or near the active site of the *E. coli* ICL, based on chemical modification by DEPC and protection of inhibition by the substrates. However, work done here in Glasgow has indicated that only His 306 is modified by DEPC and that it is unlikely that this residue is at the active site (Rua, Robertson and Nimmo, submitted). This does not however rule out the possibliity that there is a His at or near the active site. It is of note that His266 has been proposed as the site of phosphorylation of ICL in *E. coli*(Matsuoka and McFadden, 1988). There are four conserved His residues between all the ICL sequences, at positions 184, 197, 356 and 399. Comparison of ICL with the ICDH sequence reveals that of the His residues conserved between the ICLs, only His197 is also conserved in ICDH. Since His197 is very close to the active site Cys195, it seems likely that this residue is also at or near the active site. Also, the corresponding residue in ICDH, His193 is part of the clasp-like domain and as such is very close to the isocitrate binding site in a second subunit. Further indication that a His residue is at the active site came from the pH studies with the mutant ICLs. Studies with the wild type ICL suggested that there were two ionising groups affecting K_m with pK_as of very similar values within the pH range 6.8-8.5 and as such the values for each group could not be determined. However the mutants had only one ionising group which affected K_m with a pK_a of 7.8, which is close to the value usually found for His residues in enzymes.

It has been shown that the substrate analogue, itaconate epoxide, inactivates ICL from *Citrullis vulgaris* by modifying either an Asp or a Glu residue. There are 13 conserved Asp residues and 9 conserved Glu residues between the ICL sequences. When compared to the ICDH sequence, only Asp31, Glu208 and Glu 260, corresponding to Asp35, Glu204 and Glu256 in the ICDH sequence are conserved. However, there are two conservative changes in the ICDH sequence which has Glu residues at positions 274 and 371 which correspond to

residues Asp294 and Asp409 in the *E. coli* ICL sequence. Considering the structure of ICDH, Glu208 could be close to the isocitrate binding site of ICL.

Another approach to identify residues in ICL which may bind isocitrate is to look for conservation of residues which bind isocitrate in ICDH. The residues which bind isocitrate in ICDH have been identified and are shown in figure 5.1. When aligned with the *E. coli* ML308 ICL sequence, there are distinct similarities between the two enzymes. Ser113 in ICDH forms a hydrogen bond with the C5 carboxyl group of isocitrate. In *E. coli* ICL, Ser113 corresponds to Arg123, which may form an electrostatic interaction with the same group on the substrate. In the other ICLs, Ser113 corresponds to a His residue, which if protonated, could form a hydrogen bond or an electrostatic interaction with the C5 carboxyl. Asn115 in ICDH corresponds to Asn125 in *E. coli* ICL, although in the other ICL sequences, the same residue corresponds to an Asp. Arg119 in ICDH forms two hydrogen bonds with two carboxyl groups on the substrate. The corresponding residue in *E. coli* ICL is Arg129 and a His in the other enzymes, which if protonated, may form hydrogen bonds in the same manner. Arg129 in ICDH does not have a corresponding residue in any of the ICL sequences as there is a gap in the ICL sequence spanning β sheet-F (residues 124-132) in the ICDH structure. Asp311 in ICDH corresponds to Lys331 in the *E. coli* ICL sequence although at the same position in *G. hirsutum* and *R. communnis* ICL, there is a Glu, in *B. napus*, there is an Gln and in *C. tropicalis*, there is an Asp. Therefore, at this position, the possible nucleophile is conserved except in the *E. coli* enzyme. Asp 283 in the ICDH sequence, which coordinates the Mg^{2+} in the active site that is mainly contributed by second subunit, corresponds to Gln303 in the *E. coli* ICL sequence, a Glu in the plant enzymes and an Asp in yeast ICL. Asp307 in the ICDH sequence coordinates Mg^{2+} and corresponds to Asn327 in the *E. coli* ICL, while in the other ICL enzymes, Asp307 corresponds to either an Ala or a Gly. The ICDH Arg153 corresponds to either a Val or Ile in the ICL sequences. Lys230 in ICDH forms a hydrogen bond with the carboxyl group at the C3 position of isocitrate bound at the second subunit. This residue corresponds to Asp234 in the *E. coli* ICL sequence and the same residue in all the other ICL sequences. Tyr160 in ICDH forms a hydrogen bond to the same carboxyl group as Lys230 in the other subunit. In the *E. coli* ICL, this residue corresponds to Phe162 which is

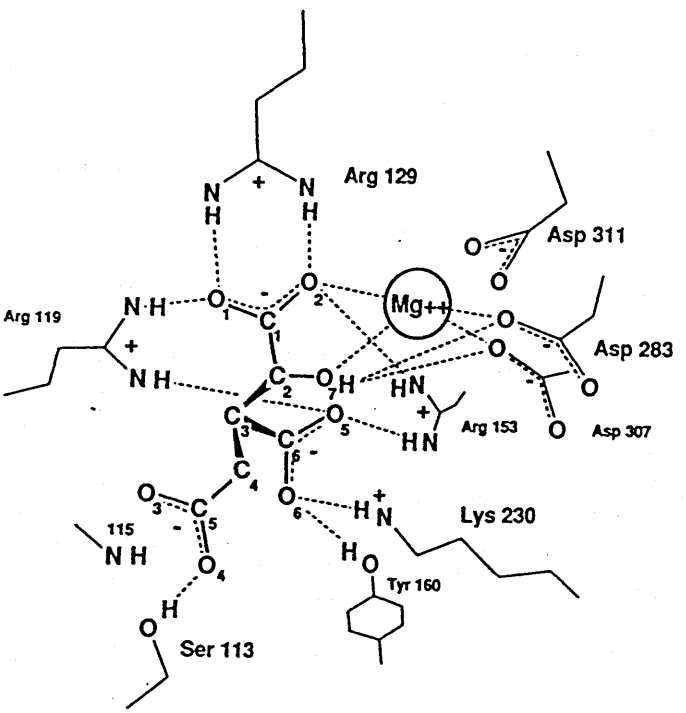


Figure 5.1: Binding Of Mg^{2+} Isocitrate In The Active Site Of ICDH

Taken from Hurley *et al.* (1990a). Hydrogen bonds and salt bridges between the enzyme and substrates are indicated by dashed lines. Residues of the second subunit are indicated by a prime. Water molecules are not shown.

conserved in all other ICL sequences.

Thus it appears that between *E. coli* ICDH and ICL, the following residues in the ICL sequence are conserved or there are conservative changes with the charge remaining the same: Arg123, Asn115 and Arg119. In the other ICL sequences, the charge at Arg119, Asp311 and Asp283 in ICDH is conserved but the –NH group of Asn115 is substituted by a carboxyl group of Asp.

Both of the above approaches have identified several posible active site residues. Until the three dimensional structure of ICL has been determined, the actual structure and composition of the active site will not be known. However, site directed mutagenesis could be used to replace these residues and so determine whether they are at the active site or not.

The second approach also reveals possible reasons as to the difference in dissociation constants between ICL and ICDH. For example, the negative charge at Asp283 in ICDH is replaced in *E. coli* ICL by a Gln, Asp311 is replaced by a Lys which has the opposite charge and the hydroxyl of Tyr160 in ICDH which forms a hydrogen bond with the substrate is replaced in all the ICL enzymes by a Phe which has no hydroxyl group. Site directed mutagenesis of the above residues in ICL replacing them with the corresponding residues in ICDH may result in ICL enzymes which have a lower K_m for isocitrate than the wild type enzyme and so may be able to compete for the common substrate more efficiently.

5.4 Importance Of Sequence Downstream Of *aceK*:

Despite the fact that pCL8, a subclone of pCL1000, contains the entire coding region of *aceK*, no proteins were found to co-migrate with ICDH kinase/phosphatase on SDS gels after maxicell labelling. This was attributed to a sharp downshift in expression of *aceK* with respect to *aceA* caused by the presence of a stem-loop structure in the *aceA-aceK* intercistronic region, found from the sequence of this region (Klumpp *et al.*, 1988; Cortay *et al.*, 1988). However, it seems more likely that the stem-loop prevent degradation of mRNA transcribed from *aceA* since deletion of the *aceA-aceK* intercistronic region results in a decrease in expression of *aceA* (Klumpp *et al.*, 1988; section 1.3.6). In addition pEM9 does express detectable levels of ICDH kinase/phosphatase. The HindIII site at the end of the insert in

pEM901 corresponds to the HindIII site at the end of pCL1000. Assuming that the functional maps of pEM901 and pCL8 are similar, the fact that pEM901, like pCL8, cannot express ICDH kinase/phosphatase indicates that DNA downstream of the HindIII site is essential for expression of ICDH kinase/phosphatase. One possibility is that the HindIII site lies upstream of, or is included in a stem-loop which prevents digestion of the mRNA from the 3' end, similar to that proposed in the *aceA-aceK* intercistronic region. Indeed, when the *aceB* gene was used as a probe in a northern blot of *E. coli* RNA, it was found that the probe hybridised to a 10kb transcript, more than 3kb longer than is required to encode all three genes in the *ace* operon. Since the transcription startsite has been located immediately upstream of *aceB* (Chung *et al.*, 1988), this suggests that a large region downstream of *aceK* is also transcribed. This region has recently been sequenced and an open frame was shown to be present in the opposite orientation to the *ace* operon (Galinier *et al.*, 1991; section 1.3.2). However, this sequence has not been analysed for the presence of any possible REP or other stem-loop structures.

5.5 Future Work:

The work carried out in this thesis has provided several directions for future work on understanding the reaction mechanism of ICL from *E. coli* ML308. The determination of the derived amino acid sequence will greatly aid the determination of the three dimensional structure of this enzyme. In conjunction with the three dimensional structure, knowledge of the primary structure allows the generation of a number of mutant enzymes which will effect the activity of ICL. Therefore, the determination of the three dimensional structure of ICL is a major priority.

Attempts have been made to crystallize ICL purified from *E. coli* KAT-1/pEM9. However, it was found that during crystallization, proteolytic degradation had occurred, resulting in non-uniform crystals which were not suitable for X-ray diffraction. Using ICL purified from *E. coli* JE10/pICL-1, crystals were formed, but unfortunately, these were not suitable for X-ray diffraction either. Another approach was taken in that ICL which had been modified using 3-BrP was crystallized giving crystals which were morphologically similar to the native crystals although the modified crystals are not disordered. In addition, crystals of

ICL have been grown in the presence of heavy atoms. These crystals have also been shown to be suitable for X-ray diffraction. Work is now underway to determine the three dimensional structure of ICL from *E. coli* (Abeysinghe *et al.*, 1991). However, even without knowledge of the three dimensional structure, site directed mutagenesis could still be done, replacing some of the residues mentioned in the previous section.

Of particular interest is the suggestion that one of the four His residues conserved in all ICLs may be located at the active site and affect the K_m for isocitrate, and perhaps the k_{cat} also. These residues are obvious targets for further site directed mutagenesis.

The suggestion that ICDH and ICL share an isocitrate binding motif is also of interest. It might be possible to obtain further information on the isocitrate binding site of ICL by random mutagenesis (Oliphant and Struhl, 1989). Strains containing ICL but no ICDH kinase/phosphatase activity are unable to grow, or grow very slowly on acetate as the sole carbon source because their intracellular isocitrate concentration is too low. It should therefore be possible to select for ICLs with a reduced K_m for isocitrate by virtue of an increased growth rate on acetate. This would allow identification of residues that affect the binding of isocitrate. It would also have another advantage. It would become possible to study the phosphorylation of ICDH in a strain in which the isocitrate concentration did not have to be increased by as much as in the wild type to allow sufficient flux through the glyoxylate bypass. The behaviour of the phosphorylation system in such a strain would be of considerable interest. It seems likely that mutagenesis along these lines will prove invaluable in the analysis of regulatory systems in the future.

References

Abeysinghe, S.I.B., Baker, P.J., Rice, D.W., Rodgers, H.E., Stillmann, T.J., Ko, Y.H., McFadden, B.A. and Nimmo, H.G. (1991) J. Mol. Biol. **220**, 13-16

Anderson, S. (1981) Nucleic Acids Res. **9**, 3015-3027

Ashworth, J.M. and Kornberg H.L. (1963) Biochem. Biophys. Acta **73**, 519-522

Atomi, H., Ueda, M., Hikida, M., Hishida, T., Teranishi, Y. and Tanaka, A. (1990) J. Biochem **107**, 262-266

Bachmann, B.J. and Low, K.B. (1980) Microbiol. Rev. **44**, 1-56

Barnes, W.M. (1983) Nucleic Acids Res. **11**, 349-368

Barnes, W.M., Bevan, M. and Son, P.H. (1983) in "Methods in Enzymology **101**" 98-122 (Eds. Wu, R., Grossman, L. and Moldave, K.. Academic Press, New York.

Bautista, J., Satrusteui, J. and Machado, A. (1979) FEBS Lett. **105**, 333-336

Beeching J.R. and Northcote, D.H. (1987) Plant Mol. Biol. **8**, 471-475

Belasco, J.G. and Chen, C.A. (1988) Gene **72**, 109-117

Bennett, P.M. and Holms, W.H. (1975) J. Gen. Microbiol. , 37-51

Berg, P. (1956) J. Biol. Chem. **222**, 991-1013

Bigelow, C.C. (1967) J. Theor. Biol. **16**, 187-211

Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. **7**, 1513-1523

Borst, P. (1986) Biochem. Biophys. Acta **866**, 179-203

Borthwick, A.C., Holms, W.H. and Nimmo, H.G. (1984a) Eur. J. Biochem. **141**, 393-400

Borthwick, A.C., Holms, W.H. and Nimmo, H.G. (1984b) Biochem. J. **222**, 797-804

Borthwick, A.C., Holms, W.H. and Nimmo, H.G. (1984c) FEBS Lett. **174**, 112-115

Bradford, M.M. (1976) Anal. Biochem. **72**, 248-254

Brice, C.B. and Kornberg, H.L. (1968) J. Bacteriol. **96**, 2185-2186

Britten, R.J. (1954) Science **119**, 578

Brown, T.D.K., Jones-Mortimer, M.C. and Kornberg, H.L.(1977) J. Gen. Microbiol. **102**, 327-336

Byrne, C., Stokes, H.W. and Ward, K.A. (1988) Nucleic Acids Res. **16**, 9342

Calos, M.P. (1978) Nature (London) **274**, 762-765

Campbell, J.J.R., Smith R.A. and Eagles, B.A. (1953) Biochem. Biophys. Acta **11**, 594

Celenza, J.L. and Carlson, M. (1986) Science **233**, 1175-1180

Chell, R.M., Sundaram., T.K. and Wilkinson, A.E. (1978) Biochem. J. **173**, 165-177

Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. **47**, 45-148

Chou, T.C. and Lipmann, F. (1952) J. Biol. Chem. **196**, 89-103

Chung, T., Klumpp, D.J. and LaPorte, D.C. (1988) J. Bacteriol. **170**, 386-392

Cionni, M., Pinzauti, G. and Vanni, P. (1981) Comp. Biochem. Physiol. **708**, 1-26

Cohn, E.J. and Edsall, J.T. (1943) Proteins, Amino Acids and Peptides 370-381. Reinhold (New York)

Colona, W.J. and McFadden, B.A. (1975) Arch. Biochem. Biophys. **170**, 608-619

Comai, L., Dietrich, R.A., Maslyar, D.J., Baden, C.S. and Harada, J.J. (1989) The Plant Cell **1**, 293-300

Conder, M.J., Ko, Y. and McFadden, B.A. (1988) Prepar. Biochem. **18**, 431-442

Cortay, J.C., Bleicher, F., Duclos, B., Cenatiempo, Y., Gautier, C., Prato, J.L. and Cozzone, A.J. (1989) Biochimie **71**, 1043-1049

Cortay, J.C., Bleicher, F., Rieul, C., Reeves, H.C. and Cozzone, A.J. (1988) J. Bacteriol. **170**, 89-97

Cortay, J.C., Negre, D., Garlinier, A., Duclos, B., Perriere, G. and Cozzone, A.J. (1991) EMBO J. **10**, 675-679

Cortay, J.C., Reeves, H,C. and Cozzone, A.J. (1986) Current Microbiol. **13**, 251-254

Dagert, M. and Ehrlich, S.D. (1979) Gene **6**, 23-28

Davis, B.J. (1964) Ann. New York Acad. Sci. **121**, 404-427

Dawson, R.M.C., Elliot, D.C., Elliot, W.H. and Jones, K.M. (1986) Data For Biochemical Research (Third Edition). Oxford Science Publications.

Dean, A.M. and Koshland, D.E.Jr. (1990) Science **249**, 1044-1046

Dean, A.M., Lee, M.H.I. and Koshland, D.E.Jr. (1989) J. Biol. Chem. **264**, 20482-20486

DeBoer, H., Comstock, L.J. and Vasser, M. (1983) Proc. Natl. Acad. Sci. USA **80**, 21-25

Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387-395

DiRusso, C.C. and Nunn, W.D. (1985) J. Bacteriol. **161**, 583-588

Dixon, G.H. and Kornberg, H.L. (1959) Biochem. J. **72**, 3P

Dixon, M. (1953) Biochem. J. **55**, 161-170

Doolittle, R.F., Johnson, M.S., Hussain, I., Van Houten, B., Thomas, D.C. and Sancar, A. (1986) Nature (London) **323**, 451-453

Duckworth, M.L. (1981) Ph.D. Thesis, University of Manitoba, Canada

El-Mansi, E.M.T., MacKintosh, C., Duncan, K., Holms, W.H. and Nimmo, H.G. (1987) Biochem. J. **242**, 661-665

El-Mansi, E.M.T., Nimmo, H.G. and Holms, W.H. (1985) FEBS Lett. **193**, 251-255

Falmange, P., Vanderwinkel, E. and Wiame, J.M. (1965) Biochem. Biophys. Acta **99**, 246-258

Fersht, A. (1977) Enzyme Structure And Mechanism (second edition) W.H. Freeman and Company

Fisher, H.F. (1964) Proc. Natl. Acad. Sci. USA **51**, 1285-1291

Frevert, J. and Kindl, H. (1978) Eur. J. Biochem. **92**, 35-43

Frischauf, A.M., Garoff, H. and Lehrach, H. (1980) Nucleic Acids Res. **8**, 5541-5549

Galassi, C., Giachetti, E., Pinzauti, G. and Vanni, P. (1988) Phytochemistry **27**, 1303-1307

Garland, D. and Nimmo, H.G. (1984) FEBS Lett. **165**, 259-263

Garlinier, A., Bleicher, F., Negre, D., Perriere, G., Duclos, B., Cozzone, A.J. and Cortay, J.C. (1991) Gene **97**, 149-150

Garlinier, A., Negre, D., Cortay, J.C., Marcandier, S., Maloy, S.R. and Cozzone A.J. (1990) Nucleic Acids Res. **18**, 3656

Garnak, M. and Reeves, H.C. (1979a) Science **203**, 1111-1112

Garnak, M. and Reeves, H.C. (1979b) J. Biol. Chem. **254**, 7915-7920

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. **120**, 97-120

Giachetti, E., Pinzauti, G., Bonaccorsi, R., Vincenzini, M.T. and Vanni, P. (1987) Phytochemistry **26**, 2439-2446

Giachetti, E., Pinzauti, G., Vincenzini, M.T. and Vanni, P. (1982) Ita. J. Biochem. **31**, 81-89

Giatechetti, E., Pinzauti, G., Bonaccorsi, R. and Vanni, P. (1988) Eur. J. Biochem. **172**, 85-91

Goldbeter, A. and Koshland, D.E.Jr. (1981) Proc. Natl. Acad. Sci. USA **78**, 6840-6844

Goldsmith, E.J., Sprang, S.R., Hamlin, R., Xoung, N.H. and Fletterick (1989) Science **245**, 528-532

Gould, S.J., Keller, G.-A. and Subramani, S. (1988) J. Cell Biol. **107**, 897-905

Gribskov, M., Devereux, J. and Burgess, R.R. (1984) Nucleic Acids Res. **12**, 539-549

Hanks, S.K., Quinn, A.M. and Hunter, T. (1988) Science **241**, 42-52

Hatch, F.T. (1965) Nature (London) **206**, 777-778

Heidecker, G., Messing, J. and Gronenborn, B. (1980) Gene **10**, 69-73

Henikoff, S. (1984) Gene **28**, 351-359

Holms, W.H. (1986) Curr. Topics In Cell. Reg. **28**, 69-105

Holms, W.H. and Bennett, P.M. (1971) J. Gen. Microbiol. **65**, 57-68

Hoyt, J.C. and Reeves, H.C. (1988) Biochem. and Biophys. Res. Commun. **153**, 875-880

Hoyt, J.C., Robertson, E.F., Berlyn, K.A. and Reeves, H.C. (1988) Biochem. Biophys. Acta **966**, 30-35

Huang, K-P. and Cabib, E. (1974) J. Biol. Chem. **249**, 3815-3857

Hurley, J.H., Dean, A.M., Koshland, D.E.Jr. and Stroud, R.M. (1991) Biochemistry **30**, 8671-8678

Hurley, J.H., Dean, A.M., Sohl, J.L., Koshland, D.E. Jr. and Stroud, R.M. (1990a) Science **249**, 1012-1016

Hurley, J.H., Dean, A.M., Thorsness, P.E., Koshland, D.E. Jr. and Stroud, R.M. (1990b) J. Biol. Chem. **265**, 3599-3602

Hurley, J.H., Thorsness, P.E., Ramalingam, V., Helmers, N.H., Koshland, D.E.Jr. and Stroud, R.M. (1989) Proc. Natl. Acad. Sci. **86**, 8635-8639

Igamberdiev, A.V. and Zemlyanukin, A.A. (1987) Biochemistry (USSR) **52**, 1104-1111

Jameel, S., El-Gul, T. and McFadden, B.A. (1984) Phytochemistry **23**, 2753-2759

Jameel, S., El-Gul, T. and McFadden, B.A. (1985) Arch, Biochem. Biophys. **236**, 72-81

Johanson, R.A. and Reeves, H.C. (1977) Biochem. Biophys. Acta **483**, 24-34

Johanson, R.A., Hill, J.M. and McFadden, B.A. (1974a) Biochem. Biophys. Acta **364**, 327-340

Johanson, R.A., Hill, J.M. and McFadden, B.A. (1974b) Biochem. Biophys. Acta **364**, 341-352

John, P.C.L. and Syrett, P.J. (1967) Biochem. J. **105**, 409-416

Jordan-Starck, T.C. and Rodwell, V.W. (1989) J. Biol. Chem. **264**, 17919-17923

Kamel, M.Y. and Fahmy, A.S. (1982) Comp. Biochem. Physiol. **72B**, 107-115

Kaplan, N.O. and Lipmann, F. (1948) Federation Proc. **7**, 163

Khan, F.R. and McFadden, B.A. (1982) Plant Physiol. **70**, 943-948

Khan, F.R., Saleemuddin, M., Siddiqi, M. and McFadden, B.A. (1977) Arch. Biochem. Biophys. **183**, 13-23

Klionsky, D.J., Skalnik, D.G. and Simoni, R.D. (1986) J. Biol. Chem. **261**, 8096-8099

Klumpp, D.J., Plank, D.W., Bowdin, L.J., Stueland, C.S., Chung, T. and LaPorte, D.C. (1988) J. Bacteriol. **170**, 2763-2769

Ko, Y.H. and McFadden, B.A. (1990) Arch. Biochem. Biophys. **278**, 373-380

Ko, Y.H., Vanni, P., Munske, G.R. and McFadden, B.A. (1991) Biochemistry **30**, 7451-7456

Kornberg, H.L. (1958) Biochem. J. **68**, 535-542

Kornberg, H.L. (1963) Biochem. Biophys. Acta **73**, 517-522

Kornberg, H.L. (1966) Biochem. J. **99**, 1-11

Kornberg, H.L. and Madsen, N.B. (1958) Biochem. J. **68**, 549-557

Kornberg, H.L. and Quayle, J.R. (1958) Biochem. J. **68**, 542-549

Kornberg, H.L., Phizackery, P.J.R. and Sadler, J.R. (1960) Biochem. J. **77**, 438-445

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

Lamb, J.E., Riezman, H., Becker, W.H. and Leaver, J. (1978) Plant Physiol. **62**, 754-760

LaPorte, D.C. and Chung, T. (1985) J. Biol. Chem. **260**, 15291-15297

LaPorte, D.C. and Koshland, D.E.Jr. (1982) Nature (London) **300**, 458-460

LaPorte, D.C. and Koshland, D.E.Jr. (1983) Nature (London) **305**, 286-290

LaPorte, D.C., Stueland, C.S. and Ikeda, T.P. (1989) Biochimie **71**, 1051-1057

LaPorte, D.C., Thorsness, P.E. and Koshland, D.E.Jr. (1985) J. Biol. Chem. **260**, 10563-10568

LaPorte, D.C., Walsh, K. and Koshland, D.E.Jr. (1984) J. Biol. Chem. **259**, 14068-14075

Lazarow, P.B. and Fujiki, Y. (1985) Ann. Rev. Cell Biol. **1**, 489-530

Levine, S.R., Ardeshir, F. and Ames, G.F. (1980) J. Bacteriol. **143**, 1081-1085

Lewendon, A., Murray, I.A., Shaw, W.V., Gibbs, M.R. and Leslie, A.G.W. (1990) Biochemistry **29**, 2075-2080

Liu, J., Albers, M.W., Chen, C-M., Schreiber, S.L. and Walsh, C.T. (1990) Proc. Natl. Acad. Sci. USA **87**, 2304-2308

Locht, C., Lobet, Y., Feron, C., Gieplak, W. and Keith, J.M. (1990) J. Biol. Chem. **265**, 4552-4559

Lowry, O.H., Carter, J., Ward J.B. and Glaser, L. (1971) J. Biol. Chem. **246**, 6511-6521

MacDonald, M.J. and Chang, C.-M. (1985) Mol. Cell. Biochem. **68**, 115-120

MacKintosh, C. (1987) Ph.D Thesis, University Of Glasgow, U.K.

MacKintosh, C. and Nimmo, H.G. (1986) Biochem. Soc. Trans. **14**, 320-321

MacKintosh, C. and Nimmo, H.G. (1988) Biochem. J. **250**, 25-31

Malhorta, O.P. and Srivastava, P.K. (1982) Arch. Biochem. Biophys. **214**, 164-171

Malloy, P.J., Reeves, H.C. and Spiess, J (1984) Curr. Microbiol. **11**, 37-42

Maloy, S.R. and Nunn, W.D. (1982) J. Bacteriol. **149**, 173-180

Maloy, S.R., Bohlander, M. and Nunn, W.D. (1980) J. Bacteriol. **143**, 720-725

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Marchalonis, J.J. and Weltman, J.K. (1971) Comp. Biochem. Physiol. **38B**, 609-625

Marr, J.J. and Weber, M.M. (1969) J. Biol. Chem. **243**, 4973-4979

Matsuoka, M. and McFadden, B.A. (1988) J. Bacteriol. **170**, 4528-4536

Matsuoka, M. Himeno, T. and Aiba, S. (1984) J. Bacteriol. **157**, 899-908

McCarthy, J.T. and Charles, A.M. (1973) Can. J. Microbiol. **19**, 513-519

McFadden, B.A. and Howes, W.V. (1963) J. Biol. Chem. **238**, 1737-1742

McFadden, B.A., Rao, G.R., Cohen, A.L. and Roche, T.E. (1968) Biochemistry **7**, 3574-3582

McKee, J.S., Hlodan, R. and Nimmo, H.G. (1989) Biochimie **71**, 1059-1064

Melgar, E. and Goldthwaite, D.A. (1968) J. Biol. Chem. **243**, 4409-4416

Messing, J., (1983) in "Methods in Enzymology **101**" 20-78 (Eds. Wu, R., Grossman, L. and Moldave, K.. Academic Press, New York.

Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. **9**, 309-321

Morikawa, M., Izui,K., Taguchi, M. and Katzuki, H. (1980) J. Biol. Chem. **243**, 5385-5395

Murray, N.E., Brammar, W.J. and Murray, K. (1977) Molec. Gen. Genet. **150**, 53-61

Nakamaye, K. and Eckstein, F. (1986) Nucleic Acids Res. **14**, 9679-9698

Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. **48**, 443-453

Negre, D., Cortay, J.C., Old, I.A., Garlinier, A., Richaud, C., Saint-Girons, I. and Cozzone, A.J. (1991) Gene **97**, 29-37

Nimmo, G.A. and Nimmo, H.G. (1984) Eur. J. Biochem. **141**, 409-414

Nimmo, G.A., Borthwick, A.C., Holms, W.H. and Nimmo, H.G. (1984) Eur. J. Biochem. **141**, 401-408

Nimmo, H.G. (1986) Biochem. J. **234**, 317-323

Nimmo, H.G., Borthwick, A.C., El-Mansi, E.M.T., Holms, W.H., MacKintosh, C. and Nimmo, G.A. (1987) Biochem. Soc. Symp. **54**, 93-101

Nimmo, H.G., Douglas, F., Kleanthous, C., Campbell, D.G. and MacKintosh, C. (1989) Biochem. J. **261**, 431-435

Nunn, W.D., Giffin, K., Clark, D. and Cronan, J.E.Jr. (1983) J. Bacteriol. **154**, 554-560

Nunn, W.D., Simons, R.W., Egan, P.A. and Maloy, S.R. (1979) J. Biol. Chem. **254**, 9130-9134

Oliphant, A.R. and Struhl, K. (1989) Proc. Natl. Acad. Sci. USA **86**, 9094-9098

Olson, J.A. (1954) Nature (London) **174**, 695-696

Olson, J.A. (1959) J. Biol. Chem. **234**, 5-10

Overath, P., Pauli, G. and Schairder, H.U. (1969) Eur. J. Biochem. **7**, 559-574

Paden, E., Zilberstein, D. and Schuldiner, S. (1981) Biochem. Biophys. Acta **650**, 151-166

Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA **85**, 2444-2448

Pinzauti, G., Giachetti, E. and Vanni, P. (1982) Int. J. Biochem. **14**, 267-275

Pinzauti, G., Giachetti, E. and Vanni, P. (1983) Arch. Biochem. Biophys. **225**, 137-142

Pinzauti, G., Giachetti, E., Camici, G., Manao, G., Cappugi, G. and Vanni, P. (1986) Arch. Biochem. Biophys. **244**, 85-93

Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. and Surrey, S. (1982) Proc. Natl. Acad. Sci. USA **79**, 4298-4302

Ragland, T.E., Kawasaki, T. and Lowenstein, J.M. (1966) J. Bacteriol. **91**, 236-244

Reed, S.I., Hadawiger, J.A. and Lorincz, A.T. (1985) Proc. Natl. Acad. Sci. **82**, 4055-4059

Reiss, U. and Rothstein, M. (1974) Biochemistry **13**, 1796-1800

Rieul, C. Bleicher, F., Duclos, B., Cortay, J.C. and Cozzone, A.J. (1988) Nucleic Acids Res. **16**, 5689

Riezman, H., Weir, E.M., Leaver, C.J., Titus, D.E. and Becker, W.M. (1980) Plant Physiol. **65**, 40-46

Rittenhouse, J.O.W. and McFadden, B.A. (1974) Arch. Biochem. Biophys. **163**, 79-86

Roberts, L.M. and Lord, J.M. (1981) Eur. J. Biochem. **119**, 43-49

Robertson, E.F. and Reeves, H.C. (1987) Curr. Microbiol. **14**, 347-350

Robertson, E.F., Hoyt, J.C. and Reeves, H.C. (1987) Curr. Microbiol. **15**, 103-105

Robertson, E.F., Hoyt, J.C. and Reeves, H.C. (1988) J. Biol. Chem. **263**, 2477-2482

Roche, T.E. and McFadden, B.A. (1969) Biochem. Biophys. Res. Commun. **37**, 239-246

Roche, T.E., McFadden, B.A. and Williams, J.O. (1971) Arch. Biochem. Biophys. **147**, 192-200

Rogers, J.E. and McFadden, B.A. (1976) Arch. Biochem. Biophys. **174**, 695-704

Rogers, J.E. and McFadden, B.A. (1977) Arch. Biochem. Biophys. **180**, 348-353

Rosenberg, H. and Court, D. (1979) Ann. Rev. Genet. **13**, 319-353

Rosenberger, R.F. (1962) Biochem. Biophys. Acta **64**, 168-170

Rua, J., DeArriaga, D., Busto, F. and Soler, J. (1990) Biochem. J. **272**, 359-367

Ruchti, M. and Windmer, F. (1984) J. Exper. Bot. **37**, 1685-1690

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

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

Sayers, J.R., Schmidt, W. and Eckstein, F. (1988) Nucleic Acids Res. **16**, 791-802

Saz, H.J. (1954) Biochem. J. **58**, xx.

Scheller, R.N., Dickerson, R.E., Boyer, H.W., Riess, A.D. and Itakura, K. (1977) Science **196**, 177-180

Schloss, J.V. and Cleland, W.W. (1982) Biochemistry **21**, 4420-4427

Shiio, I. and Ozaki, H. (1968) J. Biochem. (Tokyo) **64**, 45-53

Shiio, I. Shiio, T. and McFadden, B.A. (1965a) Biochem. Biophys. Acta **96** 114-122.

Shiio, I. Shiio, T. and McFadden, B.A. (1965b) Biochem. Biophys. Acta **96** 123-133.

Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA **71**, 1342-1346

Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell **20**, 269-281

Simons, R.W., Egan, P.A., Chute, H.T. and Nunn, W.D. (1980a) J. Bacteriol. **142**, 621-632

Simons, R.W., Hughes, K.T. and Nunn, W.D. (1980b) J. Bacteriol. **143**, 726-730

Smith T.F. and Waterman, M.S. (1981) Adv. Appl. Math. **2**, 482-489

Smith, R.A. and Gunsalus, I.C. (1954) J.Amer. Chem. Soc. **76**, 5002-5003

Smith, R.A. and Gunsalus, I.C. (1957) J. Biol. Chem. **229**, 305-319

Stadtman, E.R. and Barker, H.A. (1950) J. Biol. Chem. **184**, 769-793

Streuli, M., Krueger, N.X., Tsai, A.Y.M. and Saito, H. (1989) Proc. Natl. Acad. Sci. USA **86**, 8698-8702

Stueland, C.S., Eck, K.R., Stieglbauer, K.T. and LaPorte, D.C. (1987) J. Biol. Chem. **262**, 16095-16099

Stueland, C.S., Ikeda, T.P. and LaPorte, D.C. (1989) J. Biol. Chem. **264**, 13775-13779

Sunnarborg, A., Klumpp, D., Chung, T. and LaPorte, D.C. (1990) J. Bacteriol. **172**, 2642-2649

Tabor, S. and Richardson, C.C. (1989) Proc. Natl. Acad. Sci. USA **86**, 4076-7080

Taylor, J.W., Ott, J. and Eckstein, F. (1985b) Nucleic Acids Res. **13**, 8764-8785

Taylor, J.W., Schimdt, W., Cosstick, R., Okruszek, A. and Eckstein, F. (1985a) Nucleic Acids Res. **13**, 8749-8764

Thorsness, P.E. and Koshland, D.E.Jr. (1987) J. Biol. Chem. **262**, 10422-10425

Tsukamoto, C., Ejiri, S. and Katsumata, T. (1986) Agric. Biol. Chem. **50**, 409-416

Turley, R.B., Choe, S.M. and Trelease, R.N. (1990) Biochem. Biophys. Acta **1049**, 223-226

Uchida, M., Ueda, M., Matsuki, T., Okada, H., Tanaka, A. and Fukui, S. (1986) Agric. Biol. Chem. **50**, 127-134

Valera, I. and Nimmo, H.G. (1988) FEBS Lett. **231**, 361-365

Van Heeke, G.V. and Schuster, S.M. (1989) J. Biol. Chem. **264**, 19475-19477

Vanderwinkel, E. and DeVliegher, M. (1968) Eur. J. Biochem. **5**, 81-90

Vanderwinkel, E., Laird, P., Ramos, F. and Wiame, J.M. (1963) Biochem. Biophys. Res. Commun. **12**, 157-162

Vanni, P., Gichetti, E., Pinzauti, G. and McFadden, B.A. (1990) Comp. Biochem. Physiol. **95B**, 431-458

Vanni, P., Vincenzini, M.T., Nerozzi, F.M. and Sinha, S.P. (1979) Can. J. Biochem. **57**, 1131-1137

Vincenzini, M.T., Vanni, P., Gichetti, E., Hanozet, G.M. and Pinzauti, G. (1986) J. Biochem. **99**, 375-383

Vinopal, R.T. and Fraenkel, D.G. (1974) J. Bacteriol. **118**, 1090-1100

Volokita, M. and Somerville, C.R. (1987) J. Biol. Chem. **262**, 15825-15828

Volz, K.W., Mathews, D.A., Alden, R.A., Freer, S.T., Hansch, C., Kaufman, B.T. and Kraut, J. (1982) J. Biol. Chem. **257**, 2528-2536

Walsh, C. (1979) Enzymatic Reaction Mechanisms, W.H. Freeman and Company.

Walsh, K. and Koshland, D.E.Jr. (1984) J. Biol. Chem. **259**, 9646-9654

Walsh, K. and Koshland, D.E.Jr. (1985) Proc. Natl. Acad. Sci. USA **82**, 3577-3581

Waugh, D.F. (1954) In "Advances in Protein Chemistry" IX, 341-358, Edited by Anson, M.L., Bailey, K. and Edsall, J.T., Academic Press.

Wierenga, P.K., Swinkels, B., Michels, P.A.M., Osinga, K., Misset, O., Van Beeumen, J., Gibson, W.C., Postma, J.P., Borst, P., Opperdoes, F.R. and Hol, W.G.J. (1987) EMBO J. **6**, 215-221

Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA **80**, 726-730

Williams, J.O., Roche, T.E. and McFadden, B.A. (1971) Biochemistry **10**, 1384-1390

Wong, D.T.O. and Ajl, S.J. (1956) J. Amer. Chem. Soc. **78**, 3230-3237

Wu, R., Ruben, G., Siegel, B., Jay, E., Spielman, P. and Tu, C.D. (1976) Biochemistry **15**, 734-740

Yanish-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103-119

Zemlyanukhin, A.A., Igamberdiev, A.U. and Presnyakova, E.N. (1986) Biochemistry (USSR) **51**, 369-375

Zemlyanukhin, L.A., Igamberdiev, A.U. and Zemlyanukhin, A.A. (1984) Biochemistry (USSR) **49**, 322-328