Pyruvate Carboxylase : A Molecular Biological Study

A thesis submitted

by

ALAN IAN CASSADY, B.Sc.(Hons.), M.Sc.

To the University of Adelaide,

South Australia

for the degree of

Doctor of Philosophy.

Department of Biochemistry,

University of Adelaide,

South Australia.

December, 1987

FOR JENNY

Contents

Statement i

Acknowledgements ii

Abbreviations iii

Summary iv

1 Introduction 1

1.1 Biological and Catalytic Functions of Pyruvate Carboxylase 2

1.1.1 The Role of Pyruvate Carboxylase *In Vivo* 2

1.1.2 The Catalytic Activity of Pyruvate Carboxylase 3

1.1.3 Allosteric Activation of Pyruvate Carboxylase 4

1.2 Structural Properties of Pyruvate Carboxylase 5

1.2.1 Pyruvate Carboxylase Subunits 5

1.2.2 Structural Studies of Pyruvate Carboxylase by Electron Microscopy 6

1.2.3 Effects of Acetyl-CoA Binding on Pyruvate Carboxylase Structure 8

1.3 Structural Relationship of Pyruvate Carboxylase to Other Biotin Carboxylases 9

1.3.1 Subunit Structure of the Biotin Carboxylases 9

1.3.2 Evolutionary Relationship Between the Biotin Carboxylases . 11

1.4 Sequence Analyses of Pyruvate Carboxylase and Other Biotin Carboxylases 12

1.5 Recombinant DNA Technology 14
1.5.1 Approaches to Recombinant DNA Clone Isolation 14
1.6 Aims of the Project 16

2 Materials and Methods **17**

2.1 Materials 18
2.1.1 Antibiotics 18
2.1.2 Bacterial Strains 18
2.1.3 Cloning Vectors 19
2.1.4 Commonly Used Solutions 19
2.1.5 Enzymes and Proteins 19
2.1.6 General Chemicals and Materials 21
2.1.7 Nucleotide Triphosphates 22
2.1.8 Radiochemicals 22
2.1.9 Synthetic Oligonucleotides 22
2.1.10 Materials Used with Specific Methods 23
2.1.10.a Bacterial Culture 23
2.1.10.b Electrophoresis 23
2.1.10.c Fluorography and Scintillation Counting 24
2.1.10.d Nucleic Acid Purification 24
2.1.10.e Tissue Culture 25
2.1.10.f Cell Extracts 25
2.2 Methods 25
2.2.1 Cellular RNA Preparation 26
2.2.2 Affinity Chromatography of RNA on Oligo d(T) Cellulose . . 27
2.2.3 Sucrose Density Gradient Centrifugation of RNA 27
2.2.4 Ethanol Precipitation of Nucleic Acids 28
2.2.5 Phenol and Phenol/Chloroform Extraction of Nucleic Acids . 28
2.2.6 Restriction Endonuclease Digestion of DNA 28
2.2.7 Polyacrylamide Gel Electrophoresis of DNA 29
2.2.7.a Non-Denaturing Gels 29

2.2.7.b Denaturing Gels 29
2.2.8 Isolation of DNA from Polyacrylamide Gels 29
2.2.9 Agarose Gel Electrophoresis 30
2.2.9.a DNA Samples 30
2.2.9.b RNA Samples 30
2.2.10 Isolation of DNA from Agarose Gels 30
2.2.11 Transfer of Nucleic Acids to Nitrocellulose from Agarose Gels 31
2.2.11.a Southern Transfer of DNA to Nitrocellulose 31
2.2.11.b Northern Transfer of RNA to Nitrocellulose 31
2.2.12 SDS-Polyacrylamide Gel Electrophoresis of Proteins 31
2.2.13 Bacterial Culture Media 32
2.2.13.a Culture Media for *E. coli* Strains MC1061 and ED8799 . . . 32
2.2.13.b Culture Media for *E. coli* Strain LE392 and C600Hfl$^-$ 33
2.2.13.c Culture Media for *E. coli* Strain JM101 33
2.2.14 DNA Preparation 34
2.2.14.a Preparation of Bacteriophage M13 Single-Stranded DNA . . 34
2.2.14.b Preparation of Bacteriophage M13 RF-DNA 34
2.2.14.c Preparation of Plasmid DNA 35
2.2.14.d Small-Scale Purification of Plasmid and M13-RF DNA by CsCl Gradient Ultracentifugation 35
2.2.14.e Large-Scale Purification of Plasmid and M13-RF DNA by CsCl Gradient Ultracentifugation 36
2.2.14.f Preparation of Bacteriophage λ DNA 37
2.2.15 Preparation of Cloning Vectors 37
2.2.16 DNA Ligation 38
2.2.17 Bacterial Transformation 38
2.2.17.a Preparation of MC1061 and ED8799 Competent Cells 38
2.2.17.b Preparation of JM101 Competent Cells 39
2.2.17.c Transformations of JM101 with M13 DNA 39
2.2.17.d Transformations of MC1061 and ED8799 with Plasmid DNA 39
2.2.18 [^{32}P] Labelling of DNA 40

2.2.18.a 5′-[^{32}P] Labelling of DNA 40
2.2.18.b [^{32}P] End Labelling of DNA 41
2.2.18.c [^{32}P] Labelling of DNA by Nick-Translation 41
2.2.19 Screening of Libraries in λ Vectors 42
2.2.20 Grunstein Screening of Plasmid Clones 42
2.2.21 Hybridization Conditions 43
2.2.21.a Oligonucleotide Probes 43
2.2.21.b Other DNA Probes 43
2.2.22 Dideoxy Chain-Termination Sequencing 44
2.2.23 Sequencing Gel Electrophoresis 45
2.2.24 Computer Programs 45
2.2.25 Containment Facilities 47

3 Chicken Liver Pyruvate Carboxylase : Development of cDNA Probes and Protein Sequencing 48

3.1 Introduction 49
3.2 Specific Methods 50
3.2.1 *In Ovo* Pulse Labelling of Chicken Embryos 50
3.2.2 Preparation of Mitochondria from Embryonic Chicken Liver . 51
3.2.3 *In Vitro* Translation 51
3.2.4 Immunoprecipitation of *In Vitro* Translation Products 52
3.2.5 [^{14}C] Labelling of Marker Proteins by S-Carboxymethylation 53
3.2.6 Fluorography 53
3.2.7 RNA Dot Blot 53
3.2.8 Primer Extension cDNA Synthesis 53
3.2.8.a Primer Extensions with an [α-^{32}P] Labelled Deoxynucleotide Triphosphate 54
3.2.8.b Primer Extension with 5′ [^{32}P] Labelled B-14 Primer 54
3.2.9 Chemical Cleavage Sequencing of DNA 55
3.2.10 Proteolytic Digestion of Pyruvate Carboxylase 57
3.2.11 HPLC Separation of Proteolytic Peptides 57

3.2.12 Automated Protein Sequencing 57
3.3 Results 58
3.3.1 Pulse Labelling of Proteins in Chicken Embryos 58
3.3.2 Preparation of Embryonic Chicken Liver RNA 59
3.3.3 Size-Fractionation of Polyadenylated RNA 60
3.3.4 *In Vitro* Translation of Embryonic Chicken Liver RNA 60
3.3.5 Design of B-14 Biotin-Site Oligonucleotide 61
3.3.6 RNA Dot Blots Probed with B-14 Oligonucleotide 62
3.3.7 Northern Transfer of Chicken Liver RNA Probed with B-14 Oligonucleotide 63
3.3.8 B-14 Primer Extension Studies 63
3.3.8.a Dependence of cDNA Synthesis on Primer Concentration . . 64
3.3.8.b Optimization of Primer Extensions for Long cDNA Products 65
3.3.8.c 5′ [^{32}P] Labelled B-14 Primer Extension 66
3.3.8.d Primer Extensions using Size-Fractionated RNA 66
3.3.9 Chemical-Cleavage Sequencing of cDNA Products 67
3.3.10 Isolation and Sequencing of Peptides Fragments of Pyruvate Carboxylase 68
3.4 Discussion 70
3.4.1 Pyruvate Carboxylase Synthesis in Embryonic Chicken Liver and Isolation of Pyruvate Carboxylase mRNA 70
3.4.1.a Pulse Labelling Studies 70
3.4.1.b *In Vitro* Translation 72
3.4.2 B-14 Primer Extensions with Embryonic Chicken Liver RNA 73
3.4.3 Protein Sequencing of Pyruvate Carboxylase 75

4 Isolation and Identification of Extended Primer Selected cDNA and Genomic Clones **77**
4.1 Introduction 78
4.2 Specific Methods 79
4.2.1 Dot Benton Procedure for Bacteriophage Screening 79

4.3 Results 79
4.3.1 Preparation of B-14 Primer Extended cDNA Probes 79
4.3.2 Screening of Chicken Genomic Library with Extended Primer Probes 79
4.3.3 Dot Benton of Chicken Genomic Clones Screened with Different cDNA Probes 81
4.3.4 Restriction Mapping of Chicken Genomic Clone λPC3g . . . 82
4.3.5 Localization of Site of Hybridization of Extended Primer Probes to λPC3g 82
4.3.6 Screening of Embryonic Chicken Liver cDNA Library with Extended Primer Probes and Clone Analysis 84
4.3.7 Relationship Between λPC3g and Extended Primer Selected cDNA Clones 85
4.3.8 Northern Transfer of Chicken Liver RNA Probed with cDNA Clone pP99 85
4.3.9 Sequence Analysis of pP2 and Identification as Chicken Serum Albumin 86
4.3.10 Chicken Serum Albumin Structure and Sequence Homology with Other Species 88
4.4 Discussion 89
4.4.1 cDNA Synthesis from Chicken Serum Albumin mRNA 89
4.4.2 Non Pyruvate Carboxylase-Specific cDNA Priming by B-14 Oligonucleotide 91
4.4.3 An Alternative Approach to Pyruvate Carboxylase Clone Isolation 92

5 Isolation and Characterization of Pyruvate Carboxylase cDNA Clones **93**
5.1 Introduction 94
5.2 Specific Methods 96
5.2.1 cDNA Library Construction 96

5.2.1.a First Strand cDNA Synthesis 96
5.2.1.b Second Strand cDNA Synthesis 96
5.2.1.c End Filling 97
5.2.1.d Methylation of *Eco*RI Sites 97
5.2.1.e Cloning of cDNA 97
5.2.2 Amplification of λgt10 cDNA Libraries 98
5.2.3 Tissue Culture 98
5.2.4 RNA Dot Blots 99
5.3 Results 99
5.3.1 Sequence Analysis of Human Liver Pyruvate Carboxylase cDNA Clone 99
5.3.2 Northern Transfer of Chicken Liver RNA Probed with PC34.1 Insert 100
5.3.3 Choice of Probes to Detect Chicken Pyruvate Carboxylase Clones 102
5.3.4 Northern Transfer of Chicken Liver RNA Probed Human Propionyl-CoA Carboxylase α Subunit cDNA Clone Insert 102
5.3.5 cDNA Library Construction and Library Screening with Human Liver Pyruvate Carboxylase cDNA Clone 104
5.3.5.a Screening Plasmid cDNA Library with PC34.1 104
5.3.5.b Embryonic Chicken Liver cDNA Library Construction and Characterization 104
5.3.5.c Screening λgt10 cDNA Library with PC34.1 105
5.3.6 Screening of Chicken Genomic Libraries with the Human Pyruvate Carboxylase cDNA Clone PC34.1 106
5.3.7 Rationale for Changing Species Under Investigation 107
5.3.8 Screening of a Rat Liver cDNA Library with the Human Pyruvate Carboxylase cDNA Clone 108
5.3.9 Induction of Pyruvate Carboxylase mRNA Synthesis in 3T3-L1 Cells Demonstrated with λRL1.1 Insert Probe 108

5.3.10 Northern Transfer of Rat BRL-3A Cell RNA Probed with λRL1.1 Insert 109
5.3.11 Sequence Analysis of the λRL1.1 cDNA Clone Insert 110
5.3.12 Isolation of Longer Rat Liver Pyruvate Carboxylase cDNA Clones 111
5.3.13 Sequence Analysis of the λRL2.35 Pyruvate Carboxylase cDNA Clone 112
5.3.14 DNA Sequence Homology Between Pyruvate Carboxylase Clones from Different Species 113
5.3.15 Protein Sequence Homology Between Pyruvate Carboxylase from Different Species 113
5.3.16 Structure Prediction of Inferred Rat Liver Pyruvate Carboxylase Sequence 115
5.4 Discussion 115
5.4.1 Search for Chicken Pyruvate Carboxylase Clones 115
5.4.2 Analysis of mRNA Species in Chicken and Rat Liver 118
5.4.3 Rat Liver Pyruvate Carboxylase cDNA Clones 121

6 Isolation and 3′ End Structure of the Rat Pyruvate Carboxylase Gene 124

6.1 Introduction 125
6.2 Specific Methods 127
6.2.1 Dale Deletion Sequencing System 127
6.3 Results 127
6.3.1 Isolation of Rat Genomic Clones 127
6.3.2 Restriction Mapping of Rat Genomic Clones 128
6.3.3 Localization of Coding Region in Rat Genomic Clones 128
6.3.4 Northern Transfer of BRL-3A Cell RNA Probed with 1.4 kb *Bam*HI/*Hin*dIII Genomic Fragment 130
6.3.5 Rat Genomic Southern Probed with 1.4 kb *Bam*HI/*Hin*dIII Genomic Fragment 130

6.3.6 Sequencing of λRG1.2 1.4 kb *Bam*HI/*Hin*dIII Fragment . . . 132
6.3.7 Sequence Homology Between Rat Liver Pyruvate Carboxylase and Other Biotin Carboxylases 133
6.3.8 Intron/Exon Boundaries and Biotin Carboxylase Sequence Homology 134
6.3.9 Structural Homology Between Biotinyl Domains 135
6.4 Discussion 136
6.4.1 Rat Pyruvate Carboxylase Gene Structure 136
6.4.2 Biotin Carboxylase Biotin Domain 136
6.4.3 Relationship Between Functional Protein Regions and Gene Structure 138

7 Conclusions **142**
7.1 The Search for Chicken Liver Pyruvate Carboxylase Clones 143
7.2 The Isolation of Chicken Serum Albumin Clones 145
7.3 Rat Liver Pyruvate Carboxylase cDNA Clones and Future Experiments 145
7.4 Rat Pyruvate Carboxylase Gene Structure and Future Experiments . 146

Bibliography **i**

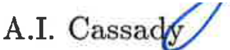
Appendix A

Appendix B

Appendix C

Statement

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief, this thesis contains no material that has been previously published or written by another person except where due reference is made in the text.

A.I. Cassady 

Acknowledgements

I wish to thank Professor W.H. Elliott for permission to undertake this project in the Department of Biochemistry, University of Adelaide.

I would like to thank my supervisor, Dr. J.C. Wallace for his advice, enthusiasm and encouragement during the course of this study. I am also very grateful to Dr. Phil Morris for many valuable discussions. I would like to thank past and present members of Labs. 109, 110 and 112 for their help and advice especially Fil Lim, Grant Booker, Kathy Salkilld, Karen Oliver, Chris Lesnikowski and David Mottershead. I would also like to thank Dr. A. Sivaprasad for sharing some of his computational skills with me and Dr. Gopesh Srivastava for his helpful suggestions.

I would like to thank and acknowledge Mark Snoswell and Joan Rowe who performed the peptide sequencing for me and Steven Rogers and Derek Skingle who synthesized oligonucleotides. I am also grateful to Yvonne Riese and Jenny Brazier for their expert technical assistance.

I am especially grateful to Jenny Cassady for her encouragement, expert preparation of the diagrams and valuable advice in the production of this thesis. Finally, the work of Debbie Moffat in typing the manuscript is appreciated.

Abbreviations

A_λ	....	absorbance at wavelength λ
BSA	....	bovine serum albumin
-CoA	....	acyl derivative of coenzyme A
DTT	....	dithiothreitol
EDTA	....	ethylene diaminetetraacetic acid
HPLC	....	high performance liquid chromatography
NTP	....	any nucleotide triphosphate
SDS	....	sodium dodecyl sulphate
Tris	....	tris(hydroxymethyl)aminomethane
UV	....	ultraviolet light

Amino acids:

A	....	Ala	....	Alanine	M	....	Met	....	Methionine
B	....	Asx	....	Asp or Asn	N	....	Asn	....	Asparagine
C	....	Cys	....	Cysteine	P	....	Pro	....	Proline
D	....	Asp	....	Aspartic acid	Q	....	Gln	....	Glutamine
E	....	Glu	....	Glutamic acid	R	....	Arg	....	Arginine
F	....	Phe	....	Phenylalanine	S	....	Ser	....	Serine
G	....	Gly	....	Glycine	T	....	Thr	....	Threonine
H	....	His	....	Histidine	V	....	Val	....	Valine
I	....	Ile	....	Isoleucine	W	....	Trp	....	Tryptophan
K	....	Lys	....	Lysine	Y	....	Tyr	....	Tyrosine
L	....	Leu	....	Leucine	Z	....	Glx	....	Glu or Gln

Summary

Pyruvate carboxylase [pyruvate:CO_2 ligase (ADP); EC 6.4.1.1] is a biotin containing enzyme which is allosterically activated. Pyruvate carboxylase catalyses the first step in the gluconeogenic pathway in the vertebrate liver and kidney and also fulfils an anaplerotic role in the tricarboxylic acid cycle. This key enzyme has been extensively studied with respect to its physical and catalytic properties not only because of its intrinsic metabolic importance but also as a model of enzymic activity and its regulation. Analysis of the relationship between the enzyme's structural and functional properties has been limited by a lack of sequence data for pyruvate carboxylase. The research reported in this thesis was directed towards the acquisition of sequence data for pyruvate carboxylase using molecular biological techniques.

Chicken liver was chosen initially as the source of pyruvate carboxylase and 17–18 day old embryonic chicken livers were found to be synthesizing significant amounts of the enzyme. RNA preparations from this source were shown by *in vitro* translation to contain pyruvate carboxylase mRNA. As a result this tissue was used as the source of pyruvate carboxylase mRNA. A synthetic oligonucleotide primer was synthesized on the basis of the known pyruvate carboxylase biotin attachment site protein sequence. The cDNA synthesized from mRNA preparations using the biotin site primer contained three specific primer extension products of 75, 220 and 320 bases. The 220 and 320 base products were shown to be synthesized from an RNA size class expected to contain the pyruvate carboxylase mRNA. The extended primer cDNAs were used as pyruvate carboxylase specific probes to libraries of chicken recombinant DNA to isolate pyruvate carboxylase clones. To aid in the identification of putative pyruvate carboxylase clones the protein sequence of eleven proteolytic peptide fragments of chicken liver pyruvate carboxylase were determined. This resulted in a total of 277 amino acids of pyruvate carboxylase sequence representing approximately 24% of the complete protein sequence.

A chicken genomic library and an embryonic chicken liver cDNA library were screened with the extended primer cDNA probes and the clones isolated were es-

tablished by hybridization to be related. Sequencing of the cDNA clone pP2, identified the clones as encoding chicken serum albumin. The complete sequence of the chicken serum pre-proalbumin coding region was determined and the 611 amino acids of the protein were inferred. An examination of the chicken pre-proalbumin cDNA sequence revealed homology to the biotin attachment site primer and provided the probable reason for the selection of the more abundant albumin clones in preference to pyruvate carboxylase clones.

The identification of a cDNA clone encoding human liver pyruvate carboxylase by other workers enabled the human pyruvate carboxylase clone PC34.1 to be used as a probe for chicken liver pyruvate carboxylase clones. PC34.1 hybridized to a 4.5 kb chicken mRNA species representing pyruvate carboxylase and was also observed to hybridize weakly to 3.3 kb and 2.2 kb mRNA species. The human propionyl-CoA carboxylase α subunit cDNA clone pPCC9-5 was found to hybridize strongly to the 3.3 kb and 2.2 kb bands. This suggested that the chicken propionyl-CoA carboxylase α and possibly β subunit mRNAs were detected by the PC34.1 probe.

Three chicken liver cDNA libraries were prepared in the λgt10 vector and screened with PC34.1. This probe failed to select any pyruvate carboxylase clones from any of the cDNA libraries or from two chicken genomic libraries. This appeared to result from the failure of these libraries to contain pyruvate carboxylase clones. In view of the inability to isolate chicken pyruvate carboxylase clones the species under investigation was changed to rat owing to the availability of rat genomic and rat liver cDNA libraries.

A rat liver cDNA library in λgt10 was screened with the PC34.1 probe and the λRL1.1 clone was isolated. The 939 bp λRL1.1 insert hybridized to a 4.2 kb mRNA species from a rat liver cell line (BRL-3A) and a 4.5 kb chicken liver mRNA species on Northern transfers, consistent with it encoding pyruvate carboxylase. λRL1.1 was identified as encoding pyruvate carboxylase by sequence analysis and included the conserved biotin attachment site sequence. This clone was used as a probe to re-screen the rat liver cDNA library and the longer λRL2.35 cDNA clone was isolated and sequenced. The total cDNA sequence for rat liver pyruvate carboxylase was 1338 bp long and contained 918 bp of coding sequence. The 306 amino acids

of predicted rat pyruvate carboxylase protein sequence showed strong homology to pyruvate carboxylase from human liver and yeast as well as a peptide from chicken liver pyruvate carboxylase. Significant sequence homology was apparent to other biotin carboxylases.

The λRL1.1 clone was used to isolate genomic clones which contain the 3′ end of the rat pyruvate carboxylase gene. These genomic clones were restriction mapped and a 1476 bp *Bam*HI/*Hin*dIII fragment containing the biotin attachment site was sequenced. Two exons were found in the sequence of this fragment. The most 3′ exon is 430 bp in length including the 3′ untranslated region and the coding region contains the biotin attachment site sequence. The more 5′ exon is 111 bp in length and is separated from the 3′ exon by a 76 bp intron. The 82 amino acids encoded by the 3′ exon are proposed to define a putative polypeptide domain which is homologous to the 82 amino acid biotin carboxyl carboxyl carrier subunit domain of acetyl-CoA carboxylase [acetyl-CoA:CO_2 ligase (ADP); EC 6.4.1.2] from *Escherichia coli.*

Chapter 1

Introduction

1.1 Biological and Catalytic Functions of Pyruvate Carboxylase

The association of biotin with enzyme catalysed carboxylation reactions was first made by Lynen *et al.* (1959) with the enzyme 3-methylcrotonyl-CoA carboxylase from *Achromobacter IVS*. The enzymes which have been found to contain biotin as a covalently attached prosthetic group have been generically termed biotin carboxylases. The members of this group of enzymes catalyse a variety of carboxylation, decarboxylation and transcarboxylation reactions and are found in both prokaryotic and eukaryotic organisms. A total of nine different biotin carboxylases have been identified and these enzymes are listed with the reactions they catalyse in Table 1.1. Several biotin enzymes have been studied in greater detail than the others owing to the importance of their role *in vivo* and because of the interesting properties which they have exhibited *in vitro*. The enzyme which has been studied extensively in this group over approximately 20 years and which will form the basis of this investigation is pyruvate carboxylase (pyruvate:CO_2 ligase (ADP); EC 6.4.1.1).

1.1.1 The Role of Pyruvate Carboxylase *In Vivo*

The synthesis of carbohydrates from tricarboxylic acid cycle and glycolytic intermediates forms the basis of gluconeogenesis. However, gluconeogenesis through the simple reversal of glycolysis is not possible because the energy requirements of the pyruvate to phospho*enol*pyruvate conversion render this reaction highly unfavourable. Pyruvate carboxylase and phospho*enol*pyruvate carboxykinase [EC 4.1.1.49] provide the means to achieve this reversal in an energetically favourable two stage reaction by catalysing the carboxylation of pyruvate to form oxaloacetate which is then converted to phospho*enol*pyruvate. Pyruvate carboxylase is therefore crucial to gluconeogenesis. In vertebrates the principal sites of gluconeogenesis are in the liver and kidney where not surprisingly the highest levels of pyruvate carboxylase activity have been found (for a review, see Wallace, 1985). Pyruvate carboxylase is also present in non-gluconeogenic tissues where its role is

Table 1.1

The Reactions Catalysed by Biotin Carboxylases Other than Pyruvate Carboxylase

The reactions catalysed by the biotin carboxylases other than pyruvate carboxylase are shown (Wood and Barden, 1977). The requirement for a divalent metal ion is indicated by M^{2+} and citrate is an allosteric activator of vertebrate acetyl-CoA carboxylase (Lane *et al.*, 1974).

Carboxylation

1. acetyl-CoA carboxylase [EC 6.4.1.2]

$$\text{acetyl-CoA} + HCO_3^- + \text{ATP} \overset{\text{citrate, }M^{2+}}{\rightleftharpoons} \text{malonyl-CoA} + \text{ADP} + P_i$$

2. geranoyl-CoA carboxylase [EC 6.4.1.5]

$$\text{geranoyl-CoA} + HCO_3^- + \text{ATP} \overset{M^{2+}}{\rightleftharpoons} \text{3-isohexenylglutaconyl-CoA} + \text{ADP} + P_i$$

3. 3-methylcrotonyl-CoA carboxylase [EC 6.4.1.4]

$$\text{3-methylcrotonyl-CoA} + HCO_3^- + \text{ATP} \overset{M^{2+}}{\rightleftharpoons} \text{3-methylglutaconyl-CoA} + \text{ADP} + P_i$$

4. propionyl-CoA carboxylase [EC 6.4.1.3]

$$\text{propionyl-CoA} + HCO_3^- + \text{ATP} \overset{M^{2+}}{\rightleftharpoons} \text{methylmalonyl-CoA} + \text{ADP} + P_i$$

5. urea carboxylase [EC 6.3.4.6]

$$\text{urea} + HCO_3^- + \text{ATP} \overset{M^{2+}}{\rightleftharpoons} \text{allophanate} + \text{ADP} + P_i$$

Decarboxylation

1. methylmalonyl-CoA decarboxylase [EC 4.1.1.41]

$$\text{methylmalonyl-CoA} + H_2O \rightleftharpoons \text{propionyl-CoA} + HCO_3^-$$

2. oxaloacetate decarboxylase [EC 4.1.1.3]

$$\text{oxaloacetate} + H_2O \rightleftharpoons \text{pyruvate} + HCO_3^-$$

Transcarboxylation

1. transcarboxylase [EC 2.1.3.1]

$$\text{methylmalonyl-CoA} + \text{pyruvate} \rightleftharpoons \text{propionyl-CoA} + \text{oxaloacetate}$$

anaplerotic, regenerating tricarboxylic acid cycle intermediates for the biosynthesis of fatty acids, cholesterol, acetyl-choline, porphyrin and several amino acids (Kornberg, 1966). The enzyme has an intra-mitochondrial location in eukaryotic species (Keech and Utter, 1963; reviewed by Wallace, 1985) with the exception of the fungal species *Saccharomyces cerevisiae* (Haarasilta and Taskinen, 1977) and *Aspergillus nidulans* (Osmani and Scrutton, 1983) where it has been shown to be cytosolic.

The crucial nature of the reaction catalysed by pyruvate carboxylase is revealed in human individuals where a mutation of the pyruvate carboxylase gene resulting in significantly reduced pyruvate carboxylase activity has been identified (Robinson *et al.*, 1984). Most affected individuals die as infants owing to the severe effects of reduced liver and brain pyruvate carboxylase activity on the central nervous system. Those individuals who have been characterized as pyruvate carboxylase deficient and yet survive are severely mentally retarded (De Vivo *et al.*, 1977; Oizumi *et al.*, 1983).

1.1.2 The Catalytic Activity of Pyruvate Carboxylase

Pyruvate carboxylase catalyses the ATP dependent carboxylation of pyruvate to form oxaloacetate as shown in Eqn. 1.1.

$$\text{pyruvate} + HCO_3^- + \text{ATP} \quad \overset{\text{acetyl-CoA, } Mg^{2+}}{\rightleftharpoons} \quad \text{oxaloacetate} + \text{ADP} + \text{Pi}$$

Eqn. 1.1

The overall reaction has a requirement for Mg^{2+} ions and is activated by acetyl-CoA. Kinetic studies of substrate and product exchange reactions (Scrutton *et al.*, 1965; McClure *et al.*, 1971; Barden *et al.*, 1972) have demonstrated that this overall reaction is in fact the sum of two partial reactions which are shown in Eqns. 1.2 and 1.3.

$$\text{Enzyme-biotin} + HCO_3^- + MgATP^{2-} \overset{\text{acetyl-CoA, } Mg^{2+}}{\rightleftharpoons} \text{Enzyme-biotin-}CO_2^- + MgADP + P_i$$

Eqn. 1.2

$$\text{Enzyme-biotin-}CO_2^- + \text{pyruvate} \rightleftharpoons \text{Enzyme-biotin} + \text{oxaloacetate}$$

Eqn. 1.3

In the first partial reaction (Eqn. 1.2) the enzyme's covalently attached biotin is carboxylated with the carboxyl group supplied by HCO_3^-. This reaction is coupled to the cleavage of $MgATP^{2-}$ to form MgADP and P_i. The reaction has a requirement for Mg^{2+} ions not only to enable formation of $MgATP^{2-}$ which is the catalytically active form (Utter and Keech, 1960) but also as free Mg^{2+} ions which appear to activate the enzyme (Keech and Barritt, 1967; Bais and Keech, 1972).

The second partial reaction (Eqn. 1.3) transfers the coupled carboxyl group from carboxybiotin to pyruvate resulting in the formation of oxaloacetate. This partial reaction regenerates non-carboxylated enzyme-biotin which may then participate in subsequent cyles of catalysis. Therefore the link between the reactions is provided by the biotinyl group has been proposed to act in a "swinging arm" fashion in moving between the first and second subsites (Barden *et al.*, 1975) in a somewhat analogous manner to the translocation of acetyl groups by lipoic acid in the pyruvate dehydrogenase complex (Ambrose and Perham, 1976). The arrangement of biotin and the subsites proposed by this model is shown in Fig. 1.1.

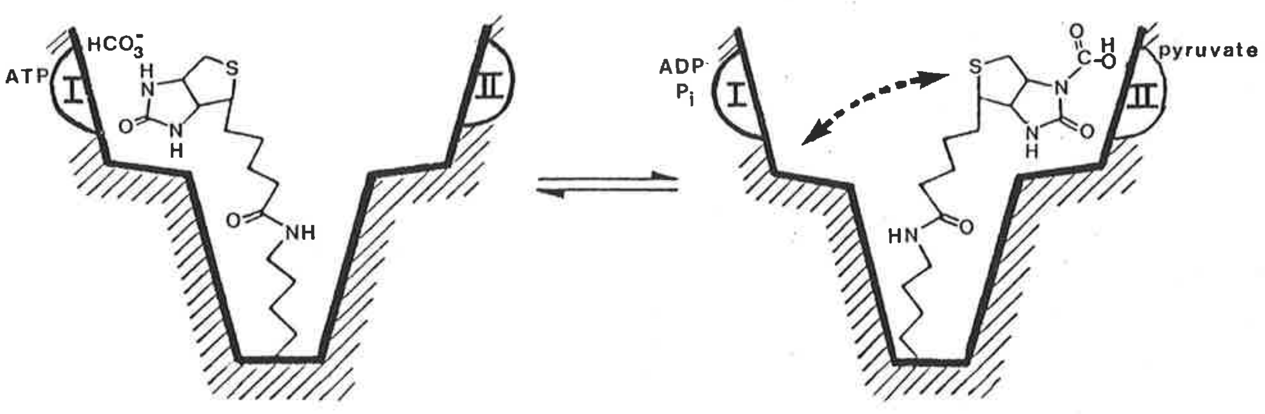
1.1.3 Allosteric Activation of Pyruvate Carboxylase

The activity of pyruvate carboxylase is modulated principally by the allosteric effector acetyl-CoA although other physiologically important acyl derivatives of CoA including propionyl-CoA and *n*-butyryl-CoA have been found to stimulate activity (Utter and Keech, 1963; for a review, see Barritt, 1985). The extent of this effect varies considerably between pyruvate carboxylase isolated from different organisms. Pyruvate carboxylase enzymes from the prokaryotes, *Pseudomonas citronellolis* (Seubert and Remberger, 1961) and *Azotobacter vinlandii* (Scrutton, 1974)

Figure 1.1

A Schematic Representation of the Active Site of Pyruvate Carboxylase

The two spatially separated catalytic subsites of pyruvate carboxylase are represented by (I) - biotin carboxylation from HCO_3^- with the associated cleavage of ATP and (II) - carboxyl transfer to pyruvate. The proposed translocation of the carboxyl group by the biotin prosthetic group between these two sites is indicated (Lynen, 1979).



ATP
HCO_3^-
I
II
S
NH
ADP
P_i
HN
pyruvate

and the eukaryotic filamentous fungus *Aspergillus niger* (Feir and Suzuki, 1969) are unaffected by acetyl-CoA. The pyruvate carboxylase enzymes from yeast (*Saccharomyces cerevisiae*) (Cazzulo and Stoppani, 1967), sheep liver (Ashman *et al.*, 1972) and rat liver (McClure *et al.*, 1971) show a small amount of acetyl-CoA-independent activity but their activity is significantly stimulated by acetyl-CoA. In contrast the enzymes from *Bacillus stearothermophilus* (Cazzulo *et al.*, 1970), *Arthrobacter globiformis* (Bridgeland and Jones, 1967) and chicken liver (Utter and Keech, 1963) show no acetyl-CoA-independent activity.

The effect of acetyl-CoA binding on pyruvate carboxylase activity has been examined at the level of the individual reaction steps by measuring the rates of exchange between isotopically labelled substrates. In the chicken and sheep enzymes the isotopic exchange between ATP and P_i was found to be dependent on the presence of acetyl-CoA (Scrutton *et al.*, 1965; Ashman *et al.*, 1972) and the isotopic exchange between pyruvate and oxaloacetate in the sheep kidney enzyme was significantly stimulated by the presence of acetyl-CoA (Ashman *et al.*, 1973). In a re-examination of the acetyl-CoA activation of reactions at both subsites Khew-Goodall (1985) found that the enhancement of activity by acetyl-CoA of the second partial reaction was an enzyme concentration effect which was abolished at high enzyme concentration. The effect of acetyl-CoA on activity therefore appears to be restricted to the first subsite. Keech (1980) suggested that that acetyl-CoA might bind across the gap separating the subsites to form a "cap" for the active site. This could be expected to alter not only the environment within the active site but also to have an effect on the structure of the subunit.

1.2 Structural Properties of Pyruvate Carboxylase

1.2.1 Pyruvate Carboxylase Subunits

The tertiary structure of pyruvate carboxylase has been examined in many species and the enzyme has almost exclusively been found to be present as a tetrameric

protein with identical subunits in the range of M_r = 100,000–150,000. This examination includes six bacterial species, two fungal species and seven vertebrate species (for a review see Wallace and Easterbrook-Smith, 1985). In most cases each subunit contains a single biotin prosthetic group and in the species which have been examined, one pyruvate binding site (Hudson *et al.*, 1975), one acetyl-CoA binding site (Ashman *et al.*, 1973; Scrutton and White, 1973; Scrutton *et al.*, 1977; Frey and Utter, 1977), one tightly bound Mn^{2+} ion (Scrutton *et al.*, 1966) or Zn^{2+} in yeast pyruvate carboxylase (Scrutton *et al.*, 1970) and one ATP binding site (Easterbrook-Smith *et al.*, 1976). An exception to this rule was found with pyruvate carboxylase from *Ps. citronellolis* in which two different types of subunits have been found. The α subunit (M_r = 65,000) contains biotin and the β subunit (M_r = 54,000) which is biotin-free (Barden *et al.*, 1975) and the subunits are arranged in an $\alpha_4\beta_4$ tertiary structure (Cohen *et al.*, 1979).

1.2.2 Structural Studies of Pyruvate Carboxylase by Electron Microscopy

Early electron microscopic studies of the structure of pyruvate carboxylase from chicken, turkey and calf liver had described a square planar tetrameric structure (Valentine *et al.*, 1966; Utter *et al.*, 1975). However, the enzyme from yeast was determined to be tetramer with the four subunits arranged at the corners of a rhombus (Valentine, 1968; Cohen *et al.*, 1979a). Subsequent studies on the enzyme from vertebrate species using highly purified preparations unequivocally demonstrated that in these species the enzyme has as a tetrahedron-like structure (Cohen *et al.*, 1979b; Goss *et al.*, 1979). A re-examination of the structure of yeast pyruvate carboxylase by Rohde *et al.* (1986) established that this enzyme shared the same structure as the vertebrate enzymes and was unlikely to have a rhomboidal structure as had been proposed by Valentine (1968) and Cohen *et al.* (1979b). The pyruvate carboxylase enzymes purified from the fungal species *Aspergillus nidulans* (Osmani *et al.*, 1984) and the bacterium *Ps. citronellolis* (Cohen *et al.*, 1979) shared this tetrahedron-like structure with the yeast and vertebrate enzymes indicating a significant degree of

evolutionary conservation of the quaternary structure (Osmani *et al.*, 1984; Cohen *et al.*, 1979).

The fine structure of the chicken, rat and sheep liver enzymes was studied further by Mayer *et al.* (1980). These workers found the enzyme to have a tetrahedron-like structure consisting of two pairs of subunits in two different planes orthogonal to each other with opposing subunit pairs interacting with each other on their convex surfaces. A clay model which represents this structure is shown in Fig. 1.2.

Mayer *et al.* (1980) also found that the preparation of chicken liver pyruvate carboxylase in the presence of acetyl-CoA resulted in an alteration in the appearance of the enzyme. In the absence of acetyl-CoA the enzyme had a "splayed-tetrameric" structure with a quite obvious cleft in each subunit whereas in the presence of acetyl-CoA the tetrahedron-like structure was more apparent. The cleft was estimated to be 1.2–2.0 nm wide and ran along the longitudinal mid-line for the full length of the subunit.

In an attempt to determine the location of the active site, Johannssen *et al.* (1983) used the protein avidin which binds biotin with high affinity and specificity ($Kd = 1 \times 10^{-15}$ M) (Green, 1963) to localize the biotin on the surface of the enzyme by electron microscopy. Johannssen *et al.* (1983) found the stoichiometry of the avidin tetramer : pyruvate carboxylase tetramer binding to be 1 : 1 and observed the presence of linear un-branched chains of avidin-pyruvate carboxylase complexes. The structural model of the pyruvate carboxylase active site which these workers proposed located the biotin at the mid-line of each subunit within 2 nm of the inter-subunit junction (Fig. 1.3). In this position the back-to-back pyruvate carboxylase dimers could each associate with an avidin tetramer and repetition of this process would result in linear chain formation. This provided evidence for the location of the biotin prosthetic group within the cleft observed by Mayer *et al.* (1980). This position was consistent with the "swinging arm" scheme for the movement of the biotin between spatially distinct subsites. Further evidence for the location of the active site in the cleft has been provided with the identification by electron microscopy of the binding to this region of pyruvate carboxylase of immunoglobulin Fab fragments. These Fab fragments were derived both from polyclonal antibodies

Figure 1.2

A Model of the Quaternary Structure of Chicken and Sheep Liver Pyruvate Carboxylase

The clay models shown here represent the quaternary structure of chicken and sheep liver pyruvate carboxylase interpreted from electron microscopic images of these enzymes by Mayer *et al.* (1980). The images shown in (1), (2), and (4) are side views of the enzyme taken from several angles whereas the images in (3) and (5) are overhead views taken from two angles.

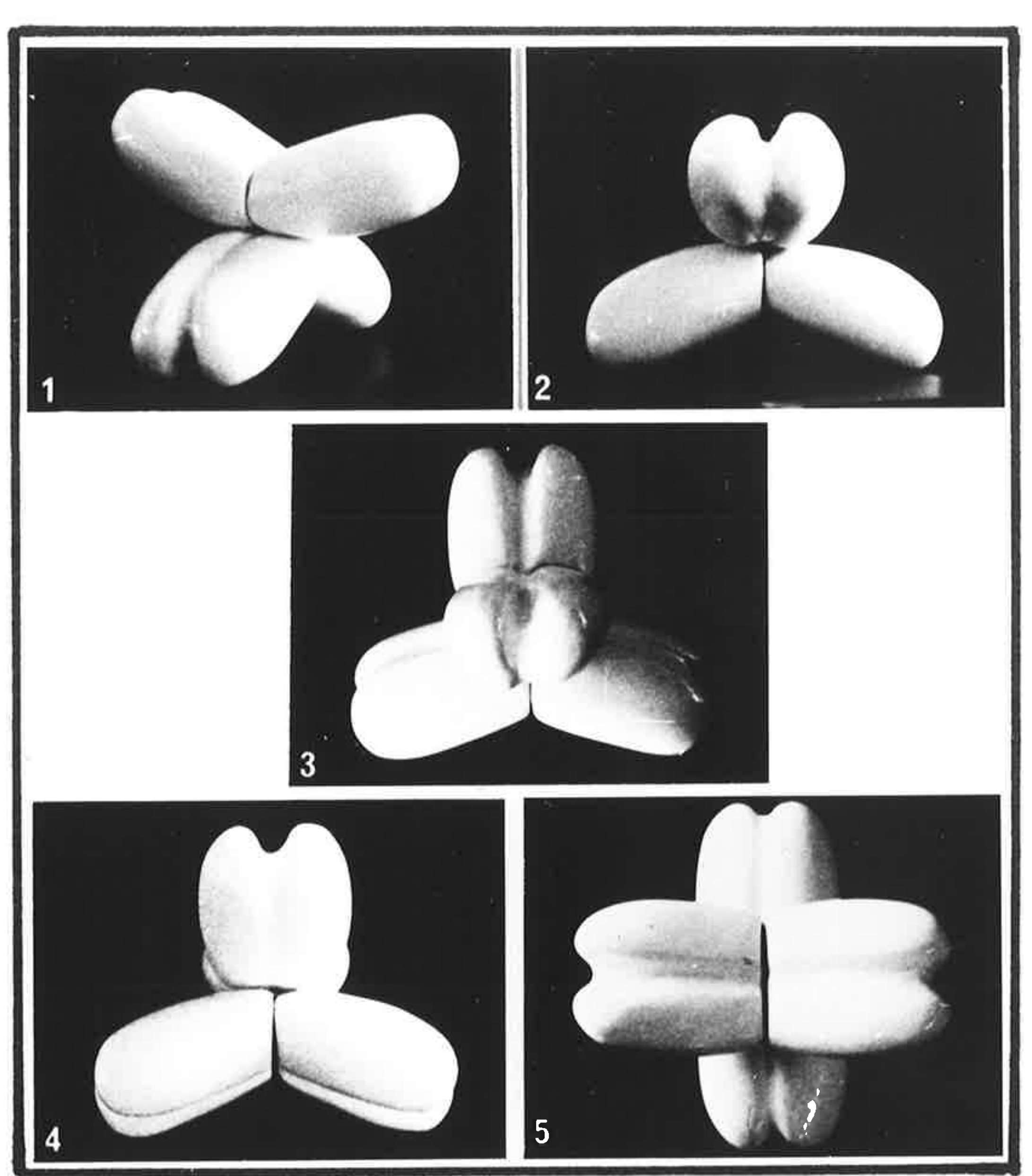
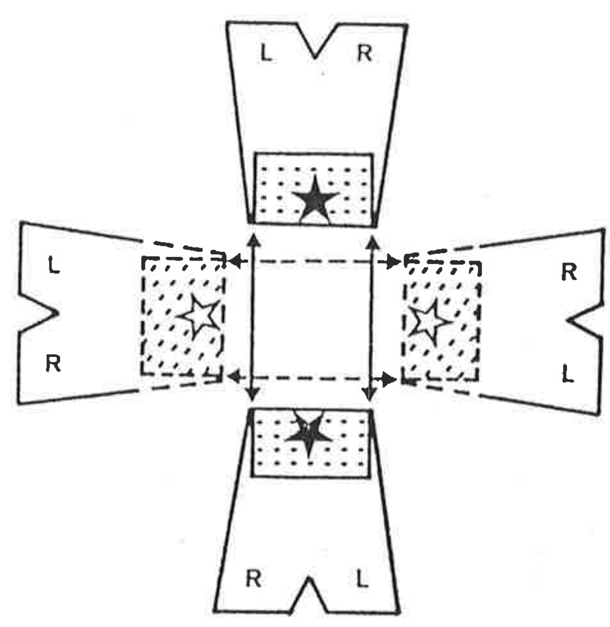

1
2
3
4
5

Figure 1.3

An Exploded Schematic View of the Pyruvate Carboxylase Tetramer

An exploded face-view of the pyruvate carboxylase tetramer with the position of the bound avidin tetramer as determined by Johannssen *et al.* (1983) indicated by shading. The position of the site of avidin binding is marked by (★) on the upper pair of pyruvate carboxylase subunits and by (☆) on the lower pair of subunits.



L
R
L
R
R
L
R
L

and from monoclonal antibodies which were directed against pyruvate carboxylase and had been demonstrated to inhibit pyruvate carboxylase activity (F. Mayer and K.J. Oliver, personal communication).

1.2.3 Effects of Acetyl-CoA Binding on Pyruvate Carboxylase Structure

Acetyl-CoA was shown to bind to pyruvate carboxylase with a stoichiometry of 1 molecule to 1 subunit (Frey and Utter, 1977) and it was demonstrated by Ashman *et al.* (1973) that a single essential lysyl residue was involved with acetyl-CoA binding. The classical theories of allosterism of Monod *et al.* (1965) and Koshland *et al.* (1966) propose a conformational change in the protein molecule in association with the binding of the effector molecule. In pyruvate carboxylase evidence for such conformational changes comes from several lines of investigation. The presence of acetyl-CoA was observed by Scrutton *et al.* (1965) to increase the rate of inactivation of pyruvate carboxylase by avidin and an increase in the sedimentation velocity of tetrameric yeast pyruvate carboxylase in the presence of acetyl-CoA was found by Taylor *et al.* (1978) implying that a conformational change had occurred. In electron microscopic studies, the presence of acetyl-CoA during the sample preparation has been found to result in both better preservation of pyruvate carboxylase structure and a "tighter" conformation of the tetramers (Mayer *et al.*, 1980; Osmani and Scrutton, 1984; Rohde *et al.*, 1986). To account for these observations Attwood *et al.* (1986) proposed that acetyl-CoA alters the equilibrium between "tight" and "loose" conformational states in favour of the "tight" state which makes the active site less accessible to the external environment. However, since homotropically cooperative binding of acetyl-CoA has not been observed (Barritt *et al.*, 1976) the conformational change must be restricted to individual subunits.

It appears likely that the structural changes associated with acetyl-CoA binding to pyruvate carboxylase have a significant effect on the conformation of the active site and thereby on the enzyme's activity. Knowledge of the relationship between enzymic activity, its regulation and protein structure is essential to an understanding

of the way pyruvate carboxylase behaves not only *in vitro* but also *in vivo*.

1.3 Structural Relationship of Pyruvate Carboxylase to Other Biotin Carboxylases

1.3.1 Subunit Structure of the Biotin Carboxylases

In general the catalytically important sites on the biotin carboxylases include an ATP binding site (with the exception of transcarboxylase), a keto acid binding site (pyruvate carboxylase and transcarboxylase) and an acyl-CoA binding site (which has a regulatory function in pyruvate carboxylase). On the basis of the subunit location of these sites and their subunit structure the biotin carboxylases may be divided into three classes.

In the first class of biotin carboxylases all these sites and the catalytic functions associated with them are localized on a single multifunctional subunit. The subunits of these enzymes are characteristically large, M_r = 110,000–125,000 for pyruvate carboxylase and M_r = 220,000–240,000 for acetyl-CoA carboxylase. This is the case with pyruvate carboxylase from all species that have been examined with the exception of pyruvate carboxylase from *Ps. citronellolis* and with acetyl-CoA carboxylase from all sources with the exception of acetyl-CoA carboxylase from *E. coli* (Alberts and Vagelos, 1972; Lane *et al.*, 1974).

In the second class of biotin carboxylases the subsites are located on two subunits of different size as exemplified by propionyl-CoA carboxylase, 3-methylcrotonyl-CoA carboxylase, geranyl-CoA carboxylase and pyruvate carboxylase from *Ps. citronellolis*. In general the α subunit which contains the biotin is of M_r = 64,000–96,000 and the β subunit which is biotin-free is of M_r = 54,000–78,000. In these enzymes the biotin has always been located on the the larger of the two subunits. The tertiary structure of these enzymes is variable with both $\alpha_4\beta_4$ (Kalousek *et al.*, 1980) and $\alpha_6\beta_6$ (Haase *et al.*, 1982) subunit arrangements being found in different species.

In studying the structure of the propionyl-CoA carboxylase from *Mycobacterium smegmatis* Haase *et al.* (1982) established that the native enzyme's $\alpha_6\beta_6$ structure consisted of two sets of $\alpha_3\beta_3$ complexes sitting on top of each other. The β subunits were arranged in a central cylindrical barrel with one α subunit attached to the external face of each β subunit on each end of the cylinder. In propionyl-CoA carboxylase from *M. smegmatis* the isolated biotinyl (α) subunit has been shown to retain the biotin carboxylase activity while the isolated β subunit has the carboxyltransferase activity (Haase *et al.*, 1982). The same activities have been described for the isolated biotinyl and non-biotinyl subunits of 3-methylcrotonyl-CoA carboxylase from *M. smegmatis* (Schiele *et al.*, 1975)

In the third class of biotin carboxylases all the subsites and the catalytic functions associated with them are localized on separate monofunctional subunits. A representative of this class which has been extensively characterized is transcarboxylase from *P. shermanii*. Transcarboxylase is a complex enzyme which is composed of 30 individual subunits of three different types and has a total $M_r \simeq 1.2\times10^6$ (for a review, see Wood and Zwolinski, 1976). The enzyme has a central cylindrical hexamer of subunits termed 12S (M_r = 60,000). Linked to this core, each through a pair of 1.3S biotinyl subunits termed 1.3S (M_r = 12,000), are six subunits termed 5S (M_r = 60,000). The 1.3S subunit has been demonstrated to bind the two types of larger subunits together through interactions in the N-terminal 26 amino acids of its sequence (Kumar *et al.*, 1982). The 1.3S subunit acts as a biotin carrier protein, the 5S subunit has the keto acid (pyruvate) binding site and the 12S subunit contains the acyl-CoA (methylmalonyl-CoA) binding site.

The other member of this class of biotin carboxylases which has three separate functional components is acetyl-CoA carboxylase from *E. coli* (Guchhait *et al.*, 1974). The enzyme is composed of a biotin carboxylase which carboxylates biotin using the reaction shown in Eqn. 1.2 and is composed of two subunits of M_r = 50,000, a carboxyltransferase which transfers the carboxyl group to acetyl-CoA and is composed of two non-identical subunits of M_r = 30,000 and M_r = 35,000 in a complex of M_r = 130,000. Also present is a biotin carrier protein of M_r = 22,500 which is present as a dimer and appears to be associated with the other components in

an active ternary complex (Guchhait *et al.*, 1974). During the process of purification the biotin carrier protein was found to be particularly susceptible to proteolysis and an additional biotin containing fragment of $M_r \simeq 10{,}000$ was routinely prepared (Guchhait *et al.*, 1974; Sutton *et al.*, 1977).

1.3.2 Evolutionary Relationship Between the Biotin Carboxylases

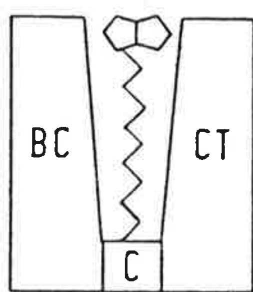
The classification of the biotin carboxylases based on the division of functional sites between subunits was described by Lynen who developed a hypothesis to account for the evolution of these enzymes (Lynen, 1975; Obermayer and Lynen, 1976; Lynen, 1979). In reviewing the structure of the biotin carboxylases Obermayer and Lynen (1976) noted that these enzymes catalyse analogous two-step reactions which utilize biotin as a carboxyl carrier between two spatially distinct subsites. The hypothesis stated that the three classes of enzymes could represent the evolutionary descendants of progessive stages in the evolution of the biotin carboxylase enzyme system. The class of enzymes with multiple monofunctional subunits would represent the precursor form of the enzymes. This form would have given rise to an intermediate form where fusion of the biotin carboxylase subunit to the biotin carrier protein had occurred. Finally, fusion of all the components of the enzyme activity would yield the class of enzymes with single multifunctional subunits, see Fig. 1.4. The hypothesis envisaged this process occurring through the fusion of the genes for the individual monofunctional proteins. Obermayer and Lynen (1976) also proposed a fourth class of enzymes, those having regulatory binding sites, such as pyruvate carboxylase but in this case no independent regulatory subunit was apparent in those enzymes with monofunctional subunits. The possibility of course exists that the function associated with subunits in these enzymes may have altered during evolution. For example, the acetyl-CoA binding site of the *E. coli* acetyl-CoA carboxylase progenitor protein may have altered from having a catalytic to a regulatory function in the evolution of multifunctional pyruvate carboxylase subunits.

The ability to test this hypothesis requires the acquisition of further structural

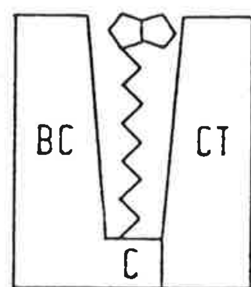
Figure 1.4

A Schematic Model of the Structural Evolution of the Biotin Carboxylases

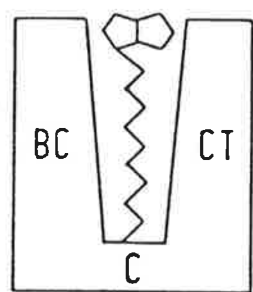
The types of structural organization and evolutionary development of the biotin carboxylases proposed by Lynen are shown (Lynen, 1975; Obermayer and Lynen, 1976; Lynen, 1979). In this model, enzymes with unifunctional subunits progressively gave rise to bifunctional and ultimately multifunctional subunits. BC = biotin carboxylase, C = biotin carrier protein, CT = carboxyl transferase. The biotinyl prosthetic group is shown bound to the biotin carrier protein.



acetyl−CoA carboxylase (E.coli)
transcarboxylase (P.shermanii)



propionyl−CoA carboxylase (vertebrate + M.smegmatis)
3−methylcrotonyl−CoA carboxylase (vertebrate + Achromobacter IVS)
pyruvate carboxylase (Ps.citronellolis)



acetyl−CoA carboxylase (vertebrate)
pyruvate carboxylase (vertebrate + yeast)

information on the biotin carboxylases. Specifically what is needed is the ability to identify the functionally important residues and regions both in the primary sequence and in the three dimensional structure of the enzyme. A comparative analysis of the sequence of the biotin carboxylases would determine if homology existed between different enzymes in the one species as well as identifying the key residues conserved in the same enzyme in different species.

To gauge the relationship between enzymes in the absence of sequence data, Goodall (1981) analysed the amino acid composition of the biotin carboxylases using the $S\Delta n$ index of protein relatedness described by Cornish-Bowden (1980). Strong correlation in composition was observed between the different subunits of individual enzymes such as propionyl-CoA carboxylase and transcarboxylase and also between pyruvate carboxylase from vertebrate sources and vertebrate acetyl-CoA carboxylase and 3-methylcrotonyl-CoA carboxylase from *Achromobacter*. However, using this method of analysis no significant correlation was apparent between the subunits with similar binding sites in different enzymes. To pursue further the problem of the extent of evolutionary relationship and the assignment of functional regions to specific residues, data on the primary sequence of the biotin carboxylases needed to be acquired and analysed.

1.4 Sequence Analyses of Pyruvate Carboxylase and Other Biotin Carboxylases

Analyses of the sequence of pyruvate carboxylase at the time this study commenced were restricted to the 19 amino acid tryptic peptides which were isolated and sequenced from around the biotin attachment site of chicken, turkey and sheep liver pyruvate carboxylase (Rylatt *et al.*, 1977). The sequences were found to be highly homologous in this region with only three amino acid substitutions having occurred in these three species. The other biotin carboxylases for which sequence data were available were the biotinyl subunits of acetyl-CoA carboxylase from *E. coli* and of transcarboxylase from *P. shermanii*. The biotin-containing M_r = 9100 fragment

isolated from acetyl-CoA carboxylase was completely sequenced from overlapping proteolytic fragments by Sutton *et al.* (1977). The 1.3S biotinyl subunit of transcarboxylase was also sequenced from overlapping proteolytic fragments by Maloy *et al.* (1979). The alignment of the sequences of these two proteins with the pyruvate carboxylase sequences determined by Rylatt *et al.* (1977) revealed several conserved sequences especially in the immediate vicinity of the biocytin residue (ϵ-N-biotinyl lysine), see Fig. 1.5. In view of the conservation of the "Ala-Met-Bct-Met" sequence in all the proteins, Rylatt *et al.* (1977) suggested the possibility that this conservation represented evidence for divergence of the region in the vicinity of the biocytin from a common evolutionary progenitor protein.

Both Sutton *et al.* (1977) and Maloy *et al.* (1979) offered an additional explanation for the sequence homology. They noted that Kosow *et al.* (1962) and McAllister and Coon (1966) had found that the holocarboxylase synthetases from rabbit, yeast and *P. shermanii* was able to biotinylate rat apo-propionyl-CoA carboxylase, apo-3-methylcrotonyl-CoA carboxylase from *Achromobacter* and apo-transcarboxylase from *P. shermanii*, hence the holocarboxylase synthetases had a very wide range of substrate enzymes. Sutton *et al.* (1977) and Maloy *et al.* (1979) suggested that conservation of the sequence around the biocytin may have been dictated by the site requirements of the synthetase rather than true sequence conservation resulting from divergence from a common ancestral protein. The extent of the sequence homology in areas of the proteins removed from the biotin attachment site could be expected to reveal the whether the homology was the result of convergent or divergent evolution. Additional sequence data for pyruvate carboxylase would also aid in the localization of the other functionally important regions in pyruvate carboxylase both through the comparison with sequences known for other biotin carboxylases and from amino acids identified by affinity labelling.

The methods which had been used to sequence other proteins had generally involved the isolation and direct protein sequencing of proteolytic fragments of the protein under investigation. However, the size of the subunits of pyruvate carboxylase rendered this an unappealing approach even though β-galactosidase [EC 3.2.1.23] from *E. coli* which has a subunit of M_r ~116,000 had been sequenced

Figure 1.5

The Amino Acid Sequence Adjacent to the Biotinyl Group in Pyruvate Carboxylase and Other Biotin Carboxylases

The amino acid sequence determined for the biotinyl peptides from sheep, chicken and turkey pyruvate carboxylase by Rylatt *et al.* (1977) is shown. The sequences are compared with the sequence adjacent to the biotinyl group of transcarboxylase from *P. shermanii* (Maloy *et al.*, 1979) and acetyl-CoA carboxylase from *E. coli* (Sutton *et al.*, 1977). Bct refers to biocytin or ϵ-*N*-biotinyl lysine.

Pyruvate Carboxylase

Sheep	1	gly	gln	pro	leu	val	leu	ser	ala	met	bct	met	glu	thr	val	val	thr	ser	pro	val	thr	glu	gly	val	arg 24
Chicken		gly	ala	pro	leu	val	leu	ser	ala	met	bct	met	glu	thr	val	val	thr	ala	pro	arg					
Turkey		gly	ala	pro	leu	val	leu	ser	ala	met	bct	met	glu	thr	val	val	thr	ala	pro	arg					

Transcarboxylase

P. shermanii	gln	thr	val	leu	val	leu'	glu	ala	met	bct	met	glu	thr	glu	ile	asn	ala	pro	thr

Acetyl-CoA Carboxylase

E. coli	asn	thr	leu	cys	ile	val	glu	ala	met	bct	met	met	asn	gln	ile	glu	ala	asp	lys

in this manner (Fowler and Zabin, 1978). The alternative approach was to use recombinant DNA technology to acquire the protein sequence through inference from the cloned DNA sequence. In addition the mRNA and gene structure of the protein may be elucidated.

1.5 Recombinant DNA Technology

Recombinant DNA technology has made the isolation and characterization of homogeneous DNA preparations possible through the ability to clone DNA into plasmid and bacteriophage DNA vectors (Sinsheimer, 1977). The sequence of isolated clones may then be rapidly determined by one of several alternative techniques (Maxam and Gilbert, 1977; Sanger *et al.*, 1977). Libraries of recombinant DNA clones representing the genomic structure of organisms have been prepared in λ bacteriophage vectors (Smithies *et al.*, 1978; Maniatis *et al.*, 1978) and cosmid vectors (Collins and Hohn, 1978; Ish-Horowicz and Burke, 1981) using randomly cleaved genomic DNA to ensure that all genes were represented. The mRNA present in individual cell types, tissues and even whole organisms could be cloned into plasmid vectors after reverse transcription into complementary DNA (cDNA) (Efstratiadis and Villa-Komaroff, 1978; Okayama and Berg, 1982). The isolation of the specific clone required for the protein under investigation has been approached in several different ways depending on the characteristics of either the mRNA of the protein or the protein itself.

1.5.1 Approaches to Recombinant DNA Clone Isolation

Many different routes have been taken by investigators in trying to isolate specific clones from libraries of recombinant DNA clones. One method which has been used successfully was to purify the mRNA and use this as a radio-labelled probe to select clones either directly or after preparation of cDNA. Usually specific characteristics of the mRNA were used to aid in its purification. For example, as the vitellogenin mRNA is significantly larger than most mRNA species it could be readily puri-

fied on the basis of its size (Deeley *et al.*, 1977). Special properties of the mRNA such as base composition may also be used. Silk fibroin mRNA from the silk-worm, *Bombyx mori*, contains about 40% G residues which causes it to be sediment anomolously on sucrose density gradient centrifugation and thereby enables it to be readily purified (Brown and Suzuki, 1974). mRNA has been readily purified in several proteins where antibodies directed against the protein have been available. In these cases antibodies were used to bind to and immunoprecipitate the nascent polypeptide chain while still attached to the polysomes containing the mRNA. In fact such an approach was successfully used to purify the mRNA for the β subunit of propionyl-CoA carboxylase from rat liver (Kraus and Rosenberg, 1982). Probes for the selection of clones have also been prepared using inducible mRNAs where clones in a cDNA library prepared from induced mRNA were screened with radio-labelled mRNA prepared both prior to and after induction of specific mRNA synthesis. In this way clones encoding induced mRNA species could readily be detected (Dworkin and Dawid, 1980)

Another successful approach has been to screen clones by binding the clone DNA onto nitrocellulose, hybidizing this with the appropriate partially purified RNA preparation, eluting the specifically bound mRNA then assaying for the specific RNA. In this way mouse β_2-microglobulin cDNA clones were isolated after the identification of an immunoprecipitated translation product from the selected mRNA (Parnes *et al.*, 1981). Similarly a human leukocyte interferon cDNA clone was isolated after the identification of expressed interferon activity in oocytes after injection of the selected mRNA (Nagata *et al.*, 1980). Because of the laborious nature of this screening procedure initial screenings were usually performed with the DNA pooled from many clones and the screening was repeated with smaller sub-groups from each pool until single positive clones were isolated.

Finally, if sequence data was available for the protein under investigation then the mRNA sequence could be inferred for regions with minimal codon redundancy, for examples areas with methionine and tryptophan residues, and a synthetic DNA oligonucleotide could be prepared which was complementary to the mRNA. This oligonucleotide either could be radio-labelled and used directly to screen a library

of clones or could be used to specifically prime the synthesis of longer cDNA probes. These approaches have been used to isolate recombinant DNA clones for many proteins including the yeast cytochrome c gene (Montgomery *et al.*, 1978), rat insulins I and II (Chan *et al.*, 1979), pig gastrin (Noyes *et al.*, 1979; Mevarech *et al.*, 1979), human fibroblast interferon (Houghton *et al.*, 1980), rabbit β-globin (Wallace *et al.*, 1981), human β_2-microglobulin (Suggs *et al.*, 1981), rat relaxin (Hudson *et al.*, 1981) and chicken histone H5 (Krieg *et al.*, 1982). The oligonucleotides used in these studies have generally been between ten and fifteen bases long, therefore only a quite short region of protein sequence was required to generate such probes given of course the presence of amino acids with suitably low codon redundancy.

The sequence of the chicken liver pyruvate carboxylase biotin attachment site shown in Fig. 1.5 contains two methionine residues in close proximity as well as several amino acids with codons with low redundancy. This sequence appeared to be appropriate for the use of the oligonucleotide strategy employed by other workers.

More recently, the development of both improved cloning vectors and procedures to synthesize longer oligonucleotide probes has greatly increased the ability to isolate low abundance cDNA clones. This will be discussed more fully in Chapter 5.

1.6 Aims of the Project

The paucity of primary sequence information for pyruvate carboxylase when this study began was a significant limitation in the further analysis of both the structure and catalytic activity of the enzyme. The aim of this project was therefore to acquire sequence data for pyruvate carboxylase from a readily available source which would provide a useful basis for the analysis of the experimental data already determined in this group and elsewhere. The techniques of recombinant DNA technology were adopted so that sequence information could obtained in a efficient manner. The precise aim of the project was therefore to use recombinant DNA technology to determine as much of the sequence of pyruvate carboxylase as possible. Once the sequence was determined this could then be related to existing structural and catalytic models of pyruvate carboxylase.

Chapter 2

Materials and Methods

2.1 Materials

The chemical reagents and materials used in this study are listed below together with their catalogue number and manufacturer. All other reagents used were of analytical grade unless otherwise specified. The water used in this study was re-distilled using a glass still from water purified by the reverse osmosis process.

2.1.1 Antibiotics

- Ampicillin : (#A-9518), Sigma Chemical Co., MO, USA.
- Chloramphenicol : (#C-0378), Sigma.
- Tetracycline : (#T-3383), Sigma.

2.1.2 Bacterial Strains

The *E. coli* strains used are listed with their genotypes:

- C600Hfl$^-$: F$^-$, *thi-1, thr-1, leuB6, lacY1, tonA21, supE44,* λ^-, *hfl*$^-$. (Stratagene Cloning Systems, CA, USA).
- ED8799 : *hsdS, (rk*$^-$*, mk*$^-$*), metB7, supE, (glnV)44, supF, (tyrT)58,* Δ*(lacZ) M15.* (Gift of Dr. S. Clarke, Biotechnology Australia, NSW, Australia).
- JM101 : Δ*(lac, pro.), supE, thi.* F$'$ *traD36, proAB, lac I*^{q} *Z*Δ*M15.* (Messing, 1979).
- LE392 : F$^-$, *hsdR514 (rk4*$^-$*, mk*$^+$*), supE44, supF58, lacY1* or Δ*(lac 1ZY)*$_6$, *galK2, galT22, metBl, trpR55,* λ^-. (Murray *et al.*, 1977).
- MC1061 : *ara D 139,* Δ*(ara, leu)7697,* Δ*lacX74, gal U*$^-$*, gal K*$^-$*, hsr*$^-$*, hsm*$^+$*, strA.* (Casadaban and Cohen, 1980).

2.1.3 Cloning Vectors

- λgt10 : (*Eco*RI digested λgt10 (imm ^{434}b527) arms, #GT10), Stratagene. (Huynh *et al.*, 1985.)
- M13mp8 and M13mp9 : Gift from Dr. A.J. Robins. (Messing and Vieira, 1982).
- M13mp18 and M13mp19 : (#UV-03 and #UV-04), Bresatec Ltd., SA, Australia. (Yanisch-Peron *et al.*, 1985).
- pBR322 : Gift of Dr. A.J. Robins. (Bolivar *et al.*, 1977).
- pUC19 : (#UV-02), Bresatec. (Yanisch-Peron *et al.*, 1985).

2.1.4 Commonly Used Solutions

- Denhardts solution : 20 mg/ml polyvinylpyrollidone, 20 mg/ml bovine serum albumin, 20 mg/ml Ficoll 400.
- PBS : 150 mM NaCl, 10 mM potassium phosphate pH 7.4.
- PSB : 100 mM NaCl, 10 mM $MgSO_4$, 10 mM Tris·HCl pH 7.5, 0.01% (w/v) gelatin.
- SSC : 150 mM NaCl, 15 mM tri-sodium citrate.
- TAE : 80 mM Tris, 50 mM acetic acid, 4 mM Na_2EDTA (pH 8.2).
- TBE : 90 mM Tris, 90 mM boric acid, 2.5 mM Na_2EDTA (pH 8.3).
- TE : 10 mM Tris pH 8.0, 0.1 mM Na_2EDTA.

2.1.5 Enzymes and Proteins

- Avian Myeloblastosis Virus reverse transcriptase (AMV-reverse transcriptase) : (#MG-101), Molecular Genetic Resources Inc., FL, USA.
- Bovine serum albumin : (nuclease-free, #BB-1), Bresatec.

- Bovine serum albumin size marker : (#A-7517), Sigma.
- Calf intestinal phosphatase : (#713 023), Boehringer Mannheim, West Germany.
- *E. coli* deoxyribonuclease I (DNase I) : (#D-4527), Sigma.
- *E. coli* DNA Polymerase I : (#DP-1), Bresatec.
- *E. coli* DNA polymerase I-large fragment (Klenow) : (#KF-1), Bresatec.
- *E. coli* β-Galactosidase size marker : (#G-8511), Sigma.
- Endoproteinase lys-C : (#476 986), Boehringer Mannheim.
- Gelatin : Davis Gelatine Co., NSW, Australia.
- Lysozyme : (#L-6876), Sigma.
- Proteinase K : (#745 723), Boehringer Mannheim.
- Restriction endonucleases : Restriction enzymes were obtained principally from New England Biolabs, Inc., MA, USA. However some enzymes were obtained from Boehringer Mannheim, International Biotechnologies Inc. (IBI), CT, USA and Toyoba Co. Ltd., Japan.
- Ovalbumin (chicken) size marker : (#A-7642), Sigma.
- Phosphorylase b (rabbit muscle) size marker : (#P-4649), Sigma.
- Ribonuclease A (RNase A) : (#R-5000), Sigma.
- Ribonuclease H (RNase H) : (#786 349), Boehringer Mannheim.
- Terminal transferase : (#27-0730-01), Pharmacia Biotechnology International AB, Sweden.
- Trypsin (1:250) : (#0152-13-1), Difco Laboratories, MI, USA.
- T4 DNA ligase : (#TL-1), Bresatec.

- T4 DNA polymerase : (#19040), IBI.
- T4 Polynucleotide kinase : (#27-0734-01), Pharmacia.

2.1.6 General Chemicals and Materials

- 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (BCIG) : (#B-4252), Sigma.
- Corex centrifuge tubes : (30 ml, #8445), Corning Ltd., England.
- Dialysis tubing : (25 mm wide, #453105), Union Carbide International Co., NY, USA.
- D,L-Dithiothreitol : (#D-0632), Sigma.
- Ficoll 400 : (#17-0400-01), Pharmacia.
- GF/A Glass fibre disks : (#1820025), Whatman Ltd., England.
- Isopropyl β-D-thiogalactopyranoside (IPTG) : (#I-5502), Sigma.
- L-Leucine : (#L-8125), Sigma.
- 2-Mercaptoethanol : (#M-6250), Sigma.
- Minifold dot blot apparatus : Schleicher and Schuell Inc., NH, USA.
- Nitrilotriacetic acid : (#N-0128), Sigma.
- Nitrocellulose : (#BA-85), Schleicher and Schuell.
- Polyallomer ultracentrifuge tubes : (10 ml, #331372), Beckman Instruments, CA, USA.
- Polycarbonate ultracentrifuge tubes : (10 ml, #3118-0010), Beckman.
- Polyvinylpyrrolidone : (#PVP-40T), Sigma.
- Quickseal ultracentrifuge tubes : (2 ml, #344625), Beckman.

- Sodium dodecyl sulphate (SDS) : (#L-4632), Sigma.
- X-Ray film : (Fuji RX), Fuji Photo Film Co. Ltd., Japan.
- X-Ray film - High sensitivity : (Kodak X-Omat XAR-5, #1651512), Eastman Kodak Ltd., NY, USA.

2.1.7 Nucleotide Triphosphates

- Disodium adenosine 5′-triphosphate (ATP) : (#A-6144), Sigma.
- Disodium 2′-deoxyadenosine 5′-triphosphate (dATP): (#D-6500), Sigma.
- Disodium 2′-deoxycytidine 5′-triphosphate (dCTP) : (#D-4760), Sigma.
- Sodium 2′-deoxyguanosine 5′-triphosphate (dGTP) : (#D-4135), Sigma.
- Sodium thymidine 5′-triphosphate (TTP) : (#T-8635), Sigma.

2.1.8 Radiochemicals

- [α-^{32}P] dATP (1800-3000 Ci/mmole) : (#ADA-1), Bresatec.
- [α-^{32}P] dCTP (1800-3000 Ci/mmole) : (#ADC-1), Bresatec.
- [α-^{32}P] dGTP (1500 Ci/mmole) : Gift of Dr. R.H. Symons.
- [γ-^{32}P] ATP (2000-3000 Ci/mmole) : (#GRA-1), Bresatec.
- L-[^{35}S] methionine (>800 Ci/mmole) : (#SJ.204), Amersham International Pty. Ltd. Co., England.

2.1.9 Synthetic Oligonucleotides

- B-14 Biotin site oligonucleotide : 5′-(GTCTCCATCTTCAT)-3′, Pharmacia.
- B-14M Mixed biotin site oligonucleotide : 5′-(GT^{T}/$_{C}$TCCAT^{T}/$_{C}$TTCAT)-3′, Bresatec.

- Dale deletion sequencing system oligonucleotides :
  5′-(CGACGGCCAGTGAATTCCCC)-3′, (#OCP-20) and
  5′-(CGACGGCCAGTGCCAAGCTTTTTTTTTTT)-3′,
  (#OCP-29), Bresatec.
- *Eco*RI Linker oligonucleotide : 5′-(GGAATTCC)-3′, (#L8-EcoRI), Bresatec.
- M13 Sequencing primer oligonucleotide :
  5′-(GTAAAACGACGGCCAGT)-3′, (#USP-17), Bresatec.
- Oligo d(T)$_{12-18}$ primer oligonucleotide : (#27-7858), Pharmacia.

2.1.10 Materials Used with Specific Methods

2.1.10.a Bacterial Culture

- Bacto-Agar : (#0140-01), Difco.
- Bacto-Casamino acids : (#0230-01-1), Difco.
- Bacto-Tryptone : (#0123-01), Difco.
- Bacto-Yeast extract : (#0127-01), Difco.
- D-Glucose : (#713), Ajax Chemicals Pty., Ltd., NSW, Australia.
- Maltose : (#M-2250), Sigma.

2.1.10.b Electrophoresis

- Acrylamide : (#A-8887), Sigma.
- Agarose : (Type I, #A-6013), Sigma.
- Ammonium persulphate : (#D3247), Ajax.
- Bromophenol blue : (#20015), British Drug Houses Chemicals Pty. Ltd., (BDH Chemicals), England.
- Coomassie brilliant blue R : Schwartz/Mann Inc., NY, USA.

- Ethidium bromide : (#E-8751), Sigma.
- Formamide : (#10326), BDH Chemicals.
- Glyoxal : (#G-5754), Sigma.
- *N,N'*-methylene-bis-acrylamide : (#M-7256), Sigma.
- *N,N,N',N'*-tetramethylenediamine (TEMED) : (#161-0801), Bio-Rad Laboratories, CA, USA.
- Tris (hydroxymethyl)-aminomethane (Tris) : (#708 976), Boehringer Mannheim.
- Xylene cyanol FF : (#X-2751), Sigma.

2.1.10.c Fluorography and Scintillation Counting

- Dimethylsulphoxide (DMSO) : (technical grade), Crown Zellerbach Corp., WA, USA.
- 2,5-diphenyloxazole (PPO) : (#D-4630), Sigma.
- Naphthalene : (technical grade, #334), Ajax.
- 1-4-bis [2-(5-phenyloxazolyl)]benzene; 2,2'-*p*-phenylene-bis(5-phenyloxazole) (POPOP) : (#P-3754), Sigma.

2.1.10.d Nucleic Acid Purification

- Brij 58 (polyoxyethylene 20 cetyl ether) : (#P-5884), Sigma.
- Cesium chloride : (Optical grade, #402301), Metalgesellschaft AG, West Germany.
- Oligo d(T) Cellulose : (#20003), Collaborative Research Inc., MA, USA.
- Phenol : (Special grade, #161-01025), Wako Pure Chemical Industries Ltd., Japan.

- Sephadex G-50 (fine) : (#17-0042-01), Pharmacia.
- Sodium azide : (#30111), BDH Chemicals.
- Sodium deoxycholate : (#5039h), Koch-Light Laboratories Ltd., England.
- Sucrose : (Nuclease-free, #5503), BRL/Life Technologies Inc., MD, USA.

2.1.10.e Tissue Culture

- Dexamethasone : (Decadron), Merck, Sharp and Dohme, NSW, Australia.
- Dulbecco's modified essential medium (DMEM) : (lyophilized powder, #430-1600), Gibco Laboratories, NY, USA.
- Foetal calf serum : (batch tested, #971 2301), Commonwealth Serum Laboratories (CSL), VIC, Australia.
- Insulin : (Actrapid, #0402-13-191-2), Novo-Industri A/S, Denmark.
- 3-Isobutyl-1-methyl-xanthine : (#I-5879), Sigma.
- Phenol red : (#20090 2S), BDH Chemicals.
- Versene buffer : (10×stock, #741 1901), CSL.

2.1.10.f Cell Extracts

- λ DNA packaging system : (Gigapack, #GP3), Stratagene.
- *In vitro* translation system : (Rabbit reticulocyte lysate, #N.90), Amersham.

2.2 Methods

Unless otherwise stated all techniques were employed using solutions which were sterilized either by autoclaving or by filtration. Glassware was sterilized by one of the following methods; (i) autoclaving, (ii) alkali washing in 0.1 M NaOH followed by thorough washing in sterile water or (iii) heating at 120°C for 12 hr in a dry-air oven.

2.2.1 Cellular RNA Preparation

RNA was prepared from liver samples and cultured cell lines using a modification of the procedure described by Chirgwin *et al.* (1979). The cell suspension from one T-75 tissue culture vessel was considered approximately equivalent to one 17 day old embryonic chicken liver.

Chicks were killed by decapitation, the liver was removed and homogenized in 15 ml of ice cold 6 M guanidine·HCl, 0.2 M sodium acetate pH 5.2, 1 mM 2-mercaptoethanol using a Dounce homogenization vessel. This was repeated until 5 livers were homogenized in the solution whereupon it was diluted with a further 15 ml of guanidine·HCl solution so that there was 10 ml of solution/g liver tissue (one 17 day old embryonic chicken liver weighs ~0.5 g). The homogenate was divided between two 30 ml Corex tubes and 7.5 ml of ethanol (at -20°C) was added. The solutions were mixed and kept for 1 hr at -20°C. The precipitate was collected by centrifugation at 12,000×g for 20 min at 4°C and was re-homogenized in 15 ml of guanidine·HCl solution. The RNA in the homogenate was ethanol precipitated again and the precipitate was resuspended in 7.5 ml of 7 M urea, 100 mM Tris·HCl, pH 8.5, 0.1 mM EDTA, 0.1% (w/v) SDS using a Dounce homogenizer. Phenol/chloroform extraction was used to remove contaminating proteins as follows: 7.5 ml of buffer saturated phenol was mixed with the homogenate using the Dounce homogenizer then 7.5 ml of chloroform was added and similarly mixed. The suspension was transferred to a Corex tube and centrifuged at 3000×g for 5 min at room temperature. The aqueous layer was extracted with sufficient water-saturated ether until the aqueous solution clarified and the suspension was centrifuged at 3000×g for 5 min at 4°C. The aqueous layer was recovered and 375 μl of 2 M potassium acetate pH 5.0 and 15 ml of ethanol (at -20°C) was added. The solution was mixed and kept overnight at -20°C.

The RNA precipitate was collected after centrifugation at 12,000×g for 20 min at 4°C and was washed with 5 ml of ice cold 2 M LiCl to extract contaminating DNA and tRNA. The RNA was again recovered after centrifugation at 12,000×g for 20 min at 4°C then washed with 5 ml of 70% (v/v) ethanol, 30% (v/v) 0.1 M

potassium acetate pH 5.0. The RNA was recovered after centrifugation at 12,000$\times g$ for 20 min at 4°C, dried under vacuum and redissolved in water. RNA samples were stored at -80°C.

2.2.2 Affinity Chromatography of RNA on Oligo d(T) Cellulose

Cellular RNA prepared as described in Section 2.2.1 was separated into polyadenylated [poly(A)$^+$] RNA and non-polyadenylated [poly(A)$^-$] RNA fractions using a modification of the procedure of Aviv and Leder (1972). RNA was diluted to a concentration of >2 mg/ml in water, heated to 65°C for 5 min and snap cooled on ice. The RNA was adjusted to the reagent concentrations of the loading buffer (0.5 M NaCl, 20 mM Tris·HCl pH 7.6, 1 mM EDTA, 0.1% (w/v) SDS) and loaded onto a 1 ml oligo d(T) cellulose column which had been equilibrated with loading buffer. The A_{254} of the column effluent was monitored with an ETP-Kortec K95 Variable Wavelength UV Detector. The RNA was loaded onto the column and the run-through represented poly(A)$^-$ RNA. The column was washed with loading buffer followed by loading buffer containing only 0.1 M NaCl until the A_{254} returned to its base value. Poly(A)$^+$ RNA was eluted with 10 mM Tris·HCl pH 7.6, 1mM EDTA, 0.5% (w/v) SDS and collected. The poly(A)$^+$ RNA was usually re-chromatographed on the oligo d(T) cellulose column to ensure complete removal of contaminating poly(A)$^-$ RNA and RNA samples were recovered by ethanol precipitation.

2.2.3 Sucrose Density Gradient Centrifugation of RNA

Poly(A)$^+$ RNA was size fractionated by centrifugation through a sucrose density gradient (Noll, 1967). Samples containing 100 μg of RNA were dissolved in 100 μl of 10 mM Tris·HCl pH 7.6, 1 mM EDTA, 0.1% (w/v) SDS, heated to 65°C and snap cooled on ice. A gradient of RNase-free sucrose was made from 5-25% (w/v) in 70% (v/v) formamide, 1 mM Tris·HCl pH 7.6, 1 mM EDTA using a linear gradient mixer and poured in 10 ml polyallomer ultracentrifuge tubes. Samples were applied to the top of the gradients and centrifuged in a Beckman SW-41 swing-out

rotor at 270,000×g for 20 hr. The gradient was fractionated using an Isco model 640 Density Gradient Fractionator and RNA was recovered from the fractions by ethanol precipitation.

2.2.4 Ethanol Precipitation of Nucleic Acids

Nucleic acid samples were mixed with a 10% volume of 3 M sodium acetate pH 5.2 and 2.5 volumes of redistilled ethanol (at -20°C). The solution was mixed, kept on dry ice for 5 min then centrifuged for 15 min in an Eppendorf centrifuge at 4°C. The supernatant was discarded and the pellet was washed with 1.5 ml of 70% (v/v) ethanol in water (at -20°C). The precipitated nucleic acid was dried under vacuum and redissolved in aqueous buffer, usually TE buffer.

2.2.5 Phenol and Phenol/Chloroform Extraction of Nucleic Acids

Nucleic acid samples were mixed thoroughly with a 0.5 volume of phenol which was saturated with 0.1 M Tris·HCl pH 8.0, 30 mM 2-mercaptoethanol, 0.1% 8-hydroxyquinoline. A further 0.5 volume of chloroform was then added and mixed. The suspension was centrifuged for 3 min in an Eppendorf centrifuge at room temperature and the upper aqueous phase was recovered. If there was a heavy precipitate at the interphase then the extraction was repeated until no precipitate was visible at the interphase.

DNA solutions from agarose gel electroelutions containing dissolved agarose were extracted as described above except that chloroform was not included.

2.2.6 Restriction Endonuclease Digestion of DNA

All restriction enzyme digestions of DNA were carried out using the conditions recommended by the manufacturer.

2.2.7 Polyacrylamide Gel Electrophoresis of DNA

2.2.7.a Non-Denaturing Gels

DNA samples in 5-10 μl of aqueous solution were mixed with a 20% volume of loading solution containing 50% (w/v) sucrose, 5 mM EDTA, 50 mM Tris·HCl pH 7.4, 1 mg/ml bromophenol blue, 1 mg/ml xylene cyanol FF and loaded onto either a 6% or 20% (w/v) polyacrylamide gel containing 1×TBE buffer. The dimensions of the gel were usually 20×20×0.05 cm or 40×20×0.05 cm.

2.2.7.b Denaturing Gels

DNA samples in 5-10 μl of aqueous solution were mixed with an equal volume of loading solution containing 95% (v/v) formamide, 10 mM EDTA, 0.2 mg/ml bromophenol blue, 0.2 mg/ml xylene cyanol FF. The DNA was heated for 2 min at 100°C and loaded onto either a 6% or 20% (w/v) polyacrylamide gel containing 50% (v/v) formamide, 1×TBE buffer. The dimensions of the gel were usually 20×20×0.05 cm.

2.2.8 Isolation of DNA from Polyacrylamide Gels

[^{32}P] labelled DNA was localized on polyacrylamide gels by autoradiography and the resulting autoradiograph was used as a template to locate the required DNA fragment on the gel. The gel containing the DNA fragment was excised and macerated using a glass rod. The [^{32}P] labelled DNA was eluted by soaking the gel fragments in 0.5 M ammonium acetate, 10 mM $MgCl_2$, 1 mM EDTA, 0.1% (w/v) SDS solution at 37°C overnight (Maxam and Gilbert, 1977). The eluate was centrifuged for 3 min in an Eppendorf centrifuge to remove gel fragments and DNA was recovered from the supernatant by phenol/chloroform extraction and ethanol precipitation.

2.2.9 Agarose Gel Electrophoresis

2.2.9.a DNA Samples

DNA samples were usually separated on horizontal submerged agarose gels. The gels varied in agarose concentration from 0.7–2.0% depending on the size of the DNA being electrophoresed. The gels contained 1×TAE or 1×TBE buffer as indicated and 0.3 μg/ml ethidium bromide. The running buffer usually also contained 0.3 μg/ml ethidium bromide.

2.2.9.b RNA Samples

RNA samples were denatured in 10 mM sodium phosphate pH 6.5, 1.0 M deionized glyoxal usually in a final volume of 14 μl for 1 hour at 50°C. A 2 μl aliquot of loading solution containing 40% glycerol, 10 mM sodium phosphate pH 6.5, 0.4% bromophenol blue, 0.4% xylene cyanol FF was added then the RNA was loaded onto an agarose gel and electrophoresed at 30 mA for 5 hr in 10 mM sodium phosphate pH 6.5 with buffer recirculation.

2.2.10 Isolation of DNA from Agarose Gels

DNA samples were localized on agarose by (i) illumination with long wave UV light after staining of the gel with 0.3 μg/ml ethidium bromide or (ii) by autoradiography if the DNA was [^{32}P] labelled. The gel containing the required DNA was excised and the DNA electro-eluted using a modification of the procedure of McDonell *et al.* (1977). A bag of dialysis tubing containing the gel slice and 500 μl of TE buffer was immersed in a tank of electrophoresis buffer. A current of ~100 mA was applied for 15-30 min until all the DNA had just migrated into the TE buffer. The buffer was removed from the bag and the DNA was recovered from it by phenol extraction and ethanol precipitation.

2.2.11 Transfer of Nucleic Acids to Nitrocellulose from Agarose Gels

2.2.11.a Southern Transfer of DNA to Nitrocellulose

DNA was transferred to nitrocellulose after agarose gel electrophoresis using the Southern transfer procedure described by Wahl *et al.* (1979). The gel was shaken gently in 250 mM HCl for 30 min to partially depurinate the DNA then the gel was neutralized by shaking in 0.5 M NaOH, 1 M NaCl for 30 min. The gel was then buffered by shaking in 0.5 M Tris·HCl pH 8.0, 3 M NaCl for 30 min and using 20×SSC as the blotting buffer, the DNA was transferred for 12–18 hr by capillary action to a sheet of nitrocellulose that had been heated for 5 min at 90°C in 20×SSC. The filter was washed in 2×SSC and the DNA was baked onto the nitrocellulose at 80°C under vacuum for 2 hr.

2.2.11.b Northern Transfer of RNA to Nitrocellulose

RNA was transferred to nitrocellulose after agarose gel electrophoresis using the Northern transfer procedure of Thomas (1980). The RNA was generally not stained with ethidium bromide prior to blotting as this was found to reduce the transfer efficiency. The RNA was transferred by capillary action for 12–18 hr, using 20×SSC as the blotting buffer, to a sheet of nitrocellulose that had been heated for 5 min at 90°C in 20×SSC. The RNA was baked onto the nitrocellulose at 80°C under vacuum for 2 hr.

2.2.12 SDS-Polyacrylamide Gel Electrophoresis of Proteins

The method used for discontinuous SDS-polyacrylamide gel electrophoresis was essentially that described by Laemmli (1970). The 20×20 cm slab gels were formed between two glass plates separated by 1.5 mm spacers at the side and the gel was sealed with yellow electrical insulation tape. The 10% (w/v) acrylamide separating gel mix was prepared by combining 12.5 ml of 375 mM Tris base, 0.2% (w/v) SDS adjusted to pH 8.8 with 14% (w/v) glycine, 6.25 ml of 39.68% (w/v) acry-

lamide, 0.32% (w/v) *N,N'*-methylene-bis-acrylamide, 6.25 ml of water and 30 μl of TEMED. Polymerization was initiated by the addition of 100 μl of 10% (w/v) ammonium persulphate and the gel mix was poured between the glass plates. A 3% (w/v) acrylamide stacking gel mix was prepared by combining 2.5 ml of 125 mM Tris·HCl pH 6.8, 0.2% (w/v) SDS, 0.5 ml of 7.5% (w/v) acrylamide, 0.2% (w/v) *N,N'*-methylene-bis-acrylamide, 1.95 ml of water and 6 μl of TEMED. Polymerization was initiated by the addition of 20 μl of 10% (w/v) ammonium persulphate and a stacking gel was poured with a well-forming comb positioned in the top of the gel.

Protein samples were mixed with 1 μl of 14.3 M 2-mercaptoethanol per 10 μl of sample and an equivalent volume of 62.5 mM Tris·HCl pH 6.8, 4% (w/v) SDS, 20% (v/v) glycerol, 0.004% (w/v) bromophenol blue and denatured by heating at 100°C for 5 min. Samples were electrophoresed at 120 V for 3 hours with 25 mM Tris base, 1% (w/v) SDS adjusted to pH 8.3 with 14% (w/v) glycine as the running buffer.

The separated proteins were stained in 0.1% (w/v) Coomassie brilliant blue R in 50% (v/v) methanol, 10% (v/v) acetic acid. The gels were destained in an aqueous solution of 5% (v/v) methanol, 10% (v/v) acetic acid.

2.2.13 Bacterial Culture Media

All culture media were prepared with glass distilled water and were autoclaved prior to use.

2.2.13.a Culture Media for *E. coli* Strains MC1061 and ED8799

Liquid Culture L-broth: 10 g of bacto-tryptone, 5 g of bacto-yeast extract, 10 g of NaCl in a final volume of 1 l.

L+amp broth: L-broth supplemented with 60 μg/ml ampicillin after autoclaving and cooling to <50°C.

L+tet broth: L-broth supplemented with 12.5 μg/ml tetracycline after autoclaving and cooling to <50°C.

Plate Culture L+amp, L+tet and L-plates were prepared using the liquid culture media with the addition of 1.5% (w/v) bacto-agar.

2.2.13.b Culture Media for *E. coli* Strains LE392 and C600Hfl$^-$

Liquid Culture NZCYM-medium: 10 g of bacto-tryptone, 5 g of bacto-yeast extract, 5 g of NaCl, 1 g of bacto-casamino acids, 2 g of $MgSO_4.7H_2O$ in a final volume of 1 l. After autoclaving and cooling to <50°C the medium was brought to 0.2% (w/v) maltose with the addition of 5 ml of 40% (w/v) maltose (sterilized by filtration through a 0.2 μm filter).

Plate Culture NZCYM Plates: NZCYM medium was autoclaved with the addition of 1.5% (w/v) bacto-agar. The NZCYM medium was brought to 0.2% (w/v) maltose after autoclaving and cooling to <50°C.

2.2.13.c Culture Media for *E. coli* Strain JM101

Liquid Culture Minimal medium: 10 g of K_2HPO_4, 4.5 g of KH_2PO_4, 1.0 g of $(NH_4)_2SO_4$, 0.5 g of trisodium citrate in a final volume of 1 l. After autoclaving and cooling to <50°C the salts were supplemented with 1 ml of 1.0 M $MgSO_4.7H_2O$, 0.5 ml of 0.2 μm filter sterilized 1.0% (w/v) thiamine·HCl, 10 ml of 0.2 μm filter sterilized 20% (w/v) D-glucose.

2×YT-medium: 16 g of bacto-tryptone, 10 g of bacto-yeast extract, 5 g of NaCl in a final volume of 1 l.

Plate Culture Minimal plates: Minimal plates were prepared by separately autoclaving sufficient salts for 1 l of medium in a volume of 500 ml and separately autoclaving 15 g of bacto-agar in 500 ml. After autoclaving, the salts and the agar solutions were mixed and after cooling to <50°C the supplements were added as described for liquid culture.

2.2.14 DNA Preparation

2.2.14.a Preparation of Bacteriophage M13 Single-Stranded DNA

A 1/100 dilution of an overnight culture of *E. coli* JM101 grown in minimal medium was prepared in 2 ml of 2×YT medium and inoculated with M13 bacteriophage from a plaque using a toothpick. The medium was incubated at 37°C with shaking for 5–6 hr and then centrifuged for 10 min at 2000 $\times g$ in a bench-top centrifuge at 4°C. The supernatant was transferred to a 2 ml tube and centrifuged for 10 min at 4°C in an Eppendorf centrifuge. A 1.5 ml fraction of the supernatant was taken for the DNA preparation and the remainder was stored at -20°C as a clone bacteriophage stock. The supernatant was transferred to a new tube and 300 μl of 2.5 M NaCl, 20% polyethyleneglycol-8000 (PEG-8000) was added. The solution was mixed, kept for 15 min at room temperature then centrifuged for 5 min at room temperature. The pellet was resuspended in 200 μl of TE buffer, phenol extracted and 150 μl of the aqueous phase was taken and the single-stranded DNA recovered by ethanol precipitation. The DNA was redissolved in 25 μl TE and stored at -20°C.

2.2.14.b Preparation of Bacteriophage M13 RF-DNA

A 1/100 dilution of an overnight culture of *E. coli* JM101 grown in minimal medium was prepared in 3 ml of 2×YT medium and inoculated with M13 bacteriophage from an M13 clone bacteriophage stock. The culture was incubated for 3 hr at 37°C with shaking and used subsequently as an inoculum for the larger culture. At the same time either 100 ml (for a small scale DNA preparation) or 500 ml (for a large-scale DNA preparation) of a 1/100 dilution of an overnight culture of *E. coli* JM101 grown in minimal medium was prepared. The culture was incubated at 37°C with shaking until the OD_{600} $\simeq$0.5 then 0.5% of the culture volume of the inoculum described above was added. The culture was incubated for 4 hr at 37°C with shaking and then taken for DNA preparation using either the procedure in Section 2.2.14.d for small-scale preparations or the procedure in Section 2.2.14.e for large scale preparations.

2.2.14.c Preparation of Plasmid DNA

For pBR322 clones a 1/100 dilution of an overnight culture of *E. coli* MC1061 was grown in either 100 ml (for small-scale preparations) or 500ml (for large-scale preparations) of L+tet broth at 37°C until the $OD_{600} \simeq 1.0$. The culture was brought to a concentration of 0.2 mg/ml chloramphenicol and incubated overnight to amplify the plasmid. The culture was taken for DNA preparation using either the procedure in Section 2.2.14.d for small-scale preparations or the procedure in Section 2.2.14.e for large scale preparations. The same procedure was used for pUC19 clones with *E. coli* ED8799 in L+amp broth.

2.2.14.d Small-Scale Purification of Plasmid and M13-RF DNA by CsCl Gradient Ultracentifugation

A modification of the CsCl ultracentrifugation procedure of Weeks *et al.* (1986) was used for small-scale DNA preparations of clones in pBR322 and pUC19 vectors and for M13 clone RF-DNA preparations.

The two 50 ml cultures prepared as described in Sections 2.2.14.b and 2.2.14.c were combined and the bacterial cells were collected by centrifugation at 4000×g for 10 min at 4°C. The cell pellet was resuspended in 4 ml of 50 mM glucose, 25 mM Tris·HCl pH 8.0, 10 mM EDTA, 2 mg/ml lysozyme and kept on ice for 20 min to lyse the cells. The suspension was mixed with 8 ml of 0.2 M NaOH, 1% SDS and kept on ice for 10 min then 5 ml of 3 M sodium acetate pH 5.0 was added and the mixture was kept on ice for a further 40 min. Cellular debris was removed by centrifugation at 27,000×g for 30 min at 4°C and the DNA present in the supernatant was recovered after phenol/chloroform extraction and ethanol precipitation.

The DNA pellet is redissolved in 720 μl of TE buffer then mixed with 120 μl of 10 mg/ml ethidium bromide. The DNA preparation was adjusted to a final density of 1.8 g/ml with 1.26 g CsCl, divided in two then layered underneath a 65% (w/v) CsCl solution in two 2 ml Quickseal ultracentrifugation tubes. The tubes were heat sealed and centrifuged at 360,000×g for 2.5 hr at 20°C in a TLA-100.2 rotor.

The top of the ultracentrifuge tube was pierced and a 24G needle was used to puncture the tube under the ethidium bromide stained DNA band which was removed into a 1 ml syringe. The DNA was collected in approximately 300 μl and the ethidium bromide was removed by extracting several times with an equal volume of isopropanol which had been saturated with 5 M NaCl, 10 mM Tris·HCl pH 8.5, 1 mM EDTA. The DNA was diluted with 2 volumes of water and recovered by ethanol precipitation. Plasmid and M13 DNA was usually stored at -80° C.

2.2.14.e Large-Scale Purification of Plasmid and M13-RF DNA by CsCl Gradient Ultracentifugation

The gentle-lysis method of DNA extraction and CsCl ultracentrifugation was used for large-scale DNA preparations of clones in pBR322 and pUC19 vectors and for M13 clone RF-DNA preparations.

The 500 ml culture prepared as described in Sections 2.2.14.b and 2.2.14.c was centrifuged at 4000$\times g$ for 10 min at 4°C. The cell pellet was resuspended in 4.7 ml of 25% (w/v) sucrose, 50 mM Tris·HCl pH 8.0, then 0.94 ml of 5 mg/ml lysozyme in 25% (w/v) sucrose, 50 mM Tris·HCl pH 8.0 was added and the suspension was kept on ice for 5 min. The lysed cells were mixed with 1.5 ml of 250 mM EDTA and kept on ice for 5 min then mixed with 7.5 ml of 1.0% Brij 58, 0.4% sodium deoxycholate, 50 mM Tris·HCl pH 8.0, 25 mM EDTA and kept on ice for 10 min. Cellular debris was removed by centrifugation at 39,000$\times g$ for 30 min at 4°C and the supernatant was recovered. The DNA preparation was brought to a concentration of 0.2 mg/ml ethidium bromide and adjusted to a final density of 0.95 g/ml with CsCl then centrifuged in 10 ml polycarbonate ultracentrifuge tubes at 50,000$\times g$ for 44 hr at 20°C.

The top of the ultracentrifuge tube was pierced and a red-hot nail was used to puncture the tube under the ethidium bromide stained circular DNA band which was collected. The ethidium bromide was removed by extraction as described in Section 2.2.14.d and the DNA was recovered by ethanol precipitation. Plasmid and M13 DNA was usually stored at -80°C.

2.2.14.f Preparation of Bacteriophage λ DNA

Bacteriophage λ DNA was prepared from the agarose overlay of plates with confluent lysis. Plates 15 cm in diameter prepared from 1.5% agarose, rather than agar, were used and the λ clone was plated at a density of 20,000–50,000 pfu per plate to ensure confluent lysis. The plates were incubated overnight at 37°C and the bacteriophage-containing overlay was recovered by scraping from the surface of the plate. The bacteriophage were extracted from the overlay by incubating in 15 ml of PSB, 50 μl of chloroform, 5 μl of 10 μg/μl RNase A, 5 μl of 10 μg/μl DNase I for 1 hr at 37°C. The overlay debris was removed by centrifugation for 15 min at 12,000×g at 4°C and the bacteriophage-containing supernatant was recovered. The bacteriophage particles were precipitated by centrifugation in 10 ml polycarbonate ultracentrifuge tubes at 140,000×g for 60 min at 4°C in a Ti50 ultracentrifuge rotor. The pellet was resuspended by soaking overnight in 1 ml PSB at 4°C followed by gentle trituration.

DNA was prepared from the bacteriophage particles as follows with care being taken to ensure that the λ DNA was not mechanically sheared. The resuspended bacteriophage heads were opened by the addition of 80 μl of 250 mM EDTA, 50 μl of 10% SDS and 2.5 μl of 20 mg/ml proteinase K and incubation for 60 min at 37°C. Protein was separated from the DNA by phenol/chloroform extraction until there is no proteinaceous precipitate at the aqueous:phenol/chloroform interphase. λ DNA was recovered by ethanol precipitation, dissolved in TE buffer and stored at -80°C after dividing into smaller aliquots.

2.2.15 Preparation of Cloning Vectors

Vector DNA was digested with the restriction enzyme (or enzymes) appropriate for the cloning of the desired DNA fragment. The vector DNA digest was electrophoresed on 0.8–1.2% (w/v) agarose gels to separate the undigested from linearized DNA and the linear DNA was extracted from the gel as described in Section 2.2.10. The vector DNA was then dephosphorylated to prevent the re-formation of vector DNA without an insert and in the ligation step. The dephosphorylation

reaction employed a modification of the procedure of Chaconas and van de Sande (1980). The following reagents were added to the DNA to be dephosphorylated: 5 μl of a mixture containing 0.5 M Tris·HCl pH 9.0, 10 mM $MgCl_2$, 1 mM $ZnCl_2$; water to a volume of 48 μl and 2 μl of 0.01 units/μl Calf Intestinal Phosphatase. The reaction mixture was mixed gently and incubated for 60 min at 37°C. The reaction was stopped by the addition of 2 μl of 30 mM nitriloacetic acid which chelates the zinc ions required by the enzyme. The reaction mixture was heated at 65°C for 15 min to inactivate the enzyme and the DNA was recovered by phenol/chloroform extraction and ethanol precipitation.

2.2.16 DNA Ligation

Double stranded DNA was ligated to (i) oligomeric linker DNA using a 5-10 fold molar excess of linkers or (ii) dephosphorylated cloning vector DNA using a 2-5 fold molar excess of fragment DNA. The reaction mixture contained the following reagents: 1 μl of 1M Tris·HCl pH 7.5, 2 μl of 100 mM $MgCl_2$, 2 μl of 10 mM DTT, 2 μl of 10 mM ATP pH 7.0, 4 μl of 25% (w/v) polyethylene glycol 8000, the DNA samples to be ligated (20 ng of vector DNA was usually present as appropriate) and water to a volume of 19.5 μl. The reagents were mixed and 0.5 μl of 1 unit/μl T4 DNA ligase was added regardless of the nature of the termini of the DNA fragments. The reaction mixture was mixed gently and incubated overnight at 14°C.

2.2.17 Bacterial Transformation

2.2.17.a Preparation of MC1061 and ED8799 Competent Cells

A culture of a 1/100 dilution of an overnight culture in 50 ml of L-broth was prepared in a 250 ml flask and incubated at 37°C with shaking for ~2 hr until the $OD_{600} \simeq 0.6$. The culture was kept on ice for 30 min then centrifuged at 4000×g for 5 min at 4°C. The cell pellet was resuspended in 25 ml of ice-cold 100 mM $MgCl_2$ then centrifuged at 4000×g for 5 min at 4°C. The cell pellet was resuspended in 2.5 ml of ice-cold 100 mM $CaCl_2$ and left on ice for 1 hr. The cells were then competent for the uptake of DNA in a transformation procedure.

2.2.17.b Preparation of JM101 Competent Cells

A culture of a 1/100 dilution of an overnight culture in 50 ml of 2×YT medium was prepared in a 250 ml flask and incubated at 37°C with shaking for ~2 hr until the $OD_{600} \simeq 0.6$. The culture was kept on ice for 30 min then centrifuged at 4000×g for 5 min at 4°C. The cell pellet was resuspended in 25 ml of ice-cold 200 mM $CaCl_2$ and kept on ice for 30 min. The cell suspension was then centrifuged at 4000×g for 5 min at 4°C. The cell pellet was resuspended in 5 ml ice-cold 80 mM $CaCl_2$ and kept on ice for 20 min. The cells were then competent for the uptake of DNA in a transformation procedure.

2.2.17.c Transformations of JM101 with M13 DNA

An aliquot of the DNA ligation reaction mixture was diluted with 100 μl TE buffer then 300 μl competent JM101 were added. The transformation solution was kept on ice for 40 min to allow cellular uptake of the DNA then the cells were heat-shocked for 2 min at 42°C. The transformation was mixed with 10 μl of 100 mM IPTG, 30 μl of 20 mg/ml BCIG dissolved in redistilled dimethylformamide, 3 ml of 2×YT soft agarose overlay (at 42°C). The mixture was poured as an overlay on a minimal plate which had been pre-warmed to 37°C. The overlay was allowed to set then the plate was incubated overnight at 37°C. Blue plaques represented non-recombinant clones with active β-galactosidase and clear (white) plaques represented clones with β-galactosidase which has been inactivated by an insertion or deletion of DNA.

2.2.17.d Transformations of MC1061 and ED8799 with Plasmid DNA

An aliquot of the DNA ligation reaction mixture was diluted with 100 μl TE buffer then 200 μl competent cells were added. The transformation solution was kept on ice for 30 min to allow cellular uptake of the DNA. The cells were heat-shocked for 2 min at 42°C then kept on ice for 30 min. To allow expression of the antibiotic resistance gene the cells were mixed with 0.5 ml of L-broth and incubated at 37°C for 45 min.

If MC1061 cells were used with the pBR322 vector then 3 ml of L-soft agarose

overlay (at 42°C) was added and the mixture was poured as an overlay on L+tet plates which had been pre-warmed to 37°C. The overlay was allowed to set and the plate was incubated overnight at 37°C.

If ED8799 cells were used with the pUC19 vector then the transformation mixture was mixed with 10 μl of 100 mM IPTG, 30 μl of 20 mg/ml BCIG dissolved in redistilled dimethylformamide and 3 ml of 2×YT soft agarose overlay (at 42°C). The mixture was poured as an overlay on an L+amp plate which had been pre-warmed to 37°C. The overlay was allowed to set and the plate was incubated overnight at 37°C. Blue colonies represented non-recombinant clones with active β-galactosidase and white colonies represented clones with β-galactosidase which has been inactivated by an insertion or deletion of DNA.

2.2.18 [^{32}P] Labelling of DNA

2.2.18.a 5′-[^{32}P] Labelling of DNA

DNA samples without a 5′ phosphate were labelled by phosphorylation of the DNA with [γ-^{32}P]ATP and T4 polynucleotide kinase using a modification of the method of Maxam and Gilbert (1980). Usually 10 μl of [γ-^{32}P]ATP (5 μCi/μl) was dried under vacuum for each 100 ng of DNA included in the reaction. The following reagents were then added: 2 μl of 1 M Tris·HCl pH 9.0, 2 μl of 100 mM $MgCl_2$, 2 μl of 100 mM DTT, 100-250 ng of 5′-dephosphorylated DNA and water to a volume of 19.5 μl. The reagents were mixed and 0.5 μl of 2 units/μl T4 polynucleotide kinase was added. The reaction mixture was gently re-mixed and incubated for 30 min at 37°C. The reaction was stopped by the addition of 1 μl of 250 mM EDTA, 0.1% (w/v) bromophenol blue and diluted with 20 μl of water.

The [^{32}P] labelled DNA was separated from unincorporated [γ-^{32}P]ATP by chromatography on a 2 ml pasteur-pipette column of Sephadex G-50 (fine). The column was run in an aqueous solution of 0.1 mM EDTA, 5 mM 2-mercaptoethanol and fractions were collected at 20 sec intervals. The [^{32}P] Cerenkov radiation in the fractions was determined using a Packard Tri Carb model 3390 liquid scintillation counter. The [^{32}P] labelled DNA usually eluted in fraction numbers 5-8 whereas

un-incorporated [γ-^{32}P]ATP eluted in fraction numbers 12-15. The DNA was usually [^{32}P] labelled to a specific radioactivity of 0.5–1.0$\times 10^8$cpm/μg DNA for 14 base oligonucleotides.

2.2.18.b [^{32}P] End Labelling of DNA

DNA fragments generated by restriction enzyme digestion and having overhanging termini were labelled by end-filling with α-[^{32}P] labelled dATP or dCTP. The choice of the [^{32}P] labelled dNTP was dictated by the presence of complementary thymidyl or guanosyl bases in the template strand. Usually 10 μl of [α-^{32}P] labelled dNTP (5 μCi/μl) was dried under vacuum for each 100 ng of DNA included in the reaction. The following reagents were then added: 1 μl of a mixture containing 100 mM Tris·HCl pH 7.5, 100 mM $MgCl_2$, 10 mM DTT; 100-250 ng of the DNA to be labelled and water to a volume of 9 μl. The reagents were mixed and 1.0 μl of 2 units/μl Klenow enzyme was added. The reaction mixture was gently re-mixed and incubated for 30 min at room temperature. The reaction mixture was diluted with 30 μl of water and was then phenol/chloroform extracted. The [^{32}P] labelled DNA was separated from unincorporated α-[^{32}P] labelled dNTP by chromatography on Sephadex G-50 as described in Section 2.2.18.a.

2.2.18.c [^{32}P] Labelling of DNA by Nick-Translation

Double-stranded DNA of >500 bp long was [^{32}P] labelled by nick-translation using a modification of the procedure described by Rigby *et al.* (1977) to a specific radioactivity of 0.5–1.0$\times 10^8$cpm/μg of DNA. Usually 10 μl of either [α-^{32}P]dATP or [α-^{32}P]dCTP (5 μCi/μl) was dried under vacuum for each 100 ng of DNA included in the reaction. The following reagents were then added: 4 μl of a mixture containing 125 μM of each of dATP, dCTP, dGTP and TTP (minus the dNTP which was α-[^{32}P] labelled), 250 mM Tris·HCl pH 7.6, 50 mM $MgS0_4$, 0.5 mM DTT, 250 μg/ml gelatin; 100-500 ng of the DNA to be labelled and water to a volume of 15 μl. The reagents were mixed and 5 μl of an enzyme mixture containing 5 units of *E. coli* DNA polymerase I, 40 pg DNase I in 40 mM potassium phosphate pH 7.0, 1 mM $MgCl_2$, 0.2 mM DTT, 50% (v/v) glycerol, 200 μg/ml gelatin was added. The reaction mixture was mixed gently and incubated for 90 min at 14°C. The labelling reaction mix was diluted with 20 μl of water and was

then phenol/chloroform extracted. The [^{32}P] labelled DNA was separated from unincorporated α-[^{32}P] labelled dNTP by chromatography on Sephadex G-50 as described in Section 2.2.18.a.

2.2.19 Screening of Libraries in λ Vectors

Essentially the same procedure as described by Benton and Davis (1977) was used for the screening of clones in any of the λ vectors. The host strains used were *E. coli* LE392 for λCharon 4A and λEMBL3 and *E. coli* C600Hfl$^-$ for λgt10.

A 1/100 dilution of the appropriate host cell strain was grown in NZCYM+0.2% maltose medium until the $OD_{600} \simeq 1.0$. An aliquot representing 25,000–50,000 pfu of the λ library bacteriophage stock in 100 μl of PSB was mixed with 500 μl of the host cell culture and incubated for 20 min at 37°C. The 9 ml NZCYM+0.2% maltose soft agarose overlay at 42°C was added to the cells and this was poured as an overlay onto a 15 cm diameter NZCYM plate. The plate was incubated for 12–18 hr at 37°C then cooled for 1 hr at 4°C to increase the rigidity of the overlay.

The plate was overlayed with a nitrocellulose disk for 2 min and the orientation of the disk was marked with ink. The disk was removed and the procedure repeated with a duplicate nitrocellulose disk for 3 min. The disks were placed on a sheet of 3MM paper saturated with 3.0 M NaCl, 0.5 M NaOH for 2 min and then were transferred to a sheet saturated with 1.5 M NaCl, 0.5 M Tris·HCl pH 8.0 for 2 min. The disks were washed with 2×SSC and the DNA was baked onto the filters at 80°C under vacuum for 2 hr. Filters were then hybridized to the probe DNA using the conditions described in Section 2.2.21.

2.2.20 Grunstein Screening of Plasmid Clones

Recombinant plasmid clones in pBR322 and pUC19 vectors were screened for hybridization to specific DNA probes using a modification of the procedure of Grunstein and Hogness (1975). Colonies of tetracycline resistant pBR322 clones and ampicillin resistant pUC19 clones were transferred by sterile toothpick to master L+Tet and L+Amp plates respectively. The colonies were also transferred to a

boiled sheet of nitrocellulose marked with a grid pattern and overlayed on a similar antibiotic plate. The plates were incubated overnight at 37°C and the master plates were then stored at 4°C as a stock of these clones. Chloramphenicol amplification of plasmid numbers in bacterial colonies prior to transfer to nitrocellulose was not used. DNA from the colonies grown on the nitrocellulose filters was denatured by sitting the filter on 3MM paper saturated with 0.5 M NaOH, 1.5 M NaCl for 5 min. The filter was then transferred to a second sheet of 3MM paper saturated with neutralizing solution containing 1.5 M NaCl, 0.5 M Tris·HCl pH 8.0 and left for 5 min. The filter was briefly rinsed with 2×SSC then baked under vacuum for 2 hr at 80°C. Filters were then hybridized to the probe DNA using the conditions described in Section 2.2.21.

2.2.21 Hybridization Conditions

2.2.21.a Oligonucleotide Probes

Oligonucleotide DNA probes that were 5′ [^{32}P] labelled (Section 2.2.18.a) were hybridized to nucleic acids bound onto nitrocellulose filters using a modification of the conditions of Szostak *et al.* (1979). The filter was prehybridized for 2 hr at 37°C in a solution containing 5×SSC, 5×Denhardts solution and 100 μg/ml of sonicated salmon sperm DNA prepared as described by Maniatis *et al.* (1982). The prehybridization was performed in a sealed plastic bag using a sufficient volume of prehybridization solution so that the filter was adequately wetted. The prehybridization solution was discarded and was replaced by the probe DNA dissolved in a small volume of the same solution. The filter was hybridized to the probe DNA for 18–24 hr at 37°C. The hybridization solution was discarded and the filter was usually washed in 1×SSC, 0.1% SDS at 37°C then autoradiographed.

2.2.21.b Other DNA Probes

[^{32}P] Labelled DNA probes were hybridized to nucleic acids bound onto nitrocellulose filters using similar conditions to those described by Maniatis *et al.* (1982). The filter was prehybridized for 2 hr at 42°C in a solution containing 50% formamide,

5×SSC, 5×Denhardts solution, 50 mM potassium phosphate pH 6.8, 25 μg/ml sonicated salmon sperm DNA and 1.0 % glycine. The prehybridization was performed in a sealed plastic bag using a sufficient volume of prehybridization solution so that the filter was adequately wetted. The prehybridization solution was discarded and was replaced by the probe DNA dissolved in a small volume of a solution containing 50% formamide, 5×SSC, 1×Denhardts solution, 20 mM potassium phosphate pH 6.8, 50 μg/ml sonicated salmon sperm DNA. The hybridization solution containing the probe was heated at 100°C for 5 min prior to adding to the filter. The filter was hybridized to the probe DNA for 18–24 hr at 42°C. The hybridization solution was discarded and the filter was washed at the stringency stated for the given probe then autoradiographed.

In library screenings using either 5′ [^{32}P] labelled or nick-translated probes $> 1 \times 10^6$cpm of probe was used for each filter.

2.2.22 Dideoxy Chain-Termination Sequencing

Sequencing was carried out using single stranded M13 clone DNA (Messing *et al.*, 1981) using a modification of the procedure described by Sanger *et al.* (1977).

A 6 μl aliquot of the single-stranded M13 DNA template was mixed with 1 μl of 100 mM Tris·HCl pH 8.0, 100 mM $MgCl_2$, 1 μl of 2.5 ng/μl M13 sequencing primer and 2 μl of water. The solution was heated at 100°C in a heating block which was then allowed to cool to 30°C over 30 min to permit annealing of the primer to the DNA. A 2 μl aliquot of [α-^{32}P]dATP (5 μCi/μl) was dried under vacuum and the annealed single-stranded DNA-primer solution was added and mixed.

A set of 4 tubes containing the nucleotide triphosphate master-mixes was prepared as follows:

"A" tube: 2 μl of A° containing 200 mM dCTP, dGTP, TTP in 1×TE buffer and 2 μl of 0.1 mM ddATP.

"C" tube: 2 μl of C° containing 15 μM dCTP, 200 μM dGTP, TTP in 1×TE buffer and 2 μl of 0.08 mM ddCTP.

"G" tube: 2 μl of G° containing 15 μM dGTP, 200 μM dCTP, TTP in 1×TE buffer and 2 μl of 0.03 mM ddGTP.

"T" tube: 2 μl of T° containing 15 μM TTP, 200 μM dCTP, dGTP in 1×TE buffer

and 2 μl of 0.05 mM TTP.

A 2 μl aliquot of the [α-^{32}P]dATP/annealed DNA mixture was added to 2μl of each master-mix together with 0.5 μl of 2 units/μl Klenow enzyme. The reaction mixtures were mixed and incubated for 15 min at either 37°C or 50°C, then 1 μl of Chase solution containing 0.25 mM of each of dATP, dCTP, dGTP, TTP was added. After incubation for a further 10 min, 4 μl of loading solution containing 95% (w/v) formamide, 10 mM EDTA, 10 mM NaOH, 1 mg/ml bromophenol blue, 1 mg/ml xylene cyanol FF was added and the DNA was heat-denatured by incubation at 100°C for 5 min prior to electrophoresis on a sequencing gel as described below.

2.2.23 Sequencing Gel Electrophoresis

Dideoxy chain-termination sequencing reactions were electrophoresed on 6% polyacylamide gels containing 8 M urea and 1×TBE buffer. The gels used had the dimensions, 20×20×0.025 cm or 40×20×0.025 cm depending on the number of reactions to be electrophoresed. Denatured samples were loaded under a current of 2 mA and the gels were run at a current such that the gels were kept at 60-70°C (usually 20 mA for 20 cm gels and 40 mA for 40 cm gels). After electrophoresis was complete the urea in the gel was removed and the DNA was fixed by washing with 10% (v/v) glacial acetic acid, 15% (v/v) ethanol in water (500 ml for 20 cm gels, 1 l for 40 cm gels). The gel was dried onto the glass plate for at least 30 min at 110°C and after cooling to room temperature was covered in plastic wrap then autoradiographed overnight at room temperature.

2.2.24 Computer Programs

These programs listed below were used with a Digital Equipment Corp., VAX 11/750 computer running on a VMS 4.5 operating system unless otherwise stated.

- ALIGNSIG – A program by I. Dodd using the Doolittle parameters (Kyte and Doolittle, 1982) for optimal protein alignments. The statistical significance of a match is determined by comparison with the randomized sequence alignment.

- ANALYSEQ – A suite of DNA and protein sequence analysis programs used for listing, restriction enzyme searches, translation, composition calculation.

- CHOUDOT – A program which predicts turn propensity, secondary structure, charge and hydropathy (Novotný *et al.*, 1984). The hydropathy parameters of Kyte and Doolittle (1982) were used.
- DELPHI – A protein secondary structure prediction program which uses the Garnier method (Garnier *et al.*, 1978).
- DIAGON – A program for comparing and aligning DNA and protein sequences which plots the output (Staden, 1982).
- HPLOT – A program to produce charge and hydrophilicity profiles of protein sequence (Kanehisa, 1983).
- MATCH – A program to search for homologous sequences in databases of DNA and protein sequence (Wilbur and Lipman, 1983).
- SEQA – A global sequence alignment program (Kanehisa, 1985).
- SEQCOLOUR – A program to format DELPHI and HPLOT program outputs for colour ink-jet printing.
- SEQH – A program which searches for local DNA sequence homology (Kanehisa, 1985).
- SEQHP – A program which searches for local protein sequence homology (Kanehisa, 1985).
- SIZING – A program to determine the size of DNA restriction fragments by comparison with a parabola fitted to the mobility of DNA standards (Duggleby *et al.*, 1981). This program was run on a Basis Medfly Microcomputer.
- TESTCD – A program which predicts the coding potential of regions of DNA sequence which was written and kindly donated by I. Dodd. This program was run on a Disc Computer Systems DS-23 microcomputer.

2.2.25 Containment Facilities

All manipulations involving recombinant DNA were carried out in accordance with the regulations and with the approval of the Recombinant DNA Monitoring Committee of the Australian Academy of Science and the University Council of the University of Adelaide.

Chapter 3

Chicken Liver Pyruvate Carboxylase : Development of cDNA Probes and Protein Sequencing

3.1 Introduction

To provide the basis for a better understanding of the structure of pyruvate carboxylase, knowledge of the enzyme's amino acid sequence is essential. Prior to the advent of recombinant DNA technology the determination of the sequence of a protein required the direct sequencing of isolated proteins or peptides. The large size of pyruvate carboxylase would render this an exceedingly difficult and time-consuming process. However, since recombinant DNA technology offers the ability to rapidly determine the sequence of cDNA or genomic clones and therefore quite readily infer the complete amino acid sequence of the protein it provides a means of more rapidly determining the sequence of pyruvate carboxylase.

The approach taken by many workers in isolating recombinant DNA clones has initially involved direct sequencing of part of the protein being studied so that the mRNA sequence may be deduced from the amino acid sequence and selected areas with low codon redundancy identified. Using this sequence data together with the preferred codon usage for the species under investigation synthetic oligonucleotides complementary to the mRNA have been designed. Oligomers such as this have been used either directly as probes to screen libraries for clones or as primers to make extended cDNA probes.

The amino acid sequence of pyruvate carboxylase from several species (Fig. 1.5, Rylatt *et al.*, 1977) offered the starting point for the use of a similar strategy in the isolation of clones for pyruvate carboxylase. The initiation of the study was dependent on the selection of the most appropriate source of both pyruvate carboxylase and its mRNA for use in the isolation and/or generation of cDNA clones.

A review of the amount of pyruvate carboxylase activity prepared from different vertebrate sources revealed that the livers of late-embryonic chickens and the mouse fibroblast cell line 3T3 yield the highest levels of pyruvate carboxylase activity per gram wet weight of tissue (Wallace, 1985). In selecting the most appropriate source of the pyruvate carboxylase to be used in this study consideration was given to the amount of information available for the enzyme from various sources. Chicken

liver and sheep liver have been the sources of pyruvate carboxylase that has been characterized to the greatest extent in this research group, therefore the enzyme from embryonic chicken liver was a strong candidate for use in this study.

In a study of the differentiation of the mouse fibroblast cell line 3T3-F442A, Angus *et al.* (1981) showed that the synthesis of pyruvate carboxylase was dependent on the level of pyruvate carboxylase mRNA. If control of the synthesis of pyruvate carboxylase resides at the level of mRNA synthesis in the embryonic chicken liver it is probable that the elevated level of pyruvate carboxylase activity is linked directly to an elevated rate of pyruvate carboxylase synthesis and ultimately to an elevated level of pyruvate carboxylase mRNA. In view of this possibility, embryonic chicken livers were chosen as the source of pyruvate carboxylase mRNA and the studies described in this chapter are restricted to chicken liver pyruvate carboxylase.

This investigation described in this Chapter includes;

(1) an examination of the development of the embryonic chicken liver to establish the time when the maximum amount of pyruvate carboxylase mRNA was present,

(2) a study of the suitability of a biotin site synthetic oligonucleotide for use as both a primer for specific cDNA synthesis and as a direct probe for pyruvate carboxylase mRNA,

(3) the generation of a much larger amount of pyruvate carboxylase amino acid sequence. This sequence may be used as a source of sequence data for the synthesis of additonal oligonucleotides and to provide a larger pyruvate carboxylase sequence data base with which to compare the sequence of isolated clones in order to verify their identify.

3.2 Specific Methods

3.2.1 *In Ovo* Pulse Labelling of Chicken Embryos

Chicken eggs were incubated at 37°C (28°–31°C wet bulb thermometer) and turned at 12 hourly intervals. The air sac was visualized by candling the eggs and the shell and the centre of the air sac was pierced with a sterile scissor point. A sterile 50

μl autozero was used to deliver 50 μl of phosphate buffered saline (PBS; 150 mM NaCl, 10 mM potassium phosphate pH 7.2) containing 100 μCi of L-[^{35}S]methionine onto the surface of the air sac membrane. The holes were sealed with tape and the eggs were incubated for a period of 15 min prior to sacrifice of the embryos.

3.2.2 Preparation of Mitochondria from Embryonic Chicken Liver

Embryos were rapidly removed from eggs and decapitated. The livers were carefully excised and immediately chilled in 50 ml of ice-cold 250 mM sucrose, 5 mM Tris·HCl pH 7.6. After gently blotting the livers dry they were weighed on pre-weighed pieces of aluminium foil. The livers were transferred to a Potter-Elvehjem homogenization vessel with four volumes of ice-cold 250 mM sucrose, 5 mM Tris·HCl pH 7.6 and homogenized with five strokes of a motor-driven teflon pestle set at low speed.

The homogenate was transferred to a 50 ml Corex tube and centrifuged in a Sorvall SS-34 rotor at 500$\times g$ for 5 min at 4°C to pellet the nuclei and cellular debris. The supernatant was decanted into another 50 ml Corex tube and centrifuged at 12,000$\times g$ for 20 min at 4°C. A pasteur pipette was used to remove the supernatant containing a cloudy suspension of ruptured mitochondria and the mitochondrial pellet was resuspended in 1 ml of 5 mM Tris·HCl pH 7.6. The mitochondrial suspension was transferred to a 50 ml round-bottomed flask and freeze-dried overnight. The mitochondria were resuspended in 500 μl of 5 mM Tris·HCl pH 7.6 prior to SDS-polyacrylamide gel electrophoresis (Section 2.2.12).

3.2.3 *In Vitro* Translation

RNA samples were translated *in vitro* using rabbit retriculocyte lysate prepared by Amersham International using a modification of the procedure of Pelham and Jackson (1976). The lysate was stored in 50 μl aliquots at -80°C and thawed on ice immediately prior to use. L-[^{35}S]Methionine was used as the labelled amino acid and sufficient was dried under vacuum to give 1μCi/μl in the translation reaction mixture. The dried L-[^{35}S]methionine was resuspended in sufficient rabbit reticu-

locyte lysate for all reactions. The translations were usually carried out in a 10 μl reaction volume containing 80% (v/v) lysate with L-[^{35}S]methionine, 100 mM L-leucine supplement (the lysate is depleted in both L-methionine and L-leucine so that either may be used as the labelled amino acid) and 100 ng/μl RNA sample. The translations were incubated for 90 min at 30°C then kept at 4°C prior to SDS-polyacrylamide gel electrophoresis (Section 2.2.12).

3.2.4 Immunoprecipitation of *In Vitro* Translation Products

The immunoprecipitation of proteins involved the formation of an antibody-antigen complex which was then bound to the protein A present on *Staphylococcus aureus* cells and collected by centrifugation (Kessler, 1981). Protein A purified rabbit anti-chicken liver pyruvate carboxylase antibody was prepared from homogeneous pyruvate carboxylase (by SDS-polyacrylamide gel electrophoresis) and was kindly supplied by J. Brazier. The antibody-antigen complex was formed in a reaction mixture containing 10 μl of the protein sample, e.g. the *in vitro* translation reaction mixture, 100 ml of 10×IP buffer (500 mM Tris·HCl pH 8.0, 1.5 M NaCl, 10% (v/v) Triton X-100, 1% (w/v) sodium deoxycholate, 10 mM EDTA), 50 μl of antibody preparation and H_2O to a final volume of 1.0 ml in a 1.5 ml tube. This solution was mixed gently and left for 30 min at 20°C then overnight at 4°C. The antibody/antigen complex was adsorbed onto protein A-bearing *S. aureus* cells by the addition of 100 μl of 10% (w/v) *S. aureus* cell suspension in 1×IP buffer and this was kept on ice for 1 hr with inversion at 15 min intervals to prevent settling out of the cells. The immunoprecipitation mixture was centrifuged for 2 min in an Eppendorf microfuge and the supernatant was discarded. The cells were washed by resuspending in 1 ml of 1×IP buffer, 0.1% (w/v) BSA and then centrifuged for 2 min in an Eppendorf microfuge. The supernatant was discarded and this washing step was repeated. The cells were finally resuspended in 25 μl of H_2O prior to SDS-polyacrylamide gel electrophoresis.

3.2.5 [^{14}C] Labelling of Marker Proteins by S-Carboxymethylation

[^{14}C] Labelled protein markers including β-galactosidase, phosphorylase b, bovine serum albumin and ovalbumin as well as authentic chicken liver pyruvate carboxylase were kindly prepared by J. Brazier. The labelling was performed using iodo[2-^{14}C]acetic acid with the S-carboxymethylation procedure of Sutton *et al.* (1977).

3.2.6 Fluorography

[^{14}C] and [^{35}S] Labelled proteins were detected after electrophoresis on polyacrylamide gels using a modification of the fluorographic method of Bonner and Laskey (1974). Gels were soaked for 2 hr in 10% (v/v) acetic acid, 10% (v/v) methanol to fix the proteins then in dimethylsulphoxide (DMSO) for 1 hr to remove the water. The DMSO was replaced with 1% (w/v) PPO, 20% (w/v) naphthalene in DMSO and the gels were soaked for 2 hr. The DMSO solution was removed and the gels were washed under running water for 1 hr to precipitate the PPO/naphthalene in the gel matrix. The gels were covered with plastic film and dried onto 3MM paper using a gel drying apparatus.

3.2.7 RNA Dot Blot

RNA samples were dissolved in 2×SSC and spotted onto a sheet of air-dried nitrocellulose which had been boiled for 5 min in water and washed in 2×SSC. The filter was baked for 2 hr at 80°C under vacuum prior to hybridization with the probe DNA.

3.2.8 Primer Extension cDNA Synthesis

Primer extension reactions using the B-14 biotin attachment site primer were performed with a modification of the conditions described by Krieg *et al.* (1982). The primer was used either unlabelled in the presence of [α-^{32}P] labelled deoxynucleotide triphosphate or as 5′ [^{32}P] labelled B-14 with unlabelled deoxynucleotide

triphosphates.

3.2.8.a Primer Extensions with an [α-^{32}P] Labelled Deoxynucleotide Triphosphate

Except where otherwise indicated 1 μg of RNA and 100 ng of B-14 primer was used in the primer extension reaction. [α-^{32}P]dATP, [α-^{32}P]dATP and [α-^{32}P]dGTP were used independently as the source of [^{32}P] label and in each case the other deoxynucleotide triphosphates in the reaction mixture were unlabelled. The reaction mixture was prepared in the following manner. A 10 μl aliquot of the [α-^{32}P] labelled deoxynucleotide triphosphate (5 μCi/μl) was dried under vacuum in a 1.5 ml tube and the other reagents were added to give a final volume of 20 μl containing 50 mM Tris·HCl pH 8.0, 50 mM KCl, 1 mM DTT (except where otherwise indicated), 7 mM $MgCl_2$, 0.4 mM of the same deoxynucleotide triphosphate as the label and 1 mM of the other the deoxynucleotide triphosphates (except where otherwise stated). The reaction mixture was brought to 65°C for 5 min to remove RNA secondary structure and was then rapidly cooled on ice to prevent re-annealing. Avian myelobastosis virus (AMV) reverse transcriptase (20 units) was added and the reaction incubated for 90 min at 37°C. The reaction was stopped by the addition of 20 μl of 1 M NaOH and the RNA was hydrolysed by incubation for 15 min at 65°C. The reaction mixture was neutralized by the addition of 20 μl of 1 M HCl and 20 μl 1 M Tris·HCl pH 7.6 and the cDNA was recovered by ethanol precipitation.

3.2.8.b Primer Extension with 5′ [^{32}P] Labelled B-14 Primer

Primer extensions using 5′ [^{32}P] labelled B-14 primer were performed with essentially the same conditions as described in Section 3.2.8.a however no [α-^{32}P] labelled deoxynucleotide triphosphate was present. All four deoxynucleotide triphosphates were present at 1 mM in the reaction mixture. The B-14 primer (100 ng) was initially labelled using polynucleotide kinase and [γ-^{32}P]ATP (Section 2.2.18.a) and unincorporated [γ-^{32}P]ATP was separated from the 5′ [^{32}P] labelled B-14 by chromatography on a 1 ml Sephadex G-50 column prior to the primer extension. The

5′ [^{32}P] labelled B-14 was dried under vacuum and included in the reaction as the sole [^{32}P] label. The cDNA was recovered from the reaction as described in Section 3.2.8.a.

3.2.9 Chemical Cleavage Sequencing of DNA

The chemical cleavage method of sequencing DNA was performed essentially as described by Maxam and Gilbert (1977) with the addition of the thymidine specific cleavage reaction (T*) of Rubin and Schmid (1980). Extreme care was taken when handling dimethyl sulphate and hydrazine and all cleavage reactions were performed in a fume hood.

The 5′ [^{32}P] labelled single stranded DNA was dissolved in 35 μl of H_2O and divided between the five cleavage reactions as indicated.

G>A Cleavage Reaction

To 5 μl of 5′ [^{32}P] labelled DNA was added 200 μl of 50 mM sodium cacodylate pH 8.0, 10 mM $MgCl_2$, 0.1 mM EDTA and 1 μl of 99% (v/v) (10.7 M) dimethyl sulphate. The reagents were mixed and left for 2 min at 20°C. The reaction was then stopped by the addition of 50 μl of 3 M sodium acetate pH 6.0, 2.5 M 2-mercaptoethanol, 1 mM EDTA, 0.1 mg/ml *E. coli* tRNA.

A+G Cleavage Reaction

To 10 μl of 5′ [^{32}P] labelled DNA was added 25 μl of 98% (v/v) (26 M) formic acid. The reagents were mixed and left for 2 min at 20°C whereupon the reaction was stopped by the addition of 250 μl of 0.3 M NaCl, 0.1 mM EDTA, 25 μg/ml *E. coli* tRNA.

T+C Cleavage Reaction

To 10 μl of 5′ [^{32}P] labelled DNA was added 10 μl of H_2O followed by 35 μl of 95% (v/v) (30 M) hydrazine. The reagents were mixed and left for 6 min at 20°C whereupon the reaction was stopped by the addition of 250 μl of 0.3 M sodium acetate pH 6.0, 0.1 mM EDTA, 25 μg/ml *E. coli* tRNA.

C Cleavage Reaction

To 5 μl of 5′ [^{32}P] labelled DNA was added 10 μl of 5 M NaCl and 30 μl of 95%

(v/v) (30 M) hydrazine. The reagents were mixed and left for 12 min at 20°C. The reaction was then stopped by the addition of 250 μl of 0.3 M sodium acetate pH 6.0, 0.1 mM EDTA, 25 μg/ml *E. coli* tRNA.

T* Cleavage Reaction

To 5 μl of 5′-labelled DNA was added 20 μl of 20 μg/ml (126.5 μM) $KMnO_4$. The reagents were mixed and left for 15 min at 20°C. The reaction was stopped by the addition of 10 μl of 100% (v/v) allyl alcohol and the mixture was dried under vacuum.

The DNA in the G>A, A+G, T+C and C cleavage reactions were precipitated by the addition of 750 μl of ice-cold ethanol and the samples were centrifuged for 15 min at 4°C in an Eppendorf microfuge. The supernatants were discarded and the pellets redissolved in 250 μl of 0.3 M sodium acetate pH 6.0, 0.1 mM EDTA, 25 μg/ml *E. coli* tRNA. The DNA samples were ethanol precipitated again, washed with 1.0 ml of ice-cold ethanol and dried under vacuum. The pellets from all the cleavage reactions were resuspended in 100 μl of 1 M piperidine, heated at 90°C for 15 min then chilled on ice. The samples were transferred to new 1.5 ml tubes and dried under vacuum. The DNA was resuspended in 50 μl of H_2O and dried then resuspended in 20 μl H_2O and dried again. The DNA samples were resuspended in 4 μl of formamide loading solution containing 90% (v/v) formamide, 0.01% (w/v) bromophenol blue, 0.01% (w/v) xylene cyanol, 0.1 mM EDTA and heated at 100°C for 5 min prior to loading onto a sequencing gel.

Sequencing gels containing 20% (w/v) polyacrylamide, 1×TBE buffer, 8 M urea were prepared by mixing 48 g of urea, 10 ml of 10×TBE buffer, 40 ml of 48.25% (w/v) acrylamide, 1.75% (w/v) *N*-*N′*-methylene-bis-acrylamide, H_2O to a volume of 100 ml and were polymerized by the addition of 80 μl of TEMED and 200 μl of 10% (w/v) ammonium persulphate.

The gel mix was poured between 20 cm x 40 cm glass plates separated by 0.5 mm spacers and a comb. The gel was pre-electrophoresed at 40 mA, 1 kV for 30 min in 1×TBE buffer then the denatured sequencing reactions were loaded and the gel was run under the same conditions until the bromophenol blue reached the bottom

of the gel. The plates were separated and the gel was covered with plastic film and autoradiographed at -80°C.

3.2.10 Proteolytic Digestion of Pyruvate Carboxylase

Pyruvate carboxylase was purified from chicken livers according to the method of Goss *et al.* (1979) and was kindly supplied by J. Brazier. The enzyme was digested with the serine protease endoproteinase lys-C which specifically cleaves proteins on the C-terminal side of accessible lysyl residues (Jekel *et al.*, 1983). The 1.0 ml digest included 513 μg of purified PC with a specific activity of 50 units/mg protein, 5 M urea, 20 mM Tris·HCl pH 8.5 and 5 μg of endoproteinase lys-C (specific activity = 30 units/mg protein). The digest was incubated for 2 hr at 20°C whereupon a further 5 μg of endoproteinase lys-C was added and the reaction incubated for a further 18 hr at 20°C.

3.2.11 HPLC Separation of Proteolytic Peptides

Peptides produced by the digestion of chicken liver PC with endoproteinase lys-C were separated using a Hewlett-Packard HP104A HPLC. A 400 μl aliquot of the digest was loaded onto a Brownlee Aquapore RP300 guard cartridge C18 reverse phase column (3 cm $\times$ 0.46 cm internal diameter) and a linear gradient of acetonitrile from 0-55% (v/v) in aqueous 0.1% (v/v) trifluoroacetic acid was applied to the column at 1.0 ml/min. The eluate was monitored at 215 nm and 280 nm and fractions were collected.

3.2.12 Automated Protein Sequencing

Purified peptides were spotted onto either acid-etched glass fibre disks or polybrene coated disks and the disks were loaded into an Applied Biosystems model 470A gas phase protein sequencer (Hunkapillar *et al.*, 1983). The phenylthiohydantoin (PTH) amino acid derivatives resulting from the Edman degradation of protein samples were analysed using an on-line system developed in the Protein Chemistry Unit, Biochemistry Department, University of Adelaide. This system includes an on-line

Brownlee microbore column, a Brownlee MPLC micropump, a Valco electrically actuated HPLC injector and a Waters model 441 UV detector interfaced directly to the sequencer and a Vista model 401 data station. This enabled completely automatic analysis of PTH-amino acids and data collection (M. Snoswell, personal communication).

3.3 Results

3.3.1 Pulse Labelling of Proteins in Chicken Embryos

An investigation of the rate of pyruvate carboxylase synthesis was made by labelling chicken embryos of different ages with a pulse of L-[^{35}S]methionine and examining the extent of labelling of pyruvate carboxylase during this period. Embryos ranging in age from 10–18 days were labelled with L-[^{35}S]methionine as described in Section 3.2.1 and mitochondria were prepared from their livers. Mitochondrial proteins were separated by SDS-polyacrylamide gel electrophoresis and the gel was stained for protein (Section 2.2.12). Proteins labelled by the *in ovo* pulse of L-[^{35}S]methionine were detected by fluorography (Section 3.2.6).

Mitochondrial protein preparations were electrophoresed together with an authentic chicken liver pyruvate carboxylase preparation and pyruvate carboxylase present in the preparations was identified by its co-migration with this marker. The density of staining of the separated mitochondrial proteins with Coomassie brilliant blue R indicated the total amount of each protein present (Fig. 3.1a). The level of the band assigned to pyruvate carboxylase was quite low from day 10 to day 14 although on day 11 there appeared to be a significantly greater amount. From day 16 to day 18 the amounts present in this band were higher, with day 17 having the peak amount.

A guide to the rate of synthesis of pyruvate carboxylase during the period of pulse-labelling is provided by the fluorograph of [^{35}S] labelled proteins shown on the gel in Fig. 3.1b. The position of the stained pyruvate carboxylase shown in Fig. 3.1b is marked on the fluorograph. Very significant amounts of synthesis of the

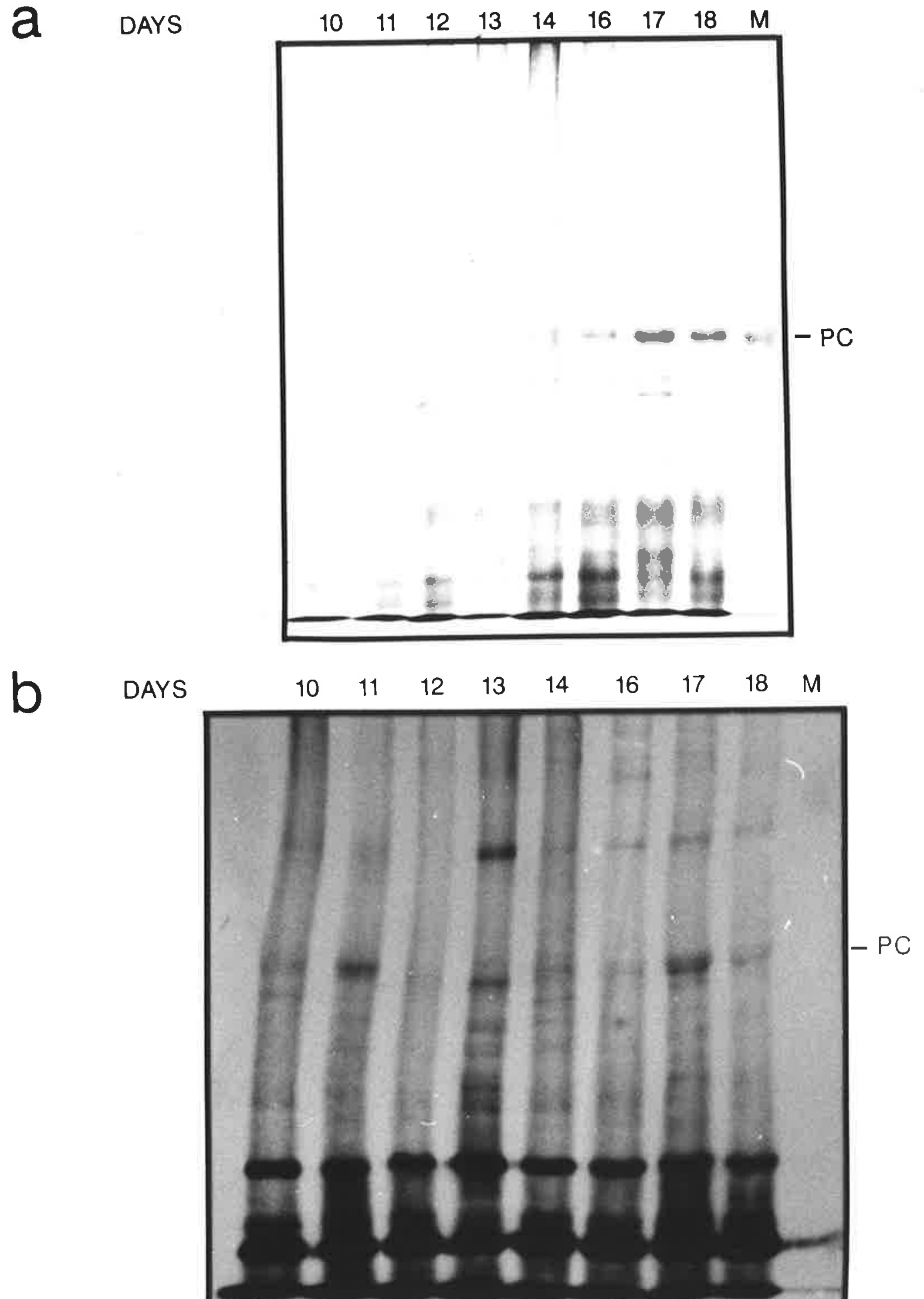
Figure 3.1

Electrophoresis of Embryonic Chicken Liver Proteins Labelled In Ovo with a Pulse of L-[^{35}S]Methionine

Embryonic chickens from 10–18 days old were labelled in ovo with a pulse of 100 μCi of L-[^{35}S]methionine. After 15 min of labelling the embryos were killed and their livers were excised. Mitochondrial proteins were prepared from the livers and the amount of L-[^{35}S]methionine incorporated was determined by liquid scintillation counting. Aliquots of each preparation containing 10,000 cpm of [^{35}S] together with a marker track containing 12 μg of purified chicken liver pyruvate carboxylase were electrophoresed on a discontinuous SDS-polyacrylamide gel (Section 2.2.12) The gel was stained with Coomassie brilliant blue R (Section 2.2.12) then was treated for fluorography (Section 3.2.6). The ages of the embryos used in days are shown above the tracks and the position of the pyruvate carboxylase marker is indicated by (PC).

a Coomassie brilliant blue R stained mitochondrial proteins from embryos of indicated age.

b Fluorograph of L-[^{35}S]methionine labelled mitochondrial proteins from embryos of indicated age.


a
DAYS
10
11
12
13
14
16
17
18
M
PC
b
DAYS
10
11
12
13
14
16
17
18
M
PC

pyruvate carboxylase band occurred on day 11 then decreased on day 12. The rate of synthesis of this band remained low on day 13 however on day 14 it increased again and remained at a relatively constant level through to day 18.

It was apparent on the fluorograph in Fig. 3.1b that two bands were present at approximately the position of the authentic PC marker. The density of protein staining in Fig. 3.1a made resolution of two bands difficult. The lower of the two bands on the fluorograph corresponded more closely to the position of the authentic pyruvate carboxylase marker. Srivastava *et al.* (1983) described a precursor form of pyruvate carboxylase of M_r = 117,000 and processed pyruvate carboxylase of M_r = 115,000 and the size difference between these two bands appears consistent with this size difference although accurate assignment of M_r values was impossible owing to the absence of protein molecular weight markers on this gel. The co-ordinated level of expression of these two bands together with their size indicated that both represented pyruvate carboxylase with the upper band being the unprocessed pyruvate carboxylase precursor described by Srivastava *et al.* (1983).

3.3.2 Preparation of Embryonic Chicken Liver RNA

The preparation of intact, translationally active mRNA requires that all efforts be taken to ensure that endogenous and exogenous agents that would degrade RNA, especially ribonucleases, are inactivated or separated from the RNA. The RNA purification described by Chirgwin *et al.* (1979), which employs the powerful protein denaturant guanidine·HCl, was chosen as the preferred method of RNA preparation. The livers of 17 day old embryonic chickens were used as the source of RNA and generally a yield of 2 mg RNA/g tissue (wet weight) was obtained. RNA was separated into polyadenylated and non-polyadenylated fractions by affinity chromatography using oligo (dT) cellulose. The polyadenylated RNA represented 1–2% of the total RNA in most preparations.

3.3.3 Size-Fractionation of Polyadenylated RNA

Estimates of the subunit molecular weight of chicken liver pyruvate carboxylase vary from M_r = 112,000–125,000 (J. Brazier, personal communication, Barden *et al.*, 1975). Taking 110 to be the average amino acid molecular weight, the number of amino acid residues present in each subunit can be estimated to be from 1000–1200. The mRNA encoding the subunits must therefore contain at least 3000 bases and may be significantly longer than this if both 3′ and 5′ non-coding regions are considered. This makes the pyruvate carboxylase mRNA larger than most mRNA species present and as such, separation on the basis of size was investigated as a means of enrichment of the pyruvate carboxylase mRNA.

Embryonic chicken liver poly(A)$^+$ RNA was centrifuged through a sucrose density gradient and the RNA was eluted. The absorbance of the eluate was monitored at 280 nm and the gradient elution profile is shown in Fig. 3.2a together with the profile of sucrose-density gradient profile of poly(A)$^-$ RNA (Fig. 3.2b). The poly(A)$^+$ RNA was not contaminated to any significant extent with the 18S and 28S non-polyadenylated ribosomal RNA peaks that are evident in the poly(A)$^-$ RNA profile. The poly(A)$^+$ RNA was either collected as fractions or the position of the 18S peak in the poly(A)$^-$ RNA was chosen as a reference point to divide the RNA into two approximately equal fractions i.e. <18S and >18S poly(A)$^+$ RNA. The RNA in these fractions were separately recovered by ethanol precipitation.

3.3.4 *In Vitro* Translation of Embryonic Chicken Liver RNA

Studies to establish the presence of pyruvate carboxylase mRNA in RNA preparations from embryonic chicken liver RNA were undertaken using the rabbit reticulocyte lysate *in vitro* translation system described by Pelham and Jackson (1976). This system was reported by these workers to result in greater amounts of full-length translation products of long mRNAs than the wheat germ extract translation system (Flick *et al.*, 1978).

SDS-polyacrylamide gel electrophoresis of the translation products from 17 day old chicken embryo RNA demonstrated the presence of a translation product which

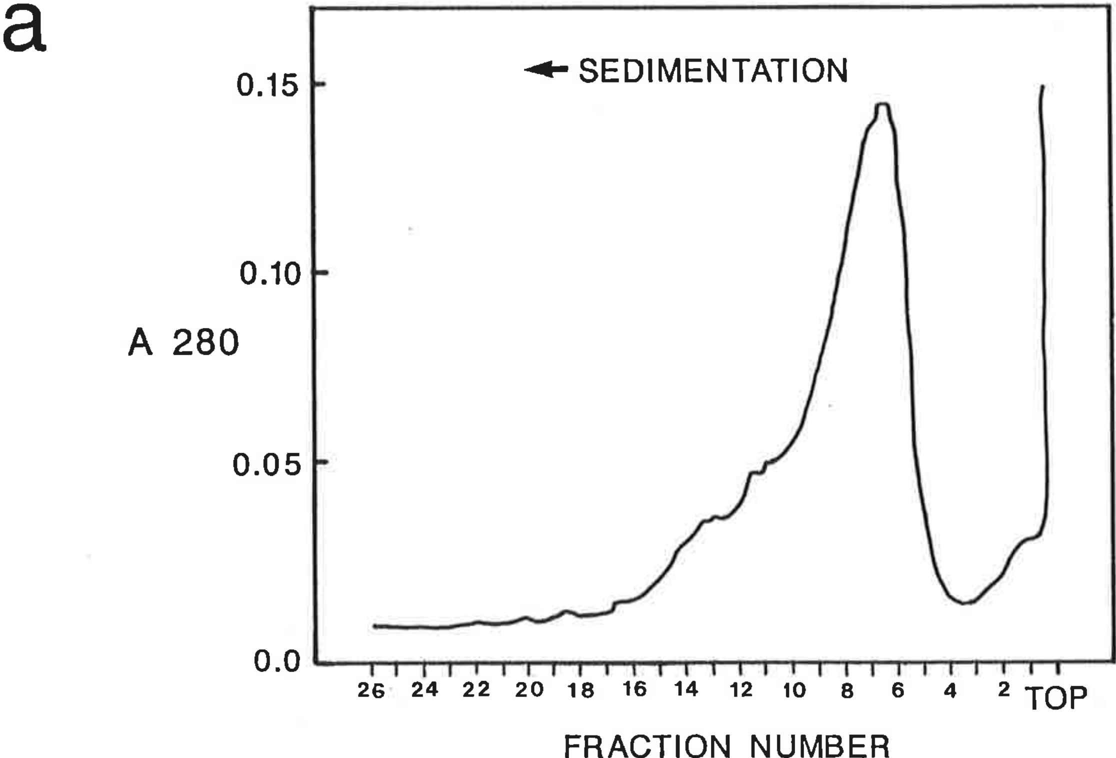
Figure 3.2

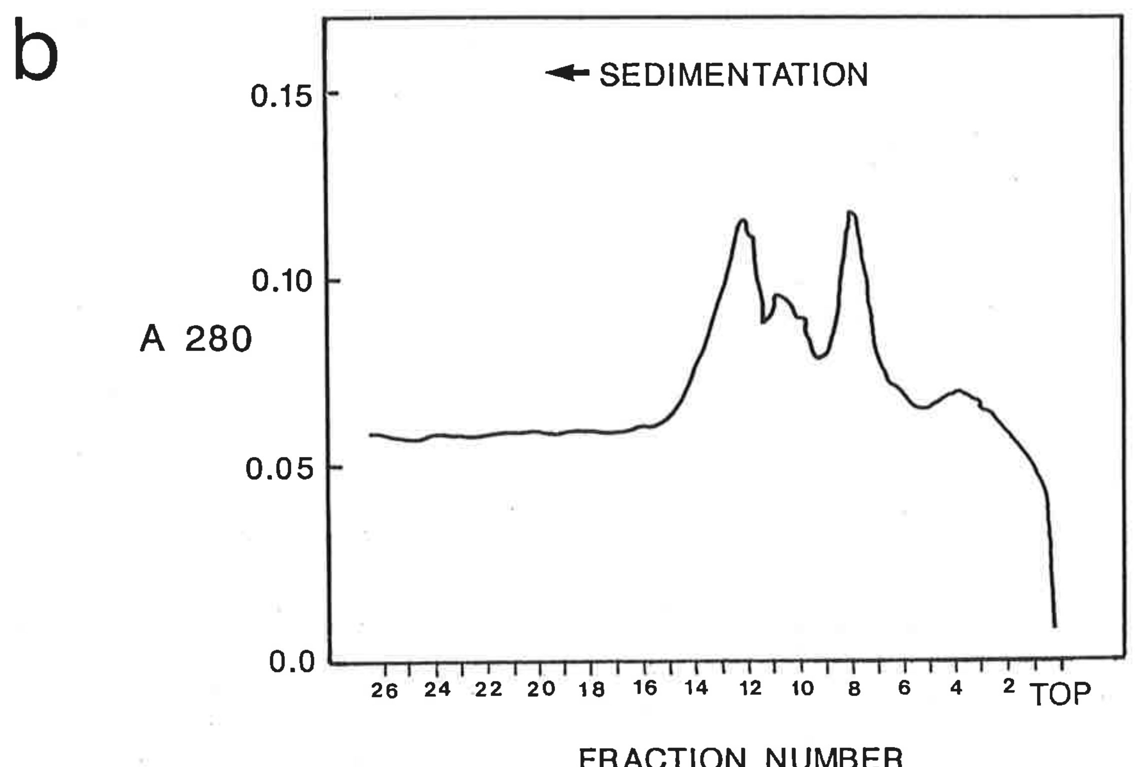
Sucrose Density Gradient Centrifugation of Embryonic Chicken Liver RNA

Aliquots of 100 μg of poly(A)$^+$ RNA and poly(A)$^-$ RNA prepared from 17 day old embryonic chicken liver (Sections 2.2.1 and 2.2.2) were dissolved in 100 μl of 10 mM Tris·HCl pH 7.6, 1 mM EDTA, 0.1% (w/v) SDS. The RNA was denatured by heating at 65°C for 5 min and was then snap cooled on ice. The samples were applied to the top of 10 ml linear gradients from 5–25% (w/v) sucrose in 70% (v/v) formamide, 1 mM Tris·HCl pH 7.6, 1mM EDTA and centrifuged in a Beckman Model L8 ultracentrifuge with a Beckman SW-41 rotor at 270,000$\times g$ for 20 hr at 4°C. The gradients were eluted using an Isco model 640 density gradient fractionator and the absorbance of the eluate was monitored at 280 nm instead of 254 nm due to interference from formamide absorbance at 254 nm, see Section 2.2.3.

a Elution profile of sucrose density gradient centifugation of 17 day old embryonic chicken liver poly(A)$^+$ RNA.

b Elution profile of sucrose density gradient centifugation of 17 day old embryonic chicken liver poly(A)$^-$ RNA.


a
← SEDIMENTATION
0.15
0.10
A 280
0.05
0.0
26
24
22
20
18
16
14
12
10
8
6
4
2
TOP
FRACTION NUMBER


b
← SEDIMENTATION
0.15
0.10
A 280
0.05
0.0
26
24
22
20
18
16
14
12
10
8
6
4
2
TOP
FRACTION NUMBER

co-migrated with an authentic pyruvate carboxylase marker (Fig. 3.3). Translation of this product was dependent upon the presence of exogenous polyadenylated RNA. Compelling evidence for the identification of the M_r = 110,000 translation product as pyruvate carboxylase was provided by immunoprecipitation of the product by anti-chicken liver pyruvate carboxylase polyclonal antibody (Fig. 3.3).

The low level of the putative pyruvate carboxylase translation product was surprising in view of the significant amount of synthesis of pyruvate carboxylase evident in the fluorograph of pulse labelled 17 day old embryonic chicken liver proteins shown in Fig. 3.1b. The possibility that the rabbit reticulocyte translation system was translating long mRNAs inefficiently was examined by translating CMV RNAs. Strain Q of CMV has four RNAs and translation of these RNAs (kindly supplied by K.J. Gordon) yields products of M_r = 105,000, M_r = 120,000, M_r = 34,000 and M_r = 24,500 for RNA1, RNA2, RNA3 and RNA4 respectively (Schwinghamer and Symons, 1977). Significant amounts of the translation products corresponding to these sizes were observed after electrophoresis of the translation products from the rabbit reticulocyte system (Fig. 3.3). This confirmed that the translational system used would adequately translate large mRNAs species.

The immunoprecipitation of embryonic chicken liver poly(A)$^+$ RNA translation products resulted in the precipitation of products of M_r = 65,000, M_r = 42,000 and M_r = 35,000 in addition to the M_r = 110,000 band already described. In fact these bands were substantially more intense than the high molecular weight band. It is probable that these proteins share antigenic determinants with pyruvate carboxylase and may represent either incomplete translation products or related biotin carboxylases such as propionyl-CoA carboxylase which has previously been shown to cross-react with anti-chicken liver pyruvate carboxylase antibodies (K.J. Oliver, personal communication).

3.3.5 Design of B-14 Biotin-Site Oligonucleotide

The amino acid sequence of a tryptic peptide containing the site of attachment of biotin to chicken liver pyruvate carboxylase was determined by Rylatt *et al.* (1977),

Figure 3.3

Electrophoresis of Chicken Liver RNA *In Vitro* Translation Products and Immunoprecipitates

RNA prepared from 17 day old embryonic chicken livers was separated into poly(A)$^+$ RNA and poly(A)$^-$ RNA fractions by affinity chromatography on oligo d(T) cellulose (Section 2.2.2). CMV RNA was kindly supplied by K. Gordon. Translations were carried out with the indicated amount of RNA (Section 3.2.3). Translation products were immunoprecipitated using polyclonal anti-chicken liver pyruvate carboxylase antibody which was purified by protein A affinity chromatography (Section 3.2.4). Protein markers (β-galactosidase M_r = 116,000, phosphorylase b M_r = 94,000, bovine serum albumin M_r = 68,000, ovalbumin M_r = 45,000) and purified chicken liver pyruvate carboxylase were labelled by S-carboxymethylation (Section 3.2.5) and 5000 cpm of each [^{14}C] labelled protein was loaded as size markers. The samples were electrophoresed on an SDS-polyacrylamide gel (Section 2.2.12) and treated for fluorography (Section 3.2.6). The M_r values of the markers are shown and the position of pyruvate carboxylase is indicated by (PC).

Track M : [^{14}C] Labelled S-carboxymethylated protein markers;

Track 1 : Translation control, 0 μg RNA,

Track 2 : Translation control, 0.5 μg CMV RNA,

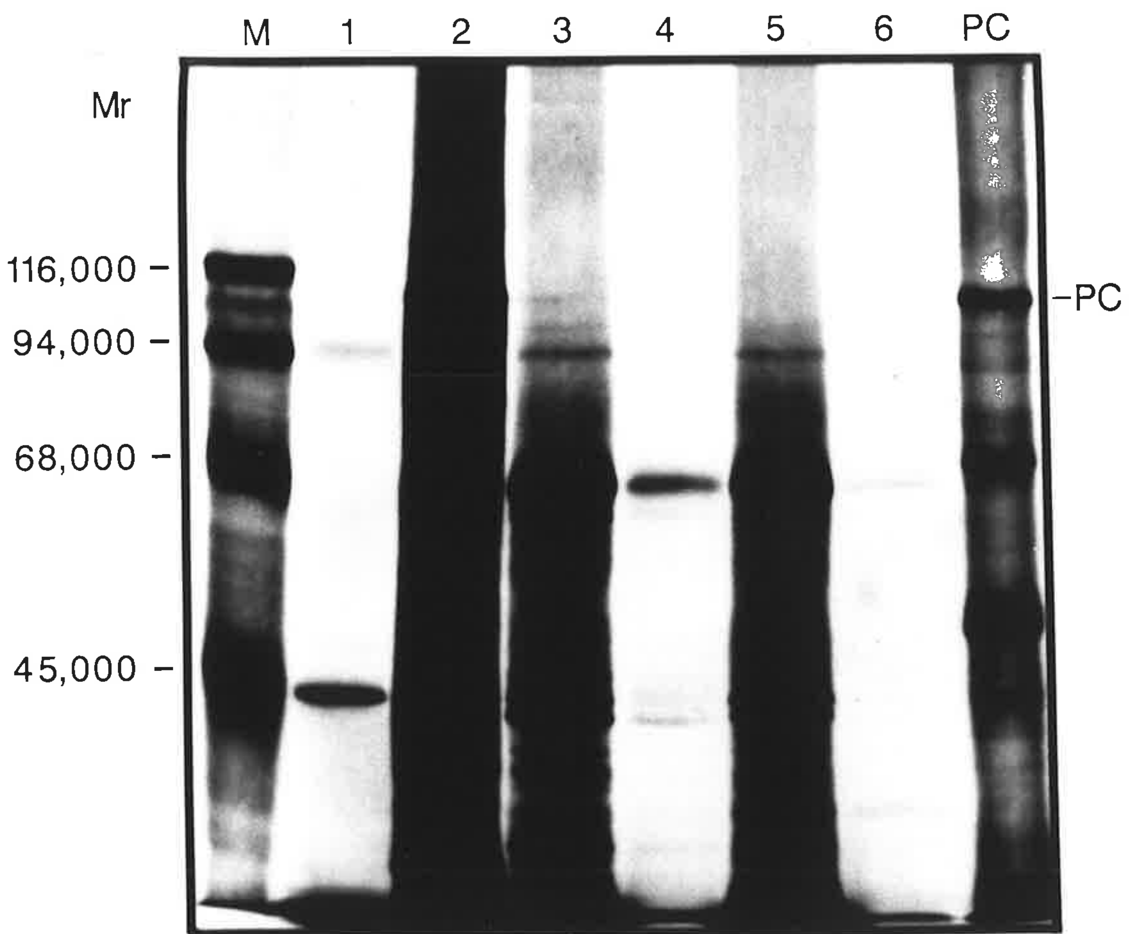
Track 3 : Translation, 1.0 μg 17 day old embryonic chicken liver poly(A)$^+$ RNA,

Track 4 : Immunoprecipitation of translation, 1.0 μg 17 day old embryonic chicken liver poly(A)$^+$ RNA,

Track 5 : Translation, 1.0 μg 17 day old embryonic chicken liver poly(A)$^-$ RNA,

Track 6 : Immunoprecipitation of translation, 1.0 μg 17 day old embryonic chicken liver poly(A)$^-$ RNA,

Track PC : [^{14}C] Labelled S-carboxymethylated chicken liver pyruvate carboxylase.


M
1
2
3
4
5
6
PC
Mr
116,000
94,000
68,000
45,000
PC

see Fig. 3.4. The sequence around the lysine residue to which biotin is covalently attached (biocytin) contains two methionines, an amino acid which is encoded in the mRNA by a non-redundant codon (Fig. 3.4). The knowledge of this sequence was used in conjunction with codon usage data for the chicken proteins to design a synthetic 14 base oligonucleotide. The rather limited codon usage data available for chicken at this time (1982) indicated strong preference in usage for the "AAG" codon for lysine and the "GAG" codon for glutamic acid (Table 3.1). The 14 base oligomer, 5′-GTCTCCATCTTCAT-3′, termed B-14, was synthesized (Section 2.1.9) and incorporated the preferred codons indicated in Table 3.1. B-14 was synthesized for use both as a pyruvate carboxylase specific primer for cDNA syntheses and as a probe for screening clones present in cDNA and genomic libraries.

3.3.6 RNA Dot Blots Probed with B-14 Oligonucleotide

The ability of the biotin attachment site oligomer B-14 to hybridize to pyruvate carboxylase mRNA was determined by using labelled B-14 to probe a dot blot of chicken RNA preparations. Poly(A)$^-$ RNA and poly(A)$^+$ RNA were prepared from 17 day old embryonic chicken liver and the poly(A)$^+$ RNA was fractionated by sucrose density gradient centrifugation into <18S and >18S fractions. Total RNA from whole 4 day old chicken embryos was prepared and kindly supplied by R.P. Harvey. The B-14 oligomer was 5′ [^{32}P] labelled using [γ-^{32}P]ATP and polynucleotide kinase as described in Section 2.2.18.a. The dot blot was hybridized with the B-14 probe at 30°C (Section 2.2.21.a) and washed in 0.5×SSC, 0.1% (w/v) SDS at 20°C then autoradiographed (Fig. 3.5).

The B-14 probe did not hybridize to a significant extent to poly(A)$^-$ RNA nor to total RNA from 4 day old embryonic chickens. However, B-14 did hybridize to both <18S and >18S poly(A)$^+$ RNA with the strongest response from the >18S poly(A)$^+$ RNA fraction. These data describe the hybridization of the B-14 oligomer to a polyadenylated RNA that is >18S by sucrose density gradient centrifugation and is not present to a significant extent in early embryonic development. This description was consistent with the available data for chicken liver pyruvate carboxylase mRNA.

Figure 3.4

Design of B-14 Biotin Site Synthetic Oligonucleotide

The amino acid sequence of the biotin attachment site of chicken liver pyruvate carboxylase (Rylatt *et al.*, 1977) is shown together with its inferred mRNA sequence. The deduced cDNA sequence of a selected 14 base region which contained only two sets of two-base redundancies, one in a lysine codon and the other in a glutamate codon is presented. The preferred codon usage for lysine and glutamate in chickens is "AAG" and "GAG" respectively (Table 3.1) and these codons were selected for incorporation in the cDNA oligonucleotide, termed B-14. The sequence which was chosen for the B-14 oligonucleotide is shown and where redundancies occurred in the mRNA codons the selected base is italicized. The lysine to which biotin is attached is marked by (*).

Protein									
NH_2	....	ala	met	lys*	met	glu	thr	....	COOH
mRNA									
5′	....	GC(U/C/A/G)	AUG	AA(A/G)	AUG	GA(A/G)	AC(U/C/A/G)	....	3′
cDNA									
3′	....		TAC	TT(T/C)	TAC	CT(T/C)	TG	....	5′
B-14									
3′	....		TAC	TT*C*	TAC	CT*C*	TG	....	5′

Table 3.1

Codon Usage for Lysine and Glutamic Acid in from Recombinant DNA Sequence of Chicken Proteins

The preferred codon usage for lysine and glutamate in the chicken protein sequences available at this time of this study are shown. The proteins include α-globin (Richards and Wells, 1980), β-globin (Richards *et al.*, 1979), Histones H2A and H2B (Harvey *et al.*, 1982) and Histone H5 (Krieg *et al.*, 1982).

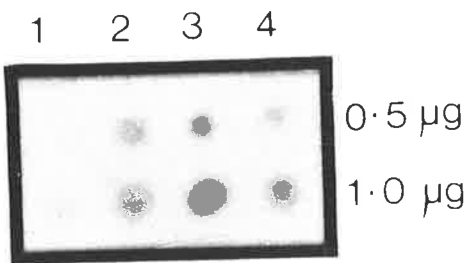
Amino Acid	Codon	α-globin	β-globin	Histone H2A	Histone H2B	Histone H5	Total	%
lys	AAA	2	1	0	0	4	7	5.7
lys	AAG	10	9	14	42	40	115	94.3
glu	GAA	1	4	0	0	1	6	16.7
glu	GAG	5	3	6	14	2	30	83.8

Figure 3.5

Dot Blot of Chicken Liver RNA Probed with B-14 Oligonucleotide

RNA was prepared from the livers of 17 day old embryonic chickens and from total 4 day old embryonic chickens (Section 2.2.1). The 17 day old embryonic chicken liver RNA was separated into poly(A)$^+$ RNA and poly(A)$^-$ RNA fractions by affinity chromatography on oligo d(T) cellulose (Section 2.2.2). The 17 day old embryonic chicken liver poly(A)$^+$ RNA was then separated by sucrose density gradient centrifugation into $<$18S and $>$18S size classes (Section 2.2.3). Aliquots representing 0.5 μg and 1.0 μg of the RNA samples were spotted onto nitrocellulose (Section 3.2.7). The blot was probed with 5′ [^{32}P] labelled B-14 oligonucleotide (Section 2.2.18.a) using the conditions described in Section 2.2.21.a. The filter was washed in 0.5$\times$SSC, 0.1% (w/v) SDS at 20°C and autoradiographed overnight at -80°C with an intensifying screen.

Dot 1 : 17 day old embryonic chicken liver $>$18S poly(A)$^-$ RNA.
Dot 2 : 17 day old embryonic chicken liver $<$18S poly(A)$^+$ RNA,
Dot 3 : 17 day old embryonic chicken liver $>$18S poly(A)$^+$ RNA,
Dot 4 : 4 day old total embryonic chicken RNA.


1
2
3
4
0·5 µg
1·0 µg

3.3.7 Northern Transfer of Chicken Liver RNA Probed with B-14 Oligonucleotide

The sizes of the mRNA species to which the biotin site-specific primer B-14, hybridized were determined by using 5′ [^{32}P] labelled B-14 to probe a Northern transfer of 17 day old embryonic chicken liver RNA. Poly(A)$^+$ RNA and poly(A)$^-$ RNA were separated on a 1.5% (w/v) agarose gel and transferred to nitrocellulose as described in Section 2.2.11.b. The B-14 oligomer was 5′ [^{32}P] labelled (Section 2.2.18.a) and hybridized under the conditions used for short oligomers (Section 2.2.21.a). Due to the low stringency of the washing step, non-specific hybridization occurred with the 18S and 28S ribosomal RNA bands in the poly(A)$^-$ RNA track, see Fig. 3.6. The <18S and >18S poly(A)$^+$ RNA samples contained negligible amounts of these ribosomal RNAs when gels were examined after ethidium bromide staining and therefore any hybridization which occurred was unlikely to be due to this non-specific effect.

In the <18S poly(A)$^+$ RNA track a single band hybridized and this was slightly greater in size than the 18S ribosomal RNA marker. In the >18S poly(A)$^+$ RNA track four bands hybridized; one of the same size as the band detected in the <18S RNA track and two bands between the 18S and 28S marker. These results also demonstrate the effective separation of RNA into size classes by the sucrose density gradient centrifugation.

3.3.8 B-14 Primer Extension Studies

The successful isolation of clones encoding pig gastrin (Noyes *et al.*, 1979), rat insulins I and II (Chan *et al.*, 1979), rat relaxin (Hudson *et al.*, 1981), chicken histone H5 (Krieg *et al.*, 1982) and human histocompatibility antigens HLA-DR α and HLA-DR β (Das *et al.*, 1983) relied on the use of synthetic oligonucleotides to prime the synthesis of cDNAs from selected RNA templates. The extended primer cDNAs were isolated and used as more specific and sensitive probes than the short oligmers themselves in screening cDNA and genomic libraries for cloned complementary sequences.

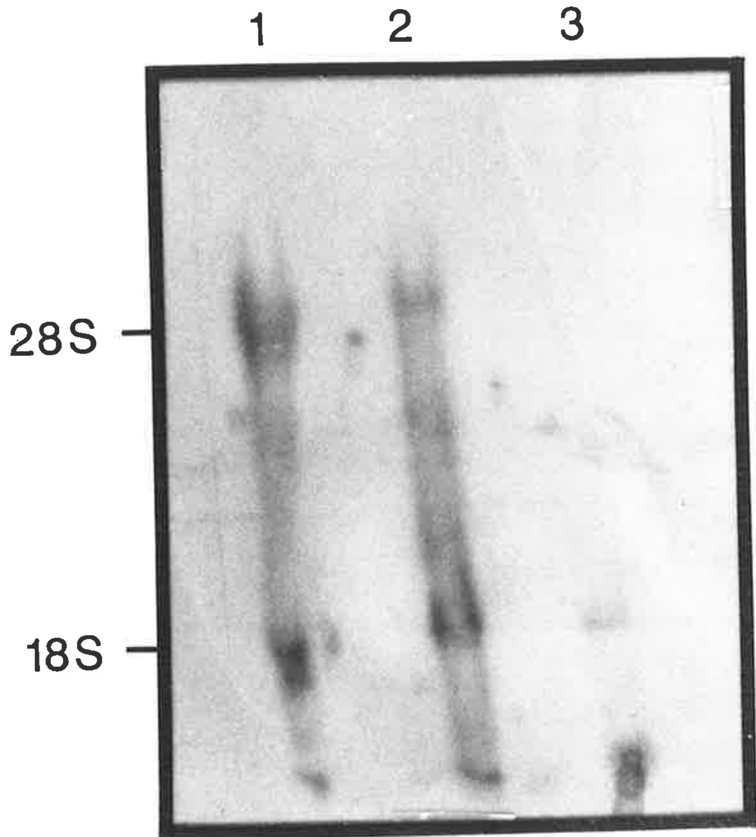
The oligonucleotide B-14, which was complementary to the inferred mRNA se-

Figure 3.6

Northern Transfer of Chicken Liver RNA Probed with B-14 Oligonucleotide

RNA prepared from 17 day old embryonic chicken livers (Section 2.2.1) was separated into poly(A)$^+$ RNA and poly(A)$^-$ RNA fractions by affinity chromatography on oligo d(T) cellulose (Section 2.2.2). The poly(A)$^+$ RNA was separated by sucrose density gradient centrifugation into <18S and >18S size classes (Section 2.2.3). Aliquots of 20 μg of RNA were denatured by glyoxalation and electrophoresed on a 1.0% agarose, 10 mM sodium phosphate pH 6.5 gel for 4.5 hr at 35 mA, 120 V with recirculation of the 10 mM sodium phosphate pH 6.5 running buffer (Section 2.2.9.b). The gel was stained with 30 μg/ml ethidium bromide then transferred to nitrocellulose. The blot was probed with 5′ [^{32}P] labelled B-14 oligonucleotide (Section 2.2.18.a) using the conditions described in Section 2.2.21.a. The filter was washed in 1×SSC, 0.1% SDS at 37°C and autoradiographed for 18 hr at -80°C with an intensifying screen. The positions of the 18S and 28S ribosomal RNA bands were determined from the ethidium bromide stained gel and are indicated.

Track 1 : 20 μg of 17 day old embryonic chicken liver poly(A)$^-$ RNA,
Track 2 : 20 μg of 17 day old embryonic chicken liver >18S poly(A)$^+$ RNA,
Track 3 : 20 μg of 17 day old embryonic chicken liver <18S poly(A)$^+$ RNA.


1
2
3
28S
18S

quence of the biotin attachment site of pyruvate carboxylase, was used in a similar manner to prime cDNA synthesis. The location of the biotin attachment site in chicken liver pyruvate carboxylase was unknown and therefore no prediction could be made for the length of a specifically primed transcript. In addition the presence of secondary structures in the RNA template may result in premature termination of transcription (Efstratiadis and Villa-Komaroff, 1979). This would result in the accumulation of cDNAs of defined length and in the appearance of a band when the primer extension reaction was separated by polyacrylamide gel electrophoresis.

3.3.8.a Dependence of cDNA Synthesis on Primer Concentration

In choosing appropriate conditions for performing B-14 primer extension reactions, the conditions of Krieg *et al.* (1982) were adopted as a starting point. Initially the concentration of the B-14 primer was varied as a molar ratio of the approximate pyruvate carboxylase mRNA concentration. The concentrations of primer to be used were calculated using the following equation (Eqn. 3.1).

$$Ap = \frac{P \times Ar \times Lp \times E}{100 \times Lr}$$

Eqn. 3.1

Where Ap = amount of B-14 primer used in the primer extension (in ng), P = percentage of the total poly(A)$^+$ RNA represented by pyruvate carboxylase mRNA, Ar = amount of poly(A)$^+$ RNA used in the primer extension (in ng), Lp = length of the B-14 primer (in bases), Lr = length of the pyruvate carboxylase mRNA (in bases) and E = molar excess of B-14 primer to the pyruvate carboxylase mRNA. Two unknown values were assumed in calculating the amount of primer to be used, firstly the percentage of the poly(A)$^+$ RNA represented by the pyruvate carboxylase mRNA was taken to be 0.5% and secondly, the size of the pyruvate carboxylase mRNA was taken to be 3300 bases. In these primer extensions the molar excess of B-14 primer to pyruvate carboxylase mRNA was varied through 2-fold, 10-fold, 100-fold and 1000-fold.

At the lowest primer concentration of a 2-fold molar excess, labelled cDNA was present as a "smear" of material from 500 bases downwards and no bands were

visible (see Fig. 3.7). Presumably this "smear" was due to a combination of factors, namely (i) incomplete transcription of the RNA template, (ii) self-priming by the RNA and (iii) random priming by the oligomer. At 100-fold molar excess of primer a band was visible at 75 bases, this band increased in intensity at 1000-fold molar excess and two lower molar weight bands became apparent. A very faint band at approximately 220 bases also appeared in the 1000-fold molar excess reaction. The size of the cDNA bands was determined using the SIZING program (Section 2.2.24). The significance of these bands was not known however at least a 1000-fold molar excess of primer to RNA were used subsequently for B-14 primer extensions.

3.3.8.b Optimization of Primer Extensions for Long cDNA Products

The effect of other factors on the synthesis of extended B-14 primer cDNAs was examined by varying several parameters of the reaction. The absence of primer from the reaction produced a "smear" with no specific bands (Fig. 3.8) similar to the result when a 2-fold molar excess of primer was used. The absence of RNA template resulted in no labelled products as expected (Fig. 3.8). Doubling the concentrations of either the [α-^{32}P] deoxynucleotide triphosphate or the concentration of the unlabelled deoxynucleotide triphosphates both resulted in a significant increase in total cDNA synthesis (Fig. 3.8).

Interestingly, when [α-^{32}P]dATP was substituted for [α-^{32}P]dCTP as the source of [^{32}P] label the 75 base band disappeared. This suggested that the RNA template for the 75 base extension product had a low uridinyl residue content but was comparatively enriched in guanosinyl residues. However, synthesis of the 220 base extension product was unaffected by the choice of either dCTP or dATP as the labelled deoxynucleotide triphosphate.

These data indicate that the 75 base and 220 base cDNA products are synthesized after priming by the B-14 oligomer on different RNA templates. If this is so then the primer would seem to be less specific for the pyruvate carboxylase mRNA than desired.

Figure 3.7

Electrophoresis of cDNA Primed with B-14 Oligonucleotide at Various Ratios of Primer : RNA

Primer extension cDNA syntheses were performed using 1 μg of 17 day old embryonic chicken liver >18S poly$(A)^+$ RNA and the B-14 oligonucleotide as the primer. Both [α-^{32}P]dCTP and [α-^{32}P]dGTP were present in the cDNA syntheses. The amount of B-14 was varied from 46.6 pg to 23.3 ng representing a 2-fold to 100-fold molar excess of B-14 to the pyruvate carboxylase mRNA as calculated using Eqn. 3.1 and the assumptions described in Section 3.3.8.a. The reactions contained 50 mM Tris·HCl pH 8.0, 50 mM KCl, 7 mM $MgCl_2$, 1 mM DTT, 0.5 mM dATP, 0.5 mM TTP and 14 units of AMV reverse transcriptase in a final volume of 20 μl. The cDNA was isolated from the reaction mixture and electrophoresed on a 6% (w/v) polyacrylamide, 50% (v/v)formamide, 1×TBE gel then autoradiographed. The sizes of the marker DNA bands are shown in bases.

Track 1 : 2-fold molar excess of B-14,

Track 2 : 10-fold molar excess of B-14,

Track 3 : 100-fold molar excess of B-14,

Track 4 : 1000-fold molar excess of B-14,

Track M : 5′ [^{32}P] labelled pBR322/*Hin*fI markers.

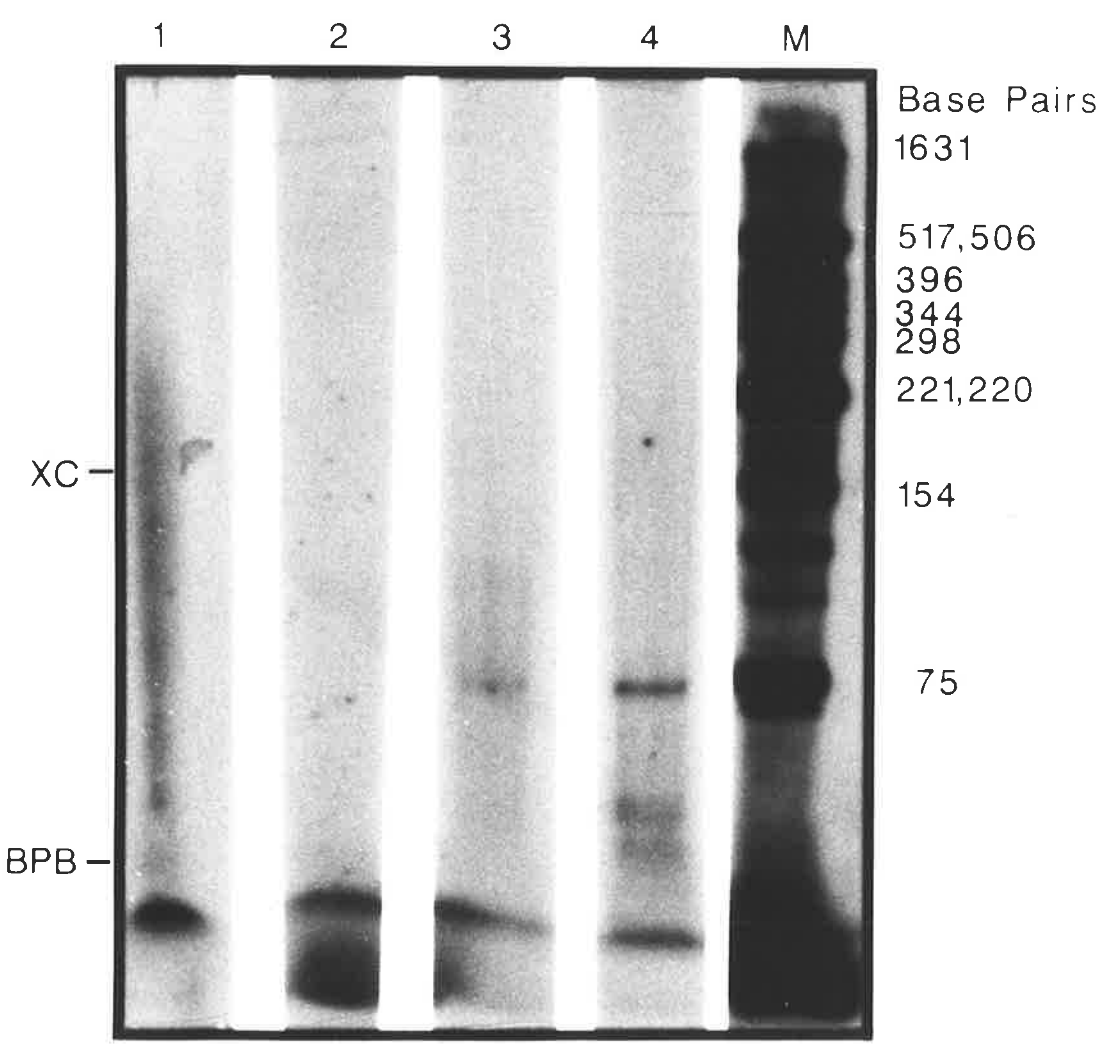

1
2
3
4
M
Base Pairs
1631
517,506
396
344
298
221,220
154
75
XC
BPB

Figure 3.8

Electrophoresis of cDNA Synthesized Under Various Conditions and Primed with B-14 Oligonucleotide at Various Ratios of Primer : RNA

Primer extended cDNA syntheses were performed using 0.5 μg of 17 day old embryonic chicken liver poly(A)$^+$ RNA and the B-14 oligonucleotide. The standard reaction mixture contained 10 μl of [α-^{32}P]dCTP (dried under vacuum), 50 mM Tris·HCl pH 8.0, 50 mM KCl, 7 mM $MgCl_2$, 1 mM DTT and 7 units of AMV reverse transcriptase in a final volume of 20 μl (Section 3.2.8.a). Unless otherwise stated the concentration of the unlabelled deoxynucleotide triphosphates was 0.5 mM and 100 ng of B-14 was present in each reaction. The cDNA was isolated from the reaction mixture and electrophoresed on a 6% polyacrylamide, 8 M urea, 1×TBE gel then autoradiographed. The sizes of the termination products are marked on the figure in bases.

Track 1 : Standard reaction with no B-14 primer,

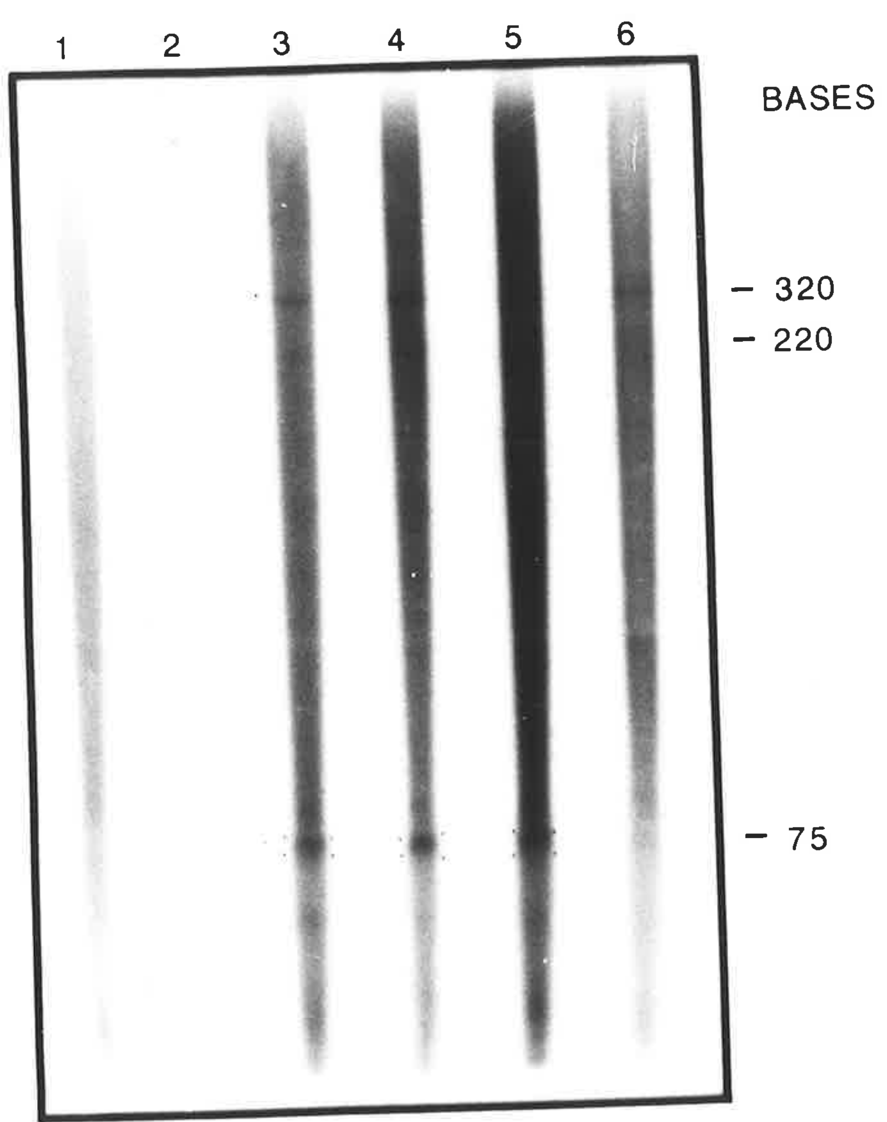
Track 2 : Standard reaction with no RNA,

Track 3 : Standard reaction,

Track 4 : Standard reaction with 2×[α-^{32}P]dCTP,

Track 5 : Standard reaction with 2× unlabelled deoxynucleotide triphosphates,

Track 6 : Standard reaction with [α-^{32}P]dATP.


1
2
3
4
5
6
BASES
320
220
75

3.3.8.c 5′ [^{32}P] Labelled B-14 Primer Extension

The nature of the bands observed in B-14 primer extensions using embryonic chicken liver poly(A)$^+$ RNA was not known and the possibility that they represented a common non-specific priming event rather than specific priming by B-14 could not be excluded. To determine if these bands were specifically primed by B-14 the oligomer was 5′ [^{32}P] labelled with [γ-^{32}P]ATP and polynucleotide kinase (Section 2.2.18.a) and the primer extension was performed as described in Section 3.2.8.b. The result, shown in Fig. 3.9 produced a "smear" of labelled cDNAs with several bands which were determined by the SIZING program (Section 2.2.24) to be 75 bases, 220 bases and 320 bases long. The agreement between the sizes of the bands in this experiment and those previously observed provides strong evidence that these products are specific B-14 primer extensions.

3.3.8.d Primer Extensions using Size-Fractionated RNA

The possibility that the cDNA bands observed in the primer extension reactions shown in Fig. 3.8 resulted from the transcription of different RNA templates was examined by performing primer extensions with RNA from different size classes. Poly(A)$^+$ RNA prepared from 17 day old embryonic chick liver was separated by size using sucrose density gradient centifugation as described in Section 2.2.3. The size classes represented by the RNA fractions were inferred from the analysis of the elution profile of an RNA preparation containing 18S and 28S ribosomal RNA markers which was co-centrifuged with the poly(A)$^+$ RNA. The fractions used in primer extension reactions with the B-14 oligonucleotide were the same as those derived from the sucrose density gradient centrifugation shown in Fig. 3.2. Fraction numbers 9, 10, 11 and 12 were taken and 3 μg of RNA was transcribed into cDNA as described in Section 3.2.8.a. The cDNA products were recovered from the reaction mixture and electrophoresed on an 8% (w/v) polyacrylamide, 8 M urea, 1×TBE gel then autoradiographed (see Fig. 3.10).

Numerous termination products were evident in the cDNA synthesized with fraction no. 9 RNA and the 220 base and 320 base termination products were

Figure 3.9

Electrophoresis of cDNA Primed with 5′ [^{32}P] Labelled B-14 Oligonucleotide

Primer extension cDNA synthesis was performed using 10 μg of 17 day old embryonic chicken liver poly(A)$^+$ RNA with 200 ng of 5′ [^{32}P] labelled B-14 oligonucleotide. The B-14 was labelled using [γ-^{32}P]ATP and polynucleotide kinase (Section 2.2.18.a) and was incorporated in a reaction mixture containing 50 mM Tris·HCl pH 8.0, 50 mM KCl, 7 mM $MgCl_2$, 1 mM DTT 1 mM each of dATP, dCTP, dGTP, TTP and 18 units of AMV reverse transcriptase in a final volume of 50 μl. The cDNA was isolated from the reaction mixture and electrophoresed on a 6% polyacrylamide, 8 M urea, 1×TBE gel then autoradiographed. The positions of the 75, 220 and 320 base primer extension products are marked by (▸).

Track 1 : B-14 Primer extended cDNA,

Track M : 5′ [^{32}P] labelled pBR322/*Hin*fI markers.

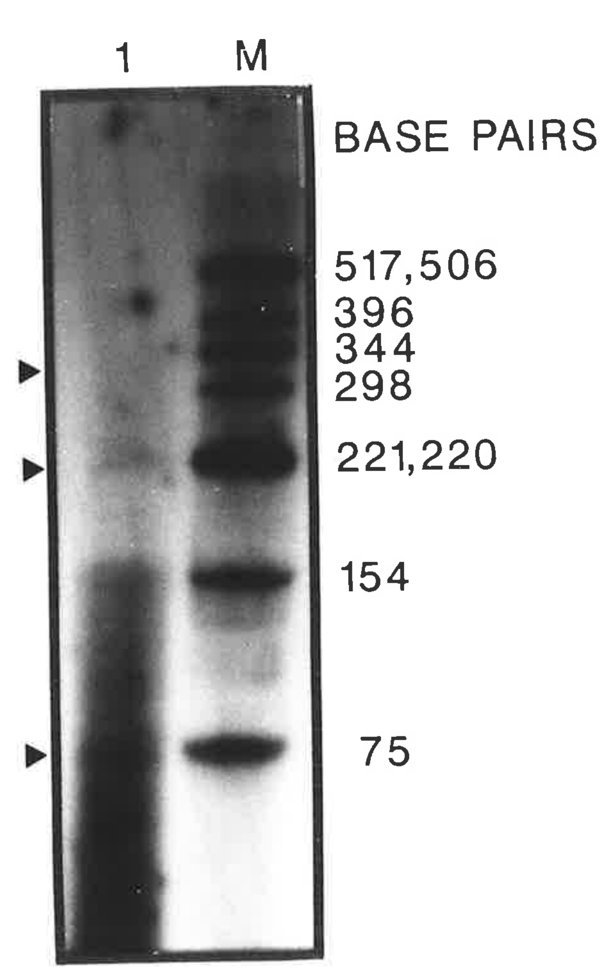
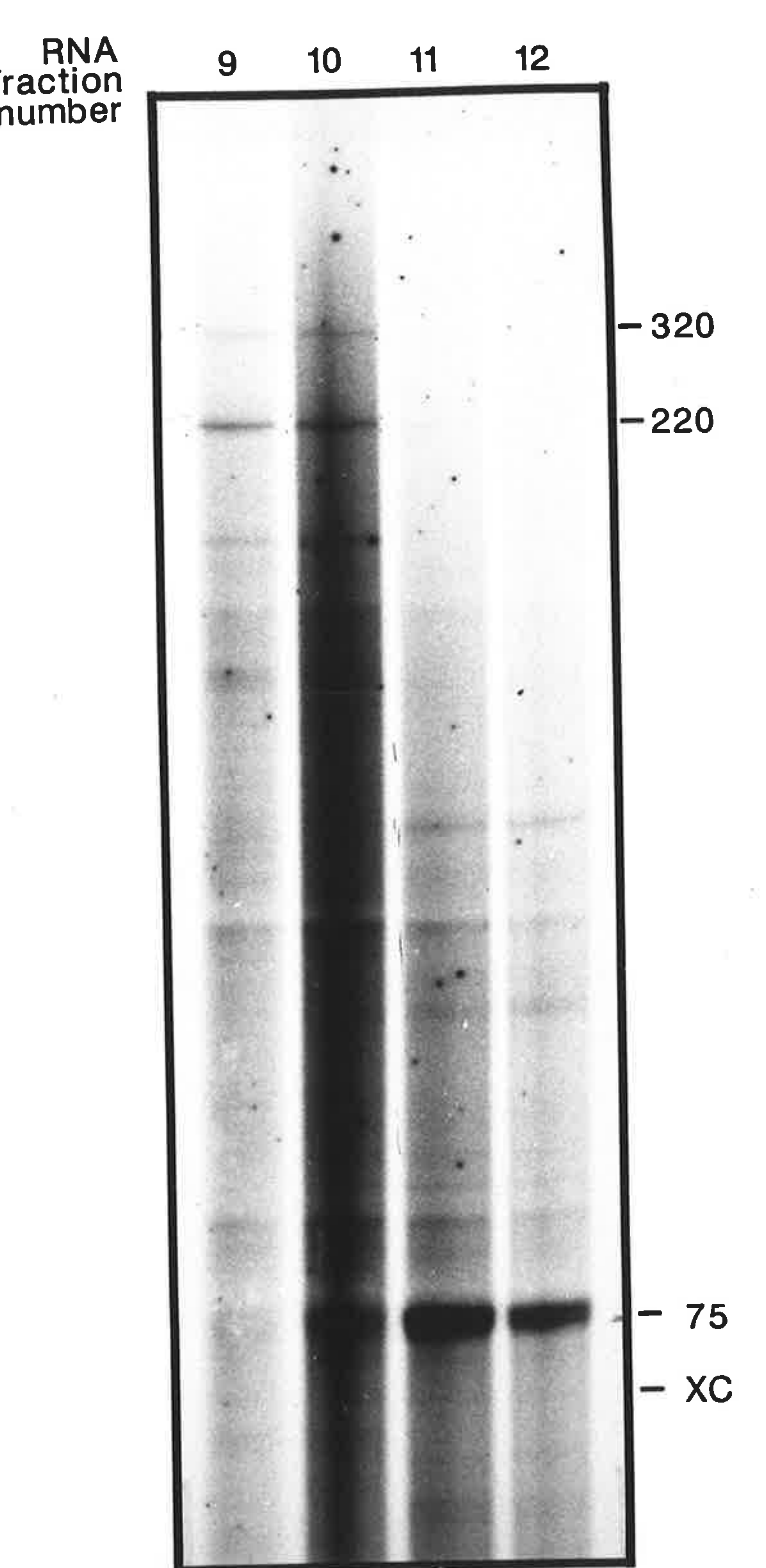

1
M
BASE PAIRS
517,506
396
344
298
221,220
154
75

Figure 3.10

Electrophoresis of cDNA Synthesized from Size Fractionated Poly(A)$^+$ RNA and Primed with B-14 Oligonucleotide

Primer extension cDNA syntheses were performed using 3 μg aliquots of the fractions of 17 day old embryonic chicken liver poly(A)$^+$ RNA which was size fractionated by sucrose density gradient centrifugation (Section 2.2.3). The reaction mixtures contained [α-^{32}P]dCTP, 1 mM dATP, 1 mM dGTP, 1 mM TTP, 200 ng of B-14 oligonucleotide, 50 mM Tris·HCl pH 8.0, 50 mM KCl, 7 mM $MgCl_2$, 1 mM DTT and 7 units of AMV reverse transcriptase in a final volume of 20 μl. The primer extensions were carried out as described in Section 3.2.8.a and the cDNA was isolated, electrophoresed on an 8% (w/v) polyacrylamide, 8 M urea, 1×TBE gel then autoradiographed.

The RNA fractions shown in the figure are the same as those described in Fig. 3.2. Where Fr. no. 9 contains RNA of about 18S, Fr. no. 10 and Fr. no. 11 are intermediate between 18S and 28S and Fr. no. 12 contains RNA of slightly <28S. The sizes of the cDNA termination products are shown on the side of the autoradiograph in bases.


RNA
Fraction
number
9
10
11
12
320
220
75
XC

clearly present but the 75 base product was absent. The RNA used in this primer extension reaction was about 18S. In fraction no. 10, which was slightly >18S in size, the 220 and 320 base cDNA products were still present and the 75 base product appeared. When the size of the mRNA increased to be intermediate between 18S and 28Sin fraction no. 11, the 320 base product disappeared, the 220 base product was diminished and the 75 base product appeared in significant amounts. In the highest molecular weight RNA class tested, that is RNA of slightly <28S, both the 220 base and 320 base products completely disappeared whereas the 75 base product was very intense.

Taken together these results indicated that the higher molecular weight cDNA products of 220 base and 320 bases were transcribed from RNA in the range from 2 kb to 5 kb by inference from the comparison of RNA molecular weights and S values described by McMaster and Carmichael, (1977). The 75 base product results from priming on RNA from approximately 4 kb to >5 kb. This confirms that different RNA templates are being primed on by B-14 as suggested by the difference in the amount of the cDNA products depending upon which deoxynucleotide triphosphate was labelled. The 200 base and 320 base products were synthesized in an RNA size class consistent with the expected size of the pyruvate carboxylase mRNA (3000-3500 bases) whereas the 75 base product was synthesized from a significantly higher RNA size class than that expected for the pyruvate carboxylase mRNA.

3.3.9 Chemical-Cleavage Sequencing of cDNA Products

Unequivocal identification of the B-14 primed cDNA products was obtained by directly sequencing the isolated single-stranded cDNA using the chemical cleavage method described by Maxam and Gilbert (1977). This method requires the DNA to be labelled at the 5′ end prior to sequencing so therefore primer extension reactions were performed where the sole label present was located on the 5′-phosphate of the B-14 primer. The reaction mixture was run on a 6% (w/v) polyacrylamide gel and the 75 base and 220 base cDNA products were located by autoradiography (insufficient amounts of the 320 base cDNA could be obtained for sequencing). The

cDNAs were isolated from gel slices and sequencing was performed as described in Section 3.2.8. The products of the base-specific cleavage reactions were separated on a denaturing 20% (w/v) polyacrylamide gel.

The 75 base cDNA was partially sequenced (Fig. 3.11a) and contained the sequence of the B-14 primer at its 5′ end showing conclusively that it was a specific primer extension. The sequence of the extended DNA (Fig. 3.11b) however was markedly at variance to the sequence required upon translation to yield the known amino acid sequence of the biotin attachment site of chicken liver pyruvate carboxylase (Fig. 3.11c). This established that the 75 base cDNA was not primed from the pyruvate carboxylase mRNA. The sequence of 220 base fragment could not readily be determined but the sequence of the B-14 primer was identifiable at the 5′ end of the sequence. This established that this was also a specific extension of the B-14 sequence. In addition, these results confirmed that the sequence of the B-14 primer supplied by the manufacturer was the same as the sequence requested.

3.3.10 Isolation and Sequencing of Peptides Fragments of Pyruvate Carboxylase

At the time that this investigation was undertaken amino acid sequence data for chicken liver pyruvate carboxylase was restricted to 19 amino acids from a biotin containing tryptic peptide (Fig. 1.5, Rylatt *et al.*, 1977). This small stretch of amino acid sequence represented a very limited data base for the purpose of identifying putative recombinant pyruvate carboxylase clones by comparison of authentic protein sequence to conceptually translated DNA sequence. It was apparent that an increased amount of amino acid sequence for pyruvate carboxylase would increase the likelihood of rapidly finding a matching sequence in a DNA clone and thereby identifying the clone. In addition, greater amounts of pyruvate carboxylase sequence would enable further areas to be searched for regions of low codon redundancy which could be used to design alternative oligonucleotide primers/probes.

Peptides were prepared from purified chicken liver pyruvate carboxylase by firstly cleaving the enzyme with endoproteinase lys-C and secondly by separat-

Figure 3.11

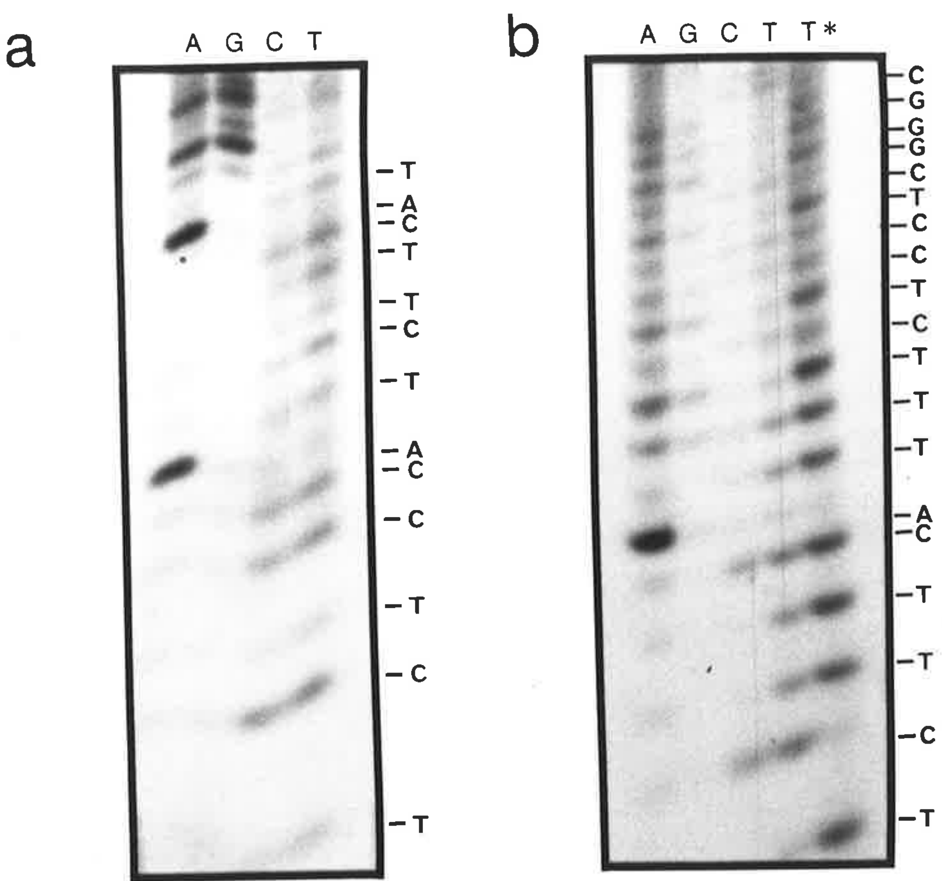
Chemical-Cleavage Sequencing of B-14 Primer Extended cDNA

The B-14 oligonucleotide was 5′ labelled using polynucleotide kinase and [γ-^{32}P]ATP and this was used to prime the synthesis of cDNA from 18 μg of 17 day old embryonic chicken liver >18S poly(A)$^{+}$ RNA. The cDNA was electrophoresed on a 6% (w/v) polyacrylamide gel and the 75 base cDNA band isolated from the gel. This cDNA product was sequenced using the chemical cleavage technique described in Section 3.2.9 and electrophoresed on a 20% (w/v) polyacrylamide, 8 M urea, 1×TBE gel then autoradiographed.

a Autoradiograph of the chemical cleavage sequencing gel of the 75 base cDNA product showing the sequence of the B-14 primer. The sequence read fom the autoradiograph is shown at the side of the gel and unfortunately the initial "G" was not visible having presumably run off the gel.

b Autoradiograph of the chemical cleavage sequencing gel of the 75 base cDNA product showing the sequence of the extended cDNA. The sequence read fom the autoradiograph is shown at the side of the gel

c Comparison of the inferred translation of part of the 75 base cDNA sequence to the amino acid sequence around the biotin attachment site of chicken liver pyruvate carboxylase.



c

5'- (G)TCTCCATCTTCATTTCTCCTCGGGC-3'	Sequence
3'- (C)AGAGGTAGAAGTAAAGAGGAGCCCG-5'	Opposite DNA strand
5'- GCCCGAGGAGAAAUGAAGAUGGAGA(C)-3'	Inferred mRNA
ala- arg-gly -glu-met-lys- met-glu-	Inferred protein

ing peptides present in the digest by HPLC (Section 3.2.11). The absorbance of the column eluate was monitored at both 215 nm and 280 nm and the resulting separation profiles are shown in Figs. 3.12a and 3.12b. Absorbance at 215 nm is characteristic of the peptide bond, tyrosine, tryptophan, phenylalanine, histidine and methionine and as such may be used to monitor for the presence of protein in general. A significant number of peptide peaks are apparent in Fig. 3.12a and they are generally well separated from each other. As might be expected the number of peaks is markedly reduced where the A_{280} was monitored (Fig. 3.12b). Since only the aromatic amino acids, tyrosine and tryptophan absorb strongly at 280 nm only those peptides that contain these amino acids will yield peaks in the A_{280} profile.

The column eluate was collected as fractions and the fraction numbers associated with various peaks are shown in Fig. 3.12a. Peptide fractions were sequenced using the automated Edman degradation and data collection system described in Section 3.2.12. Amino acids which were assigned after examination of the HPLC separation of PTH-amino acid residues derived from each cycle of the sequencer's operation are shown in Fig. 3.13. Although both primary and secondary sequences were obtained from several peptide fractions only the primary sequence which could be unequivocally assigned is shown. The yield of each PTH-amino acid derivative was calculated from the peak height of the derivative by comparison with the peak height of the the appropriate standard PTH-amino acid derivative after separation by HPLC as described in Section 3.2.12. The the formula described in Eqn. 3.2 was used to calculate this yield.

$$Au = \frac{Hu \times As \times F}{Hs \times I}$$

Eqn. 3.2

Where Au = the amount of the unknown PTH-amino acid derivative (in pmol), Hu = the height of the unknown PTH-amino acid derivative peak on the HPLC profile (in mm), As = the amount of the standard PTH-amino acid derivative (in pmol), F = an absorbance attenuation correction factor to bring the standard and unknown profiles to the same scale, Hs = the height of the standard PTH-amino acid derivative peak (in mm), I = the percentage of the cleavage product from each sequencer cycle injected onto the HPLC. The As value was usually 50 pmol and the

Figure 3.12

HPLC Separation of Peptides from the Proteolytic Digestion of Chicken Liver Pyruvate Carboxylase

Purified chicken liver pyruvate carboxylase (513 μg) was digested in 1.0 ml of 5 M urea, 20 mM Tris·HCl pH 8.5 with 5 μg of 30 units/mg endoproteinase lys-C for 2 hr at 20°C and with a further 5 μg of endoproteinase lys-C for 18 hr at 20°C (Section 3.2.10). A 400 μl aliquot of the reaction mixture was loaded onto a Brownlee Aquapore RP300 guard-cartridge C18 reverse phase column attached to a Hewlett-Packard model HP104A HPLC. The column was run at 1.0 ml/min with a linear gradient of acetonitrile in aqueous 0.1% (v/v) trifluoroacetic acid applied from 0% (v/v) at 0 min to 20% (v/v) at 5 min and then from this point to 55% (v/v) at 100 min. The absorbance of the eluate was monitored at 215 nm and 280 nm in milli-absorbance units (mAu) and fractions were collected (Section 3.2.11). The absorbance is shown in the elution profile as a solid line while the gradient of acetonitrile (%B) is shown as a dotted line. The numbers of the fractions associated with peaks at 215 nm are shown.

a Elution profile of the HPLC of a chicken liver pyruvate carboxylase proteolytic digest monitored at 215 nm.

b Elution profile of the HPLC of a chicken liver pyruvate carboxylase proteolytic digest monitored at 280 nm.

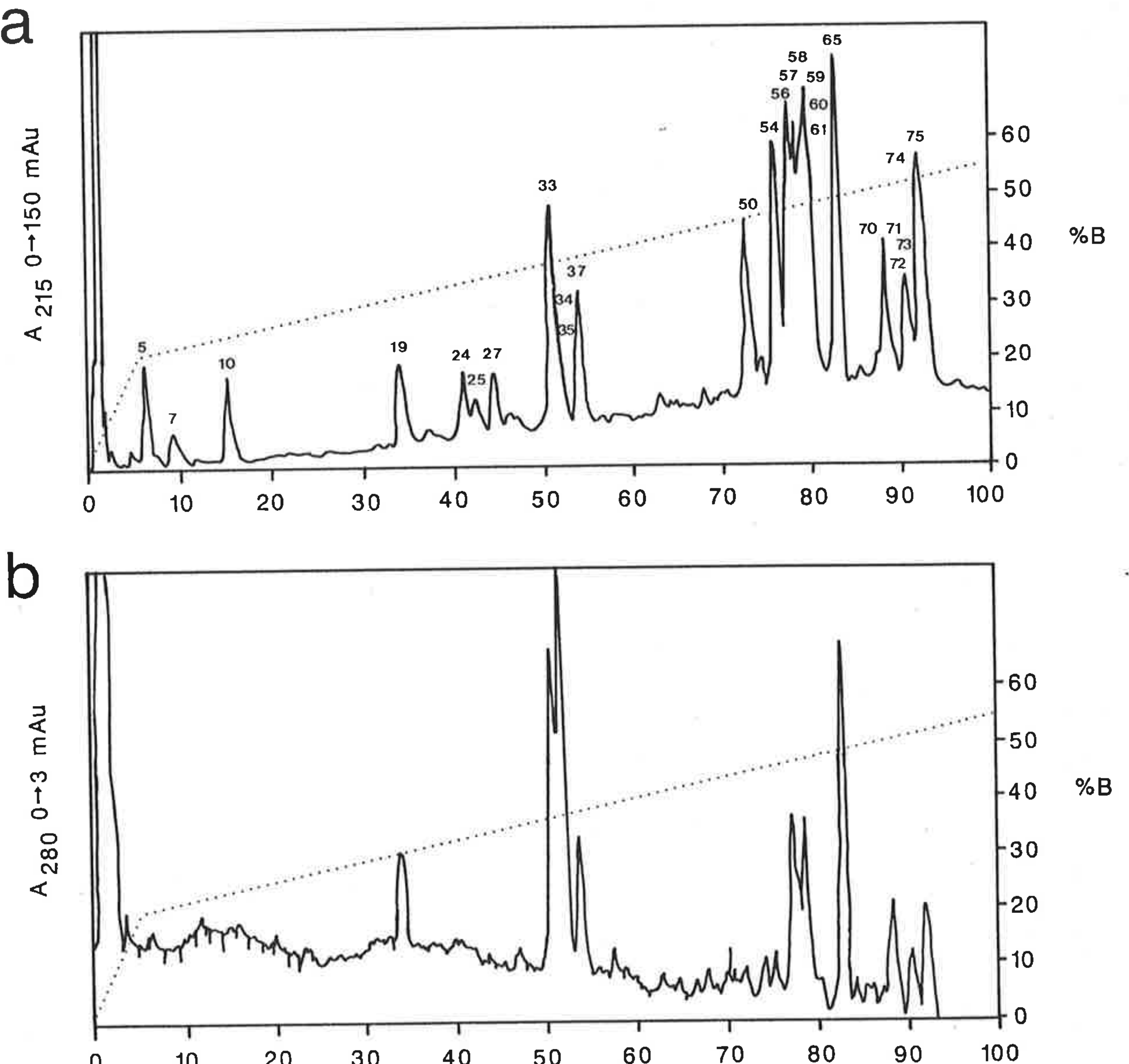
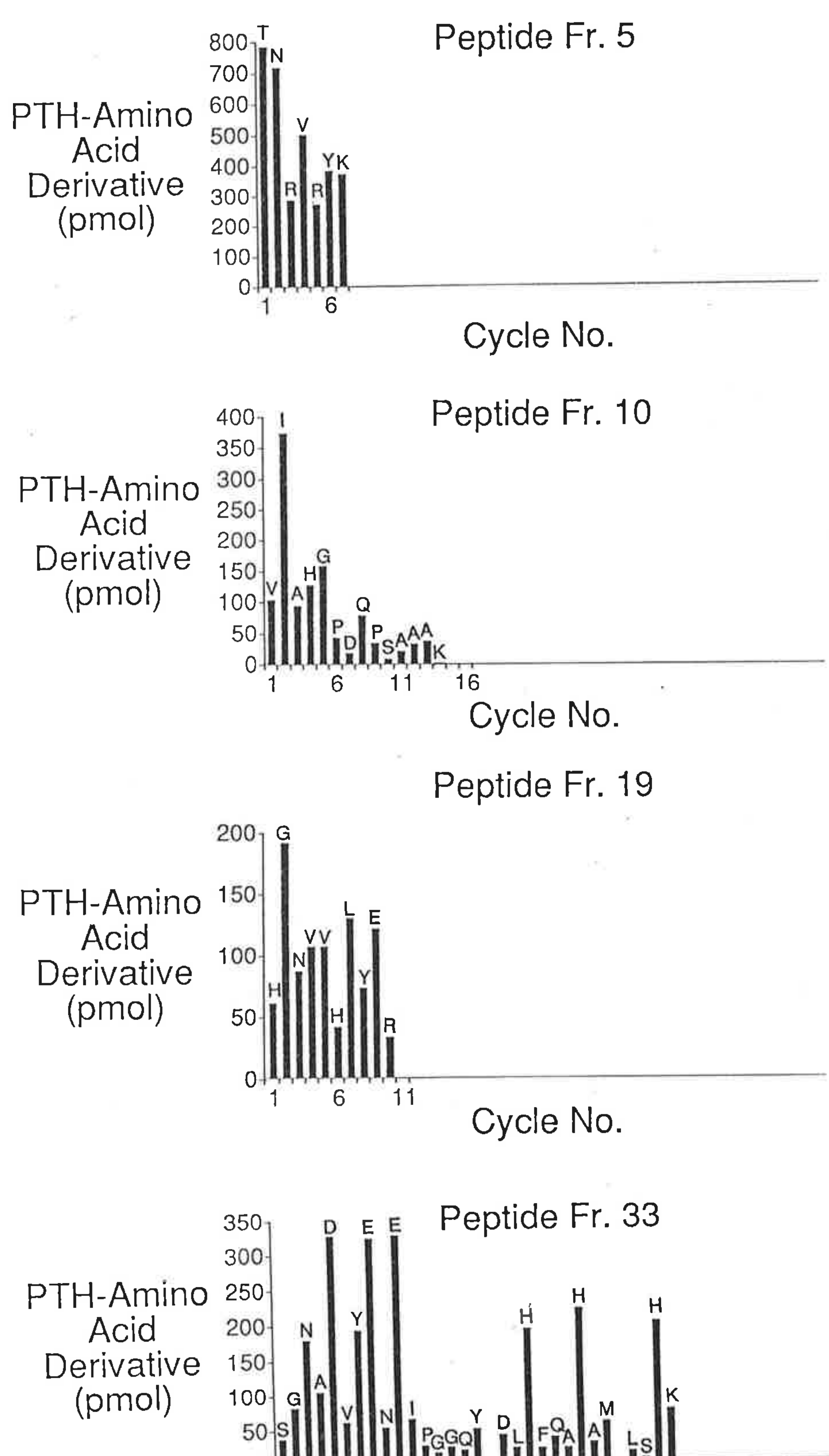

a
A 215 0→150 mAu
%B
5
7
10
19
24
25
27
33
34
35
37
50
54
56
57
58
59
60
61
65
70
71
72
73
74
75
b
A 280 0→3 mAu
%B

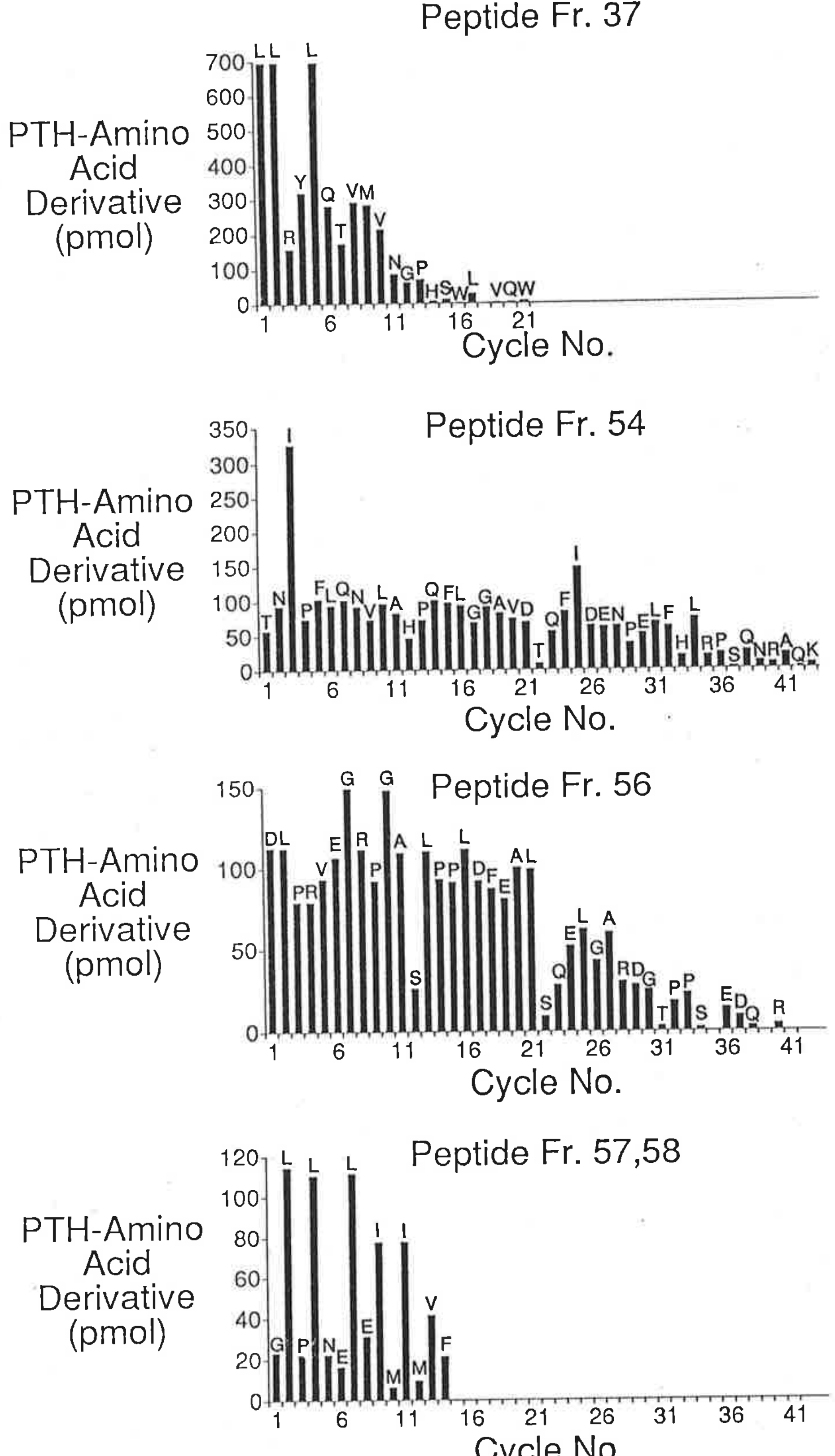
Figure 3.13

Recovery of PTH-Amino Acid Derivatives for the Residue Identified from Each Cycle of Protein Sequencing

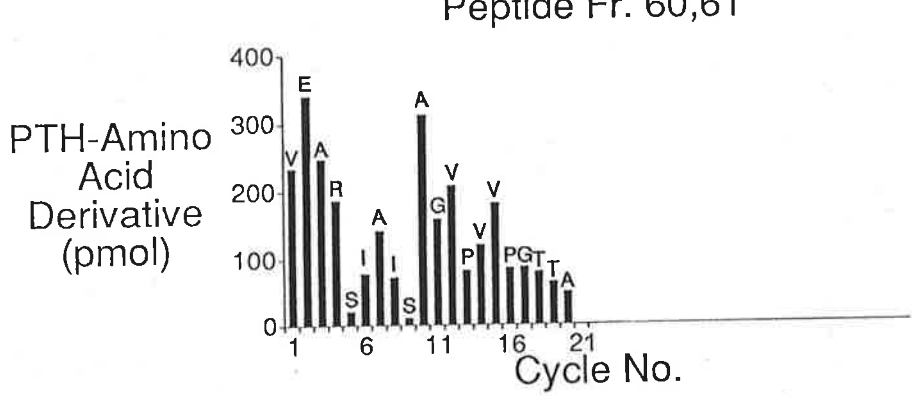
The yield of the PTH-amino acid derivative for the residue identified from each cycle of protein sequencing was calculated as described in Section 3.3.10. The yields are plotted as histograms for the sequence of each peptide isolated from the fractionation shown in Fig. 3.12. The amount of PTH-amino acid derivative is given in pmol and the amino acid which was identified on each cycle is shown above the bar in the single letter amino acid code.



Peptide Fr. 5
PTH-Amino Acid Derivative (pmol)
800
700
600
500
400
300
200
100
0
T N R V R Y K
1
6
Cycle No.
Peptide Fr. 10
PTH-Amino Acid Derivative (pmol)
400
350
300
250
200
150
100
50
0
V I A H G P D Q P S A A A K
1
6
11
16
Cycle No.
Peptide Fr. 19
PTH-Amino Acid Derivative (pmol)
200
150
100
50
0
H G N V V H L Y E R
1
6
11
Cycle No.
Peptide Fr. 33
PTH-Amino Acid Derivative (pmol)
350
300
250
200
150
100
50
0
S G N A D V Y E N E I P G G Q Y D L H F Q A H A M L S H K
1
6
11
16
21
26
31
Cycle No.



Peptide Fr. 37
PTH-Amino Acid Derivative (pmol)
Cycle No.
Peptide Fr. 54
PTH-Amino Acid Derivative (pmol)
Cycle No.
Peptide Fr. 56
PTH-Amino Acid Derivative (pmol)
Cycle No.
Peptide Fr. 57,58
PTH-Amino Acid Derivative (pmol)
Cycle No.


Peptide Fr. 60,61
PTH-Amino Acid Derivative (pmol)
Cycle No.
V E A R S I A I S A G V P V V P G T T A
0
100
200
300
400
1
6
11
16
21

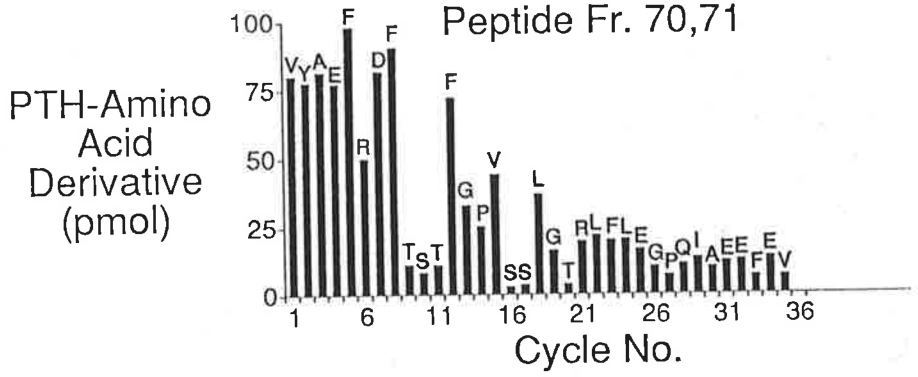

Peptide Fr. 70,71
PTH-Amino Acid Derivative (pmol)
Cycle No.
V Y A E F R D F T S T F G P V S S L G T R L F L E G P Q I A E E F E V
0
25
50
75
100
1
6
11
16
21
26
31
36

Figure 3.14

Protein Sequence of Endoproteinase Lys-C Digestion Fragments of Chicken Liver Pyruvate Carboxylase

The protein containing fractions isolated from the HPLC separation of an endoproteinase lys-C digest of purified chicken liver pyruvate carboxylase were applied to either glass fibre disks or polybrene coated disks. The disks were loaded into an Applied Biosystems 470A gas phase protein sequencer and the resulting PTH-amino acid derivatives were analysed using the on-line system described in Section 3.2.12. The peptide fraction numbers refer to the fractions containing 215 nm absorbing material which were indicated in Fig. 3.12. The amino acids that absorb strongly at 280 nm are underlined and (X) indicates residues which could not be identified.

Peptide Fr. 5 thr asn arg val arg tyr lys

Peptide Fr. 10 val ile ala his gly pro asp gln pro ser ala ala ala lys

Peptide Fr. 19 his gly asn val val his leu tyr glu arg

Peptide Fr. 33 ser gly asn ala asp val tyr glu asn glu ile pro gly gly gln tyr x asp leu his phe gln ala his ala met x leu ser his lys

Peptide Fr. 37 leu leu arg tyr leu gln thr val met val asn gly pro his ser trp leu x val gln trp

Peptide Fr. 54 thr asn ile pro phe leu gln asn val leu ala his pro gln phe leu gly gly ala val asp thr gln phe ile asp glu asn pro glu leu phe his leu arg pro ser gln asn arg ala gln lys

Peptide Fr. 56 asp leu pro arg val glu gly arg pro gly ala ser leu pro pro leu asp phe glu ala leu ser gln glu leu gly ala arg asp gly thr pro pro ser x glu asp gln x arg

Peptide Fr. 57,58 gly leu pro leu asn glu leu glu ile met ile met val phe

Peptide Fr. 60,61 val glu ala arg ser ile ala ile ser ala gly val pro val val pro gly

Peptide Fr. 65 ala asp glu ala tyr leu val gly arg gly leu pro pro val gln ala tyr leu his val pro asp ile cys arg val ala arg glu asn ala val asp ala ile his pro gly tyr gly phe leu ser glu arg ala x phe ala gln

Peptide Fr. 70,71 val tyr ala glu phe arg asp phe thr ser thr phe gly pro val ser ser leu gly thr arg leu phe leu glu gly pro gln ile ala glu glu phe glu val

I value was usually 90%.

The sequences which were determined are listed in Fig. 3.14 and these sequences were assigned to particular peptide fractions in the A_{215} profile and as expected the fractions that have a significant absorbance at 280 nm were found to contain one or more tyrosine residues. Four of the sequences ended with a lysyl residue indicating that the end of the peptide had most likely been reached.

3.4 Discussion

The investigations described in this chapter had several aims, all of which were directed towards the ultimate goal of isolating and identifying recombinant DNA clones encoding pyruvate carboxylase. Briefly stated, the immediate aims were: to isolate optimal amounts of intact pyruvate carboxylase mRNA and to confirm this fact, to produce cDNA products which were specifically primed on the pyruvate carboxylase mRNA and could be used as probes and finally to increase the amount of available pyruvate carboxylase protein sequence to aid identification of pyruvate carboxylase clones.

3.4.1 Pyruvate Carboxylase Synthesis in Embryonic Chicken Liver and Pyruvate Carboxylase mRNA Isolation

3.4.1.a Pulse Labelling Studies

The demonstration of very high levels of pyruvate carboxylase activity in the livers of embryonic chickens by several workers (Arinze and Mistry, 1970, Hendrick and Moller, 1973) indicated that this was a tissue which might be a source of elevated levels of pyruvate carboxylase mRNA. The significant amount of pyruvate carboxylase present in the embryonic chicken liver in the period from 16-18 days was quite evident in the total proteins present in liver mitochondrial extracts. The increasing pyruvate carboxylase present on day 11 and from days 16-18 was reflected in the level of pyruvate carboxylase synthesis as indicated by the incorporation of

L-[^{35}S]methionine into proteins during the 15 min pulse. A total picture of the dynamics of pyruvate carboxylase synthesis and degradation could be obtained by correlation of the presented data with pyruvate carboxylase activity and with the rate of pyruvate carboxylase degradation. However, the data presented for total protein present and for protein synthesized during the pulse of L-[^{35}S] methionine are in agreement with data for enzyme activity presented by other workers. Felicioli *et al.* (1967) found that pyruvate carboxylase activity in the embryonic chicken liver increased from day 13 to a maximum on day 17 followed by a reduction prior to hatching. Arinze and Mistry (1970) in a similar study found a high level of pyruvate carboxylase activity at day 12 which decreased until the 17th day and increased again until hatching whereupon it decreased rapidly. This period correlates well with a period of high gluconeogenesis (Ballard and Oliver, 1963) and since there appears to be strong indirect evidence that pyruvate carboxylase catalysis is a rate-limiting step in gluconeogenesis (for a review see Barritt *et al.*, 1976) it is not surprising that elevated levels of pyruvate carboxylase occur at this time.

In contrast, Hendrick and Moller (1973) in a study of pyruvate carboxylase activity in the embryonic chicken found an essentially stable level of pyruvate carboxylase activity during the period from 11 days to hatching. These authors argued that the peaks of activity observed by Felicioli *et al.* (1967) and Arinze and Mistry (1970) were artefactual owing to variability in the retention of pyruvate carboxylase activity in the extraction media they used. The results presented here do not lend support to the data and explanations of Hendrick and Moller (1973).

It was apparent that unprocessed pyruvate carboxylase precursor may have been present in the pulse-labelled mitochondrial preparation. The presence of this form of pyruvate carboxylase suggested that the mitochondrial preparation may have been contaminated with cytosolic proteins since the pyruvate carboxylase precursor is the cytosolic form of pyruvate carboxylase (Srivastava *et al.*, 1983). However the precursor form is present in an approximately equivalent amount to the processed pyruvate carboxylase in each preparation and such a gross contamination appears unlikely. Alternatively the pyruvate carboxylase precursor present in these preparations may represent the pyruvate carboxylase which had bound to the mitchondrion

and/or had been transported into the mitochondria but had not been processed. Studies of the processing by mitochondrial extracts of the *Neurospora*-F_1F_0 ATPase subunit 9 precursor (Schmid *et al.*, 1984) and the bovine Cytochrome P-450 (SCC) precursor (Ou *et al.*, 1986) indicated that only about 50% of the precursor form was processed in 30 min. Therefore the 15 min incubation period may have been insufficient time for the complete processing of the pyruvate carboxylase precursor prior to chilling in the extraction buffer and probable inactivation of the processing protease.

3.4.1.b *In Vitro* Translation

The ability of the described method for RNA purification and fractionation to yield intact pyruvate carboxylase mRNA was established by identifying pyruvate carboxylase in *in vitro* translations of the RNA. However, several rather unexpected results were observed in these experiments. The amount of the full length pyruvate carboxylase translation product was very small and this was shown not to be a result of the inability of the translation system to yield full length products. The possibility exists that a substantial loss of pyruvate carboxylase mRNA may have occurred during the preparation of the RNA. Cathala *et al.* (1983) have shown that selective losses of high molecular weight RNA may occur with the method of RNA preparation used in this study (Chirgwin *et al.*, 1979). Nevertheless, agarose gel electrophoresis of total embryonic chicken liver RNA showed approximately equivalent amounts of the 18S and 28S ribosomal RNAs indicating that of high molecular weight RNA had not been selectively lost at this stage (data not shown).

An alternative explanation for the low levels of pyruvate carboxylase translation product could be that translation of pyruvate carboxylase mRNA *in vivo* is under the control of additional factors not present in the rabbit reticulocyte lysate. Credence is given to this hypothesis by the extremely low levels of pyruvate carboxylase translation product resulting from the translation of RNA from differentiated mouse fibroblast 3T3-F442A cells which have a 20-fold greater amount of pyruvate carboxylase mRNA than the undifferentiated cells (Angas *et al.*, 1981). At present however no other data is available to support this hypothesis.

The immunoprecipitation of significant amounts of low molecular weight translation products by the anti-chicken liver pyruvate carboxylase antibody was also not expected. This antibody preparation has been demonstrated to bind to chicken liver propionyl-CoA carboxylase (Mottershead *et al.*, 1984) and therefore the possibility that the immunoprecipitated translation products represent propionyl-CoA carboxylase or other biotin carboxylases cannot be excluded. The levels of activity of acetyl-CoA carboxylase and propionyl-CoA carboxylase in the chicken embryo are exceedingly low (Arinze and Mistry, 1970; data for the other biotin carboxylase, 3-methylcrotonyl-CoA carboxylase, was not presented by these authors) therefore it is probable that the mRNA levels of these enzymes are also low and would not be expected to yield bands as intense as those observed. If this is so then the immunoprecipitation of these bands may result from the interaction of the antibody with prematurely terminated pyruvate carboxylase translation products or by a low-level cross-reaction between the antibody and some of the more abundant non-biotin carboxylase translation products. However, the nature of the bands and the cause of their immunoprecipitation was not further investigated.

3.4.2 B-14 Primer Extensions with Embryonic Chicken Liver RNA

The isolation of clones encoding chicken liver pyruvate carboxylase required a DNA probe which could specifically bind to the pyruvate carboxylase mRNA. The biotin-attachment site oligonucleotide B-14 was found to hybridize strongly to an RNA preparation which would be expected to contain the pyruvate carboxylase mRNA. This provided the basis for an examination of the cDNA synthesized when B-14 was used to prime cDNA extension reactions in order to produce a longer, more specific cDNA probe. The synthesis of cDNA using site-specific oligonucleotide primers has resulted in several instances in the appearance of cDNA bands of defined length, when the products were analysed by polyacrylamide gel electrophoresis (Noyes *et al.*, 1982 and Das *et al.*, 1983). These cDNA bands were produced by the termination of cDNA synthesis at specific points on the RNA either prematurely or through

the end of the RNA template having been reached. Specific premature termination may occur when for example secondary structure occurs in the RNA which inhibits the activity of the reverse transcriptase (Efstratiadis and Villa-Komaroff, 1979).

Given that the location of the biotin attachment site was not known for chicken liver pyruvate carboxylase no reasonable prediction could be made for the length of the cDNA products primed by the B-14 oligonucleotide. In the protein sequence of both the *E. coli* acetyl-CoA carboxylase biotin carboxyl carrier subunit (Sutton *et al.*, 1977) and the *P. shermanii* 1.3 SE biotinyl subunit (Maloy *et al.*, 1979) the biocytin residue was 35 amino acids from the C-terminus. This suggested that evolutionary conservation of this position may have occurred and if the biocytin were similarly positioned in chicken liver pyruvate carboxylase then it would be unlikely that significant amounts of a cDNA product representing the 5′ end of the mRNA would be observed. In fact termination products of 75 bases, 220 bases and 320 bases were reproducibly obtained.

A very high molar excess of the B-14 primer to pyruvate carboxylase mRNA was required to produce significant amounts of the cDNA bands, however the value used was consistent with the order of primer molar excess required by Krieg *et al.* (1982) to produce Histone H5-specific cDNA. Actually the assumptions used in the calculation of the molar excess (Eqn. 3.1) were quite conservative estimates and therefore the value of the molar excess factor (E) may have been quite inaccurate.

The presence of three major cDNA species suggested either than B-14 was priming on different RNA species or that a common non-specific priming event was occurring. The absence of any of these products when no B-14 was added to the primer extension and their presence when 5′ [^{32}P] labelled B-14 was used as the sole label confirmed that the three cDNA bands *were* specific B-14 extensions. However, the differential synthesis of the 75 base, 220 base and 320 base bands depending on the concentration of the deoxynucleotide triphosphates supported the proposition that priming was occurring on different RNAs. The possibility of B-14 annealing with the biotin attachment sites sequence of the mRNAs of the other biotin carboxylases present in embryonic chicken liver was not unreasonable given the sequence homology apparent in the enzymes for which sequence data is available (Sutton

et al., 1977, Maloy *et al.*, 1979). In view of low adundancies of these enzymes in embryonic chicken liver as discussed above, primer extension from B-14 annealed to the mRNAs of these enzymes was not expected to be significant.

The analysis of the cDNA products primed on size fractionated RNA and the sequence analysis of the 75 base cDNA established that this product was not primed from the pyruvate carboxylase mRNA. The size class of the RNA on which the 75 base cDNA was synthesized would be expected to contain the acetyl-CoA carboxylase mRNA which is estimated to be >6 kb (Beaty and Lane, 1982). The sequence of chicken liver acetyl-CoA carboxylase, which recently became available (Takai *et al.*, 1987), is however not homologous to the sequence of the 75 base cDNA.

The 220 base and 320 base cDNA products were synthesized at a similar level relative to each other when the parameters of the cDNA synthesis reaction including the size class of the RNA, were varied. This suggested that the RNA templates for each were very similar or identical and that the 320 base cDNA was possibly an extension of the 220 base cDNA. The observation that the RNA size class where both cDNAs were maximally synthesized was expected to contain an enriched level of pyruvate carboxylase mRNA suggested that the pyruvate carboxylase mRNA may have been the template. These 220 base and 320 base products were not investigated further but provided the basis for subsequent studies with chicken genomic and chicken liver cDNA libraries.

3.4.3 Protein Sequencing of Pyruvate Carboxylase

The amount of primary sequence data of chicken liver pyruvate carboxylase at the time this study began was limited to the 19 amino acids around the site of biotin attachment described by Rylatt *et al.* (1977). As a result of this paucity of protein sequence one of the on-going aims of the project was the accumulation of a significant amount of amino acid sequence for chicken liver pyruvate carboxylase. The 277 amino acid residues in 11 peptides presented here represent the first significant amount of sequence data for pyruvate carboxylase from all sources. The total number of amino acids present in these peptide sequences represents ~24% of the

total number of residues present in chicken liver pyruvate carboxylase if the protein is assumed to contain 1100–1200 amino acids. Additional peptide sequence has also been obtained for chicken liver pyruvate carboxylase using the endoproteinase clostripain, which cuts on the C-terminal side of arginyl residues (Keil, 1977), to generate peptides for sequencing (H. Plaggemars and M. Snoswell, personal communication). One peptide overlapped the sequence of endoproteinase lys-C peptide Fr. no. 54 and thereby confirmed the sequence of this peptide.

Chapter 4

Isolation and Identification of Extended Primer Selected cDNA and Genomic Clones

4.1 Introduction

The overall aim of this project was the determination of the primary sequence of pyruvate carboxylase through the isolation and sequencing of recombinant DNA clones. The characteristics of the extended B-14 primer cDNA products described in Chapter 3 appeared to be consistent with their synthesis from pyruvate carboxylase mRNA and therefore it was decided to use these isolated cDNA fragments to screen libraries of chicken recombinant DNA.

The chicken DNA libraries readily available at the time the investigation was initiated were restricted to a chicken genomic library prepared by Dodgson *et al.* (1979). In view of the successful isolation of clones from this library in the past by other workers (Heilig *et al.*, 1980) and in the absence of a chicken liver cDNA library, the genomic library was screened. The option of making a chicken liver cDNA library was not exercised until later (see Chapter 5) after a chicken liver cDNA prepared by Hobbs *et al.* (1986) had been thoroughly studied.

The work presented here describes the isolation and characterization of clones from this chicken genomic library and the subsequent isolation of cDNA clones from an embryonic chicken liver cDNA library. Evidence was found for sequence homology between the cDNA and genomic clones probes using hybridization analysis. The cDNA clones were identified as encoding chicken serum albumin by comparison with the known sequence from the 5′ region of the chicken serum albumin gene. The complete sequence of the coding region of the chicken serum albumin cDNA clone was determined and very strong homology was found between the amino acid sequence inferred from the cDNA sequence and the sequence of serum albumins from other species. Pyruvate carboxylase clones were not isolated with the cDNA probes described in Chapter 3 but analysis of the serum albumin sequence revealed the probable reason for the selection of albumin clones by the B-14 primer extended cDNA probe.

4.2 Specific Methods

4.2.1 Dot Benton Procedure for Bacteriophage Screening

Individual λCharon 4A clones were screened using a modification of the procedure described by Benton and Davis (1977). A suspension of 0.1 ml of an overnight culture of *E. coli* LE392 was prepared in 3 ml of NZCYM medium, 0.7% agarose at 42°C, poured onto a 10 cm diameter NZCYM plate and allowed to set. A grid pattern was marked on the back of the plate and aliquot from 1-4 μl of each λCharon 4A clone stock was spotted in duplicate onto the lawn of *E. coli*. The plate was incubated overnight then transferred to nitrocellulose and treated for binding of the λ DNA as described by Benton and Davis (1977), see Section 2.2.19.

4.3 Results

4.3.1 Preparation of B-14 Primer Extended cDNA Probes

The B-14 primer extended cDNA described in Section 3.3.8 was prepared from 18 μg of 17 day old embryonic chicken liver poly(A)$^+$ RNA using the conditions described in Section 3.3.8.b with [α-^{32}P]dCTP as the label. The cDNA was electrophoresed on an 8% polyacrylamide, 8 M urea, 1$\times$TBE gel (Section 2.2.23) and autoradiographed. Significant amounts of the 75 base, 220 base and 320 base cDNA termination products were formed (Fig. 4.1a) and these three bands were eluted from the gel (Section 2.2.8). The purity of the 75, 220 and 320 base bands was examined by polyacrylamide gel electrophoresis as descibed above and they did not appear to be contaminated by other cDNA to a significant extent (Fig. 4.1b).

4.3.2 Screening of Chicken Genomic Library with Extended Primer Probes

In view of the apparently co-ordinated synthesis of the 200 base and 320 base cDNAs and the lack of knowledge about their RNA templates it was decided to use them

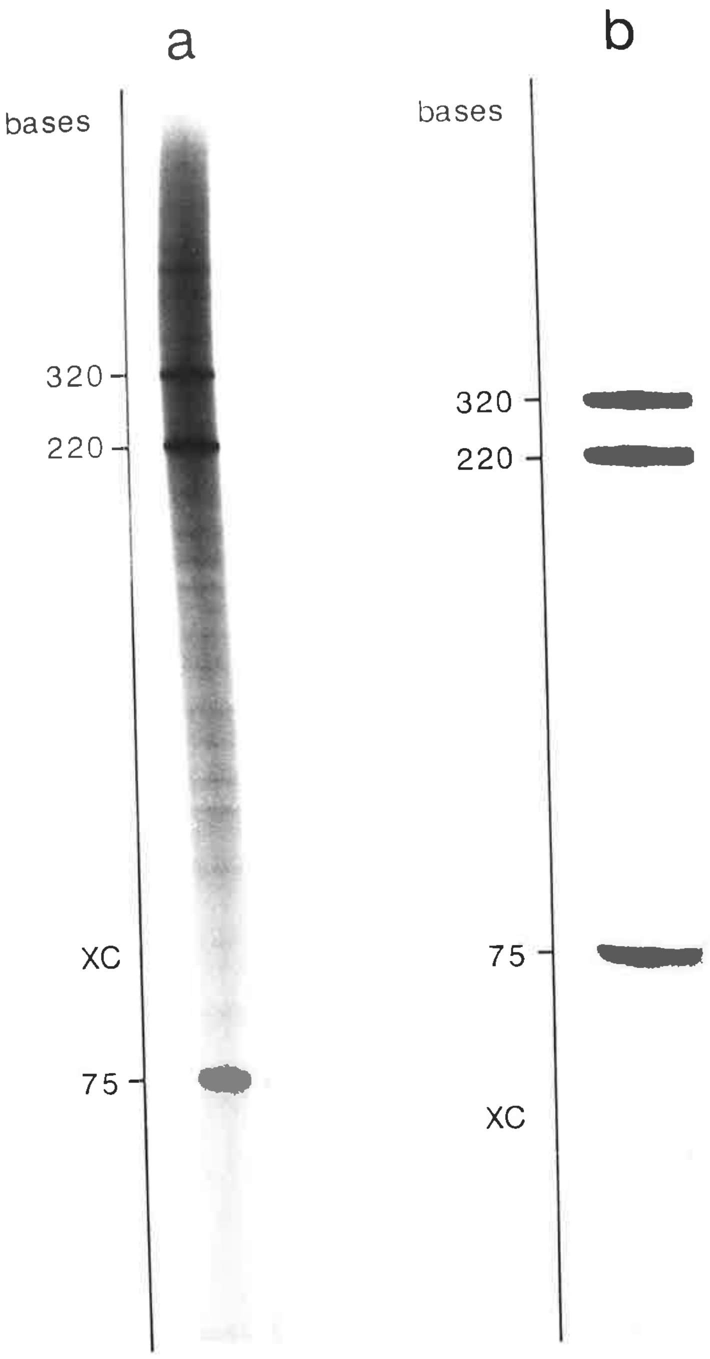
Figure 4.1

Preparative B-14 Primer Extension

The synthesis of B-14 primer extended cDNA from >18S 18 day old embryonic chicken liver poly(A)$^+$ RNA was performed in 6 lots of 20 μl reactions each containing 50 mM Tris·HCl pH 8.0, 50 mM KCl, 7 mM $MgCl_2$, 1 mM DTT, 1 mM each of dATP, dGTP, TTP, 0.4 mM dCTP and 36 units of AMV reverse transcriptase (Section 3.2.8.a). The cDNA was isolated from the reaction mixture and electrophoresed on an 8% polyacrylamide, 8 M urea, 1×TBE gel (Section 2.2.23) then autoradiographed. The 75, 220 and 320 base termination products were located, eluted from the gel slice (Section 2.2.8) and an aliquot of each purified cDNA was taken and re-mixed. This mixed cDNA product aliquot was electrophoresed on an 8% polyacrylamide, 8 M urea, 1×TBE gel and autoradiographed. The sizes of the cDNA products are shown at the side of each autoradiograph in bases.

a Total cDNA synthesized,

b Re-mixed aliquot of 75, 220 and 320 base cDNA products.



a
b
bases
320
220
XC
75
bases
320
220
75
XC

together as a mixed probe. A chicken genomic library in the λCharon 4A vector prepared by Dr. J.D. Engels and colleagues (Dodgson *et al.*, 1979) was screened using this probe mixture. Clones in this library contain chicken genomic DNA inserts of 15–22 kb and since the complexity of the chicken genome is 1.85×10^9 bp (Rosen *et al.*, 1973) at least 1×10^5 clones needed to be screened to be equivalent to one chicken genome if one assumes an average insert length of 19 kb. In fact 3×10^5 pfu were plated, equivalent to three chicken genomes, to ensure adequate representation of all genes and the clones were screened as described in Section 2.2.19. The probability of a particular DNA sequence being present in the clones screened may be calculated using the equation described below (Eqn. 4.1) which re-expresses the equation described by Maniatis *et al.* (1982) in terms of probability,

$$P = 1 - (1 - f)^N$$

Eqn. 4.1

where $P =$ the probability of a given piece of DNA being represented, $f =$ the fractional proportion of the genome in a single recombinant and N is the number of recombinants screened. Using the parameters described above $P = 0.954$ so that there was a 95.4% probability of the pyruvate carboxylase gene being present in the clones screened.

The filters were probed with the [^{32}P] labelled mixed 220 base and 320 base primer-extended cDNAs prepared as described in Section 4.3.1 and hybridized at 37°C (Section 2.2.21.b). The filters were washed in 2×SSC, 0.1% SDS at 30°C then autoradiographed.

The first round of screening yielded 37 duplicate positive signals of varying intensities, rather more than would be anticipated for a single copy gene. There were several possible explanations for the large numbers of clones detected; (i) there are multiple copies of the pyruvate carboxylase gene in the chicken genome, (ii) amplification of the library has produced multiple copies of identical clones, (iii) the probe is detecting a large number of overlapping clones, (iv) the probe is detecting more than one gene. However, since the likelihood of any of these possibilities could not be estimated a representative sample of nine clones were selected from the 37 for isolation including clones with a variety of signal intensities.

Second round screening of these nine clones produced seven sets of duplicate signals, the isolation of the other two clones was not pursued. The signal intensity of these seven clones was equivalent which suggested that the variation previously observed was

Figure 4.2

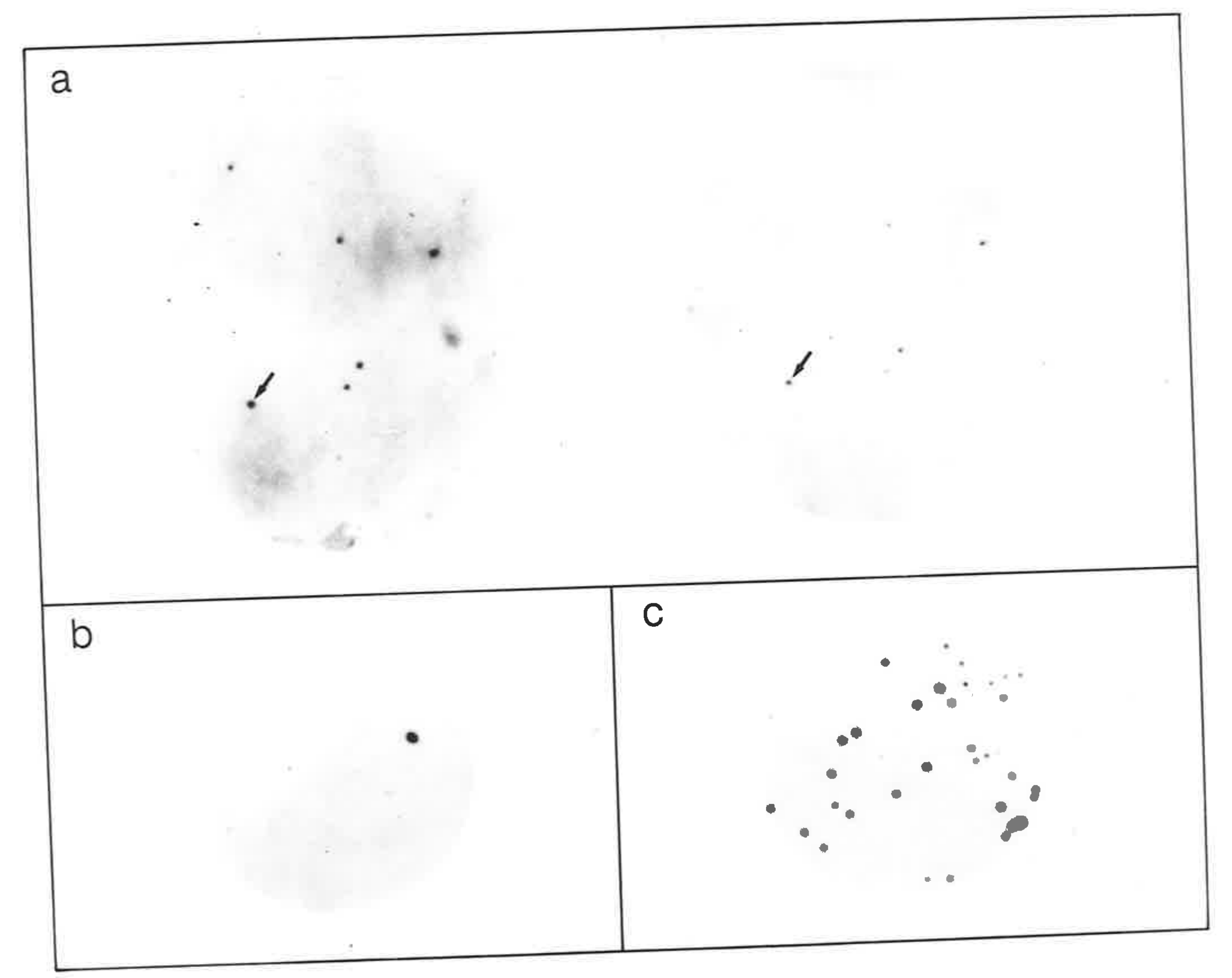
Isolation of the Chicken Genomic Clone λPC3g

A chicken genomic library in the λCharon 4A vector was plated on 15 cm diameter plates at the indicated density and transferred to nitrocellulose as described in Section 2.2.19. The DNA bound onto the filter was hybridized to the [^{32}P] labelled 220 and 320 base extended primer cDNAs (Section 2.2.21.b). The filter was washed in 2$\times$SSC, 0.1% SDS at 30°C then autoradiographed. The second and third rounds of screening were performed using 10 cm diameter plates however the screening procedure was the same as that used for the first round.

a Duplicate filters from the first round library screening. The plate contained 6×10^4 pfu and the duplicate positive plaques representing λPC3g are marked with arrows.

b Filter of second round screening of λPC3g. The plate contained approximately 5×10^2 pfu.

c Filter of third round screening of λPC3g. The plate contained 37 pfu and all clones hybridized to the probe.


a
b
c

artefactual. After a third round of screening, individual isolated plaques could be selected which gave strong hybridization signals and a homogeneous bacteriophage stock was therefore prepared for each clone. The clones were given the following names λPC1f, λPC2b, λPC2d, λPC3f, λPC3g, λPC4g, λPC5d and an example of the screening and isolation of clone λPC3g is shown in Fig. 4.2.

4.3.3 Dot Benton of Chicken Genomic Clones Screened with Different cDNA Probes

The isolated chicken genomic clones were screened by the "dot benton" method (Section 4.2.1) to ensure that they hybridized to the primer extended cDNA and also to establish which of the cDNA probes they hybridized to most strongly. The bacteriophage stocks were spotted onto lawns of *E. coli* LE392 on 4 plates, transferred to nitrocellulose then hybridized to the cDNA probes individually. The B-14 oligonucleotide was 5′ [^{32}P] labelled as described in Section 2.2.18.a and hybridized to the filter using the conditions in Section 2.2.21.a. The B-14 primer extended cDNAs were synthesized and isolated as described in Section 4.3.1 and hybridized to the filters at 37°C (Section 2.2.21.b).

λCharon 4A was used as a negative control for the hybridization signal and failed to hybridize to any of the probes (Fig. 4.3). All the isolated clones hybridized strongly to both the 220 base and 320 base B-14 primed cDNA products which confirmed that the plaques isolated from the third round of screening were those giving positive signals. This result also indicated strong sequence homology both between the two probes and between the selected clones. The clones hybridized very weakly to the 75 base B-14 primed cDNA product which was not unexpected since this probe had previously been demonstrated to be primed on a different mRNA to the 220 base and 320 base cDNAs. Surprisingly, none of the clones hybridized to the B-14 oligonucleotide which was used to prime the synthesis of the cDNA probes. One explanation for this result could have been that hybridization of the cDNA probes was occurring through the extended part of the cDNA alone, therefore the genomic clones may have only partly overlapped the cDNA probes. To establish the location of the region in the inserts of the λ clones which hybridized to the cDNA probes it was firstly necessary to map the restriction sites present in the inserts.

Figure 4.3

Dot Benton of Chicken Genomic Clones Probed with B-14 and Extended Primer cDNAs

A dot Benton analysis of λ clones was performed as described in Section 4.2.1. Bacteriophage stocks were spotted in duplicate onto a lawn of *E. coli* LE392 cells on four sets of plates then incubated at 37°C overnight to allow plaque formation. The bacteriophage DNA was transferred to nitrocellulose which were hybridized at 37°C to the indicated probes (Section 2.2.21.b), washed in 2×SSC, 0.1% SDS at 30°C and autoradiographed.

Dot 1 : λCH01, a Histone H1 clone as a negative control,

Dot 2 : λPC4g,

Dot 3 : λPC3f,

Dot 4 : λPC3g,

Dot 5 : λPC5d,

Dot 6 : λPC1F,

Dot 7 : λPC2d,

Dot 8 : λPC2b

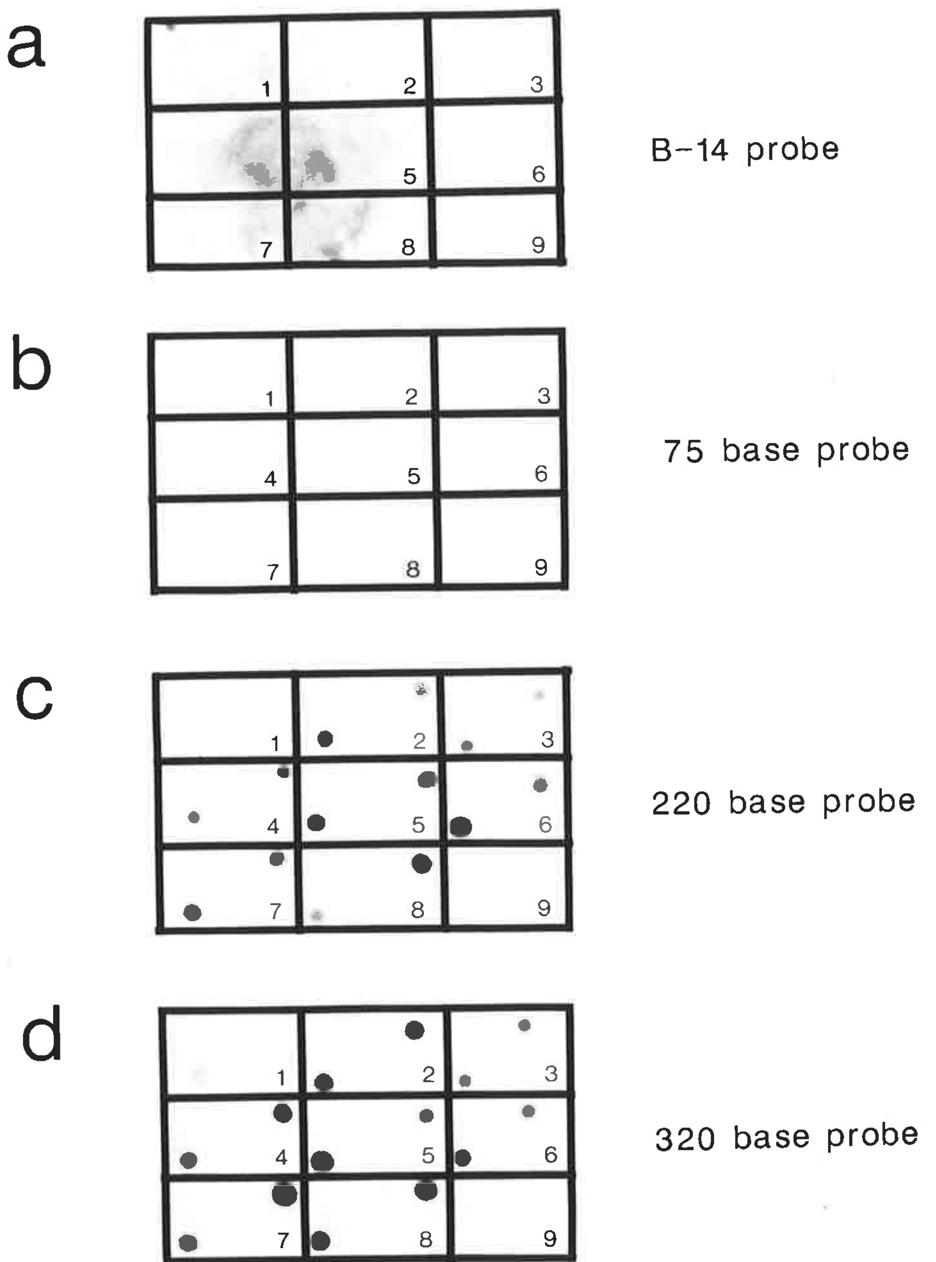
Dot 9 : an unknown negative clone from the second round screening as a negative control.

a Dot benton probed with 5′[^{32}P] labelled B-14,

b Dot benton probed with 75 base primer extended cDNA,

c Dot benton probed with 220 base primer extended cDNA,

d Dot benton probed with 320 base primer extended cDNA.


a
1
2
3
5
6
7
8
9
B-14 probe
b
1
2
3
4
5
6
7
8
9
75 base probe
c
1
2
3
4
5
6
7
8
9
220 base probe
d
1
2
3
4
5
6
7
8
9
320 base probe

4.3.4 Restriction Mapping of λPC3g Chicken Genomic Clone

DNA was prepared for each clone using the method described in Section 2.2.14.f and the genomic DNA inserts were excised with *Eco*RI. The restriction fragments were separated by agarose gel electrophoresis the resulting pattern of fragments indicated that all 7 clones had identical inserts (data not shown). Multiple copies of the same clone such as this may occur in a library by selectively enhanced replication of a clone during amplification of the library. In view of this, only one clone, λPC3g, was selected for further investigation.

λPC3g DNA was digested with several restriction enzymes including, *Eco*RI, *Bam*HI, *Hin*dIII and *Sma*I both singly and in combination, in order to generate a map of the restriction sites present in the insert. The digested DNA was recovered from the reaction mixtures after phenol extraction and ethanol precipitation and was electrophoresed on a 0.7% (w/v) agarose, 1×TAE gel (Section 2.2.9.a). The gel was stained with 0.3 μg/ml ethidium bromide and photographed under UV illumination. An example of an ethidium bromide stained gel of these DNA digests is shown in Fig. 4.4a. The size of the restriction fragments was determined using the SIZING program (Section 2.2.24). After additional combinations of restriction digests were performed a restriction map of λPC3g was constructed (Fig. 4.5) and the total length of the insert was found to be 17.7 kb.

4.3.5 Localization of Site of Hybridization of Extended Primer Probes to λPC3g

The site of hybridization of the cDNA probes to the λPC3g chicken genomic DNA insert was studied by hybridizing the cDNA probes to restriction digests of λPC3g DNA after transfer to nitrocellulose. The ethidium bromide stained gel shown in Fig. 4.4a was transferred to nitrocellulose (Section 2.2.11.a) and the Southern transfer was probed with a mixture of the labelled 220 base and 320 base primer extended cDNAs prepared as described in Section 4.3.1. The filter was hybridized

Figure 4.4

Restriction Enzyme Analysis and Southern Transfer of λPC3g DNA

a Restriction enzyme analysis of λPC3g DNA. Aliquots of 3.0 μg of λPC3g DNA were digested with 10 units of the indicated restriction enzyme in single digests and in double digests by *Eco*RI followed by redigestion with 10 units of the second enzyme. The fragments were recovered after phenol extraction and ethanol precipitation and electrophoresed on a 0.7% agarose, 1×TAE gel (Section 2.2.9.a). The gel was stained with 0.3 μg/ml ethidium bromide and photographed under UV illumination.

Track 1 : λPC3g/*Eco*RI/*Bam*HI,
Track 2 : λPC3g/*Eco*RI/*Sma*I,
Track 3 : λPC3g/*Eco*RI/*Hin*dIII,
Track 4 : λPC3g/*Bam*HI,
Track 5 : λPC3g/*Sma*I,
Track 6 : λPC3g/*Hin*dIII,
Track 7 : λPC3g/*Eco*RI,
Track M1 : λ_{wt}/*Eco*RI/*Hin*dIII size marker,
Track M2 : λ_{wt}/*Hin*dIII size marker.
Track M3 : λ_{wt}/*Eco*RI size marker,

b Southern transfer of λPC3g restriction digests probed with primer extended DNA. The gel shown in Fig. 4.4a was transferred onto nitrocellulose (Section 2.2.11.a) and hybridized (Section 2.2.21.b) with [^{32}P] labelled 220 base and 340 base cDNA products. The filter was washed in 2×SSC, 0.1% SDS at 30°C and autoradiographed. The size of the markers are indicated at the side of the autoradiograph in kb.

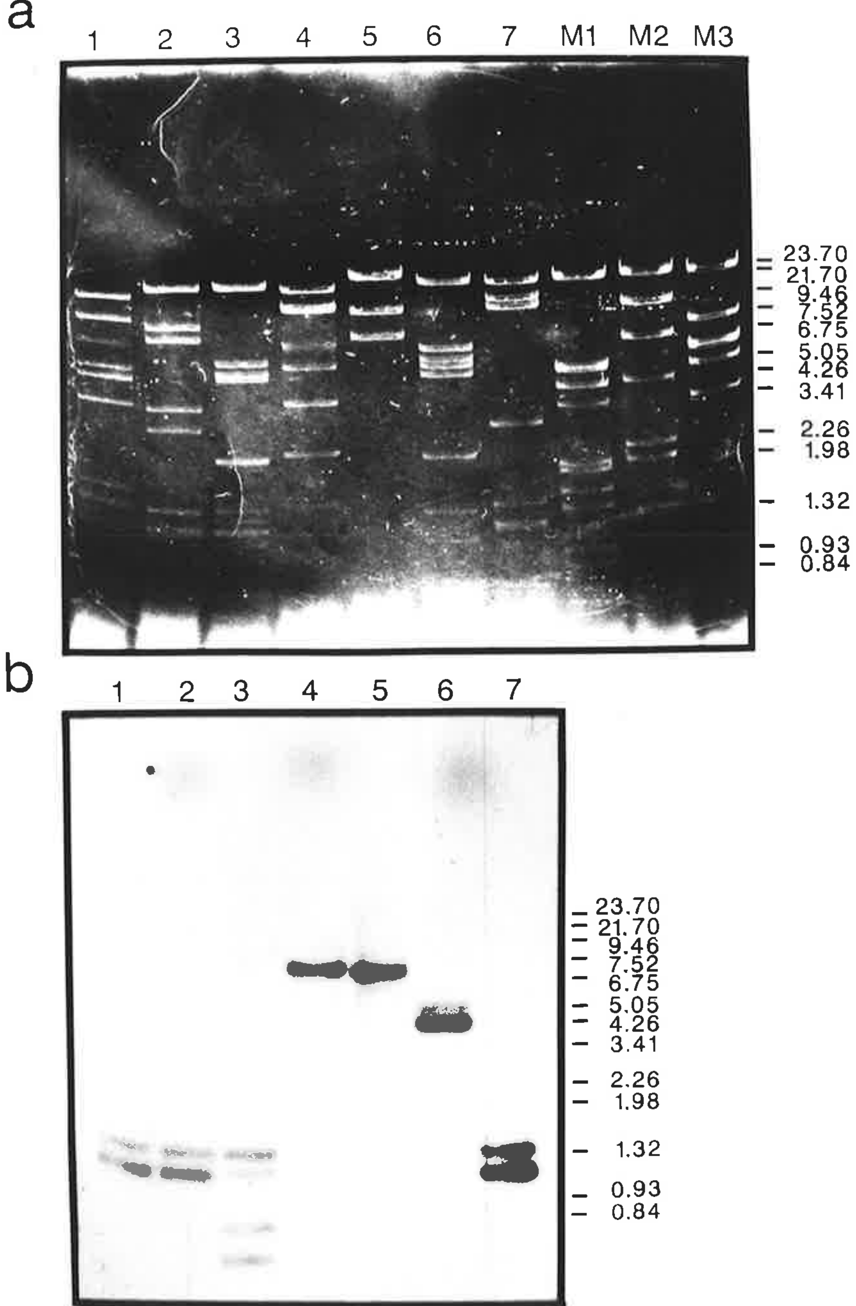
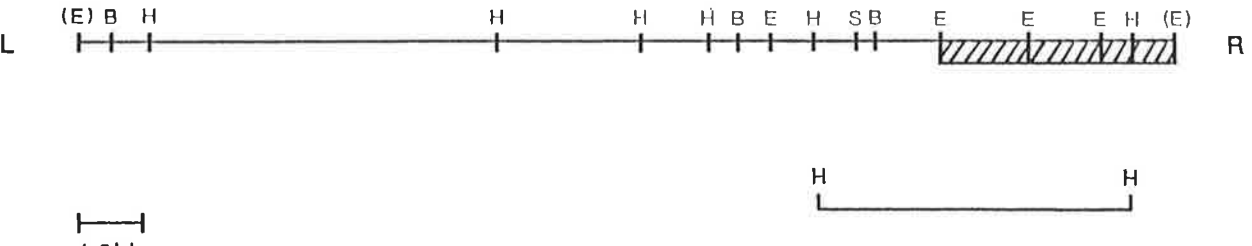

a
1
2
3
4
5
6
7
M1
M2
M3
23.70
21.70
9.46
7.52
6.75
5.05
4.26
3.41
2.26
1.98
1.32
0.93
0.84
b
1
2
3
4
5
6
7
23.70
21.70
9.46
7.52
6.75
5.05
4.26
3.41
2.26
1.98
1.32
0.93
0.84

Figure 4.5

Restriction Map of λPC3g Insert DNA

Restriction enzyme sites were located in the λPC3g insert using the information in Figs. 4.4a and 4.4b, together with additional double digests. The length of the insert was 17.7 kb and the fragments which hybridized to the 220 base and 340 base cDNA products are indicated by the boxed crosshatching. The extent and location of the 5.3 kb *Hin*dIII fragment which was used as a probe is marked by the line underneath the map. E = *Eco*RI, B = *Bam*HI, H = *Hin*dIII, S = *Sma*I and L and R refer to the left and right arms of λCharon 4A respectively.



L
(E) B H
H
H
H B E H S B
E
E
E H (E)
R
H
H
1.0kb

to this mixed probe (Section 2.2.21.b) and was washed in 2×SSC, 0.1% SDS at 30°C and autoradiographed (Fig. 4.4b).

The size of the bands which hybridized to the probes was compared with the bands present on the ethidium bromide stained gel (Fig. 4.4a) and with the location of these bands on the restriction map of λPC3g. This revealed that the four restriction fragments which hybridized to the extended primer probe; a 1.4 kb *Eco*RI fragment, a 1.25 kb *Eco*RI fragment, a 0.5 kb *Eco*RI/*Hin*dIII fragment and a 0.7 kb *Hin*dIII/*Eco*RI fragment were adjacent to each other at one end of the insert (Fig. 4.5).

The position of the site of hybridization of the primer extended cDNA on the end of the insert lent support to the hypothesis that the λPC3g insert only partially overlapped the extended portion of the primer extended cDNAs and not the primer itself. In addition the hybridization of these quite small cDNA probes over fragments totalling about 3.8 kb suggested either that the sequence of the extended cDNA was interrupted in the genome by one or more introns or that the cDNA probes were hybridizing to repeated homologous regions in the gene.

Preliminary sequence analysis was performed on the restriction fragments which hybridized to the cDNA probes after subcloning into M13 and sequencing the inserts using the dideoxy chain-termination system (Sections 2.2.22 and 2.2.23). Computational analysis of the sequence data, using the TESTCD program (Section 2.2.24) to search for potential coding regions strongly indicated the presence of non-coding DNA. This supported the conclusion that intronic sequences were present in these restriction fragments which had been suggested by the hybridization data. Further computer analysis of the sequence data which used the ANALYSEQ program (Section 2.2.24) to translate in all six reading frames, failed to reveal any homology to the known amino acid sequence of chicken pyruvate carboxylase (Rylatt *et al.*, 1977).

The difficulties associated with the sequencing of large regions of these genomic clones in order to find areas of coding potential was not considered sufficiently productive to be warranted. This conclusion was reinforced when a cDNA library prepared from 18 day old embryonic chicken liver poly(A)$^+$ RNA became available.

The focus of this study was then shifted to an investigation of the cDNA clones present in this library.

4.3.6 Screening of Embryonic Chicken Liver cDNA Library with Extended Primer Probes and Clone Analysis

An embryonic chicken liver cDNA library was prepared and kindly supplied by L. Mattschoss and A. Hobbs (Hobbs *et al.*, 1986). The library was prepared from the livers of 18 day old embryonic chickens after the administration of the porphyrinogenic drugs: 2-allyl-2-isopropylacetamide (AIA) and 3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC) to the embryo *in ovo*. Total poly(A)$^+$ RNA was used for the cDNA synthesis and the double tailing procedure (Land *et al.*, 1981) was used to clone double stranded cDNA of >2 kb into pBR322. The ligated recombinant plasmids were transformed into *E. coli* MC1061 and a total of 1100 recombinants were obtained.

The library was plated and screened using the Grunstein procedure (Grunstein and Hogness, 1975) described in Section 2.2.20. The mixture of the [^{32}P] labelled 220 base and 320 base B-14 primer extended cDNAs were hybridized to the filters at 37°C using the conditions described in Section 2.2.21.b. A total of 124 clones yielded positive signals on the first round of screening and this was reduced to 96 upon the second round of screening with the same probe (Fig. 4.6a). The number of clones hybridizing to the primer extended cDNAs represent 8.7% of the total library. In spite of the fact that the cDNA used to make the library was size-selected and therefore would be representative of a reduced percentage of the total number of mRNA species, the abundance of the positive clones was extremely high. Restriction analysis with *Pst*I was performed on the DNA prepared from several of the clones indicated in the legend to Fig. 4.6. The restriction fragments were separated by agarose gel electrophoresis and a determination of the size of the fragments with the SIZING program (Section 2.2.24) showed that most positive clones contained a 1.0 kb *Pst*I fragment in their inserts (Fig. 4.7). It was therefore considered likely that the clones encoded the same cDNA sequence.

Figure 4.6

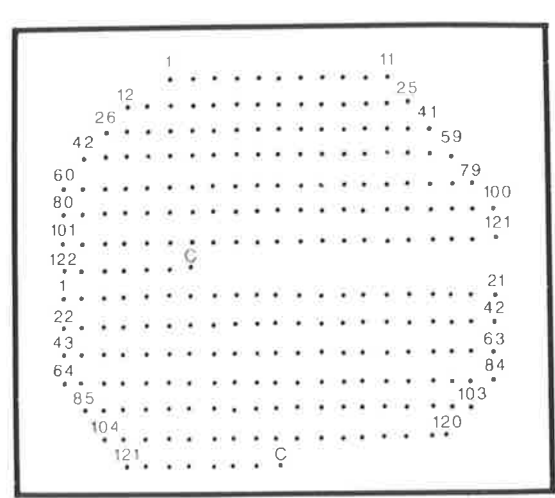
Grunstein of Extended Primer Selected cDNA Clones Hybridized to Several Probes

Clones selected from an embryonic chicken liver cDNA plasmid library were transferred by toothpick in duplicate onto a nitrocellulose disk overlaying an L+12.5 μg/ml tetracycline antibiotic plate. The plate was incubated overnight at 37°C and the filter bearing the colonies was removed and treated to fix DNA using the Grunstein procedure (Section 2.2.20). The filter was hybridized to several probes sequentially with thorough washing between probings to ensure that there was no signal carryover. The plan of the colonies is shown and the numbering bears no relationship to the numbers in the plasmid names. The clones which were selected for further characterization included pP1 (position 4), pP2 (position 17), pP6 (position 43), pP7 (position 49), pP18 (position 108), pP38 (position 80), pP99 (position 123) and pP100 (position 125). C refers to a host cell clone carrying the pBR322 vector with no insert.

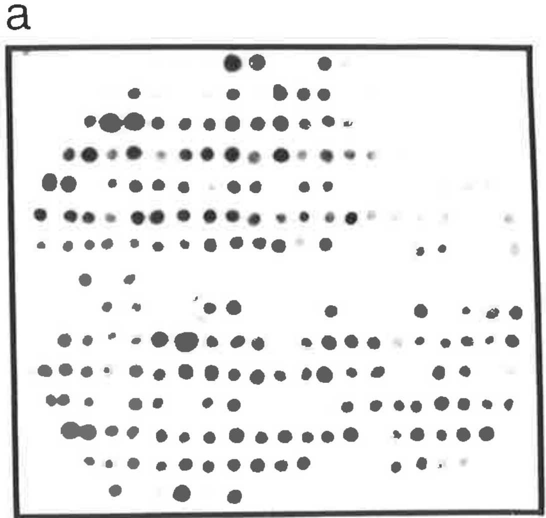
a Grunstein filter was probed with the mixed 220 and 320 base primer extended cDNAs at 37° (Section 2.2.21.b). The filter was washed in 2×SSC, 0.1% SDS at 65°C.

b Grunstein filter was probed with the [^{32}P] labelled 5.3 kb *Hin*dIII fragment from λPC3g (Section 2.2.21.b). The filter was washed in 1×SSC, 0.1% SDS at 65°C.

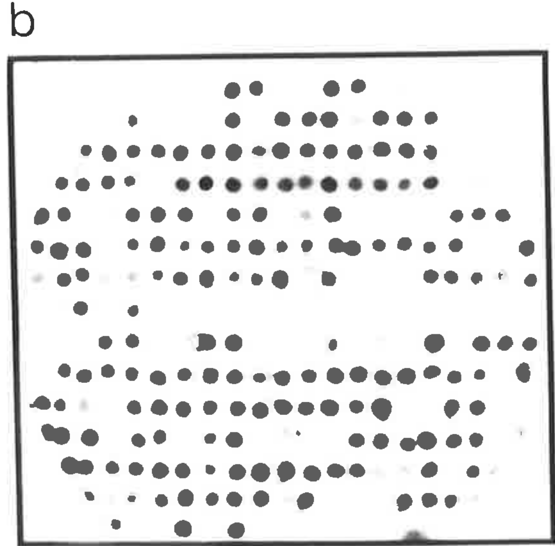
c Grunstein filter was probed with the [^{32}P] labelled 1.0 kb *Pst*I fragment from pP38 (Section 2.2.21.b). The filter was washed in 0.5×SSC, 0.1% SDS at 65°C.



a



b



c

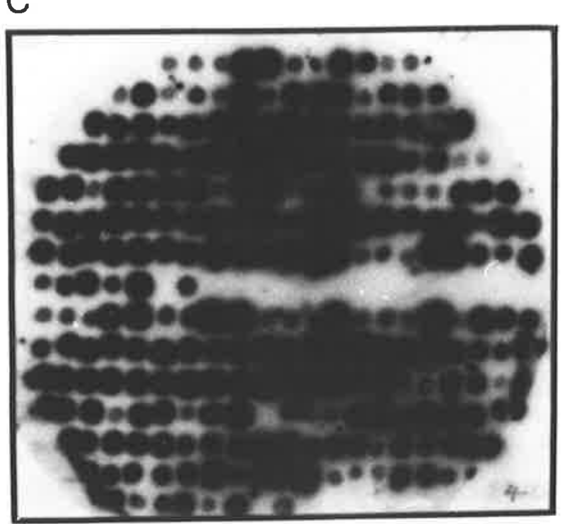
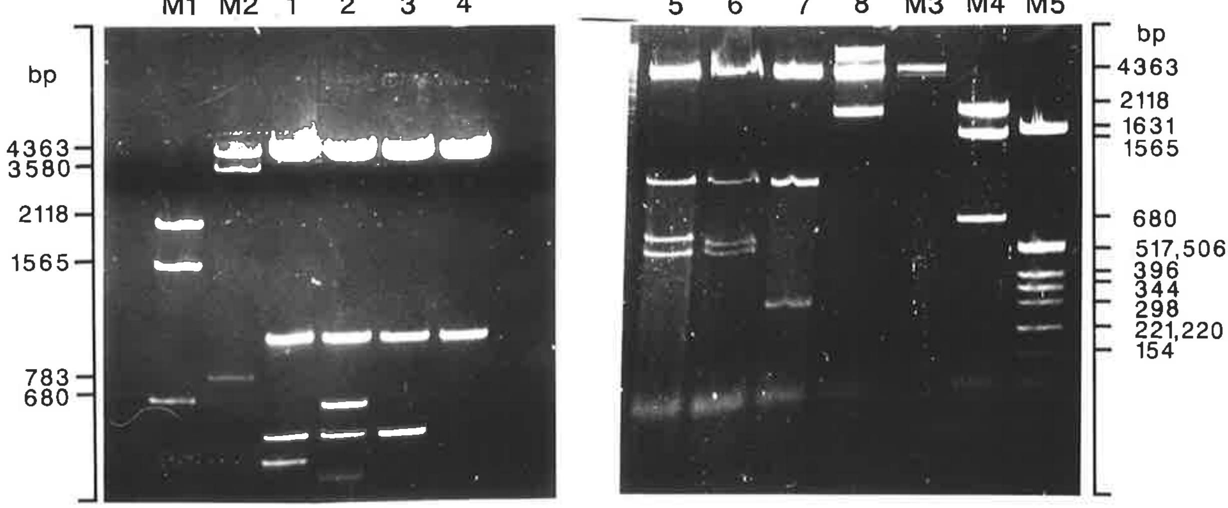


Figure 4.7

Electrophoresis of Restriction Digests of Extended Primer Selected cDNA Clones

Aliquots of 2 μg of several extended primer selected cDNA clones were digested with 10 units of *Pst*I for 2 hr then phenol extracted and ethanol precipitated. The digested clone DNAs were electrophoresed on two 2% (w/v) agarose, 1×TAE gels at 30 mA then stained with 30 μg/ml ethidium bromide (Section 2.2.9.a). The gels were destained with water and photographed under short-wave UV illumination. The sizes of the marker DNA bands are indicated in base pairs.

Track M1 : pBR322/*Rsa*I size marker,
Track M2 : pBR322/*Hin*dIII + pBR322/*Pst*I/*Hin*dIII size markers,
Track 1 : pP1/*Pst*I,
Track 2 : pP2/*Pst*I,
Track 3 : pP7/*Pst*I,
Track 4 : pP18/*Pst*I,
Track 5 : pP6/*Pst*I,
Track 6 : pP38/*Pst*I,
Track 7 : pP99/*Pst*I,
Track 8 : pP100/*Pst*I,
Track M3 : pBR322/*Pst*I size marker,
Track M4 : pBR322/*Rsa*I size marker,
Track M5 : pBR322/*Hin*fI size marker.


M1
M2
1
2
3
4
5
6
7
8
M3
M4
M5
bp
4363
3580
2118
1565
783
680
bp
4363
2118
1631
1565
680
517,506
396
344
298
221,220
154

4.3.7 Relationship Between λPC3g and Extended Primer Selected cDNA Clones

It was important to establish whether the genomic clones selected by the primer extended cDNA probe were related to these cDNA clones. This possibility was tested by using a fragment of the genomic clone λPC3g which hybridizes to the primer extended cDNA to probe the cDNA clones. A 5.3 kb internal *Hin*dIII fragment was chosen since it contains most of the area which hybridizes to the primer extended cDNA (Fig. 4.5). The 5.3 kb fragment was isolated and purified (Sections 2.2.9.a and 2.2.10) from a *Hin*dIII digest of λPC3g and was labelled by nick-translation (Section 2.2.18.c). The labelled 5.3 kb fragment was then used to re-probe the Grunstein filter which was used for the second round screening of the cDNA clones.

The λPC3g 5.3 kb *Hin*dIII fragment hybridized strongly under conditions of high washing stringency to the 96 cDNA clones previously selected by the primer extended cDNA (Fig. 4.6b). This confirmed the relationship between the cDNA clones indicated by restriction analysis and established that the genomic and cDNA clones have a very high degree of sequence homology. The 1.0 kb *Pst*I fragment was isolated from one clone termed pP38 (see the legend to Fig. 4.6), [^{32}P] labelled by nick-translation (Section 2.2.18.c) and used to re-probe the Grunstein filter. The result shown in Fig. 4.6c indicates that the 96 cDNA clones previously selected by the primer extended cDNA have all or part of this 1.0 kb *Pst*I fragment in common. Therefore these clones represent multiple copies of the same mRNA but with inserts of varying size.

4.3.8 Northern Transfer of Chicken Liver RNA Probed with cDNA Clone pP99

To determine if the extended primer selected cDNA clones encoded pyruvate carboxylase the size of the mRNA to which they hybridized was determined by using one of the cDNA clones to probe a Northern transfer of chicken liver RNA. Poly(A)$^+$

RNA and Poly(A)$^-$ RNA prepared from 18 day old embryonic chicken liver and poly(A)$^+$ RNA prepared from 4 day old hatched chicken liver (Sections 2.2.1 and 2.2.2) was denatured by glyoxalation and separated by electrophoresis on a 0.8% (w/v) agarose gel (Section 2.2.9.b). The RNA was transferred to nitrocellulose (Section 2.2.11.b) and the filter was hybridized to the nick-translated clone pP99 DNA. This clone was chosen for use because it appeared to be representative of the other clones because (i) it hybridized to the primer-extended cDNA probes, the genomic clone 5.3 kb *Hin*dIII fragment and the 1.0 kb *Pst*I restriction fragment (Fig. 4.6) and (ii) it was demonstrated to contain the 1.0 kb *Pst*I restriction fragment (Fig. 4.7).

The pP99 clone hybridized very strongly to a single poly(A)$^+$ RNA species which was determined by use of the SIZING program (Section 2.2.24) to be 2.1 kb long (Fig. 4.8). The intensity of the hybridization signals indicated that the the level of the mRNA was greater in the embryonic chicken RNA. Since the size of the chicken liver pyruvate carboxylase mRNA must be at least 3.0 kb it was therefore unlikely that this clone or any of the other related cDNA clones represented pyruvate carboxylase. However, the immunological cross-reactivity between the biotin carboxylases demonstrated by Mottershead *et al.* (1984) and the sequence homology apparent between the biotin attachment sites (Rylatt *et al.*, 1977; Sutton *et al.*, 1977; Maloy *et al.*, 1979) suggested the possibility that these clones may code for another biotin carboxylase such as propionyl-CoA carboxylase which based on the size of the α subunit might be expected to have a mRNA of about this size (the size of the α subunit mRNA from any source had not as yet been determined). To establish whether this was the case the DNA sequence of the cDNA clones was determined.

4.3.9 Sequence Analysis of pP2 and Identification as Chicken Serum Albumin

The cDNA clone pP2 selected with extended primer cDNA (see Fig. 4.6) and shown by restriction analysis to have the longest insert (Fig. 4.7) was chosen for sequence

Figure 4.8

Northern Transfer of Chicken Liver RNA Probed with pP99 DNA

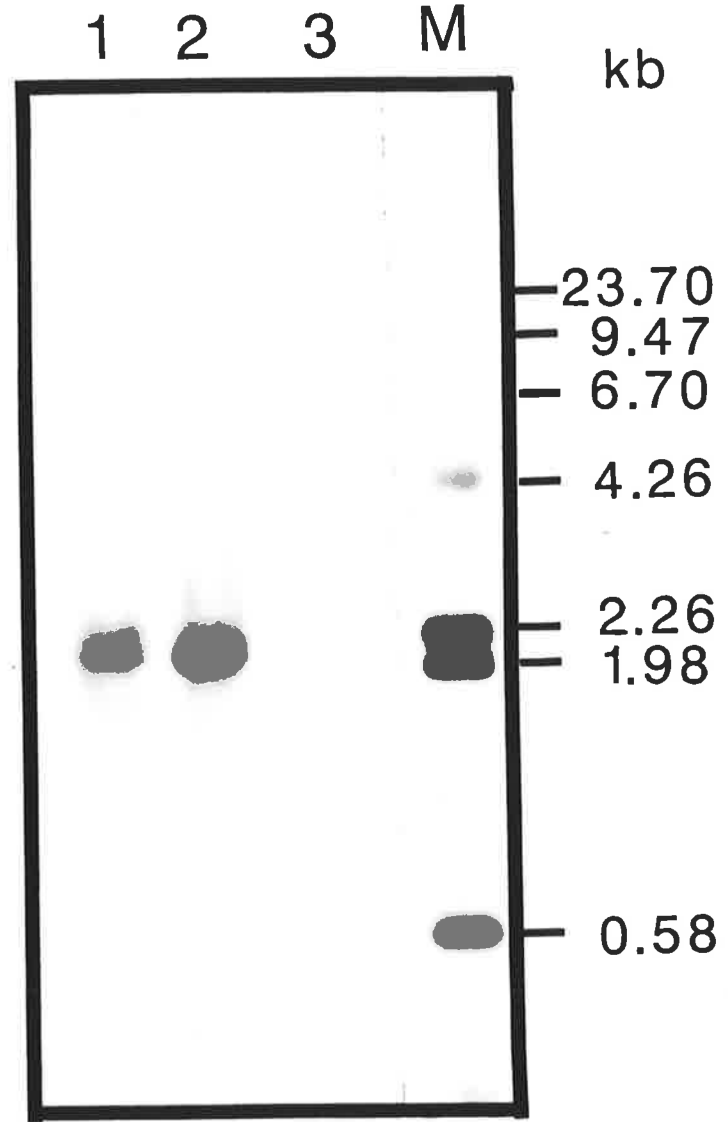
RNA prepared from 17 day old embryonic and 4 day old hatched chicken livers (Section 2.2.1) was separated into poly(A)$^+$ RNA and poly(A)$^-$ RNA fractions by affinity chromatography on oligo(dT) cellulose (Section 2.2.2). Aliquots of 10 μg of the RNA samples were denatured by glyoxalation and electrophoresed on a 1.0% agarose, 10 mM sodium phosphate pH 6.5 gel for 5 hr at 50 mA with recirculation of the running buffer (Section 2.2.9.b). The RNA was transferred to nitrocellulose (Section 2.2.11.b) and the filter was hybridized to pP99 clone DNA (Section 2.2.21.b) which was [^{32}P] labelled by nick-translation (Section 2.2.18.c). The filter was washed in 1×SSC, 0.1% SDS at 37°C and autoradiographed. The sizes of the marker bands are indicated in kb.

Track 1 : 4 day old hatched chick liver poly(A)$^+$ RNA,

Track 2 : 18 day old embryonic chicken liver poly(A)$^+$ RNA,

Track 3 : 18 day old embryonic chicken liver poly(A)$^-$ RNA,

Track 4 : λ/*Hin*dIII size markers which were [^{32}P] end labelled and glyoxalated prior to electrophoresis.


1
2
3
M
kb
23.70
9.47
6.70
4.26
2.26
1.98
0.58

analysis. Subfragments of the insert were prepared using the restriction enzymes *Pst*I, *Hin*dIII and *Dra*I and these were cloned into M13 bacteriophage vectors (Sections 2.2.15 and 2.2.16). The fragments were sequenced using the dideoxy chain-termination sequencing system (Sections 2.2.22 and 2.2.23) employing the sequencing strategy shown in Fig. 4.9.

To aid in the identification of this clone, the pP2 DNA sequence was compared with all the DNA sequences present in the GenBank (Release 35.0, 1-Aug-1985) database using the MATCH program (Section 2.2.24). Sequence identity of 93 contiguous bases was found between the 5′ end of the chicken serum albumin gene and the pP2 sequence (Fig. 4.10). This degree of homology was sufficient to unequivocally identify pP2 as encoding chicken serum albumin. Since the sequence used in the comparison was obtained from one end of the 1.0 kb *Pst*I fragment which is common to most of the extended primer selected cDNA clones, then all of the selected clones probably represent chicken serum albumin cDNA clones (see Figs. 4.6c and 4.7). Partial sequence data obtained from pP6, pP38 and pP99 confirmed that there was sequence identity between these clones and pP2.

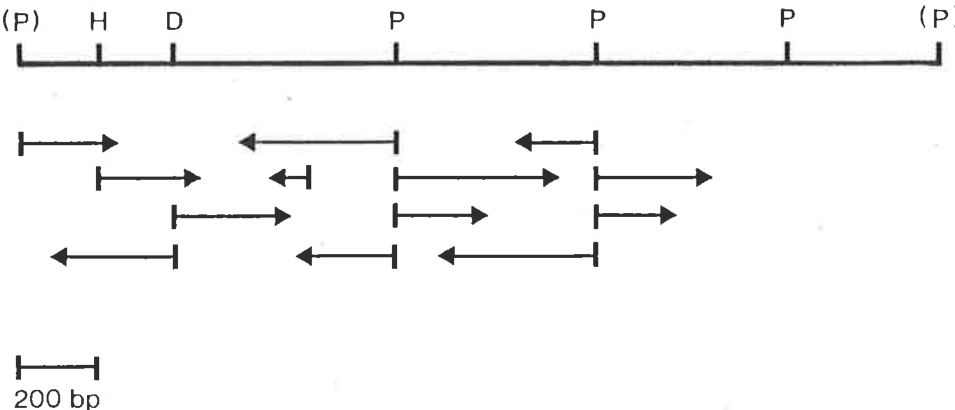
The total sequence determined for pP2 is shown in Fig. 4.11 together with the complete inferred amino acid sequence of chicken serum albumin. The published sequence data for chicken serum albumin (Haché *et al.*, 1983) is restricted to the data shown in Fig. 4.10. Therefore the data provided here is the first report of a substantial amount of coding sequence data for chicken serum albumin.

Evidence for the identification of the genomic clones including λPC3g encoding the chicken serum albumin gene is limited to the strong cross-hybridization which is apparent between the 5.3 kb genomic fragment and the chicken serum albumin cDNA clones. However, the restriction map of the 5′ end of the chicken serum albumin gene described by Haché *et al.* (1983) shows a striking similarity to the map of λPC3g especially with respect to the *Eco*RI sites (Fig. 4.12). Haché also localizes the exons within each of the *Eco*RI fragments confirming the Southern transfer results showing that all fragments hybridize to the extended primer probes. This data lends strong support to the identification of λPC3g as containing the 5′ end of the chicken serum albumin gene.

Figure 4.9

Restriction Map and Sequencing Strategy of pP2

The restriction sites used in the sequence analysis of the pP2 clone insert are marked. Restriction fragments from pP2 were subcloned into M13mp8 and mp9 (Messing and Vieira, 1982) and were sequenced using the dideoxy chain termination method (Sections 2.2.22 and 2.2.23). The 5′ end of the each M13 clone is indicated underneath the map and the extent and orientation of each sequence obtained is indicated by the arrow. D = *Dra*I, H = *Hin*dIII, P = *Pst*I and the bracketed *Pst*I sites are in the pBR322 vector.



(P)
H
D
P
P
P
(P)
200 bp

Figure 4.10

Sequence Homology Between the 5′ End of the Chicken Serum Albumin Gene and pP2.

The sequence of the 5′ end of chicken serum albumin gene including the most 5′ exon (Haché *et al.*, 1983) was compared to the sequence of the pP2 clone using the SEQH program (Section 2.2.24). Chicken serum albumin exonic sequence is shown in upper case and residue 1 represents the cap site. Sequence homology is indicated by (*).

```
                      -260      -250      -240      -230      -220      -210      -200      -190      -180
Chicken Albumin Gene  gaactccatgtttgtaatttgataaggaaatttacgtaaactttatctgtttggaatgctggaaggttcaacctttcaggacaaaacg

                      -170      -160      -150      -140      -130      -120      -110      -100      -90
Chicken Albumin Gene  gtctagaggaaacataataaacaaataatatcaaattcattttggcaaggagcactccaagttttgtaaatcttcagaccaaatggaa

                      -80       -70       -60       -50       -40       -30       -20       -10       +1
Chicken Albumin Gene  atctacaattgccagctctacaggtaatgttttacagaagcagtcagtaaaggtatataagaaaatgatttccctcaatcatcctagca

                      10        20        30        40        50        60        70        80        90
Chicken Albumin Gene  tttttgaataatttagcccacatcataatctgcagccATGAAGTGGGTAACATTAATTTCATTCATTTTCCTCTTCAGTTCAGCAACATC
                                                   *************************************************************
pP2                                                ctgcagccATGAAGTGGGTAACATTAATTTCATTCATTTTCCTCTTCAGTTCAGCAACATC
                                                           10        20        30        40        50        60

                      100       110       120       130       140       150       160       170       180
Chicken Albumin Gene  CAGGAATCTGCAAAGATTTGCTCGTGATGCAGgtaagaaaatactaattaattgtccttatttttctttgttatactgacttaatgacttc
                      ********************************
pP2                   CAGGAATCTGCAAAGATTTGCTCGTGATGCAG
                              70        80        90

                      190       200       210       220       230       240       250       260       270
Chicken Albumin Gene  aatgtgatttgatggaaagattggcttgaaaagaaactgatgagggatctgaggggatggatgacatcagacgatacgtacatagtttgg

                      280       290       290
Chicken Albumin Gene  attctttagcactattagacagtacatggca
```

Figure 4.11

Sequence of Chicken Pre-Proalbumin Inferred from the pP2 cDNA Clone

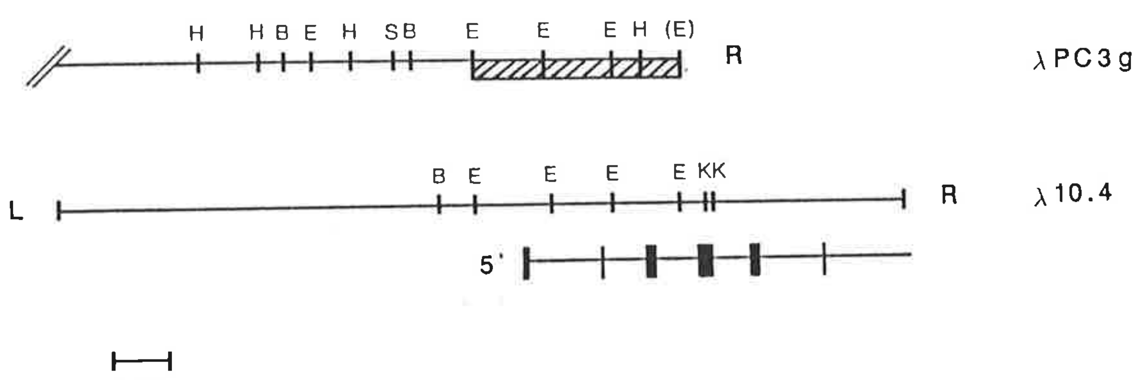
The sequence which was determined for the pP2 cDNA clone insert is shown together with the amino acid sequence deduced using the ANALYSEQ program (Section 2.2.24). Residue 9 represents the start of the initiation codon and residue 1833 the start of the termination codon. The *Pst*I site at the 5′ end of the cDNA insert is indicated. The extent of the coding region establishes that the chicken pre-proalbumin contains 611 amino acids.

```
        met lys trp val thr leu ile ser phe ile phe leu phe ser ser ala thr ser arg asn leu gln arg phe ala arg asp ala glu his lys ser glu ile ala his arg
CTGCAGCCATGAAGTGGGTAACATTAATTTCATTCATTTTCCTCTTCAGTTCAGCAACATCCAGGAATCTGCAAAGATTTGCTCGTGATGCAGAGCACAAGAGTGAAATTGCCCATCGCT
        10        20        30        40        50        60        70        80        90        100       110       120

tyr asn asp leu lys glu glu thr phe lys ala val ala met ile thr phe ala gln tyr leu gln arg cys ser tyr glu gly leu ser lys leu val lys asp val val asp leu ala
ACAATGATTTGAAAGAAGAAACATTTAAGGCAGTTGCCATGATCACATTTGCCCAGTATCTCCAGAGGTGCTCTTATGAAGGACTGTCTAAGCTTGTGAAGGATGTTGTTGATCTGGCAC
        130       140       150       160       170       180       190       200       210       220       230       240

gln lys cys val ala asn glu asp ala pro glu cys ser lys pro leu pro ser ile ile leu asp glu ile cys gln val glu lys leu arg asp ser tyr gly ala met ala asp cys
AAAAATGTGTAGCCAATGAAGATGCTCCTGAATGCTCAAAACCACTGCCTTCCATTATCCTGGATGAAATCTGCCAAGTGGAAAAGCTCCGTGACTCTTATGGTGCAATGGCCGACTGCT
        250       260       270       280       290       300       310       320       330       340       350       360

cys ser lys ala asp pro glu arg asn glu cys phe leu ser phe lys val ser asn gln thr ser phe ser his thr lys asp gln leu leu asn val ile cys gln glu tyr gln asp
GTAGCAAAGCTGATCCTGAAAGAAATGAGTGTTTCCTGTCATTTAAAGTTTCCAACCAGACTTCGTTCAGCCATACCAAAGACCAGCTTCTGAATGTGATATGCCAGGAATACCAGGACA
        370       380       390       400       410       420       430       440       450       460       470       480

asn lys cys ser phe leu gly his phe ile tyr ser val ala arg arg his pro phe leu tyr ala pro ala ile leu ser phe ala val asp phe glu his ala leu his ser cys cys
ACAAGTGTTCATTTCTGGGACATTTCATCTATTCTGTTGCAAGAAGACACCCCTTCTTGTATGCCCCTGCAATCCTTAGTTTTGCTGTTGATTTTGAGCATGCACTTCACAGCTGTTGCA
        490       500       510       520       530       540       550       560       570       580       590       600

lys glu ser asp val gly ala cys leu asp thr lys glu ile val met arg glu lys pro arg glu cys val ser lys gln tyr phe cys gly ser ser ser ser ser glu arg val phe
AAGAGAGTGATGTCGGTGCTTGCCTGGACACCAAGGAAATTGTTATGAGAGAAAAGCCAAGGGAGTGTGTGAGCAAGCAGTATTTTTGTGGATCTTCAAGCAGTTCGGAGAGAGTTTTCC
        610       620       630       640       650       660       670       680       690       700       710       720

gln ala arg gln leu ile tyr leu ser gln lys tyr pro lys ala pro phe ser glu val ser lys phe cys thr cys ser ile gly val his lys glu cys cys glu gly asp met val
AAGCACGACAACTTATTTACCTAAGCCAAAAATACCCCAAGGCTCCATTCTCAGAGGTTTCTAAATTTTGTACATGTTCTATCGGCGTCCACAAAGAGTGCTGTGAAGGGGACATGGTGG
        730       740       750       760       770       780       790       800       810       820       830       840

glu cys met asp asp met ala arg met met ser asn leu cys ser gln gln asp val phe ser gly lys ile lys asp cys cys glu lys pro ile val glu arg ser gln cys ile met
AGTGCATGGATGACATGGCACGTATGATGAGCAATCTGTGCTCTCAACAAGATGTTTTCTCAGGTAAAATCAAAGACTGCTGTGAGAAGCCTATTGTGGAACGAAGCCAGTGCATTATGG
        850       860       870       880       890       900       910       920       930       940       950       960

glu ala glu phe asp glu lys pro ala asp leu pro ser leu val glu lys tyr ile glu asp lys glu val cys lys ser phe glu ala gly his asp ala phe met ala glu phe val
AGGCAGAATTTGATGAGAAACCTGCAGATCTGCCTTCATTAGTTGAAAAGTACATAGAAGATAAGGAAGTGTGTAAAAGTTTTGAAGCAGGCCACGATGCATTCATGGCAGAGTTCGTTT
        970       980       990       1000      1010      1020      1030      1040      1050      1060      1070      1080

tyr glu tyr ser arg arg his pro glu phe ser ile gln leu ile met arg ile ala lys gly tyr glu ser leu leu glu lys cys cys lys thr asp asn pro ala glu cys tyr ala
ATGAATACTCACGAAGACACCCTGAGTTCTCCATACAGCTTATTATGAGAATTGCCAAAGGATATGAATCACTTCTGGAAAAGTGCTGCAAAACTGATAACCCTGCTGAGTGCTACGCAA
        1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200

asn ala gln glu gln leu asn gln his ile lys glu thr gln asp val val lys thr asn cys asp leu leu his asp his gly glu ala asp phe leu lys ser ile leu ile arg tyr
ATGCTCAAGAGCAACTGAACCAACATATCAAAGAAACTCAGGATGTTGTGAAGACAAACTGTGATCTTCTCCATGACCATGGCGAGGCAGACTTCCTCAAGTCCATCCTGATCCGCTACA
        1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320

thr lys lys met pro gln val pro thr asp leu leu leu glu thr gly lys lys met thr thr ile gly thr lys cys cys gln leu gly glu asp arg arg met ala cys ser glu gly
CTAAGAAAATGCCTCAAGTACCAACTGATCTCCTGCTTGAAACTGGAAAGAAAATGACAACTATTGGTACTAAGTGCTGCCAGCTTGGTGAAGACAGACGCATGGCTTGTTCTGAGGGTT
        1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440

tyr leu ser ile val ile his asp thr cys arg lys gln glu thr thr pro ile asn asp asn val ser gln cys cys ser cys ser gln leu tyr ala asn arg arg pro cys phe thr
ATCTGAGCATTGTGATTCATGATACGTGCAGGAAACAGGAGACCACACCTATAAATGACAACGTTTCACAATGCTGCAGCTGCAGTCAGCTCTATGCTAACAGAAGACCATGTTTCACTG
        1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560

ala met gly val asp thr lys tyr val pro pro pro phe asn pro asp met phe ser phe asp glu lys leu cys ser ala pro ala glu glu arg glu val gly gln met lys leu leu
CTATGGGAGTAGATACCAAATATGTTCCTCCACCATTTAATCCTGATATGTTCAGCTTTGATGAAAAATTGTGCAGTGCTCCTGCTGAAGAACGAGAAGTAGGCCAGATGAAATTGCTAA
        1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680

ile asn leu ile lys arg lys pro gln met thr glu glu gln ile lys thr ile ala asp gly phe thr ala met val asp ser ala ala glu ser asp ile asn thr cys leu glu lys
TCAACCTCATTAAACGCAAGCCCCAGATGACAGAAGAACAAATAAAGACAATTGCTGATGGTTTCACTGCCATGGTTGACAGTGCTGCAGAGTCGGACATCAATACATGCTTGGAGAAGA
        1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800

arg cys gln leu ile val gln ala glu pro his *
GGTGCCAACTAATAGTCCAAGCAGAGCCACATTAGGATTGGTGCTTAAGAACGGTTGCCAGTGAGAA
        1810      1820      1830      1840      1850      1860   1867
```

Figure 4.12

Comparison of Restriction Maps of a Chicken Serum Albumin Genomic Clone and λPC3g

The end of the λPC3g clone adjacent to the λCharon 4A right arm is shown aligned with the restriction map of the λ10.4 chicken serum albumin clone described by Haché *et al.* (1983). The exons localized by Haché *et al.* (1983) are shown underneath the λ10.4 map as solid boxes along a line which begins at the 5′ end of the gene. The fragments in λPC3g which hybridized to the extended primer probe are indicated by hatched boxes. L and R refer to the left and right arms of the λ vectors respectively.


H
H B E H SB E E E H (E)
R
λPC3g
B E E E E KK
L
R
λ10.4
5'
1.0kb

4.3.10 Chicken Serum Albumin Structure and Sequence Homology with Other Species

The coding sequence of the chicken serum albumin cDNA clone pP2 was analysed on the basis of its content using the ANALYSEQ program (Section 2.2.24) and the results are shown in Table 4.1 and Table 4.2. A codon usage table was derived from the sequence and the percentage content of each base together with the percentage content of A+T and G+C bases was also calculated. The codon usage was not markedly different from the usage compiled from 33 chicken genes by Maruyama *et al.* (1986). However, the G+C content was quite low which was undoubtedly contributed to by the heavy usage of A+T rich codons for the acidic amino acids aspartic acid and glutamic acid which abound in serum albumin. The hydrophobicity value which was determined indicates that the protein is highly charged as would be expected for a protein that attains very high plasma concentration ($\sim$30 mg/ml) in many species (Peters, 1985). The length of the chicken serum pre-proalbumin polypeptide chain is very close to pre-proalbumins from other species which vary in length from 608–610 amino acids and hence the predicted protein molecular weight is also similar to the molecular weight of human, rat and bovine pre-proalbumin (Peters, 1985). The inferred composition of chicken serum pre-proalbumin agrees very well with the experimentally determined composition of the processed albumin described by Peters (1970) if one considers the percentage composition rather than the amino acid numbers (Table 4.2).

The complete sequence of the coding region of cDNA clones for human serum albumin (Lawn *et al.*, 1981; Dugaiczyk *et al.*, 1982) and rat serum albumin (Sargent *et al.*, 1981) have been reported as has been the essentially complete protein sequence of bovine serum albumin (Brown, 1975; Reed *et al.*, 1976; Brown, 1976) determined by direct protein sequencing. The sequences of these serum albumins were compared with the chicken serum albumin sequence using the ALIGNSIG program (Section 2.2.24), see Fig. 4.13. The homology between the sequence of the chicken and mammalian proteins is quite high, varying from 43.0% for bovine, 43.2% for rat and 46.5% for human. The homology extends to conservation of the

Table 4.1

Analysis of Chicken Serum Albumin cDNA Clone and Inferred Protein by Content

The ANALYSEQ program (Section 2.2.24), was used to analyse the content of the coding region from the sequence of the pP2 cDNA clones shown in Fig. 4.11.

a A table of codon usage is shown. Amino acids are represented by the single letter code and are shown beside their codons with the number of times they occur.

b The base content of the sequence is shown. The percentage content of each base is given with the percentage content of A+T and G+C bases.

c The total numbers of codons and amino acids, the estimated protein molecular weight and the estimated protein hydrophobicity are shown.

a

F	TTT	13	S	TCT	9	Y	TAT	11	C	TGT	14
F	TTC	18	S	TCC	5	Y	TAC	9	C	TGC	26
L	TTA	2	S	TCA	11	*	TAA	0	*	TGA	0
L	TTG	6	S	TCG	1	*	TAG	1	W	TGG	1
L	CTT	9	P	CCT	14	H	CAT	8	R	CGT	3
L	CTC	8	P	CCC	3	H	CAC	5	R	CGC	4
L	CTA	3	P	CCA	10	Q	CAA	19	R	CGA	4
L	CTG	15	P	CCG	0	Q	CAG	16	R	CGG	0
I	ATT	15	T	ACT	9	N	AAT	9	S	AGT	8
I	ATC	13	T	ACC	3	N	AAC	8	S	AGC	8
I	ATA	6	T	ACA	11	K	AAA	24	R	AGA	12
M	ATG	20	T	ACG	1	K	AAG	25	R	AGG	4
V	GTT	17	A	GCT	15	D	GAT	26	G	GGT	7
V	GTC	3	A	GCC	9	D	GAC	14	G	GGC	4
V	GTA	6	A	GCA	18	E	GAA	32	G	GGA	8
V	GTG	9	A	GCG	0	E	GAG	19	G	GGG	1

b

T	C	A	G	A+T	G+C
25.98%	20.70%	30.94%	22.39%	56.92%	43.09%

c

Total Number of Codons	=	612
Total Number of Amino Acids	=	611
Protein Molecular Weight	=	69687
Hydrophobicity	=	-45.01

Table 4.2

Comparison of Inferred and Experimentally Determined Amino Acid Composition of Chicken Serum Albumin

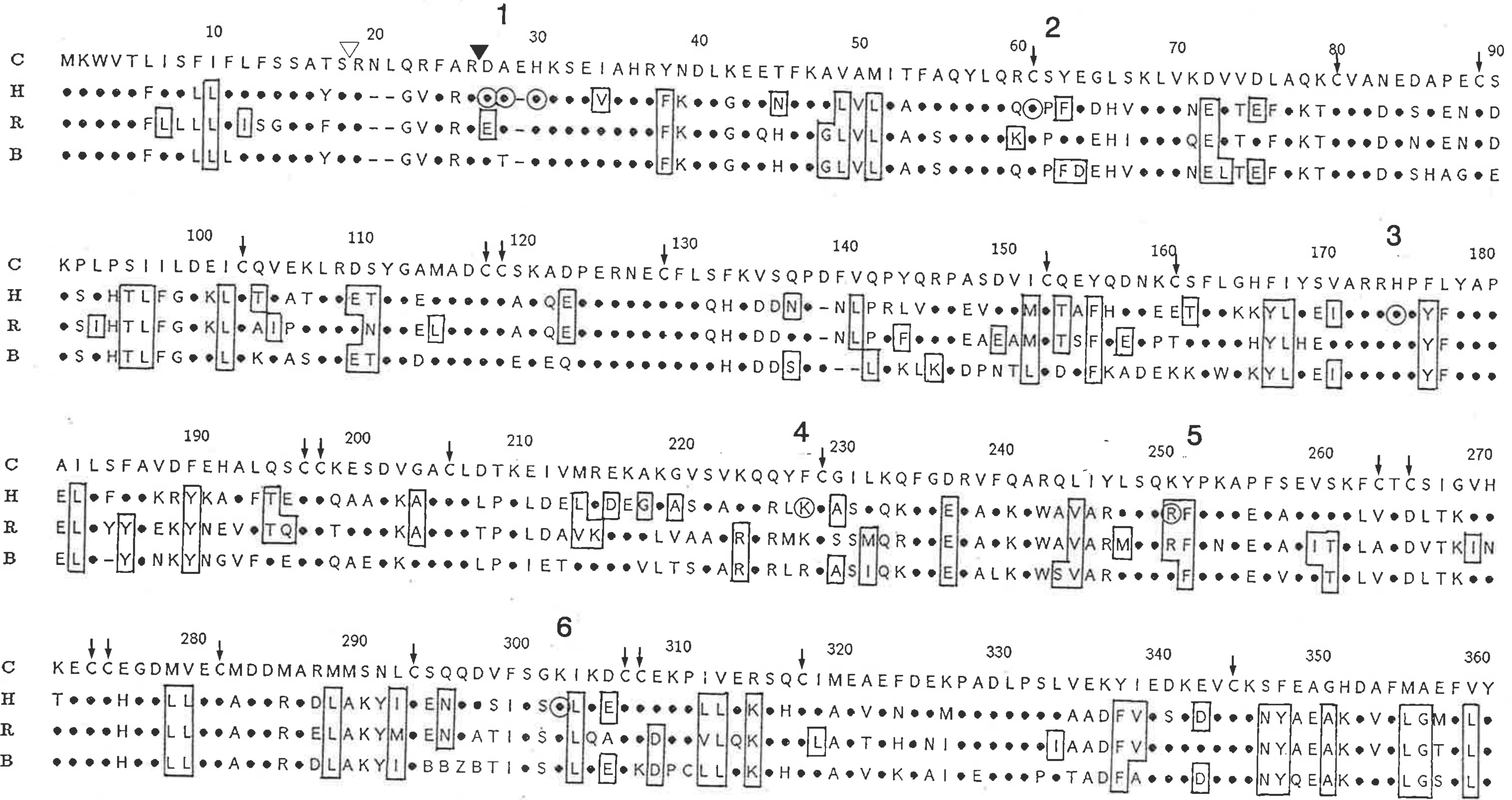
The total inferred amino acid content of the pP2 coding region determined with the ANALYSEQ program (Section 2.2.24) was compiled to show the amino acid composition of chicken serum albumin both in terms of numbers of amino acids and percentage of all amino acids. This data is compared with the composition determined by Peters (1970) who used acid hydrolysis of the protein followed by amino acid analysis of the hydrolysate. The aspartic acid and asparagine composition was given by Peters (1970) as a combined value (Asx), as was the glutamic acid and glutamine composition (Glx) and the asparagine and glutamine composition (Amino-NH_3).

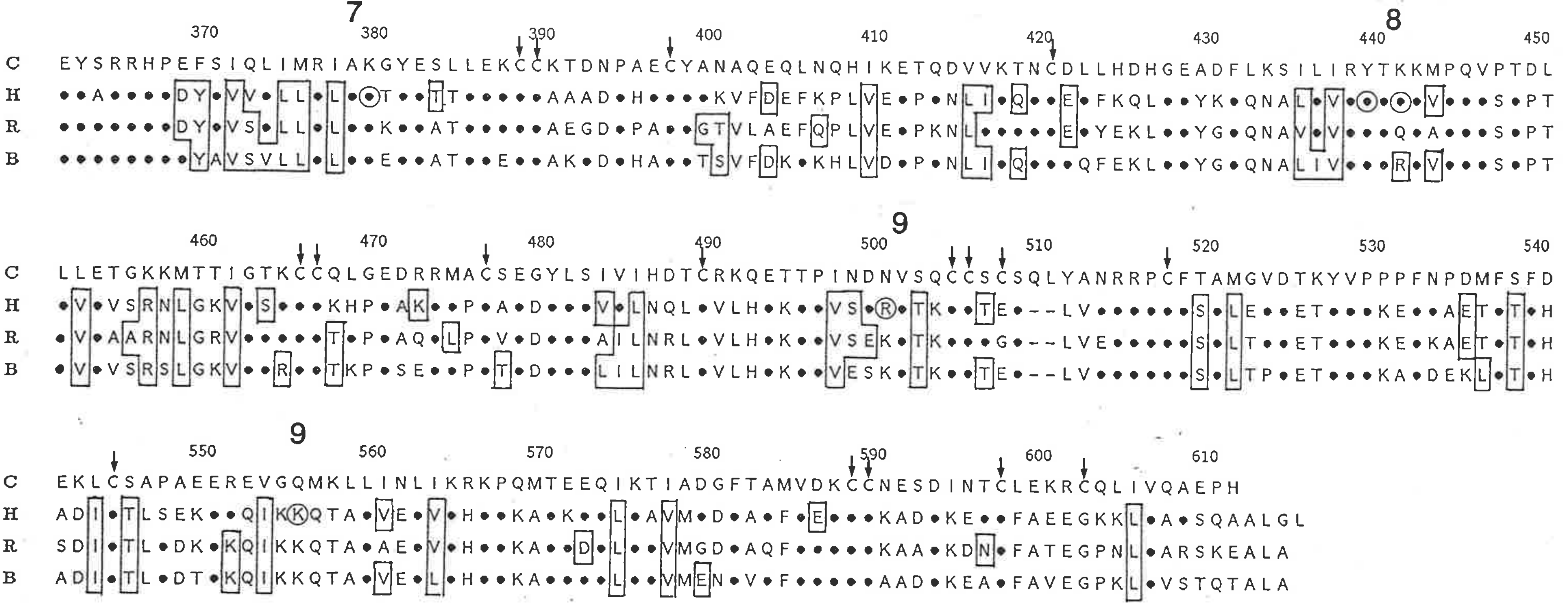
Amino Acid	Inferred Composition		Composition (Peters, 1970)	
Ala	40	6.8%	41	7.1%
Arg	24	4.1%	25	4.3%
Asn	16	2.7%	–	–
Asp	40	6.8%	–	–
Asx		(9.5%)	56	9.7%
Cys	40	6.8%	33	5.7%
Gln	34	5.8%	–	–
Glu	51	8.7%	–	–
Glx		(14.5%)	78	13.6%
Gly	20	3.4%	23	4.0%
His	13	2.2%	13	2.3%
Ile	32	5.5%	31	5.4%
Leu	40	6.8%	39	6.8%
Lys	48	8.2%	43	7.5%
Met	19	3.2%	18	3.1%
Phe	27	4.6%	28	4.9%
Pro	27	4.6%	26	4.5%
Ser	38	6.5%	42	7.3%
Thr	22	3.8%	23	4.0%
Trp	0	0%	0	0%
Tyr	20	3.4%	19	3.3%
Val	34	5.8%	37	6.4%
TOTAL	585	100%	575	100%
Amide-NH_3			(43)	(7.5%)

Figure 4.13

Sequence Homology Between Chicken Pre-Proalbumin and Pre-Proalbumin from Other Species

The primary sequence of chicken serum pre-proalbumin (**C**) inferred from the DNA sequence of the cDNA clone pP2 was aligned with the sequences of human (**H**), rat (**R**) and bovine (**B**) pre-proalbumin using the ALIGNSIG program (Section 2.2.24). The Z scores which represent the number of standard deviations greater than random sequence homology between the sequences were calculated using 20 randomizations of the data ($p < 0.001$). The sites of cleavage of the pre-proalbumin and proalbumin in the proteins from these species are indicated by (▽) and (▼) respectively. Cysteinyl residues are marked with an arrow and residues which have been demonstrated to have functional significance in human serum albumin are ringed and numbered (Ledden and Feldhoff, 1983). The ligands which bind to these proposed physiological binding sites are as follows; (1) Cu(II) and Ni(II), (2) free thiol which binds heavy metals such as Hg(II), (3) tertiary fatty acids, (4) acetylsalicylate, (5) bilirubin, (6) quaternary fatty acids, (7) secondary fatty acids, (8) tryptophan and thyroxine and (9) primary fatty acids. Sequence identity is indicated by (•) and conservative substitutions are boxed.





The Z scores for the alignment of the chicken pre-proalbumin sequence with the following sequences are : Human pre-proalbumin = 66.64, Rat pre-proalbumin = 60.35, Bovine pre-proalbumin = 46.26.

positions of 85% of the cysteinyl residues in chicken serum albumin with respect to the other three species. The disulphide bridges which form in the folded protein essentially dictate the three domain tertiary structure of the protein. The amino acids which have been implicated in the binding to rat and human serum albumins of primary, secondary and tertiary fatty acids, bilirubin and acetylsalicylate (Reed *et al.*, 1975; Ledden and Feldhoff, 1982; Feldhoff and Ledden, 1983) are indicated in Fig. 4.13 also appear to be conserved in chicken serum albumin.

The cleavage point for chicken proalbumin was determined by Rosen and Geller (1977) and is marked in Fig. 4.13 as is the cleavage point for chicken pre-proalbumin (Compere *et al.*, 1981). Both these cleavage points were determined by direct N-terminal protein sequencing of the purified precursor forms of chicken albumin. In the pre-proalbumin sequence, Compere *et al.*, 1981) assigned asparagine as residue no. 2, whereas Haché *et al.* (1983) in sequencing the chicken pre-proalbumin gene determined it to be a tryptophan. The cDNA sequence analysis presented here confirmed that a tryptophan residue is present in this position, in common with pre-proalbumin from other species.

4.4 Discussion

The identification of the extended primer cDNA selected clones as encoding chicken serum albumin raised the following question: Why should the biotin attachment site oligonucleotide, B-14, prime cDNA synthesis on the serum albumin mRNA and yield specific termination products?

4.4.1 cDNA Synthesis from Chicken Serum Albumin mRNA

An examination of the chicken serum albumin sequence shown in Fig. 4.11 for homology to the B-14 and the structure of the serum albumin mRNA provided a possible explanation. Only one area of significant homology was found and this contained 9 contiguous base matches and a possible U-G base pair. The 9 bases

which matched are 5′-AUGAAGAUG-3′ which in the pyruvate carboxylase mRNA translate to "Met-Lys-Met" in the protein sequence, which is the sequence of the highly conserved biotin attachment site sequence. However, the reading frame used in the albumin mRNA will not result in translation of the bases to yield this sequence and hence there is no site for biotinylation in chicken serum albumin.

The sequence to which B-14 is homologous is 289 bases from the cap site which had been identified by Haché *et al.* (1983) from the chicken albumin genomic sequence. Extension of the B-14 primer from the indicated site in Fig. 4.14a to the cap site of the mRNA would yield a 302 base cDNA product. Haché *et al.* (1983) also proposed the presence of stem-loop structures which had sequence homology to the ribosomal 18S RNA in the chicken, human and rat serum albumin mRNAs. The bases involved in forming the stem-loop in the chicken serum albumin mRNA were 9–65 as marked in Fig. 4.14a and the resulting stem loop structure is shown in Fig. 4.14b. The presence of such a secondary structure in the mRNA is precisely the circumstance which would result in the termination of cDNA synthesis by reverse transcriptase (Villa-Komaroff, 1978; Efstratiadis and Villa-Komaroff, 1979).If this termination occurred on the 3′ side of the stem at the site indicated in Fig. 4.14b then the resulting cDNA product would be 237 bases long. The sizes of these 237 and 302 base theoretical cDNA products are consistent with the 220 and 320 base cDNA products empirically found if the error involved in sizing these bands on the polyacrylamide gels is considered. Production of both species may have resulted from the presence of a mixed population of denatured and non-denatured albumin mRNA. The stem loop indicated for the chicken serum albumin mRNA has a lower ΔG° and is significantly longer than the structures proposed by Haché *et al.* (1983) for the human and rat albumin mRNAs. As such the proposed chicken albumin stem-loop structure may have been partially resistant to the non-rigorous RNA denaturation step used prior to the cDNA synthesis.

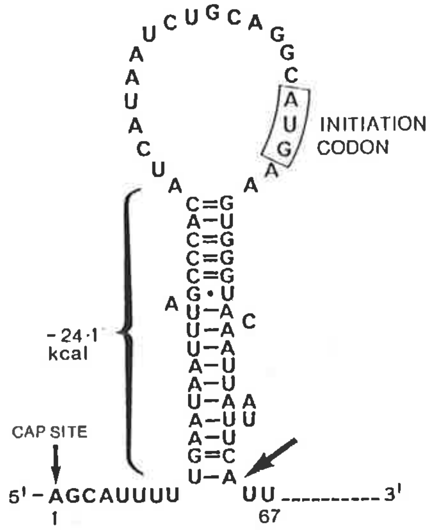
Figure 4.14

Structure of the 5′ End of Chicken Serum Albumin mRNA and B-14 Sequence Homology

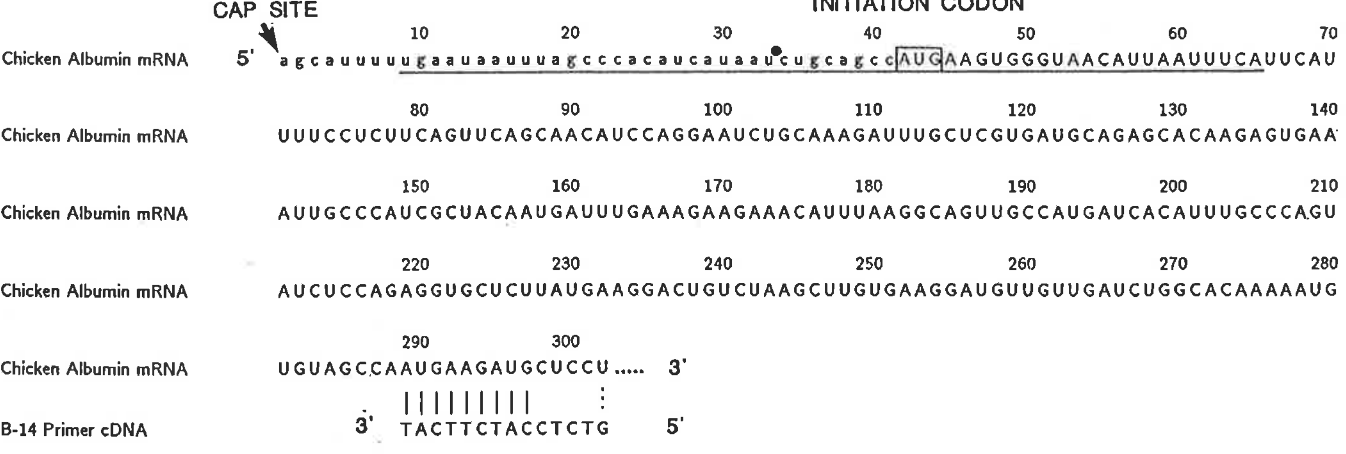
a The structure of a stem-loop structure proposed by Haché *et al.* (1983) is shown. The cap site and the initiation codon are indicated. The $\Delta G°$ value of the stem is shown together with the possible site of premature termination of cDNA synthesis which is marked by an arrow.

b The putative priming site of the B-14 oligonucleotide on residues near the 5′ end of the chicken serum albumin mRNA is shown. The B-14 primer is shown annealed to the mRNA by 9 base pairs and a U-G base pair. The mRNA sequence is a composite of the sequence determined by Haché *et al.* (1983) and the cDNA clone sequence determined in this study. The junction of these two sequences is indicated by a dot. The residues involved in the stem-loop structure proposed by Haché *et al.* (1983) are underlined and the cap site and initiation codon are marked.

a



b



4.4.2 Non Pyruvate Carboxylase-Specific cDNA Priming by B-14 Oligonucleotide

The very low hybridization signal obtained when the extended primer selected cDNA and genomic clones were probed with B-14 suggests that annealing of B-14 to the chicken serum albumin mRNA may be an inefficient and unstable event. However, in an experiment which measured the fraction of *in vitro* translation products from total adult chicken liver RNA represented by serum albumin, the albumin mRNA was found to be 15% of all mRNA present (Gordon *et al.*, 1978). Further, Morris *et al.* (1984) found that 17–18 day old embryonic chicken livers contain 2-3 fold higher levels of albumin mRNA than 4 day old hatched chicken livers which was confirmed by the Northern analysis shown in Fig. 4.8. The level of embryonic chicken liver serum albumin mRNA is therefore very high and may be greater than 100 fold more than the expected level of pyruvate carboxylase mRNA (<0.5%). Therefore, even if annealing of the B-14 primer to the albumin mRNA was relatively inefficient owing to the presence of only 9 hybridizing bases, significantly greater amounts of the albumin cDNA may be synthesized than the cDNA primed on the pyruvate carboxylase mRNA. The amount of B-14 primer extension on the chicken serum albumin mRNA appears to have masked any cDNA specifically primed on the pyruvate carboxylase mRNA which in any case may not have yielded small, premature termination products.

Could the possibility of B-14 annealing to an abundant, non-biotin carboxylase mRNA species, have been predicted? Certainly neither the human nor the rat serum albumin sequences which both became available in 1981 prior to the commencement of this study contain homology to the B-14 oligonucleotide greater than a 7 base match which would not prime efficiently. The chicken serum albumin genomic sequence which was published in 1983 does not include the exon which contains the B-14 homology. Therefore it is unlikely that this cross-hybridization could have been foreseen.

4.4.3 An Alternative Approach to Pyruvate Carboxylase Clone Isolation

The demonstration that the B-14 primer extended cDNA probes were not primed on the pyruvate carboxylase mRNA suggested that continued use of this approach was liable to result in the generation of further non-pyruvate carboxylase specific probes. At this time the first report of the cloning of pyruvate carboxylase from any species appeared in the literature. Freytag and Collier (1984) described the isolation and sequencing of 3′ end of a cDNA clone encoding human liver pyruvate carboxylase. In view of the high degree of homology between the amino acid sequence of human liver pyruvate carboxylase and the chicken liver pyruvate carboxylase biotinyl peptide (Rylatt *et al.*, 1977) the possibility of using the human pyruvate carboxylase clone as a probe for chicken liver pyruvate carboxylase clones was investigated. This provided the starting point for the studies of chicken liver and rat liver pyruvate carboxylase described in Chapter 5.

Chapter 5

Isolation and Characterization of Pyruvate Carboxylase cDNA Clones

5.1 Introduction

In view of the establishment that the B-14 primer extended cDNA was primed on the serum albumin mRNA, a reappraisal of the use of the primer extension strategy was undertaken. Obviously any probe used to select clones in cDNA or genomic libraries should have sufficient DNA sequence homology to chicken pyruvate carboxylase to permit highly specific binding to these clones. The timely report of the isolation of a cDNA clone for human liver pyruvate carboxylase Freytag and Collier (1984) provided the information necessary for a change of approach in the investigation.

In the isolation of the pyruvate carboxylase cDNA clone phPC1, Freytag and Collier (1984) had used a 14 base oligonucleotide prepared against the inferred sheep liver pyruvate carboxylase biotin attachment site sequence (Rylatt *et al.*, 1977). In fact the sequence used was the same as that chosen for the B-14 primer. However instead of using human codon preference to select a single sequence these workers had synthesized a mixture of four oligonucleotides which represented all possible codon sequences. On their first round of cDNA library screening, Freytag and Collier plated 50,000 transformants from an adult human liver cDNA library and obtained 19 positive signals. Subsequent rounds of screening revealed that only three clones encoded pyruvate carboxylase and the other clones were not identified. This established that the abundance of pyruvate carboxylase clones was only 0.006%. Since the abundance of cDNA clones may be used as a guide to the abundance of the mRNA, this figure is considerably lower than previous estimates of the abundance of the embryonic chicken liver pyruvate carboxylase mRNA. The pyruvate carboxylase activity data for different vertebrate species compiled by Wallace (1985) gave values of 45±1 units/g (net weight) tissue at 25°C in 17 day old embryonic chicken liver and 9.5-11.3 units/g (net weight) tissue at 37°C in adult human liver. If one assumes that the enzyme activity provides a rough guide to the relative levels of pyruvate carboxylase mRNA, then embryonic chicken liver would have 8–9 fold higher levels than the adult human liver (assuming a 10% increase in enzymic activity for each °C increase in assay incubation temperature). If one uses the figures for the abundance of the human pyruvate carboxylase cDNA clones (0.006%) and

the estimated increased level of pyruvate carboxylase mRNA in embryonic chicken liver then the maximum expected level of abundance of the embryonic chicken liver pyruvate carboxylase clones is 0.042%. This is over 10 fold less than the estimates made when the study began (Chapter 3). It was therefore apparent that the embryonic chicken liver cDNA library in pBR322 containing 1100 recombinants described in Chapter 4 was probably too small to include any pyruvate carboxylase cDNA clones.

A requirement for this study to proceed would therefore be a new and much larger chicken liver cDNA library. Advances in the construction of cDNA libraries using the λgt10 and λgt11 vectors have resulted in libraries of substantially greater size than has been possible using plasmid vectors. This was made possible because the packaging of the recombinant λ DNA into infectious viral particles occurs with up to 10^3 fold greater efficiency than the standard $CaCl_2$ plasmid transformation technique (Cohen *et al.*, 1972) which routinely yields a transformation efficiency of 1×10^6 transformants/μg of DNA. The other advantage of cDNA libraries in λ vectors is that they may be readily screened at very high plaque densities and non-abundant clones may be thereby detected more efficiently. This combination of factors made the construction of cDNA libraries in λ vectors the method of choice for use in the isolation of chicken pyruvate carboxylase clones.

The screening of such recombinant DNA libraries for chicken pyruvate carboxylase clones required a probe which would selectively detect these clones. The options available in selecting a probe would be (i) to use a mixed oligonucleotide probe directed to the biotin attachment site as Freytag and Collier (1984) had successfully done, (ii) use an oligonucleotide probe directed to another region of the chicken pyruvate carboxylase sequence determined by the protein sequencing in Chapter 3, (iii) use the human pyruvate carboxylase clone described by Freytag and Collier directly as a probe. The results in this chapter describe the construction of cDNA libraries and the choice of the human pyruvate carboxylase clone PC34.1 as a probe to screen these cDNA libraries. In addition the screening of two chicken genomic libraries is described. The inability to isolate chicken pyruvate carboxylase clones led to a change in the species under investigation from chicken to rat. The

identification of rat pyruvate carboxylase cDNA clones is described as well as the characterization of the DNA sequence of these clones is presented.

5.2 Specific Methods

5.2.1 cDNA Library Construction

Chicken liver cDNA libraries were constructed and were cloned in the λgt10 vector using a modification of the procedure described by Huynh *et al.* (1985).

5.2.1.a First Strand cDNA Synthesis

The first strand cDNA synthesis was performed in a 50 μl reaction mixture containing 10 μg poly(A)$^+$ RNA, 25 ng/μl oligo d(T)$_{10}$ primer, 0.1 μg/μl BSA, 0.5 mM each of dATP, dCTP, dGTP and TTP, 50 mM Tris·HCl pH 8.3, 100 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 50 units AMV-reverse transcriptase. The reaction mixture was heated at 65°C for 3 min then cooled on ice prior to the addition of the enzyme. The reaction mixture was incubated for 50 min at 37°C then for 10 min at 43°C. The RNA/cDNA hybrids were recovered after phenol/chloroform extraction and ethanol precipitation.

5.2.1.b Second Strand cDNA Synthesis

The RNA/cDNA hybrid synthesized in the first reaction (Section 5.2.1.a) was dried after ethanol precipitation and 5 μl of [α-^{32}P] dCTP (5 μCi/μl) was added and this was also dried under vacuum. The second cDNA strand was synthesized in a 100 μl reaction mixture containing the above components as well as 100 mM KCl, 10 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 20 mM Tris·HCl pH 7.5, 30 μg/ml BSA, 100 μM DTT, 0.2 mM of each of dATP, dGTP and TTP, 3 units *E. coli* DNA polymerase I and 1 unit RNase H. The reaction mixture was incubated for 1 hr at 13°C then for 1 hr at room temperature (about 24°C). The resulting double stranded cDNA was recovered after phenol/chloroform extraction and ethanol precipitation.

5.2.1.c End Filling

The cDNA was end-filled in a 20 μl reaction mixture containing 10 mM Tris·HCl pH 7.5, 10 mM $MgCl_2$, 50 μM of each of dATP, dCTP, dGTP and TTP, 1 mM DTT and 4 units Klenow. The reaction mixture was incubated for 30 min at room temperature and the cDNA recovered after phenol/chloroform extraction and ethanol precipitation.

5.2.1.d Methylation of *Eco*RI Sites

Approximately 100 ng of the cDNA was taken and the *Eco*RI sites were methylated in a 20 μl reaction mixture containing 100 mM Tris·HCl pH 8.0, 10 mM EDTA, 80 μM S-adenosylmethionine, 100 μg/ml BSA and 10 units *Eco*RI methylase. The reaction mixture was incubated for 30 min at 37°C and the cDNA was recovered after phenol/chloroform extraction and ethanol precipitation.

5.2.1.e Cloning of cDNA

*Eco*RI linkers were phosphorylated at their 5′ ends with unlabelled ATP using the procedure described in Section 2.2.18.a and unincorporated ATP was separated by chromatography on Sephadex G-50. The phosphorylated linkers were ligated to the cDNA (Section 2.2.16) and the cDNA was recovered from the reaction mixture after phenol/chloroform extraction and ethanol precipitation. The cDNA was digested with an excess of *Eco*RI and size-fractionated on a 1.0% (w/v) agarose gel. Only cDNA >500 bp was recovered and the isolated DNA was recovered by electroelution from the gel slice (Section 2.2.10). The cDNA was then ligated into *Eco*RI digested λgt10 arms (Stratagene Cloning Systems) and the ligated cDNA/λgt10 was packaged into infectious λ virions using the Gigapack *in vitro* packaging system (Stratagene Cloning Systems). The recombinant bacteriophages resulting from this procedure could then be plated directly on *E. coli* C600Hfl$^-$ cells.

5.2.2 Amplification of λgt10 cDNA Libraries

Libraries of recombinant chicken liver cDNA in λgt10 were plated out using *E. coli* C600Hfl$^-$ cells on a sufficient number of 15 cm diameter NZCYM plates such that the entire library was present at a density of approximately 10,000 pfu/plate. The plates were incubated overnight at 37°C and the resulting plaques were observed to be discrete although touching each other. The plates were overlayed with 5 ml of PSB and stored for 1 hr at 4°C with tilting at about 10 min intervals to ensure that all plaques were adequately covered by the PSB. The bacteriophage stock was drawn off the plates using a pasteur pipette, mixed and centrifuged for 5 min at 3,000×g to remove cellular debris. The supernatant was taken and stored at 4°C with about 400 μl chloroform as an amplified library stock. The stocks of the amplified libraries had bacteriophage titres in the range 0.7–3.0×10^{10} pfu/ml.

5.2.3 Tissue Culture

3T3-L1 cells were grown in Dulbecco's modified essential medium (DMEM) with 10% foetal calf serum in T-75 tissue culture bottles. Cells were induced to differentiate in DMEM with 5% foetal calf serum for 2 days in the presence of 0.25 μM insulin, 0.5 mM methyl isobutyl xanthine and 0.25 μM dexamethasone. This medium was then replaced by DMEM with 5% foetal calf serum and 0.25 μM insulin for 3 days prior to harvesting the cells. Undifferentiated cells were maintained in DMEM with 10% foetal calf serum throughout. Cells were carefully washed with PBS solution then 2 ml of trypsin solution containing 1 mg/ml trypsin, 1×Versene buffer, 0.2 μg/ml phenol red was then added to detach the cells from the culture vessel. The trypsin solution was left for 10 min at room temperature whereupon 15 ml of 6 M guanidine·HCl RNA extraction buffer was added to the cell suspension. The preparation of RNA from the cells was carried out as described in Section 2.2.1 and 2.2.2.

5.2.4 RNA Dot Blots

RNA samples were dissolved in 100 μl of TE, 300 μl of 10×SSC, 50% (v/v) formaldehyde was added and they were denatured by heating at 65°C for 5 minutes. The RNA samples were cooled on ice and then filtered through nitrocellulose wetted with 10×SSC, using a Minifold dot blot apparatus connected to a vacuum pump. The dots were washed with 400 μl 10×SSC and the RNA was fixed onto the filter by baking for 2 hours at 80°C under vacuum.

5.3 Results

The report by Freytag and Collier (1984) of the partial sequence of a cDNA clone encoding human liver pyruvate carboxylase represented the isolation of the first recombinant DNA clone for pyruvate carboxylase from any source. Although the sequence provided by these workers was limited to 152 bp at the 3′ end of the coding sequence, the inferred translation of this sequence identified the biotin attachment site lysine (biocytin) 35 amino acids from the C-terminus. The position of the biocytin was conserved between human pyruvate carboxylase and the biotin-containing subunits of *E. coli* acetyl-CoA carboxylase (Sutton *et al.*, 1977) and *P. shermanii* transcarboxylase (Maloy *et al.*, 1979). Very strong sequence homology is apparent between human liver pyruvate carboxylase and the sequence of the biotinyl peptides of chicken, turkey and sheep pyruvate carboxylase (Rylatt *et al.*, 1977) and this is shown in Fig. 5.1. The degree of sequence homology between human liver pyruvate carboxylase and chicken liver pyruvate carboxylase suggested the possibility of using a human liver pyruvate carboxylase cDNA clone to probe for chicken pyruvate carboxylase clones.

5.3.1 Sequence Analysis of Human Liver Pyruvate Carboxylase cDNA Clone

The human liver pyruvate carboxylase cDNA clone PC34.1 was kindly supplied by Dr. R.A. Gravel (Toronto, Ontario, Canada). PC34.1 was isolated from a human

Figure 5.1

Comparison of Inferred Human Pyruvate Carboxylase Amino Acid Sequence to Chicken Pyruvate Carboxylase Biotinyl Peptide Sequence

The inferred amino acid sequence of human liver pyruvate carboxylase in the vicinity of the biotin attachment site is shown aligned with the amino acid sequence of the chicken biotinyl peptide (Rylatt *et al.*, 1977). Where there is sequence identity the residues are boxed and the lysine to which biotin is attached is boxed and marked by (*).

											*									
Human PC	Gly	Gln	Pro	Leu	Cys	Val	Leu	Ser	Ala	Met	Lys	Met	Glu	Thr	Val	Val	Thr	Ser	Pro	Met
Chicken PC	Gly	Ala	Pro	Leu	—	Val	Leu	Ser	Ala	Met	Lys	Met	Glu	Thr	Val	Val	Thr	Ala	Pro	Arg

liver cDNA library cloned into the Okayama and Berg (1982) pcD vector and includes all the cDNA present in the cDNA clone phPC1 described by Freytag and Collier (1984) with an additional 700 bp extending in the 5′ direction. The restriction map of PC34.1 is shown aligned with the map of the phPC1 clone in Fig. 5.2. The restriction maps obtained for each clone included some differerent sites, so to firmly establish the identity of the PC34.1, partial sequence analysis of the insert was undertaken.

A 390 bp *Pvu*II restriction fragment was isolated from PC34.1, subcloned into M13mp8 (Section 2.2.16) and the insert was partially sequencing the insert using the dideoxy chain-termination sequencing system (Sections 2.2.22 and 2.2.23). The extent and orientation of the sequences obtained are shown in Fig. 5.2a. The sequence shown in Fig. 5.2b contained the biotin attachment site sequence and confirmed the data of Freytag and Collier (1984). The sequence also agreed with the preliminary sequence data provided by Dr. R.A. Gravel.

In subsequent studies where the human pyruvate carboxylase cDNA from PC34.1 was used as a probe, two restriction fragments were employed and these are shown in Fig. 5.2a. One fragment of 1.8 kb was excised from the clone with *Bam*HI and contained the entire insert as well as 160 bp of the pcD vector DNA. The insert could not be excised in one piece using the cloning site enzyme *Pst*I without including any vector sequence because; (i) there are two internal *Pst*I sites in the insert and (ii) the end of the cDNA insert representing the 3′ end of the mRNA does not regenerate a *Pst*I site upon ligation of the vector and insert DNAs. The 1.1 kb fragment which was also used as a probe was excised from PC34.1 with *Stu*I and *Bgl*II. This fragment contained the 3′ end of the coding sequence as well as the biotin attachment site and was wholly internal to the insert.

5.3.2 Northern Transfer of Chicken Liver RNA Probed with PC34.1 Insert

For the human liver pyruvate carboxylase cDNA clone PC34.1 to be useful as a probe to detect chicken liver pyruvate carboxylase clones, it was necessary to demonstrate

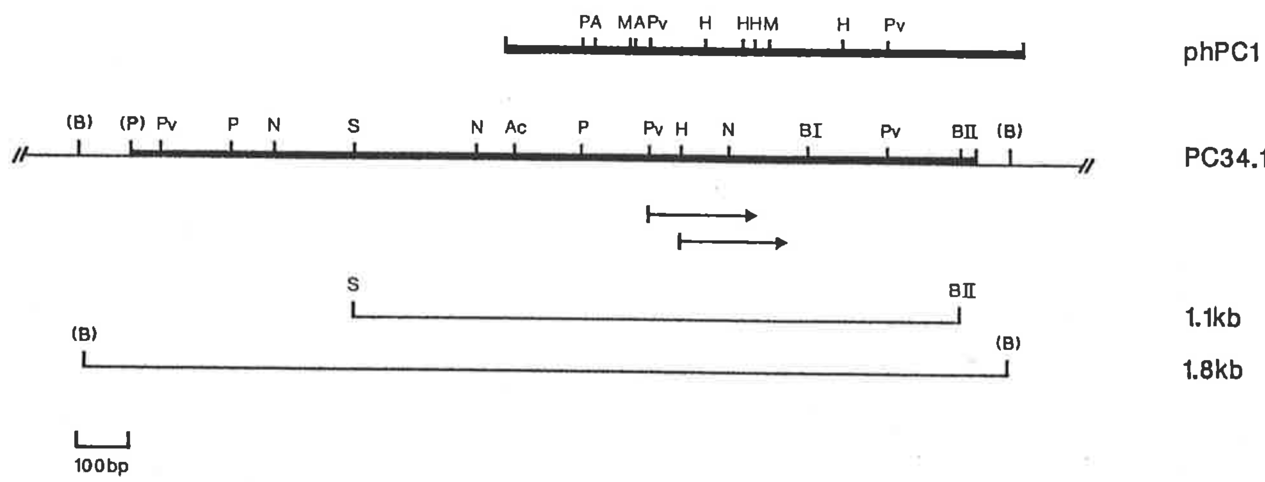
Figure 5.2

Restriction Maps and Sequence of Human Liver Pyruvate Carboxylase cDNA Clones

a The restriction map of the human pyruvate carboxylase cDNA clone phPC1 (Freytag and Collier, 1984) is shown aligned with the restriction map of the human pyruvate carboxylase cDNA clone PC34.1 kindly provided by Dr. R.A. Gravel. A 390 bp *Pvu*II fragment of PC34.1 was subcloned into M13mp8 and sequenced using the dideoxy chain termination method (Sections 2.2.22 and 2.2.23). The extent and orientation of the sequences are shown as arrows. The 1.1 kb *Stu*I/*Bgl*II fragment and the 1.8 kb *Bam*HI fragment which were subsequently used as probes are shown. A = *Alu*I, Ac = *Acc*I, B = *Bam*HI, BI = *Bgl*I, BII = *Bgl*II, H = *Hae*III, M = *Mbo*II, N = *Nco*I, P = *Pst*I, Pv = *Pvu*II, S = *Stu*I.

b Partial sequence of the PC34.1 clone insert. The sequence of 291 bp of the human pyruvate carboxylase cDNA clone insert was determined from a 390 bp *Pvu*II restriction fragment using the dideoxy chain-termination method (Sections 2.2.22 and 2.2.23). The sequence is shown together with the inferred amino acid sequence. The sequence described by Freytag and Collier (1984) is italicized and the 5′ end of this sequence is marked by (▼).

a



b

gln leu arg ser ile leu val lys asp thr gln ala met lys glu met his phe his pro lys ala leu lys asp val lys gly gln ile
CAGCTGCGGTCCATCTTGGTCAAGGACACCCAGGCAATGAAGGAGATGCACTTCCACCCCAAGGCCCTAAAGGACGTGAAGGGCCAGATC
10 20 30 40 50 60 70 80 90

gly ala pro met pro gly lys val ile asp ile lys val val ala gly ala lys val ala lys gly ▼ *gln pro leu cys val leu ser ala*
GGGGCGCCCATGCCTGGGAAGGTGATAGACATCAAAGTGGTGGCAGGGGCCAAGGTGGCCAAGGG*CCAGCCCCTGTGTGTGCTCAGTGCC*
100 110 120 130 140 150 160 170 180

met [*lys*] *met glu thr val val thr ser pro met glu gly thr val arg lys val his val thr lys asp met thr leu glu gly asp asp*
ATGAAGATGGAGACTGTGGTGACCTCACCCATGGAGGGTACTGTCCGCAAGGTTCATGTGACCAAGGACATGACACTGGAAGGTGACGAC
190 200 210 220 230 240 250 260 270

leu ile leu glu ile glu *
CTCATCCTGGAGATCGAGTGA
280 290

that PC34.1 would hybridize in a specific manner to the chicken liver pyruvate carboxylase mRNA. The 1.8 kb *Bam*HI fragment was isolated from the PC34.1 clone, [^{32}P] labelled by nick-translation (Section 2.2.18.c) and hybridized using the conditions described in Section 2.2.21.b to a Northern transfer of chicken liver RNA samples (Sections 2.2.9.b and 2.2.11.b). The filter was washed at low stringency (1×SSC, 0.1%SDS at 37°C) because the level of DNA sequence homology between the two species was unknown. To establish if any hybridization signal which might be detected was the result of hybridization through the vector sequences present in the 1.8 kb probe an identical Northern transfer of chicken liver RNA was probed with nick-translated pBR322 vector DNA under the same conditions.

As can be seen in Fig. 5.3a, PC34.1 hybridized strongly to an RNA species in both 18 day embryonic chicken liver poly(A)$^+$ RNA and 4 day old hatched chicken liver poly(A)$^+$ RNA. This band was determined to be 4.5 kb by comparison with the DNA size markers using the SIZING program (Section 2.2.24). As expected the pBR322 DNA failed to hybridize to any mRNA species (Fig. 5.3b). The level of hybridization to the post-hatching RNA preparation was considerably less than to the embryonic RNA preparation. The reduction in the level of this RNA species after hatching was consistent with the reduced level of chicken liver pyruvate carboxylase enzymic activity described by Arinze and Mistry (1970) and discussed in Chapter 3. The size of the mRNA species which hybridized was significantly greater than might be estimated (3300–3600 bases) if one only considered the size of the protein. However, the mRNA species detected by Freytag and Collier (1984) using phPC1 to probe Northern transfers of rat, baboon and human liver RNAs were 4.2 kb in all these species. As these workers found, the length of the 3′ untranslated region in the human pyruvate carboxylase mRNA was about 400 bases and this contributed to the larger than expected size of the mRNA. Therefore a size of 4.5 kb was not unreasonable for the chicken liver pyruvate carboxylase mRNA although it is larger than the mRNAs observed in these other species.

Figure 5.3

Northern Transfers of Chicken Liver RNA Probed with PC34.1 Insert and pBR322

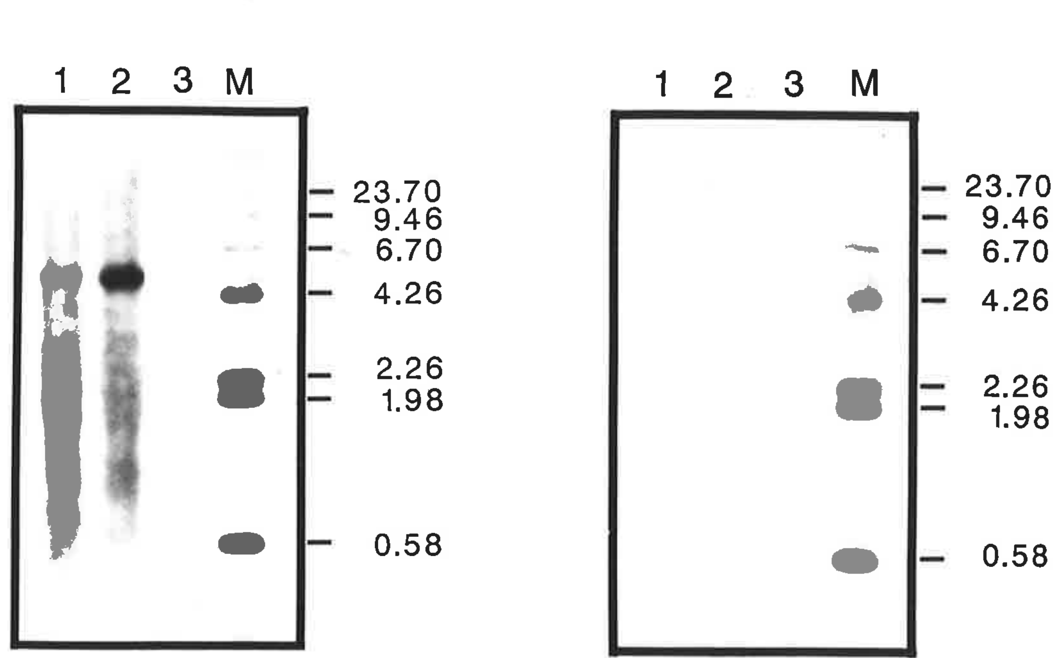
Duplicate samples of 10 μg of chicken liver poly(A)$^+$ RNA prepared from 18 day old embryonic chickens and from 4 day old hatched chickens as well as 10 μg of 18 day old embryonic chicken liver poly(A)$^-$ RNA (Sections 2.2.1 and 2.2.2) were glyoxalated and electrophoresed with glyoxalated [^{32}P] labelled (Section 2.2.18.b) DNA size markers on 1% (w/v) agarose, 10 mM sodium phosphate pH 6.5 gels (Section 2.2.9.b). The RNA was transferred to nitrocellulose using the Northern procedure (Section 2.2.11.b) and the two identical filters were hybridized to the indicated DNA probes (Section 2.2.21.b). The filters were washed in 1$\times$SSC, 0.1% SDS at 37°C and autoradiographed. The DNA marker bands are shown in kb.

Track 1 : 4 day old hatched chicken liver poly(A)$^+$ RNA,
Track 2 : 18 day old embryonic chicken liver poly(A)$^+$ RNA,
Track 3 : 18 day old embryonic chicken liver poly(A)$^-$ RNA,
Track M : [^{32}P] end filled λ_{wt} *Hin*dIII markers.

a Northern transfer probed with the 1.8 kb *Bam*HI fragment from PC34.1 which was [^{32}P] labelled by nick-translation (Section 2.2.18.c).

b Northern transfer probed with pBR322 which was [^{32}P] labelled by nick-translation (Section 2.2.18.c).

a
b


1 2 3 M
23.70
9.46
6.70
4.26
2.26
1.98
0.58
1 2 3 M
23.70
9.46
6.70
4.26
2.26
1.98
0.58

5.3.3 Choice of Probes to Detect Chicken Pyruvate Carboxylase Clones

The result of the Northern analysis described above provided a firm basis for the use of PC34.1 as a probe to detect chicken liver pyruvate carboxylase clones. The other options available were those mentioned in Section 5.1; either the use of a mixed oligonucleotide probe directed to the biotin attachment site or the synthesis of new oligonucleotide probes inferred from other regions of the protein sequence described in Chapter 3. The successful use of the mixed 14 base biotin site probe by Freytag and Collier (1984) resulted in the adoption of this type of mixed biotin site probe. A mixed 14 base oligonucleotide probe directed to the chicken liver pyruvate carboxylase biotin attachment site was synthesized as a mixture of four oligonucleotides and was termed B-14M. Since the estimated codon usage was not used to determine the sequence of this mixture of oligonucleotides, one out of the four possibilities must hybridize to the correct coding sequence for chicken pyruvate carboxylase. The option of synthesizing additional oligonucleotides directed to alternative protein sequences was not exercised in view of the specific nature of the hybidization by the PC34.1 probe.

5.3.4 Northern Transfer of Chicken Liver RNA Probed Human Propionyl-CoA Carboxylase α Subunit cDNA Clone Insert

Careful examination of the Northern transfer autoradiograph shown in Fig. 5.3a revealed several smaller bands which hybridized very weakly to the 1.8 kb probe. In view of the low stringency used in the washing of the Northern filter the possibility that cross-hybridization could have occurred to related biotin carboxylases was investigated. Reports of Northern analyses using cDNA clones for the α and β subunits of human propionyl-CoA carboxylase (Lamhonwah *et al.* 1986) and the β subunit of rat propionyl-CoA carboxylase (Kraus *et al.*, 1986a) as probes had not noted any cross-hybridization by either subunit clone to the pyruvate carboxylase

mRNA. To investigate the possibility that cross-hybridization to propionyl-CoA carboxylase mRNA was occurring in the Northern transfer shown in Fig. 5.3a, the α subunit clone pPCC9-5 was kindly supplied by Dr. R.A. Gravel (Toronto, Ontario, Canada). The study was confined to the α subunit in view of the significant amount of sequence homology apparent between pyruvate carboxylase and the propionyl-CoA carboxylase α subunit noted by Lamhonwah *et al.* (1986).

The map of pPCC9-5, which was cloned into the pcD vector, is shown in Fig. 5.4 and the 2.2 kb insert together with ~160 bp of the vector DNA was excised with *Bam*HI. The fragment was [^{32}P] labelled by nick-translation and used to probe a Northern transfer of chicken liver RNA. The filter was washed at low stringency (1×SSC, 0.1%SDS at 37°C) then autoradiographed (Fig. 5.5b). A further Northern transfer of chicken liver RNA was probed with the 1.8 kb human pyruvate carboxylase clone fragment under the same conditions (Fig. 5.5a). It was apparent that the pPCC9-5 insert fragment hybridized to three mRNA species which were determined to be 3.3 kb, 2.2 kb and 1.7 kb long using the SIZING program (Section 2.2.24). The 1.8 kb PC34.1 fragment also hybridized weakly to the 3.3 kb and 2.2 kb mRNAs. Although the human pyruvate carboxylase probe hybridized to these mRNA species the human propionyl-CoA carboxylase α subunit clone did not hybridize to a significant extent to the chicken pyruvate carboxylase mRNA. These bands appeared to be the result of specific hybridization because the pBR322 probe failed to detect non-specific hybridization to any species. The significantly increased hybridization intensity of the 3.3 kb and 2.2 kb bands observed when the propionyl-CoA carboxylase probe was used suggest that these bands may represent the mRNA species for chicken propionyl-CoA carboxylase. This appears to be the first instance of cross-hybridization of pyruvate carboxylase and propionyl-CoA carboxylase probes and their mRNAs.

Figure 5.4

Restriction Map of Human Propionyl-CoA Carboxylase α Subunit cDNA Clone

The restriction map of the human propionyl-CoA carboxylase cDNA clone, pPCC9-5 which was kindly provided by Dr. R.A. Gravel is shown (Lamhonwah *et al.*, 1986). The 2.3 kb *Bam*HI fragment which was subsequently used as a probe is indicated. AI = *Ava*II, AII = *Ava*II, B = *Bam*HI, Hf = *Hin*fI, P = *Pst*I, Pv = *Pvu*II, St = *Sst*I, Sp = *Sph*I, T = *Tth*111I, XI = *Xho*I, XII = *Xho*II.

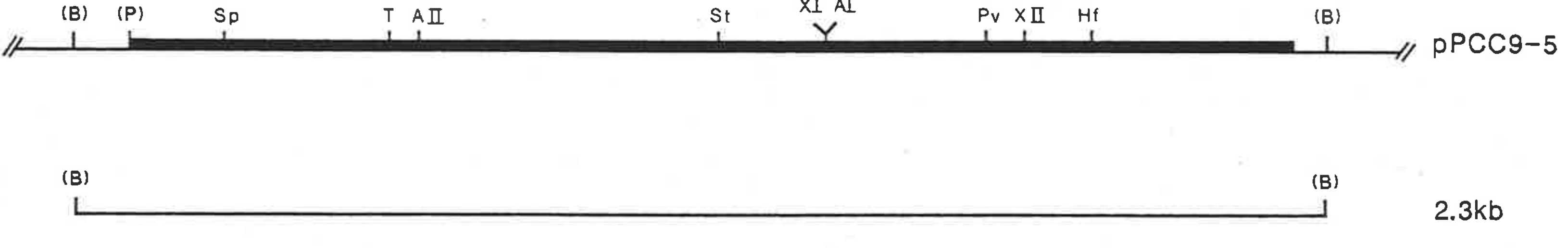

(B)
(P)
Sp
T
AII
St
XI
AI
Pv
XII
Hf
(B)
pPCC9-5
(B)
(B)
2.3kb
100bp

Figure 5.5

Northern Transfers of Chicken Liver RNA Probed with PC34.1 Insert and pPCC9-5 Insert

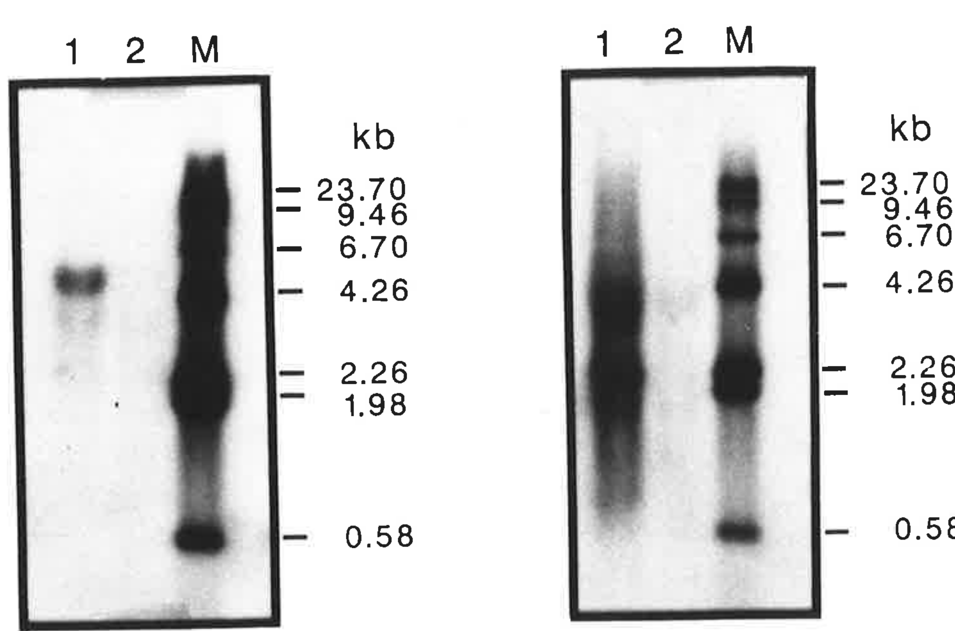
Duplicate samples of 10 μg of poly(A)$^+$ RNA and 10 μg of poly(A)$^-$ RNA prepared from 18 day old embryonic chicken liver (Sections 2.2.1 and 2.2.2) were glyoxalated and electrophoresed with glyoxalated [^{32}P] labelled (Section 2.2.18.b) DNA size markers on 1% (w/v) agarose, 10 mM sodium phosphate pH 6.5 gels (Section 2.2.9.b). The RNA was transferred to nitrocellulose using the Northern procedure (Section 2.2.11.b) and the two identical filters were hybridized to the indicated probe DNA (Section 2.2.21.b). The filters were washed in 1×SSC, 0.1% SDS at 37°C and autoradiographed. The sizes of the marker bands are shown in kb.

Track 1 : 18 day old embryonic chicken liver poly(A)$^+$ RNA,
Track 2 : 18 day old embryonic chicken liver poly(A)$^-$ RNA,
Track M : [^{32}P] end filled λ_{wt} *Hin*dIII markers.

a Northern transfer probed with the 1.8 kb *Bam*HI fragment from PC34.1 which was [^{32}P] labelled by nick-translation (Section 2.2.18.c).

b Northern transfer probed with the 2.3 kb *Bam*HI fragment from pPCC9-5 which was [^{32}P] labelled by nick-translation (Section 2.2.18.c).

a
b


1
2
M
kb
23.70
9.46
6.70
4.26
2.26
1.98
0.58
1
2
M
kb
23.70
9.46
6.70
4.26
2.26
1.98
0.58

5.3.5 cDNA Library Construction and Library Screening with Human Liver Pyruvate Carboxylase cDNA Clone

5.3.5.a Screening Plasmid cDNA Library with PC34.1

In view of the specific selection of an appropriately sized chicken liver mRNA species by the human liver pyruvate carboxylase cDNA clone PC34.1 it was decided to use this clone as a probe to screen libraries of chicken recombinant DNA for chicken liver pyruvate carboxylase cDNA clones. The plasmid cDNA library prepared from 18 day old embryonic chicken liver RNA (Hobbs *et al.*, 1986) described in Section 4.3.6 was extensively screened using the 1.1 kb fragment from PC34.1 shown in Fig. 5.2a as the probe. However, no strong hybridization signals were obtained and it therefore appeared unlikely that this library contained clones representing pyruvate carboxylase. The relatively small size of the plasmid cDNA library with only 1100 recombinants limited the clones represented in the library to those with quite high mRNA abundancies. In addition the possibility that the AIA and DDC drug treatment of the embryos used as the source of mRNA in the construction of this plasmid library described in Section 4.3.6 could have affected the level of pyruvate carboxylase mRNA could not be discounted. Further, the pyruvate carboxylase mRNA was expected to have a rather lower level of abundance in the embryonic chicken liver than originally predicted, as discussed in Section 5.1.

5.3.5.b Embryonic Chicken Liver cDNA Library Construction and Characterization

As a result of the inadequate size of the plasmid cDNA library a cDNA library was constructed using the λgt10 vector which has been used to achieve very large library sizes (Huynh *et al.*, 1985). As the drug treatment of the embryos was not required in this study, non-drug treated embryos were used as the source of RNA in constructing the λgt10 libraries. The procedure used to construct cDNA libraries was a modification of the procedure of Huynh *et al.* (1985) and is described in Section 5.2.1. RNase H was used in the second strand cDNA synthesis because it has been reported to result in the efficient synthesis of the full length of the second

cDNA strand (Okayama and Berg, 1982). The double stranded cDNA was sized on a 1% agarose gel and cDNA >500 bp was used in subsequent steps. The final packaged recombinant bacteriophages were plated on *E. coli* C600Hfl$^-$ cells and the titre of the library determined. The library initially contained 125,000 recombinants and after amplification as described in Section 5.2.2 and the titre was determined to be 3×10^{10} pfu/ml. A summary of the cDNA library construction procedure is given in Fig. 5.6.

To ensure that this library contained clones representing low to moderate abundance mRNA species a probe was prepared from a cDNA clone, pCHP3 which encodes a phenobarbital-inducible cytochrome P450 (Hobbs *et al.*, 1986). The mRNA for this cytochrome P450 is present in non-drug induced 18 day old embryonic chicken liver poly(A)$^+$ RNA at an abundance of 0.01–0.05% (A. Hansen, personal communication). The pCHP3 clone DNA was kindly supplied by Dr. A. Hobbs and a 2365 bp *Sac*I/*Nde*I insert fragment was used to screen 75,000 pfu from the amplified library (Section 2.2.19). If the clones in the library adequately represented the estimated abundance of the pCHP3-specific mRNA, 7–35 positive signals should have been detected. Two strong positive signals were obtained as well as 8 lower intensity signals (data not shown) which indicated that the library contained approximately the expected number of clones and therefore the library could be expected to include clones representing low abundance mRNAs.

5.3.5.c Screening λgt10 cDNA Library with PC34.1

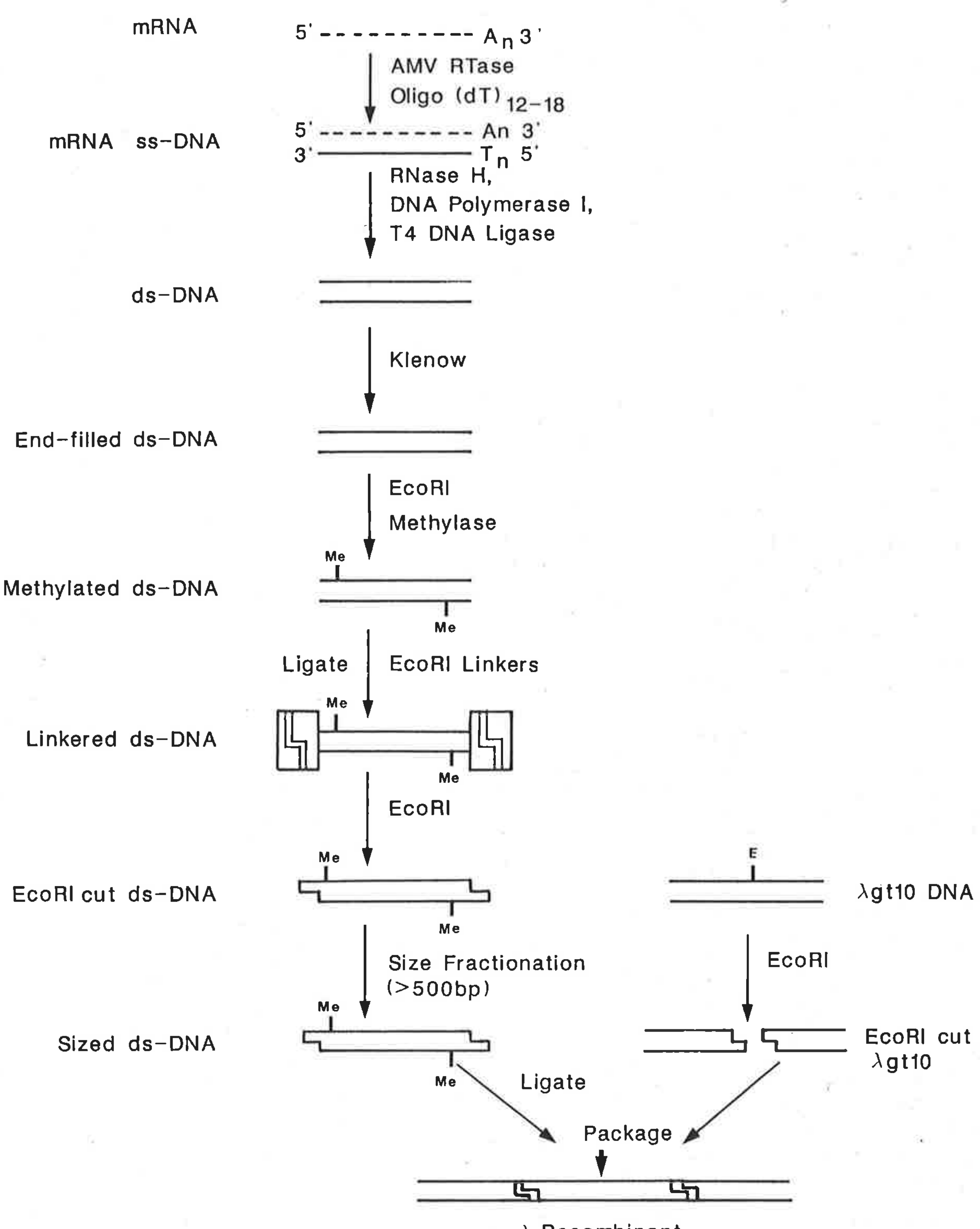
Both the 1.1 kb *Stu*I/*Bgl*II fragment and the 1.8 kb *Bam*HI fragment of the human liver pyruvate carboxylase clone PC34.1 were used as probes to screen the λgt10 cDNA library. In 6 separate screenings $> 1 \times 10^6$ pfu from the library were probed and approximately 100 low intensity signals were obtained. DNA prepared from the clones isolated from these positive signals was examined in two ways to determine if they were likely to encode pyruvate carboxylase.

Firstly, the DNA was probed with the 5′ [^{32}P] labelled mixed biotin site oligonucleotide, B-14M. None of the clones hybridized to this probe indicating that the biotin site was not present. As a second identification the inserts from the stongest hybridizing

Figure 5.6

Construction of Chicken Liver cDNA Libraries in λgt10

A flow-scheme for the construction of cDNA libraries in the λgt10 vector is shown and is modified from the scheme described by Huynh *et al.* (1985).


mRNA
5' An 3'
AMV RTase
Oligo (dT) 12-18
mRNA ss-DNA
5' An 3'
3' Tn 5'
RNase H,
DNA Polymerase I,
T4 DNA Ligase
ds-DNA
Klenow
End-filled ds-DNA
EcoRI
Methylase
Me
Methylated ds-DNA
Me
Ligate
EcoRI Linkers
Me
Linkered ds-DNA
Me
EcoRI
Me
EcoRI cut ds-DNA
Me
E
λgt10 DNA
Size Fractionation
(>500bp)
EcoRI
Me
Sized ds-DNA
Me
EcoRI cut
λgt10
Ligate
Package
λ Recombinant

clones were isolated and used to probe a Northern transfer of 18 day old embryonic chicken liver poly(A)$^+$ RNA to determine the size of the mRNA species from which they were derived. The resulting Northern transfers showed that the clones were derived from several different mRNA species all of which were <3 kb. None of the clone inserts hybridized to a 4.5 kb mRNA species and therefore it was unlikely that any of these clones encoded chicken liver pyruvate carboxylase. Presumably these clones were detected by the PC34.1 probes as a result of a low-level of sequence homology which was stable under the relatively non-stringent washing conditions employed in the screening.

It was apparent therefore that this library also did not contain clones representing chicken liver pyruvate carboxylase. To ascertain if this was a specific characteristic of this library, two additional cDNA libraries were prepared in the same manner in the λgt10 vector. One library was prepared from 18 day old embryonic chicken liver poly(A)$^+$ RNA and contained 70,000 recombinants while the second library was prepared from 4 day old hatched chicken liver poly(A)$^+$ RNA and contained 60,000 recombinants. Both this libraries were exhaustively screened with 1.8 kb fragment PC34.1 probe but no clones representing chicken liver pyruvate carboxylase were identified.

The sizes of all the cDNA libraries were such that even low abundance RNA species should have been represented therefore it was surprising that pyruvate carboxylase clones were not present. This suggested that other factors in the library construction besides mRNA abundance were responsible. In an effort to avoid the apparent lack of pyruvate carboxylase clones in these cDNA libraries the area of study turned to an examination of chicken genomic libraries.

5.3.6 Screening of Chicken Genomic Libraries with the Human Pyruvate Carboxylase cDNA Clone PC34.1

Two chicken genomic libraries were screened for the chicken pyruvate carboxylase gene. One was the library cloned in the λCharon 4A vector which was described in Chapter 4 (Dodgson *et al.*, 1979) and the other was a library kindly supplied by

Dr. J.D. Engels which was prepared from *Sau*3AI partially digested chicken DNA cloned into the *Bam*HI site of the λEMBL3 vector. A total of 8 genome-equivalents of the λCharon 4A library and 12 genome-equivalents of the λEMBL3 library were screened with probes prepared from the human pyruvate carboxylase cDNA clone PC34.1 (Section 5.3.1). No strong positive signals were obtained but a total of 25 weak positive signals were identified and these clones were isolated. The isolated clone DNAs were spotted onto nitrocellulose and the dot blots was probed with both the 1.8 kb PC34.1 fragment and with the B-14M biotin site oligonucleotide. The clones hybridized very weakly to the human pyruvate carboxylase probe under conditions of low stringency and did not hybridize to the B-14M biotin site probe. As a result, these clones were not characterized further since it appeared that even if any of the clones encoded the pyruvate carboxylase gene, the biotin attachment site sequence which would have immediately aided their identification was not present.

Why then did extensive screening of three embryonic chicken liver cDNA libraries and two chicken genomic libraries fail to isolate pyruvate carboxylase clones when the probes used were apparently able to specifically detect the chicken pyruvate carboxylase mRNA? The probable reasons are discussed in Section 5.4.1.

5.3.7 Rationale for Changing Species Under Investigation

The repeated inability to isolate chicken pyruvate carboxylase clones from both cDNA and genomic libraries indicated that continued effort with these chicken libraries was unlikely to be immediately fruitful. Therefore at this point in the study a change was made from chicken to rat as the species of pyruvate carboxylase under investigation. The vertebrate species which have been studied to a significant degree with respect to their structural and catalytic properties in addition to chicken are rat and sheep (for a review, see Wallace, 1985). The deciding factor in the choice of rat was the ready availability of rat liver cDNA libraries and genomic libraries as opposed to the paucity of sheep libraries.

5.3.8 Screening of a Rat Liver cDNA Library with the Human Pyruvate Carboxylase cDNA Clone

A cDNA library prepared from Wistar rat liver, cloned in the λgt10 vector and containing 2.2×10^7 recombinants (Aldred *et al.*, 1984) was kindly supplied by Dr. G.J. Howlett. A total of 250,000 pfu from the library were plated on *E. coli* C600Hfl$^-$ host cells and were transferred to nitrocellulose filter disks in duplicate. The 1.1 kb *Bgl*II/*Stu*I fragment of the human liver pyruvate carboxylase clone PC34.1 was [^{32}P] labelled by nick-translation (Section 2.2.18.c) and hybridized to the filters using the conditions described in Section 2.2.21.b. The filters were washed at moderate stringency (1×SSC, 0.1% SDS at 37°C) since the degree of DNA sequence homology between the two species was unknown. Four strong autoradiographic signals were obtained amongst many lower intensity signals. The four positive clones termed λRL1.1, λRL1.2, λRL1.3 and λRL1.4, were isolated after second round screening. Restriction digestion with *Eco*RI to excise the insert revealed that λRL1.1 had an insert of ~0.9 kb which was the longest insert of these clones. The λRL1.1 clone was therefore examined in greater detail to establish its identity.

5.3.9 Induction of Pyruvate Carboxylase mRNA Synthesis in 3T3-L1 Cells Demonstrated with λRL1.1 Insert Probe

Angus *et al.* (1981) demonstrated that an increase occurs in the amount of translatable mRNA for pyruvate carboxylase present in the mouse fibroblast cell line 3T3-L1 after hormonal induction of the cells to differentiate into adipocytes. This phenomenon was also observed by Freytag and Collier (1984) who used the human pyruvate carboxylase cDNA clone to probe for the pyruvate carboxylase mRNA and found a 23-fold increase in an autoradiographic signal from the differentiated RNA. To determine if the λRL1.1 clone could demonstrate this effect, 3T3-L1 cells were cultured in four T-75 tissue culture bottles as described in Section 5.2.3. The cells in two bottles were harvested without induction to differentiate (Section 5.2.3)

and total RNA was prepared which was then separated into poly(A)$^+$ and poly(A)$^-$ RNA fractions (Section 2.2.2). The cells in the other two culture bottles were induced to differenciate into adipocytes (Section 5.2.3) and poly(A)$^+$ and poly(A)$^-$ RNA was similarly prepared. Duplicate aliquots of 150 ng of each RNA sample were denatured with formaldehyde and dot blotted onto nitrocellulose (Section 5.2.4).

The λRL1.1 insert was excised from λRL1.1 with *Eco*RI and the insert was separated from the vector DNA by electrophoresis on a 1% (w/v) agarose, 1×TBE gel (Section 2.2.9.a). The insert DNA was isolated (Section 2.2.10), [^{32}P] labelled by nick-translation (Section 2.2.18.c) and used to probe the RNA dot blot. The filter was hybridized using the conditions described in Section 2.2.21.b, washed in 1×SSC, 0.1% SDS at 45°C then autoradiographed. The result shown in Fig. 5.7 clearly demonstrates the low basal level of pyruvate carboxylase mRNA in undifferentiated 3T3-L1 cells and the dramatic increase which occurs 2 days after the induction to differentiate. The ability of the λRL1.1 insert when used as a probe to demonstrate this effect in 3T3-L1 cells was consistent with it encoding rat liver pyruvate carboxylase.

5.3.10 Northern Transfer of Rat BRL-3A Cell RNA Probed with λRL1.1 Insert

Freytag and Collier (1984) established that the rat liver pyruvate carboxylase mRNA was 4.2 kb in length by using their human pyruvate carboxylase cDNA clone to probe a Northern transfer of rat liver RNA. Knowledge of the length of the rat liver pyruvate carboxylase mRNA was used when the λRL1.1 insert was used to probe a Northern transfer of RNA prepared from the rat liver cell line BRL-3A to establish if it encoded rat liver pyruvate carboxylase.

Poly(A)$^+$ RNA and poly(A)$^-$ RNA prepared from BRL-3A cells using the NP-40 procedure of Dalton *et al.* (1986) and oligo (dT) cellulose chromatography were gifts from D.G. Mottershead. RNA samples were denatured by glyoxalation and electrophoresed on a 1% agarose, 10 mM sodium phosphate pH 6.5 gel using the conditions described in Section 2.2.9.b. The RNA was transferred to a nitrocellulose

Figure 5.7

Dot Blot of 3T3-LI RNA Probed with λRL1.1 Insert

Aliquots of 300 ng of 3T3-LI mouse fibroblast cell RNA were denatured with formaldehyde then divided into two aliquots of 150 ng each. The RNA samples were suction filtered in duplicate onto nitrocellulose (Section 5.2.4) and the dot blot was probed with [^{32}P] labelled λRL1.1 insert DNA (Section 2.2.21.b). The filter was washed in 1$\times$SSC, 0.1% SDS at 45°C then autoradiographed using high-sensitivity Kodak X-Omat XAR-5 film.

Dot 1 : Undifferentiated 3T3-LI poly(A)$^-$ RNA,

Dot 2 : Differentiated 3T3-LI poly(A)$^-$ RNA,

Dot 3 : Undifferentiated 3T3-LI poly(A)$^+$ RNA,

Dot 4 : Differentiated 3T3-LI poly(A)$^+$ RNA.

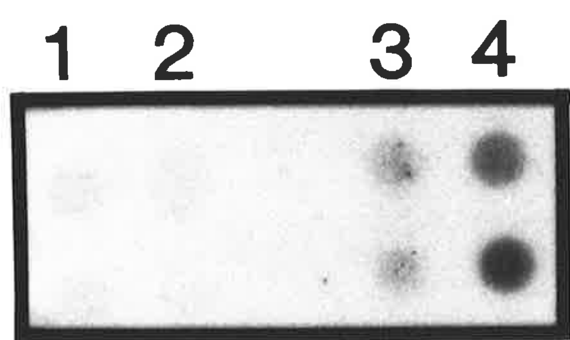
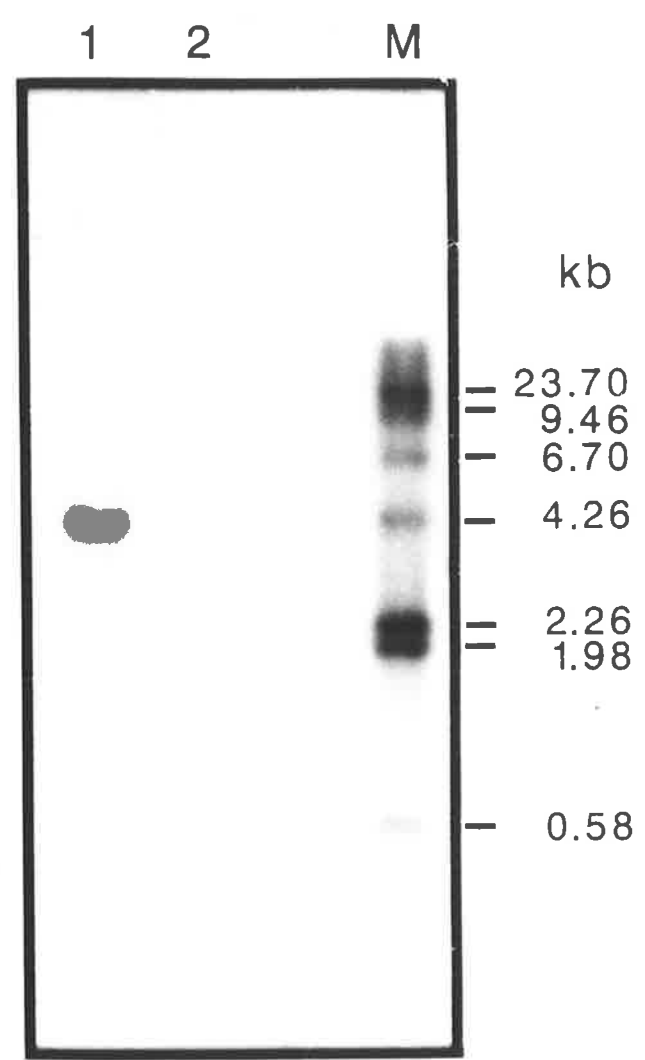

1
2
3
4

Figure 5.8

Northern Transfer of BRL-3A Cell RNA Probed with λRL1.1 Insert

Aliquots of 10 μg of poly(A)$^+$ RNA and 10 μg of poly(A)$^-$ RNA prepared from rat liver cell line BRL-3A (see Section 5.3.10) and kindly supplied by D.G. Mottershead were glyoxalated and electrophoresed with glyoxalated [^{32}P] labelled (Section 2.2.18.b) DNA size markers on a 1% (w/v) agarose, 10 mM sodium phosphate pH 6.5 gel (Section 2.2.9.b). The RNA was transferred to nitrocellulose using the Northern procedure (Section 2.2.11.b) and the filter was hybridized using the conditions described in Section 2.2.21.b to the λRL1.1 insert which was [^{32}P] labelled by nick-translation (Section 2.2.18.c). The filter was washed in 1$\times$SSC, 0.1% SDS at 65°C and autoradiographed. The sizes of the marker bands are shown in kb.

Track 1 : BRL-3A cell poly(A)$^+$ RNA,
Track 2 : BRL-3A cell poly(A)$^-$ RNA,
Track M : [^{32}P] end filled λ_{wt} *Hin*dIII markers.


1
2
M
kb
23.70
9.46
6.70
4.26
2.26
1.98
0.58

filter (Section 2.2.11.b) which was probed with [^{32}P] labelled λRL1.1 insert DNA prepared as described in Section 5.3.9. The filter was washed in 1×SSC, 0.1% SDS at 65°C then autoradiographed. The insert hybridized to a single RNA species in the BRL-3A cell poly(A)$^+$ RNA which was estimated to be 4.2 kb using the SIZING program (Section 2.2.24) (Fig. 5.8). This result supported the identification of λRL1.1 as a rat pyruvate carboxylase cDNA clone since a rat liver mRNA species of 4.2 kb was consistent with the findings of Freytag and Collier (1984).

5.3.11 Sequence Analysis of the λRL1.1 cDNA Clone Insert

The identity of λRL1.1 was unequivocally established as a rat liver pyruvate carboxylase cDNA clone by sequencing the insert. The λRL1.1 insert was digested using the restriction enzymes *Alu*I, *Nco*I, *Sau*3AI and *Rsa*I and the resulting fragments were subcloned into M13mp8 and M13mp9 and sequenced using the dideoxy chain-termination method (Sections 2.2.22 and 2.2.23). The strategy employed is shown in Fig. 5.9a and the DNA sequence of the λRL1.1 insert together with its inferred amino acid sequence is shown in Fig. 5.9b. An example of a dideoxy sequencing autoradiograph is shown in Fig. 5.10 and includes the sequence of the biotin attachment site.

The λRL1.1 clone insert contains 521 bp of coding sequence, 418 bp of 3′ non-coding sequence and a consensus polyadenylation signal, A-A-T-A-A-A, 21 nucleotides 5′ to a 54 bp long poly(A) tail. The rat pyruvate carboxylase cDNA clone 3′ untranslated region was only slightly smaller than in the human pyruvate carboxylase cDNA clone described by Freytag and Collier (1984) although these workers only gave an approximate size. The position and sequence of the polyadenylation signal in the 3′ untranslated region of the rat liver pyruvate carboxylase cDNA clones conformed to the consensus signal determined for other mRNAs (Proudfoot and Brownlee, 1976).

The overall aim of this project was to obtain clones which represented as much of the coding region of pyruvate carboxylase as possible. Since 44% of the λRL1.1 clone was 3′ non-coding sequence the rat liver cDNA library was re-screened in

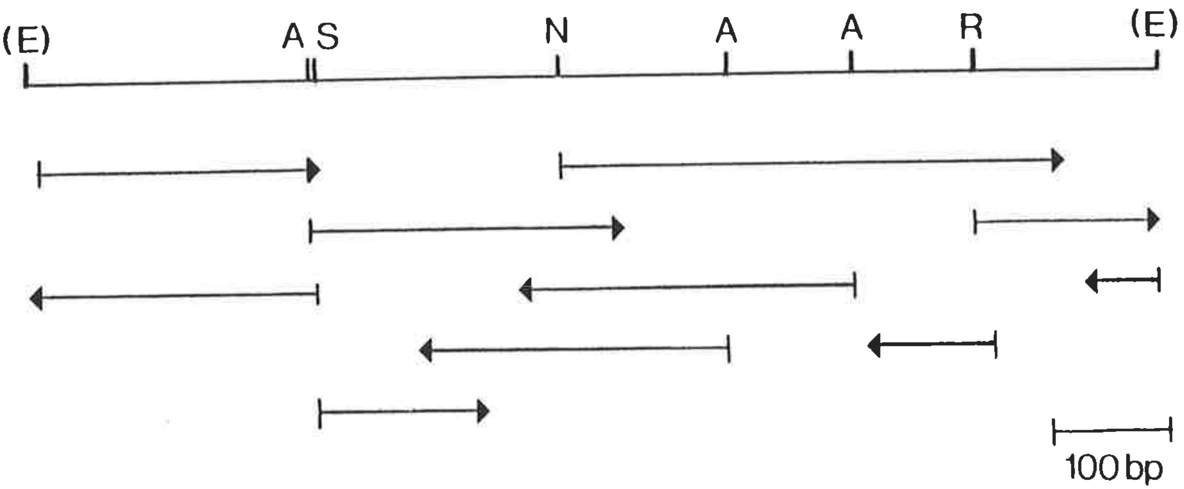
Figure 5.9

Sequencing Strategy and Sequence of λRL1.1 Insert

a Sequencing strategy of λRL1.1 insert. The λRL1.1 cDNA clone insert was digested with the restriction enzymes; *Alu*I (A), *Nco*I (N), *Rsa*I (R) and *Sau*3AI (S) and the restriction fragments generated were subcloned into either M13mp8 or M13mp9 (Section 2.2.16). These clones were sequenced using the dideoxy chain termination system (Sections 2.2.22 and 2.2.23) and the extent and orientation of each sequence is marked by an arrow. The λgt10 *Eco*RI cloning sites are indicated by (E).

b Sequence of λRL1.1 insert. The 939 bp of DNA sequence of the λRL1.1 insert is shown together with the inferred amino acid sequence. The coding region is 521 bp long and there are 418 bp of 3′ non-coding sequence including a 54 bp poly(A) tail. The termination codon is marked by (∗). The site of biotinylation is boxed and the proposed polyadenylation signal is underlined.

a



b

gly gly val pro ala ala met tyr pro asp val phe ala gln phe lys asp phe thr ala thr phe gly pro leu asp ser leu asn thr
GTGGTGGAGTTCCTGCAGCCATGTACCCTGATGTCTTTGCTCAGTTCAAAGACTTCACGGCTACCTTTGGCCCCCTGGATAGCCTGAATA
10 20 30 40 50 60 70 80 90

arg leu phe leu gln gly pro lys ile ala glu glu phe glu val glu leu glu arg gly lys thr leu his ile lys ala leu ala val
CTCGTCTCTTTCTTCAAGGACCCAAAATTGCAGAGGAGTTTGAGGTTGAGCTGGAACGGGGCAAGACCTTGCACATCAAAGCCCTGGCTG
100 110 120 130 140 150 160 170 180

ser asp leu asn arg ala ala gln arg gln val phe phe glu leu asn gly gln leu arg ser ile leu val lys asp thr gln ala met
TAAGCGACCTGAACCGTGCTGCGCAGAGGCAGGTGTTCTTTGAACTCAATGGGCAGCTTCGATCCATTCTGGTTAAAGACACCCAGGCCA
190 200 210 220 230 240 250 260 270

lys glu met his phe his pro lys ala leu lys asp val lys gly gln ile gly ala pro met pro gly lys val ile asp val lys val
TGAAGGAGATGCACTTCCATCCCAAGGCCTTGAAGGATGTGAAGGGCCAAATTGGGGCCCCTATGCCTGGGAAGGTCATAGACGTCAAGG
280 290 300 310 320 330 340 350 360

ala ala gly ala lys val val lys gly gln pro leu cys val leu ser ala met lys met glu thr val val thr ser pro met glu gly
TGGCAGCAGGAGCCAAGGTGGTTAAGGGCCAGCCCCTCTGTGTGCTCAGCGCCATGAAGATGGAGACTGTGGTGACTTCGCCCATGGAGG
370 380 390 400 410 420 430 440 450

thr ile arg lys val his val thr lys asp met thr leu glu gly asp asp leu ile leu glu ile glu *
GCACTATCCGAAAGGTTCACGTGACCAAGGACATGACTCTGGAAGGCGATGACCTCATCCTAGAGATTGAGTGATCTTACTCCAGACTGG
460 470 480 490 500 510 520 530 540

CAGCCTGGCCAACCCTACCCCAAGCCTCTCAACAGAAGCTGTGCAGCCAGGGCAGGCCCAGCGAGTACCTGAGGGCTAGGCTTGAGGTCC
550 560 570 580 590 600 610 620 630

TGTCCCATGGGCACACACACACTACCTGCAATGGCCCTCCCATTGCCCTTCAGCTATTTGTCCTTGTCTTGCTGGCAGGCAGTTCTCACA
640 650 660 670 680 690 700 710 720

TATTCTCTCTTGCCAAATAAGGGTCTACTCCTGTGGGAGACCAAGGTGTACAGTAGGTGGCCTTGTACCTGGGAGAGGGGTTCTACCTCT
730 740 750 760 770 780 790 800 810

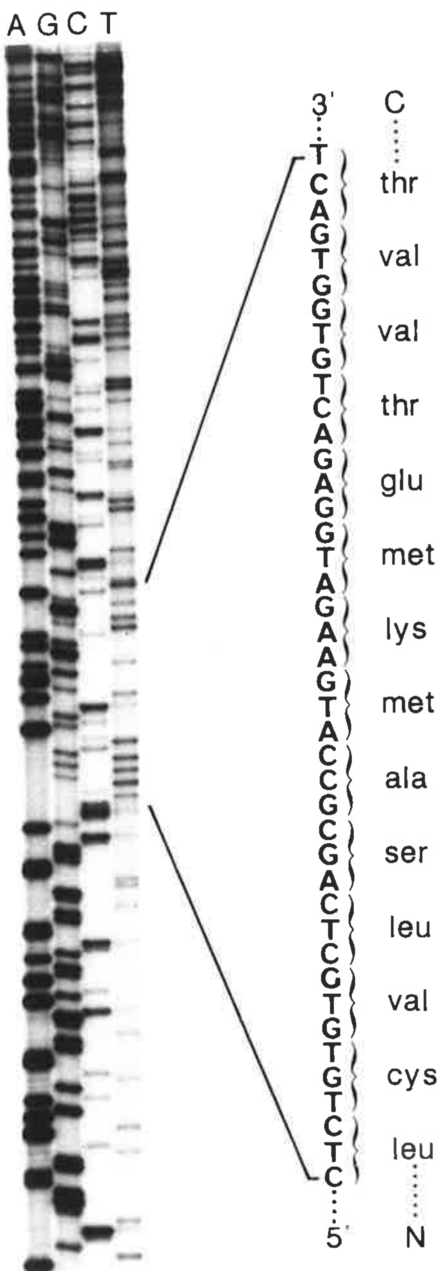
GGGGGTAGAGGGAAGAAGACCTAATTCATAGGTCCTGGGAAATTTGCTCAATAAAGTGGCCTTCCCTTGCCCTCCAAAAAAAAAAAAAAA
820 830 840 850 860 870 880 890 900

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
910 920 930 939

Figure 5.10

Dideoxy Sequencing of λRL1.1 Insert

A 342 bp *Alu*I fragment of the λRL1.1 insert was subcloned into M13mp8 and sequenced using the dideoxy chain termination system (Section 2.2.22). The sequencing reaction mixtures were electrophoresed on an 8 M urea, 6% polyacrylamide sequencing gel, fixed and autoradiographed as described in Section 2.2.23. The tracks represent the dideoxy(A, G, C, T) chain termination reactions and are labelled accordingly. The section of the autoradiograph representing the region of the clone encoding the biotin attachment site is indicated. The DNA sequence is shown at the side of the autoradiograph together with the inferred amino acid sequence.



A G C T
3'
C
T
C
A
G
T
G
G
T
G
T
C
A
G
A
G
G
T
A
G
A
A
G
T
A
C
C
G
C
G
A
C
T
C
G
T
G
T
G
T
C
T
C
5'
N
thr
val
val
thr
glu
met
lys
met
ala
ser
leu
val
cys
leu

order to isolate clones representing more of the coding region.

5.3.12 Isolation of Longer Rat Liver Pyruvate Carboxylase cDNA Clones

A total of 750,000 pfu from the rat liver cDNA library were screened using the λRL1.1 insert as the probe. The λRL1.1 insert contains the entire 3′ untranslated region of rat pyruvate carboxylase whereas the 1.1 kb human pyruvate carboxylase probe which was previously used to screen the library contains only about 100 bp of the 3′ untranslated region. Therefore the rat probe would be expected to select any short clones representing the 3′ end of the mRNA in addition to longer clones. A total of 103 positive autoradiographic signals were obtained and after a second round of screening 68 clones were isolated. To reduce the number of clones necessary to examine in order to find the clone with the insert extending the greatest furthest in the 5′ direction the clones were probed with the B-14M biotin site probe. This probe should only hybridize to clones extending at least as far as 511 bp from the 3′ end of the λRL1.1 insert. When this was done three positive signals were obtained and these clones were isolated and designated λRL2.10, λRL2.35 and λRL2.39. DNA was prepared from these positive clones in order to establish which had the longest insert.

The clone DNAs were digested with the restriction enzyme *Eco*RI to excise the inserts and the digested DNAs were separated on a 1% (w/v) agarose, 1×TBE gel (Fig. 5.11a). The DNA was transferred to nitrocellulose using the Southern transfer method (Section 2.2.11.a) then probed with the [^{32}P] labelled λRL1.1 insert. Under washing conditions of high stringency (0.5×SSC, 0.1% SDS at 65°C) all clones hybridized strongly to the λRL1.1 insert probe, in fact as strongly as the probe hybridized to to the λRL1.1 insert DNA (Fig. 5.11b). The clones were determined to have inserts of the following sizes using the SIZING program (Section 2.2.24): λRL2.10 insert = 1.30 kb, λRL2.35 insert = 1.43 kb and λRL2.39 insert = 0.95 kb. The analysis of these clones was therefore restricted to the λRL2.35 clone which had the longest insert.

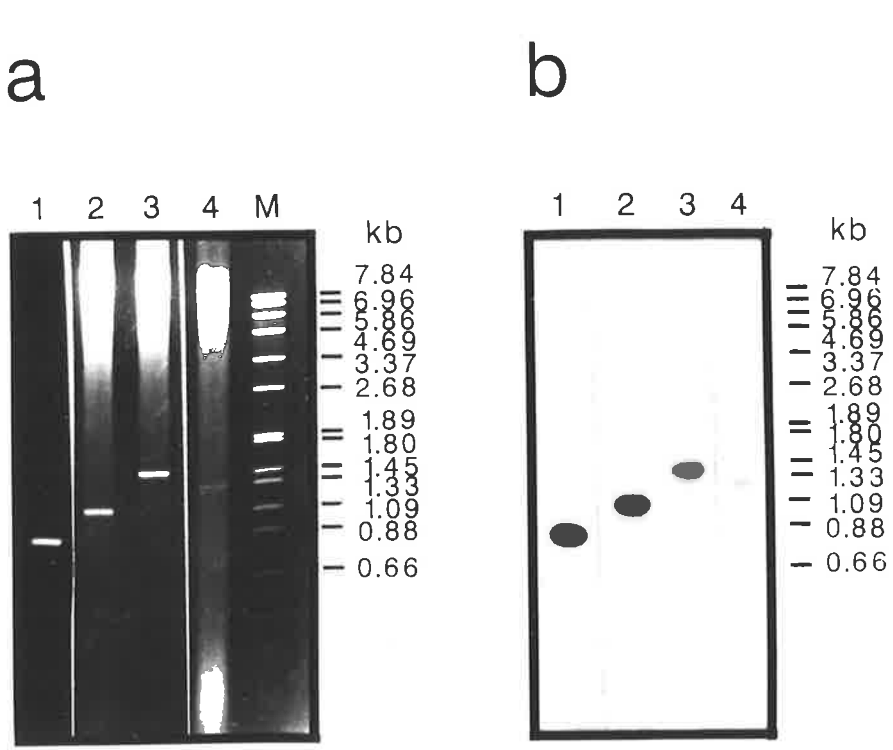
Figure 5.11

Restriction Digests of Rat Liver cDNA Clones and Southern Transfer Probed with λRL1.1 Insert

a Restriction digests of rat liver cDNA clones. Aliquots of 10 μg of the indicated rat liver cDNA clone DNAs were digested with the restriction enzyme *Eco*RI to excise the inserts. The DNA samples were electrophoresed on a 1% (w/v) agarose, 1×TBE gel, stained with 0.3 μg/ml ethidium bromide (Section 2.2.9.a) then photographed under short-wave UV illumination. The sizes of the marker DNA bands are shown in kb.

Track 1 : 200 ng of λRL1.1 insert DNA,
Track 2 : *Eco*RI digestion of 10 μg of λRL2.39 DNA,
Track 3 : *Eco*RI digestion of 10 μg of λRL2.35 DNA,
Track 4 : *Eco*RI digestion of 10 μg of λRL2.10 DNA,
Track M: *Eco*RI digested bacteriophage SPP-1 DNA size markers.

b Southern transfer of rat liver cDNA clone inserts probed with λRL1.1 insert DNA. The DNA from the gel shown in Fig. 5.11a was transferred to nitrocellulose using the Southern transfer method (Section 2.2.11.a) and the filter was probed with the λRL1.1 insert (Section 2.2.21.b) which was [^{32}P] labelled by nick-translation (Section 2.2.18.c). The filter was washed in 0.5×SSC, 0.1% SDS at 65°C then autoradiographed. The sizes of the marker DNA bands are shown in kb.


a
b
1 2 3 4 M
kb
7.84
6.96
5.86
4.69
3.37
2.68
1.89
1.80
1.45
1.33
1.09
0.88
0.66
1 2 3 4
kb
7.84
6.96
5.86
4.69
3.37
2.68
1.89
1.80
1.45
1.33
1.09
0.88
0.66

The absence of a significant number of rat liver pyruvate carboxylase cDNA clones with inserts approximately the same size as the λRL2.35 insert suggested that it was unlikely that there were clones with substantially longer inserts or full length inserts present in the library. A total of 1×10^6 pfu were screened and even though this was a very substantial number it still only represented 4.5% of the total number of clones in the library. However re-screening a large number of clones from this library was not considered likely to be successful in yielding clones with longer inserts and was therefore not proceeded with.

5.3.13 Sequence Analysis of the λRL2.35 Pyruvate Carboxylase cDNA Clone

Analysis of the λRL1.1 sequence using the ANALYSEQ program had revealed the presence of a *Pst*I site 16 bp from the 5′ end of the λRL1.1 insert. Digestion of λRL2.35 inset DNA with *Pst*I resulted in 2 fragments of 416 bp and 922 bp which were subcloned into M13mp8 and mp9 and sequenced using the dideoxy chain-termination method (Section 2.2.22). Sequencing of the 5′ end of the 922 bp fragment confirmed the sequence of the 5′ end of the λRL1.1 clone and the sequence on the 3′ end of the 416 bp fragment also overlapped the λRL1.1 sequence. The sequencing strategy of the λRL2.35 clone is shown in Fig. 5.12a and the compiled sequence of λRL1.1 and λRL2.35 is also shown in Fig. 5.12b together with the inferred amino acid sequence. The total length of the rat pyruvate carboxylase cDNA sequence was 1338 bp which contained 920 bp of the coding region and included 399 bp 5′ to the sequence of λRL1.1 (Fig. 5.12b).

A codon usage table was determined for the rat pyruvate carboxylase sequence using the ANALYSEQ program (Section 2.2.24), see Table 5.1. The results of this analysis were not significantly different to the rat codon usage table compiled by Maruyama *et al.* (1986). The base content, inferred amino acid content, protein molecular weight and the protein hydrophobicity were also determined from the cDNA sequence, using the ANALYSEQ program, see Table 5.1.

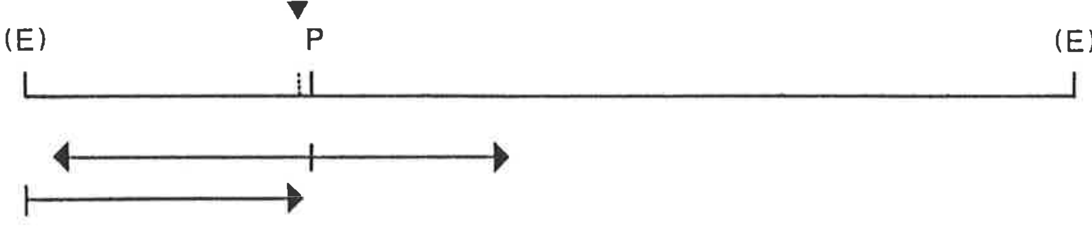
Figure 5.12

Sequencing Strategy of λRL2.35 Insert and Compiled Sequence of Rat Liver Pyruvate Carboxylase cDNA Clones

a Sequencing strategy of λRL2.35 insert. The λRL2.35 cDNA clone insert was digested with *Pst*I (P) and the resulting 416 bp and 922 bp fragments were subcloned into both M13mp8 and M13mp9 (Section 2.2.16). The resulting clones were sequenced using the dideoxy chain termination system (Sections 2.2.22 and 2.2.23) and the extent and orientation of each sequence is marked by an arrow. The position of the 5′ end of the λRL1.1 clone insert with respect to the λRL2.35 clone is marked by (▾) and a dotted line. The λgt10 *Eco*RI cloning sites are indicated by (E).

b Compiled sequence of rat pyruvate carboxylase cDNA clones. The sequences of the λRL1.1 insert and the λRL2.35 were compiled into a single contiguous sequence. The junction between the sequences of the two clones is marked by (▲) and the λRL1.1 sequence is italicized. The inferred amino acid sequence is shown and the lysine to which biotin is attached is boxed.

a



b

```
   ala leu val ala val pro lys gly leu leu trp thr thr glu val pro leu glu arg val phe asp tyr ser glu tyr trp glu gly ala   30
GGGCCCTGGTGGCTGTACCAAAGGGACTCCTCTGGACCACAGAGGTACCCCTGGAGCGTGTGTTTGACTACAGTGAGTATTGGGAAGGGG
        10        20        30        40        50        60        70        80        90
   arg gly leu tyr ala ala phe asp cys thr ala thr met lys ser gly asn ser asp val tyr glu asn glu ile pro gly gly gln tyr   60
CTCGGGGGCTGTATGCAGCCTTTGATTGCACGGCTACCATGAAGTCTGGCAACTCAGACGTGTATGAGAATGAGATCCCAGGGGGCCAGT
       100       110       120       130       140       150       160       170       180
   thr asn leu his phe gln ala his ser met gly leu gly ser lys phe lys glu val lys lys ala tyr val glu ala asn gln ser ala   90
ACACCAACCTACACTTCCAGGCCCACAGCATGGGACTTGGCTCCAAGTTCAAGGAGGTCAAGAAGGCCTATGTGGAGGCTAACCAGAGTG
       190       200       210       220       230       240       250       260       270
   gly gly pro his gln gly asp thr ile leu gln asp cys gly gly ser gly pro val his gly ala glu arg val glu pro gly arg gly  120
CTGGGGGACCTCATCAAGGTGACACCATCCTCCAAGATTGTGGGGGATCTGGCCCAGTTCATGGTGCAGAACGGGTTGAGCCGGGCAGAG
       280       290       300       310       320       330       340       350       360
   arg ser ser gly arg arg ala val leu pro pro leu cys gly gly val pro ala ala met tyr pro asp val phe ala gln phe lys asp  150
GCAGAAGCTCAGGCAGAAGAGCTGTCCTTCCCCCGCTCTGTGGTGGAGTTCCTGCAGCCATGTACCCTGATGTCTTTGCTCAGTTCAAAG
       370       380       390       ▲400      410       420       430       440       450
   phe thr ala thr phe gly pro leu asp ser leu asn thr arg leu phe leu gln gly pro lys ile ala glu glu phe glu val glu leu  180
ACTTCACGGCTACCTTTGGCCCCCTGGATAGCCTGAATACTCGTCTCTTTCTTCAAGGACCCAAAATTGCAGAGGAGTTTGAGGTTGAGC
       460       470       480       490       500       510       520       530       540
   glu arg gly lys thr leu his ile lys ala leu ala val ser asp leu asn arg ala ala gln arg gln val phe phe glu leu asn gly  210
TGGAACGGGGCAAGACCTTGCACATCAAAGCCCTGGCTGTAAGCGACCTGAACCGTGCTGCGCAGAGGCAGGTGTTCTTTGAACTCAATG
       550       560       570       580       590       600       610       620       630
   gln leu arg ser ile leu val lys asp thr gln ala met lys glu met his phe his pro lys ala leu lys asp val lys gly gln ile  240
GGCAGCTTCGATCCATTCTGGTTAAAGACACCCAGGCCATGAAGGAGATGCACTTCCATCCCAAGGCCTTGAAGGATGTGAAGGGCCAAA
       640       650       660       670       680       690       700       710       720
   gly ala pro met pro gly lys val ile asp val lys val ala ala gly ala lys val val lys gly gln pro leu cys val leu ser ala  270
TTGGGGCCCCTATGCCTGGGAAGGTCATAGACGTCAAGGTGGCAGCAGGAGCCAAGGTGGTTAAGGGCCAGCCCCTCTGTGTGCTCAGCG
       730       740       750       760       770       780       790       800       810
   met [lys] met glu thr val val thr ser pro met glu gly thr ile arg lys val his val thr lys asp met thr leu glu gly asp asp  300
CCATGAAGATGGAGACTGTGGTGACTTCGCCCATGGAGGGCACTATCCGAAAGGTTCACGTGACCAAGGACATGACTCTGGAAGGCGATG
       820       830       840       850       860       870       880       890       900
   leu ile leu glu ile glu *                                                                                            306
ACCTCATCCTAGAGATTGAGTGATCTTACTCCAGACTGGCAGCCTGGCCAACCCTACCCCAAGCCTCTCAACAGAAGCTGTGCAGCCAGG
       910       920       930       940       950       960       970       980       990
GCAGGCCCAGCGAGTACCTGAGGGCTAGGCTTGAGGTCCTGTCCCATGGGCACACACACACTACCTGCAATGGCCCTCCCATTGCCCTTC
      1000      1010      1020      1030      1040      1050      1060      1070      1080
AGCTATTTGTCCTTGTCTTGCTGGCAGGCAGTTCTCACATATTCTCTCTTGCCAAATAAGGGTCTACTCCTGTGGGAGACCAAGGTGTAC
      1090      1100      1110      1120      1130      1140      1150      1160      1170
AGTAGGTGGCCTTGTACCTGGGAGAGGGGTTCTACCTCTGGGGGTAGAGGGAAGAAGACCTAATTCATAGGTCCTGGGAAATTTGCTCAA
      1180      1190      1200      1210      1220      1230      1240      1250      1260
TAAAGTGGCCTTCCCTTGCCCTCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
      1270      1280      1290      1300      1310      1320      1330     1338
```

Table 5.1

Content Analysis of the Coding Region of the Combined Rat Liver Pyruvate Carboxylase cDNA Clone Sequence

The ANALYSEQ program (Section 2.2.24), was used to analyse the content of the coding region from the combined sequence of the rat liver pyruvate carboxylase cDNA clones shown in Fig. 5.12b.

a A table of codon usage is shown. Amino acids are represented by the single letter code and are shown beside their codons with the number of times they occur.

b The base content of the sequence is shown. The percentage content of each base is given with the percentage content of A+T and G+C bases.

c The inferred amino acid content is shown. The number of times each amino acid occurs is indicated beside the amino acid.

d The total numbers of codons and amino acids, the estimated protein molecular weight and the estimated protein hydrophobicity are shown.

a

F	TTT	7	S	TCT	2	Y	TAT	4	C	TGT	3
F	TTC	6	S	TCC	2	Y	TAC	3	C	TGC	1
L	TTA	0	S	TCA	2	*	TAA	0	*	TGA	1
L	TTG	2	S	TCG	1	*	TAG	0	W	TGG	2
L	CTT	4	P	CCT	5	H	CAT	3	R	CGT	3
L	CTC	9	P	CCC	7	H	CAC	5	R	CGC	0
L	CTA	2	P	CCA	3	Q	CAA	4	R	CGA	2
L	CTG	10	P	CCG	2	Q	CAG	9	R	CGG	3
I	ATT	4	T	ACT	5	N	AAT	3	S	AGT	2
I	ATC	5	T	ACC	8	N	AAC	4	S	AGC	5
I	ATA	1	T	ACA	1	K	AAA	4	R	AGA	4
M	ATG	10	T	ACG	2	K	AAG	18	R	AGG	1
V	GTT	7	A	GCT	10	D	GAT	6	G	GGT	3
V	GTC	5	A	GCC	11	D	GAC	9	G	GGC	13
V	GTA	3	A	GCA	6	E	GAA	5	G	GGA	7
V	GTG	12	A	GCG	1	E	GAG	17	G	GGG	8

b

T	C	A	G	A+T	G+C
21.06%	25.19%	23.45%	30.29%	44.51%	55.48%

c

Ala	28	Leu	27
Arg	13	Lys	22
Asn	7	Met	10
Asp	15	Phe	13
Cys	4	Pro	17
Gln	13	Ser	14
Glu	22	Thr	16
Gly	31	Trp	2
His	8	Tyr	7
Ile	10	Val	27

d

Total Number of Codons	=	307
Total Number of Amino Acids	=	306
Protein Molecular Weight	=	33259
Hydrophobicity	=	-26.23

5.3.14 DNA Sequence Homology Between Pyruvate Carboxylase Clones from Different Species

A comparison of the primary sequence of the rat pyruvate carboxylase clones to the sequence of human pyruvate carboxylase (Freytag and Collier, 1984) using the SEQA program (Section 2.2.24) revealed a very high degree of DNA sequence homology (89%) with substitutions in the third base of the codons representing the most common changes in the coding sequence (Fig. 5.13a). The only other species for which a significant amount of pyruvate carboxylase sequence data has become available is yeast. A yeast pyruvate carboxylase genomic clone was isolated in this laboratory using a synthetic oligonucleotide probe complementary to the biotin attachment site and was partially sequenced (Morris *et al.*, 1987). The full sequence of the coding region has recently been completed (F. Lim, personal communication) and this was also compared to the the rat pyruvate carboxylase DNA sequence using the SEQA program (Section 2.2.24). The level of sequence homology between rat and yeast was 35%, dramatically lower than the rat-human homology, as might be expected from organisms so distantly related in evolutionary terms. In localized parts of the sequence the homology was considerably greater, this was especially noticeable in the vicinity of the biotin attachment site, see Fig. 5.13b.

5.3.15 Protein Sequence Homology Between Pyruvate Carboxylase from Different Species

The inferred amino acid sequence of the rat liver pyruvate carboxylase cDNA clones reveal a very strong homology between the rat enzyme and the protein sequence determined for pyruvate carboxylase from chicken, turkey, sheep (Rylatt *et al.* 1977) and yeast (Morris *et al.*, 1987) as well as with the inferred amino acid sequence of human pyruvate carboxylase (Freytag and Collier, 1984), see Fig. 5.14a. The sequences of the chicken, turkey and sheep peptides have an alignment gap introduced into them because of the presence of a cysteinyl residue in the human and rat sequences and an alanyl residue in the yeast sequence at this point. In fact the pos-

Figure 5.13

DNA Sequence Homology Between Pyruvate Carboxylase Clones from Different Species

a The DNA sequence of λRL1.1 rat pyruvate carboxylase cDNA clone is shown aligned with the DNA sequence of PC34.1 human pyruvate carboxylase cDNA clone (this report, Lamhonwah *et al.*, 1987). The codons for the lysine to which biotin is attached are boxed and sequence identity is indicated by (*). The bases are numbered with the 5′ end of the sequence as number 1.

b The compiled DNA sequence (R) of the rat pyruvate carboxylase cDNA clones λRL1.1 and λRL2.35 is shown aligned with the DNA sequence (Y) of the yeast pyruvate carboxylase gene (F. Lim, personal communication, Morris *et al.*, 1987) using the SEQA program (Section 2.2.24). The codons for the lysine to which biotin is attached are boxed and sequence identity is indicated by (*). A 3 base alignment gap in the yeast pyruvate carboxylase sequence is indicated by dashes. The bases are numbered with the 5′ end of the sequence as number 1.

a

```
                   640       650       660       670       680       690       700       710
Rat PC     CAGCTTCGATCCATTCTGGTTAAAGACACCCAGGCCATGAAGGAGATGCACTTCCATCCCAAGGCCTTGAAGGATGTGAA
           ***** ** *****  **** ** *********** ******************** ********* * ***** *****
Human PC   CAGCTGCGGTCCATCTTGGTCAAGGACACCCAGGCAATGAAGGAGATGCACTTCCACCCCAAGGCCCTAAAGGACGTGAA
                    10        20        30        40        50        60        70        80

                   720       730       740       750       760       770       780       790
Rat PC     GGGCCAAATTGGGGCCCCTATGCCTGGGAAGGTCATAGACGTCAAGGTGGCAGCAGGAGCCAAGGTGGTTAAGGGCCAGC
           ****** ** ***** ** ************** ****** **** ****  ***** **********  **********
Human PC   GGGCCAGATCGGGGCGCCCATGCCTGGGAAGGTGATAGACATCAAAGTGGTGGCAGGGGCCAAGGTGGCCAAGGGCCAGC
                    90       100       110       120       130       140       150       160

                   800       810       820       830       840       850       860       870
Rat PC     CCCTCTGTGTGCTCAGCGCCATGAAGATGGAGACTGTGGTGACTTCGCCCATGGAGGGCACTATCCGAAAGGTTCACGTG
           **** *********** ************************** ** *********** *** **** ******** ***
Human PC   CCCTGTGTGTGCTCAGTGCCATGAAGATGGAGACTGTGGTGACCTCACCCATGGAGGGTACTGTCCGCAAGGTTCATGTG
                   170       180       190       200       210       220       230       240

                   880       890       900       910       920
Rat PC     ACCAAGGACATGACTCTGGAAGGCGATGACCTCATCCTAGAGATTGAGTGA
           ************** ******** ** *********** ***** ******
Human PC   ACCAAGGACATGACACTGGAAGGTGACGACCTCATCCTGGAGATCGAGTGA
                   250       260       270       280       290
```

b

```
         10        20        30        40        50        60        70        80        90       100       110       120
R GGGCCCTGGTGGCTGTACCAAAGGGACTCCTCTGGACCACAGAGGTACCCCTGGAGCGTGTGTTTGACTACAGTGAGTATTGGGAAGGGGCTCGGGGGCTGTATGCAGCCTTTGATTGCA
         * ** **  *  * *  *   * *                  * *    ***                 * **      *  ***   *  ** *          *  * *   **     ** *
Y CAAACTTGTTGTTTCAAGCCCAACAATTGGGTCTTGGAGAACAATGGGCCCAAACAAAAAGAGCTTACAGAGAAGCCAATTATTTATTGGGTGATATTGTCAAAGTTACCCCAACTTCGA
  3520      3530      3540      3550      3560      3570      3580      3590      3600      3610      3620      3630

        130       140       150       160       170       180       190       200       210       220       230       240
R CGGCTACCATGAAGTCTGGCAACTCAGACGTGTATGAGAATGAGATCCCAGGGGGCCAGTACACCAACCTACACTTCCAGGCCCACAGCATGGGACTTGGCTCCAAGTTCAAGGAGGTCA
   **       *     *  ** *     ** *   * **    *   * * *    *     *     * *   **   *              **        *** *  *    * *
Y AGGTCGTTGGTGATCTGGCAAAATTTATGGTCTCCAATAAATTAACTTCCGATGATGTGAGACGCCTGGCTAATTCTTTGGATTTCCCTGACTCTGTTATGGATTTCTTCGAAGGCTTAA
  3640      3650      3660      3670      3680      3690      3700      3710      3720      3730      3740      3750

        250       260       270       280       290       300       310       320       330       340       350       360
R AGAAGGCCTATGTGGAGGCTAACCAGAGTGCTGGGGGACCTCATCAAGGTGACACCATCCTCCAAGATTGTGGGGGATCTGGCCCAGTTCATGGTGCAGAACGGGTTGAGCCGGGCAGAG
                 *  *   **      *    ** *  *          * **      * * *     *   *       *  *    *    *   * *         *
Y TCGGCCAACCATATGGTGGGTTCCCAGAACCATTTAGATCAGACGTTTTAAGGAACAAGAGAAGAAAGTTGACTTGTCGTCCAGGCCTGGAACTAGAGCCATTTGATCTCGAAAAAATTA
  3760      3770      3780      3790      3800      3810      3820      3830      3840      3840      3860      3870

        370       380       390       400       410       420       430       440       450       460       470       480
R GCAGAAGCTCAGGCAGAAGAGCTGTCCTTCCCCCGCTCTGTGGTGGAGTTCCTGCAGCCATGTACCCTGATGTCTTTGCTCAGTTCAAAGACTTCACGGCTACCTTTGGCCCCCTGGATA
  *   *  **     *  ***  **    *          ** *  *  * * **      *********    ** * **    * *** ** *  * *  *  ** * ***     *   *
Y GAGAAGACTTGCAGAATAGATTTGGTGATGTTGATGAGTGCGACGTTGCTTCTTATAACATGTACCCAAGAGTTTATGAAGACTTCCAAAAGATGAGAGAAACGTATGGTGATTTATCTG
  3880      3890      3900      3910      3920      3930      3940      3950      3960      3970      3980      3990

        490       500       510       520       530       540       550       560       570       580       590    597
R GCCTGAATACTCGTCTCTTTCTTCAAGGACCCAAAATTGCAGAGGAGTTTGAGGTTGAGCTGGAACGGGGCAAGACCTTGCACATCAAAGCCCTGGCTGTAAGCGACCTGAAC---CGTG
     **    **  *   **** *        **     * * **   ** **   **** ****    * ****  ** ** **   *      *****     * ******  * **  *****
Y TATTGCCAACAAGAAGCTTTTTGTCTCCACTAGAGACTGACGAAGAAATTGAAGTTGTAATCGAACAAGGTAAAACGCTAATTATCAAGCTACAGGCTGTGGGTGATTTGAACAAAAAGA
  4000      4010      4020      4030      4040      4050      4060      4070      4080      4090      4100      4110

  600       610       620       630       640       650       660       670       680       690       700       710
R CTGCGCAGAGGCAGGTGTTCTTTGAACTCAATGGGCAGCTTCGATCCATTCTGGTTAAAGACACCCAGGCCATGAAGGAGATGCACTTCCATCCCAAGGCCTTGAAGGATGTGAAGGGCC
  * *   * **  * ** * ******  * *****  *   *  **    ****  ***   ****           *    *** *          *  **** **    * * ***    *
Y CCGGTGAAAGAGAAGTTTACTTTGATTTGAATGGTGAAATGAGAAAAATTCGTGTTGCTGACAGATCACAAAAAGTGGAAACTGTTACTAAATCCAAAGCAGACATGCATGATCCATTAC
  4120      4130      4140      4150      4160      4170      4180      4190      4200      4210      4220      4230

  720       730       740       750       760       770       780       790       800       810       820       830
R AAATTGGGGCCCCTATGCCTGGGAAGGTCATAGACGTCAAGGTGGCAGCAGGAGCCAAGGTGGTTAAGGGCCAGCCCCTCTGTGTGCTCAGCGCCATGAAGATGGAGACTGTGGTGACTT
  * ***** ** ** *** * **       *  * ** ** ** **     **** *    *    ******** **   *   **  * *********** ***** *   *  *  ***
Y ACATTGGTGCACCAATGGCAGGTGTCATTGTTGAAGTTAAAGTTCATAAAGGATCACTAATAAAGAAGGGCCAACCTGTAGCCGTATTAAGCGCCATGAAAATGGAAATGATTATATCTT
  4240      4250      4260      4270      4280      4290      4300      4310      4320      4330      4340      4350

  840       850       860       870       880       890       900       910       920       930       940       950
R CGCCCATGGAGGGCACTATCCGAAAGGTTCACGTGACCAAGGACATGACTCTGGAAGGCGATGACCTCATCCTAGAGATTGAGTGATCTTACTCCAGACTGGCAGCCTGGCCAACCCTAC
  * **    ** **    *   * * **    **  *  * *     * * ****     ***  *  *  *     * **         *          *        ****
Y CTCCATCCGATGGACAAGTTAAAGAAGTGTTTGTCTCTGATGGTGAAAATGTGGACTCTTCTGATTTATTAGTTCTATTAGAAGACCAAGTTCCTGTTGAAACTAAGGCATGAACCGGTT
  4360      4370      4380      4390      4400      4410      4420      4430      4440      4450      4460      4470

  960       970       980       990       1000      1010      1020      1030      1040      1050      1060      1070
R CCCAAGCCTCTCAACAGAAGCTGTGCAGCCAGGGCAGGCCCAGCGAGTACCTGAGGGCTAGGCTTGAGGTCCTGTCCCATGGGCACACACACACTACCTGCAATGGCCCTCCCATTGCCC
          * * *  *  * *      * *              *  *  *   * * *   **    *  **    *   *  ** *    **          *
Y AGTTCTCATTTATAATGTATAATATACCCGAATCTTATTTATTTACCTTTCCTATTTTTTGACGACCAGTAAATACTAATACATAATTAGGAACAAAAGTTAAATAAAAAAAAAAAAATA
  4480      4490      4500      4510      4520      4530      4540      4550      4560      4570      4580      4590

  1080      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190
R TTCAGCTATTTGTCCTTGTCTTGCTGGCAGGCAGTTCTCACATATTCTCTCTTGCCAAATAAGGGTCTACTCCTGTGGGAGACCAAGGTGTACAGTAGGTGGCCTTGTACCTGGGAGAGG
   *       *    **  * **            *  *  *  *  ***     *  * * * *    ** *          *  * *               * * *  *    * *
Y ATTTAACGCATCCAATTAACGTGTCCTTTTTTCATCATTAATTTATCTACTATTTCGATTTAAATTCCATATACAATAAATCCTAGATACATTCCCGAAAGTCATCTTTTAGCGAAACAT
  4600      4610      4620      4630      4640      4650      4660      4670      4680      4690      4700      4710

  1200      1210      1220      1230      1240      1250      1260      1270      1280
R GGTTCTACCTCTGGGGGTAGAGGGAAGAAGACCTAATTCATAGGTCCTGGGAAATTTGCTCAATAAAGTGGCCTTCCCTTGCCCTCC
     * **    ***   * ** ***    *    *      * *      ** ** *    **    *    * *
Y CTTCCTTGAGCTGCTAGCAGTGGGCTTAGTCCACCTGTTAGTTACTCTTGGTATACCACTAGGTCTTTTCGAGGCAGGAACATGGGT
  4720      4730      4740      4750      4760      4770      4780      4790      4800
```

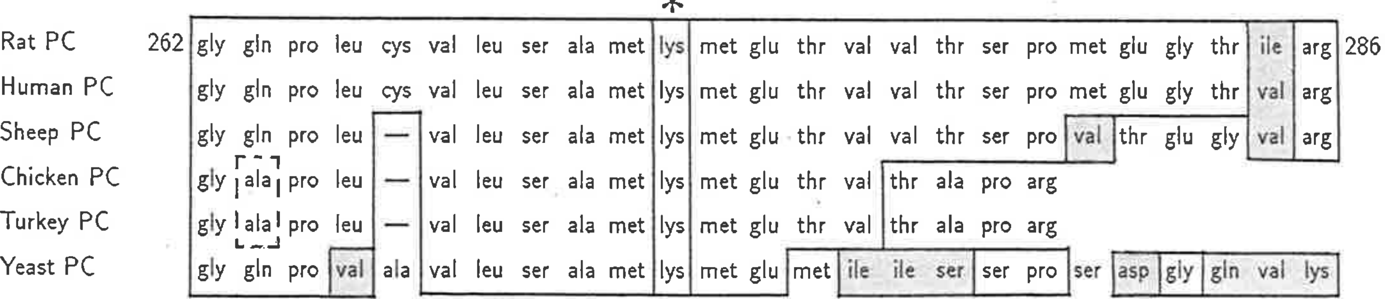
Figure 5.14

Protein Sequence Homology Between Pyruvate Carboxylase from Different Species

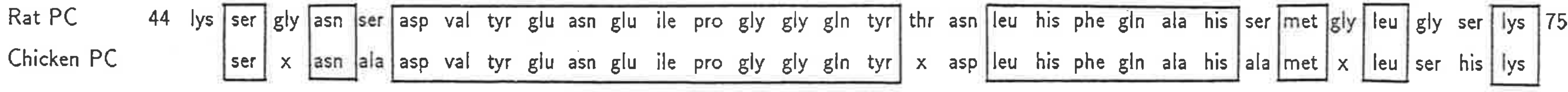
a The inferred amino acid sequence of rat liver pyruvate carboxylase is shown aligned with the sequence determined by direct protein sequencing of pyruvate carboxylase from several different species. The peptide sequence data was obtained from the following sources :- Sheep PC : sheep liver pyruvate carboxylase, Chicken PC : chicken liver pyruvate carboxylase, Turkey PC : turkey liver pyruvate carboxylase, all from Rylatt *et al.* (1977), Yeast PC : yeast pyruvate carboxylase, (Morris *et al.*, 1987). Where there is sequence identity the residues are boxed and where there is a conservative difference this is indicated by shaded boxes. The lysine to which biotin is attached is boxed and marked by (*). Non-homologous amino acids which are enclosed by homologous amino acids are boxed with a dashed line.

b The inferred amino acid sequence of rat liver pyruvate carboxylase was compared with the peptide sequence data obtained from chicken liver pyruvate carboxylase. The peptide termed Peptide Fr. 33 (Fig. 3.14) exhibited strong homology and is shown aligned with the rat liver pyruvate carboxylase sequence. Where there is sequence identity the residues are boxed and where there is a conservative difference this is indicated by shaded boxes.

a



b



sibility exists that the manual sequencing of the chicken, turkey and sheep peptides failed to identify a cysteinyl residue at that point and hence the published sequence may contain a deletion of one amino acid. The likelihood of a cysteinyl residue being present in this position in the avian enzymes is given additional credibility by the finding by Fall *et al.* (1976) that cysteine-42 which is in the immediate vicinity of the biotin attachment site of the biotinyl subunit of the *E. coli* acetyl-CoA carboxylase has an important role in maintaining the structure of the subunit (refer to Fig. 1.4 for the partial sequence of the biotinyl subunit of acetyl-CoA carboxylase).

The inferred amino acid sequence of the rat liver pyruvate carboxylase cDNA was compared to the protein sequence of the set of peptides generated from the proteolytic digestion of chicken liver pyruvate carboxylase described in Chapter 3. One chicken pyruvate carboxylase peptide identified in Fig. 3.14 as peptide Fr. 33 contained 82% homology to the rat liver pyruvate carboxylase sequence, see Fig. 5.14b.

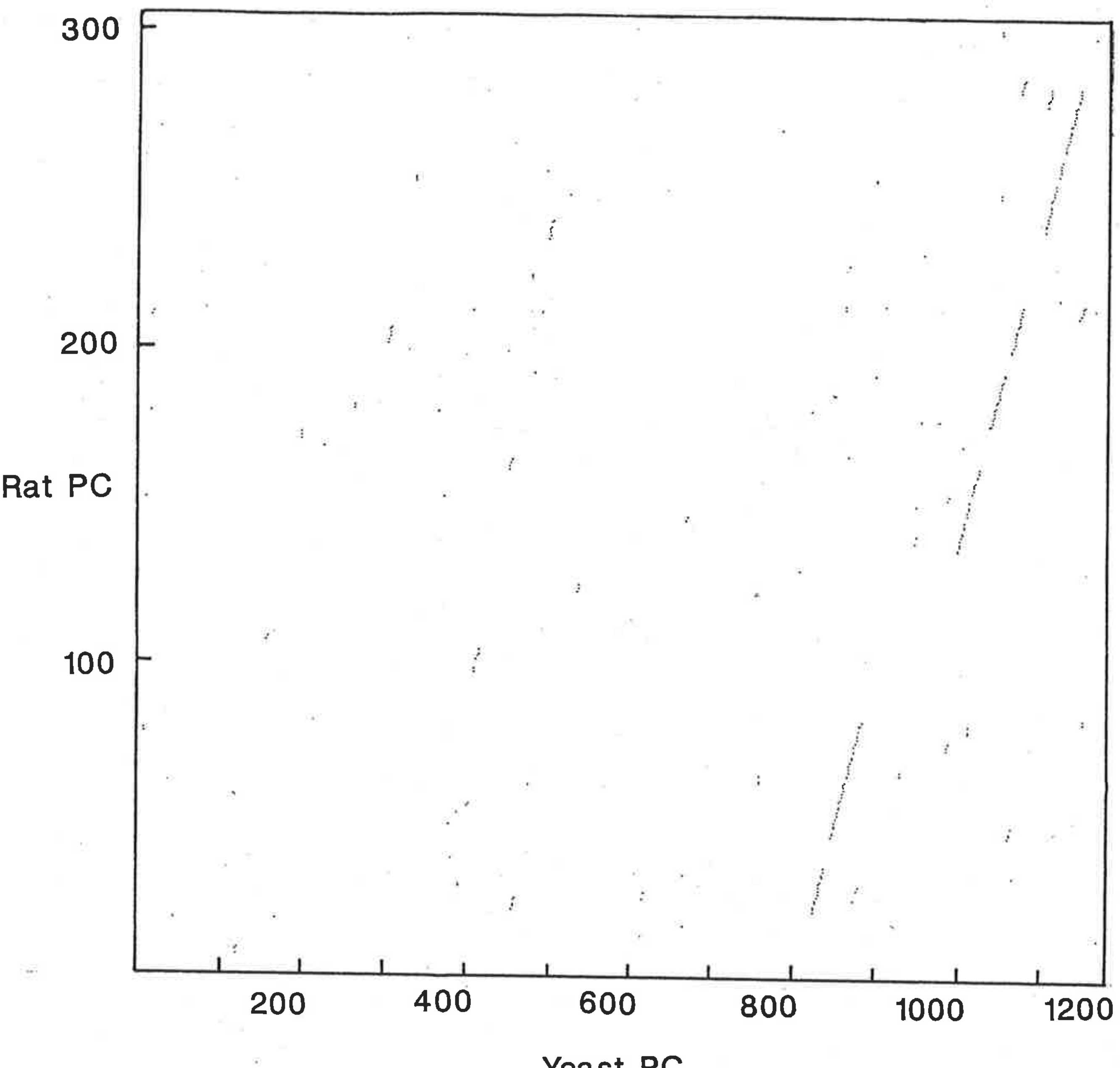
The inferred amino acid sequence of the rat liver pyruvate carboxylase was compared with the inferred amino acid sequence of the yeast pyruvate carboxylase gene (F. Lim, personal communication) using the DIAGON program (Section 2.2.24), see Fig. 5.15. Significant homology was found at the C-terminus where the biotin attachment site is located in pyruvate carboxylase from both species. There is a region of 56 amino acids in rat pyruvate carboxylase and 114 amino acids in yeast pyruvate carboxylase where there is no homology. Together with the shift in the position of the diagonal this indicated that a rearrangement of the sequence has occurred between these two species. Either a deletion has occurred in the yeast pyruvate carboxylase sequence or there has been an insertion in the rat pyruvate carboxylase sequence.

An examination of the inferred rat liver pyruvate carboxylase protein sequence failed to reveal a consensus sequences typical of ATP binding sites (Hollemans *et al.*, 1983; Higgins *et al.*, 1986). A consensus site for glycosylation (Asn-X-Ser) (Spiro, 1973) is present at amino acids 87–89 (Fig. 5.13b) although as yet there is no evidence for the glycosylation of rat pyruvate carboxylase. A triplet of amino acids, "Pro-Met-Pro", noted by Lamhonwah *et al.* (1987) as conserved between human

Figure 5.15

DIAGON Program Comparison of Rat Liver Pyruvate Carboxylase and Yeast Pyruvate Carboxylase Sequences

The inferred amino acid sequences of rat liver pyruvate carboxylase and yeast pyruvate carboxylase (F. Lim, personal communication; Morris *et al.*, 1987) were compared using the DIAGON program (Section 2.2.24). The yeast pyruvate carboxylase sequence is represented on the x-axis and the rat liver pyruvate carboxylase sequence on the y-axis. Sequence homology is indicated by a dot for a given span of amino acids. The amino acids on each axis are numbered with the most N-terminal inferred amino acid taken as number 1 and this is positioned at the origin for both sequences. The program used the proportional algorithm and the parameters used were; SPAN = 11, % SCORE = 132, * SCORE = 8.



300
200
Rat PC
100
200
400
600
800
1000
1200
Yeast PC

pyruvate carboxylase and human propionyl-CoA carboxylase was also present at the same position in rat liver pyruvate carboxylase (Fig. 5.12b).

5.3.16 Structure Prediction of Inferred Rat Liver Pyruvate Carboxylase Sequence

The structure of the inferred amino acid sequence of rat liver pyruvate carboxylase was predicted using the several different programs to provide a means of determining the validity of the prediction made with a single program. The CHOUDOT program (Section 2.2.24) was used and this predicts; turn propensity, α-helix or β-sheet propensity, charge and hydropathy using the parameters of Kyte and Doolittle (1982), see Fig. 5.16. To provide an additional prediction to increase the confidence in the prediction the DELPHI program (Section 2.2.24) was run to predict secondary structure (Fig. 5.17) and the HPLOT program (Section 2.2.24) was run to produce charge (Fig. 5.18a) and hydrophilicity profiles (Fig. 5.18b). Both secondary structure programs essentially agree in their assignment of α-helices. A prediction of an α-helical structure is made for the region in the vicinity of the biocytin residue. The hydropathy profile indicates an alternating hydrophobic/hydrophilic nature for this region consistent with a location on the surface of the protein. The DELPHI program predicts a substantially greater extent of β-sheet than the CHOUDOT program however this may be because the degree of the local propensity is not indicated by the DELPHI program. The group of amino acids, "Pro-Met-Pro" which was mentioned in Section 5.3.15 as conserved between rat pyruvate carboxylase and human pyruvate carboxylase and propionyl-CoA carboxylase were indicated by the CHOUDOT program (Fig. 5.16) as a probable turn region.

5.4 Discussion

5.4.1 Search for Chicken Pyruvate Carboxylase Clones

The requirements for a pyruvate carboxylase-specific probe and for cDNA libraries sufficiently large to adequately represent the level of pyruvate carboxylase mRNA

Figure 5.16

Protein Structure Prediction for Rat Liver Pyruvate Carboxylase Using the CHOUDOT Program

The inferred amino acid sequence of rat liver pyruvate carboxylase was analysed using the CHOUDOT program (Section 2.2.24) to predict the structure of this section of the enzyme. In each profile from the top to the bottom the following is shown: the amino acid sequence, the reverse-turn propensity (t), the α-helix (solid line) and β-sheet (dotted line) propensities (α/β), the charge profile (+ or $-$) and hydrophobicity (Hϕ). The amino acids are numbered on the bottom line with the most N-terminal inferred amino acid as number 1. The hydropathy plot was produced using the parameters of Kyte and Doolittle (1982). Three iterations of data smoothing were applied and the propensities were smoothed.

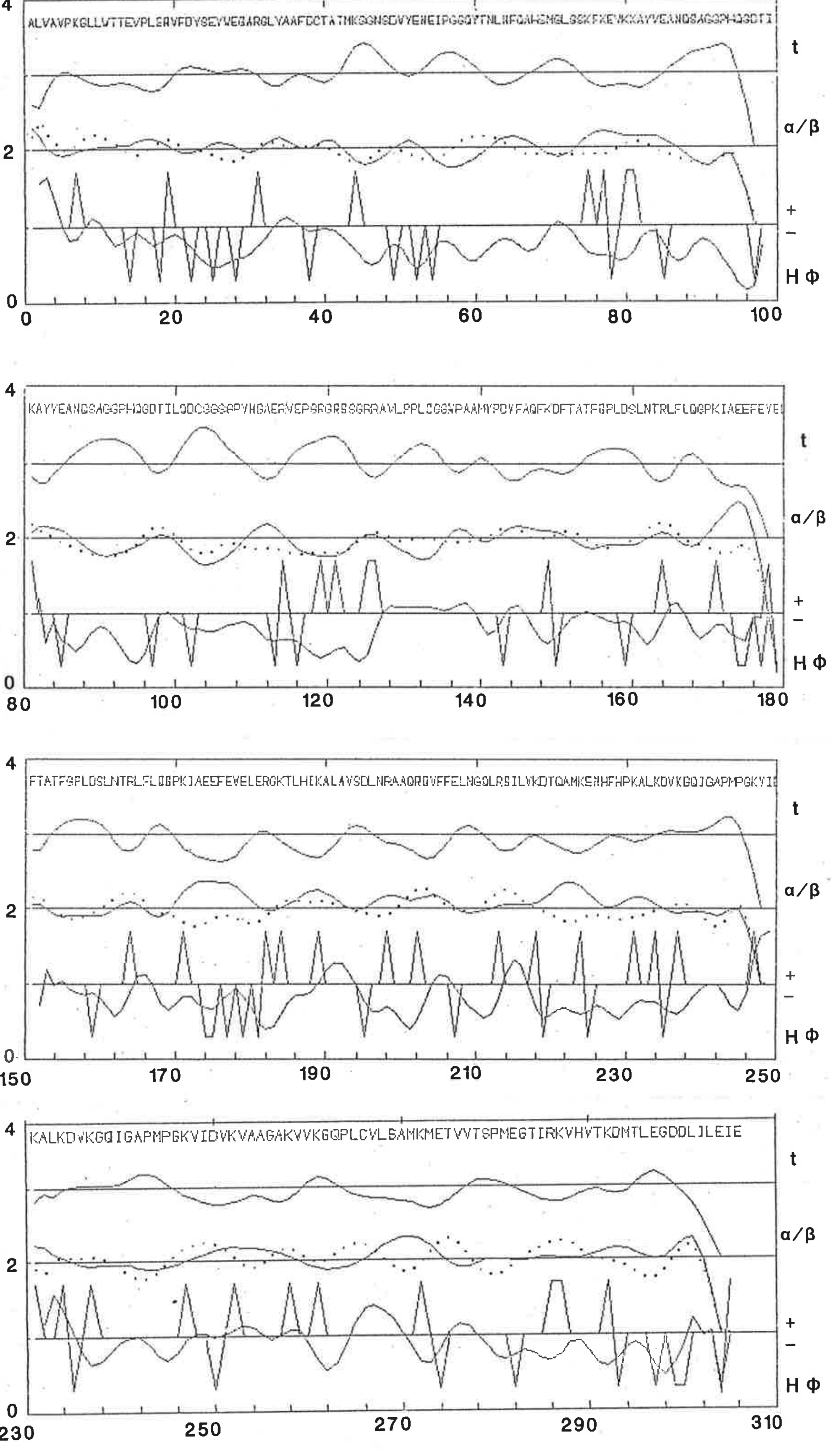
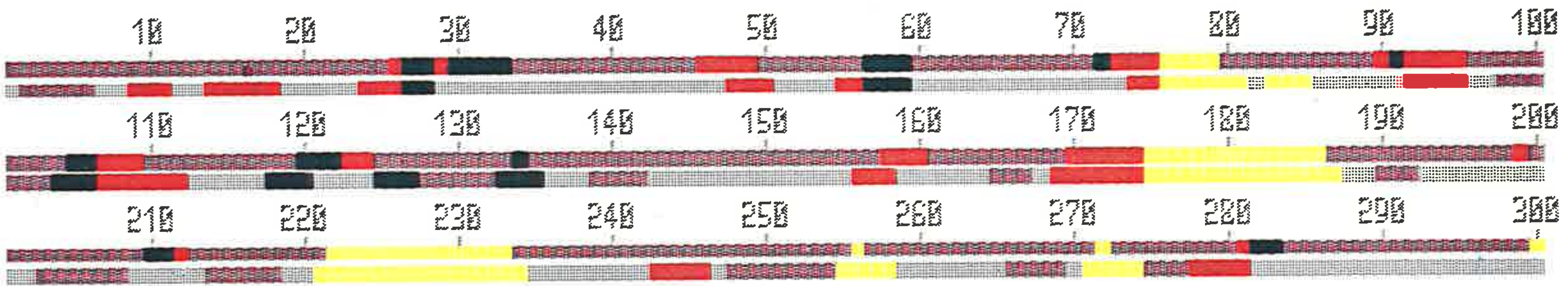

ALVAVPKGLLVTTEVPLERVFDYSEYVEGARGLYAAFDCTATMKGGNSDVYENEIPGGQYTNLHFQAHSMGLSSKFKEVKKAYVEANGSAGGPHQGDTI
KAYVEANGSAGGPHQGDTILQDCGGSSPVHGAERVEPSRGRSSGRRAVLPPLCGSNPAAMYPDVFAQFKDFTATFGPLDSLNTRLFLQGPKIAEEFEVE
FTATFGPLDSLNTRLFLQGPKIAEEFEVELERGKTLHIKALAVSDLNRAAQRGVFFELNGQLRSILVKDTQAMKENHFHPKALKDVKGQIGAPMPGKVI
KALKDVKGQIGAPMPGKVIDVKVAAGAKVVKGQPLCVLSAMKMETVVTSPMEGTIRKVHVTKDMTLEGDDLILEIE
t
α/β
+
−
H Φ
4
2
0
0
20
40
60
80
100
80
100
120
140
160
180
150
170
190
210
230
250
230
250
270
290
310

Figure 5.17

Protein Structure Prediction for Rat Liver Pyruvate Carboxylase Using the DELPHI Program

The inferred amino acid sequence of rat liver pyruvate carboxylase was analysed using the DELPHI program (Section 2.2.24) to predict the structure of this region of the enzyme. The program was run twice using the first estimate of α-helix and β-sheet content to set the Decision Constants for the second and more reliable prediction. The second prediction is shown on the upper line and the first prediction on the lower line. The SAFE parameter was set at 1.0. The program output was formatted for colour printing by Dr. A. Reisner (CSIRO Division of Molecular Biology, Sydney, NSW, Australia) using the SEQCOLOUR program (Section 2.2.24). The colour code of the local propensity is indicated under the profile.



Uncertain

Beta Sheet

Turn

Alpha Helix

Coil

Figure 5.18

Charge and Hydropathy Profiles of Rat Liver Pyruvate Carboxylase Using the HPLOT Program

The charge and hydropathy profiles for the inferred amino acid sequence of rat liver pyruvate carboxylase are shown. The profiles were determined using the HPLOT program (Section 2.2.24) with the Kyte and Doolittle (1982) parameters. The program output was formatted for colour printing by Dr. A. Reisner (CSIRO Division of Molecular Biology, Sydney, NSW, Australia) using the SEQCOLOUR program (Section 2.2.24).

a The charge profile is shown and the colour code of amino acid charge is indicated under the profile.

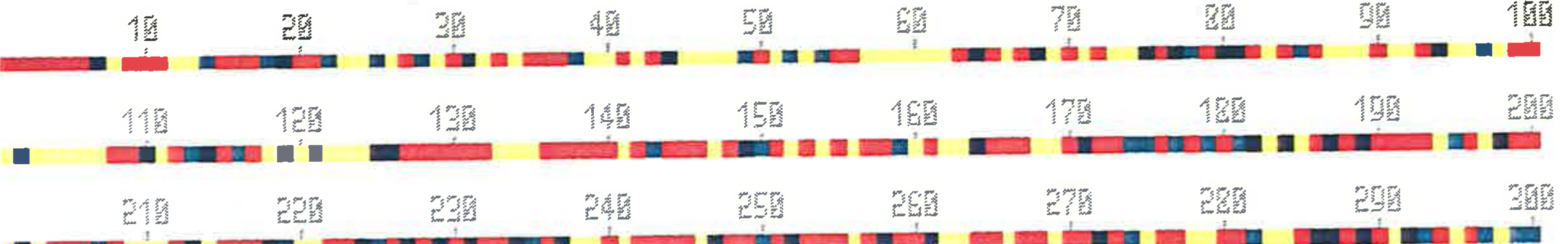
b The hydropathy profile is shown and the colour code of amino acid hydropathy is indicated under the profile.

a

Polar/neut.

Polar/neg.

Polar/pos.



b

Hydropathy - K/D

4.5 to 3.8

2.8 to 1.8

0 to -1.6

-3.2 to -4.5

10 20 30 40 50 60 70 80 90 100

110 120 130 140 150 160 170 180 190 200

210 220 230 240 250 260 270 280 290 300

were apparently met during this investigation. The human pyruvate carboxylase probe specifically hybridized to a single appropriately sized chicken liver mRNA species, the level of which changed during development in the manner predicted for the pyruvate carboxylase mRNA. The cDNA libraries which were constructed, were demonstrated to contain numbers of clones appropriate to the low-moderate abundance cytochrome P450 mRNA chosen as a control. Why then were chicken pyruvate carboxylase clones not detected in either the cDNA or genomic libraries? There are several possible explanations which I shall discuss.

Firstly, the cDNA clones representing chicken liver pyruvate carboxylase may have been incomplete and only represented the 3′ untranslated region which in human liver pyruvate carboxylase has been found to be 400 bp long (Freytag and Collier, 1984). If this was the case then the human pyruvate carboxylase probe may not have hybridized to the chicken clones since the level of sequence homology between related mRNA has been observed to be much lower in the 3′ untranslated region (Heilig *et al.*, 1980; Morris, 1984). Secondly, if only short cDNA transcripts were produced from the pyruvate carboxylase mRNA for example as a result of secondary structure in the mRNA then the pyruvate carboxylase cDNA may have been selectively lost at the size-fractionation step where only cDNA >500 bp was taken. However, both these possibilities should have been circumvented by screening the chicken genomic libraries. The coding region of the pyruvate carboxylase gene would be expected to be adequately represented in any chicken pyruvate carboxylase genomic clone and the presence of these clones would of course be unaffected by any cDNA transcriptional artefacts.

Perhaps then, the genomic libraries were similarly deficient in pyruvate carboxylase clones owing to a particular property of the chicken pyruvate carboxylase gene? Maniatis *et al.* (1978) suggested that the method of genomic library construction used for the library in the λCharon 4A vector may result in under-representation of genomic sequences rich in *Hae*III and *Alu*I sites. However the genomic library constructed in the λEMBL3 vector used a *Sau*3AI partial digest of chicken genomic DNA and this should have obviated any potential problem with the λCharon 4A library.

We are therefore left with those properties which the cDNA and genomic libraries have in common to provide explanations for this phenomenon. Simply stated these are an apparent absence of pyruvate carboxylase clones and the fact that they were cloned into λ vectors. One explanation which is consistent with these properties is that chicken pyruvate carboxylase sequences may have characteristics which reduce the efficiency of replication of λ and as a result pyruvate carboxylase clones would be "diluted out" by other clones during the library plating out and amplification steps. There are several examples of particular sequences being cloned very inefficiently. Wyman *et al.* (1984) found that 8.9% of the recombinant λ bacteriophage in a human genomic library failed to grow on the standard rec^+ host cells owing to DNA secondary structure in the inserts. Nader *et al.* (1986) described the presence of homopolymeric DNA tracts in λ genomic clones of *Physarum polycephalum* which resulted in ~80% of the clones being able to grow on rec^+ host cells. The presence of such DNA structures in the chicken pyruvate carboxylase gene would certainly be expected to affect the isolation of genomic clones but these structures are essentially confined to intronic regions and therefore they would not be expected to affect cDNA cloned in λgt10. An alternative explanation is provided by an examination of the method of construction of all these λ libraries.

In an examination of *in vitro* packaging of λ DNA, Rosenberg (1985) noted that all published methods including the commercially available packaging systems available at that time were extracts of *E. coli* K-12 cells which contain the *Eco*K restriction system. *Eco*K is a type-I restriction enzyme which recognizes and cleaves at 5′-(AACNNNNNNGTGC)-3′. Modification of this site by methylation of the adenine bases prevents restriction. The subunits of the enzyme are encoded by the genes hsdS (sequence specificity), $hsdR_k$ (restriction) and $hsdM_k$ (modification) (for a review of the properties of *Eco*K, see Bickle, 1982). Rosenberg (1985) determined the ratio of the titre of λ produced from unmodified λ DNA (which has 5 *Eco*K sites) to the titre of λ produced from modified λ DNA using 3 packaging systems; (i) the Gigapack system of Stratagene Cloning Systems, (ii) the system described by Kobayashi and Ikeda (1977) and (iii) the system prepared from $hsdr_k^-$ *E. coli* SMR10 cells (Rosenberg *et al.*, 1985). Compared with the extract from *Eco*K deficient

SMR10 cells use of the Gigapack system resulted in an 80% reduction in the titre of λ bacteriophage produced and the Kobayaski and Ikeda system resulted in a 60% reduction. This demonstrated that the *Eco*K restriction enzyme was active in the packaging extracts and any DNA with unmodified *Eco*K sites was liable to be cleaved thereby preventing this DNA from being packaged.

Both the cDNA and the genomic libraries were constructed prior the the publication of this data. The Gigapack system was used with all the cDNA libraries and *E. coli* hsdr$_k^+$ strains were used in the construction of the λCharon 4A chicken genomic library (Maniatis *et al.*, 1978). Since the λEMBL3 chicken genomic library was obtained prior to the publication of the packaging method of Rosenberg *et al.* (1985) which is the only method employing an *E. coli* hsdr$_k^-$ strain it may be assumed that an *E. coli* hsdr$_k^+$ strain was used to prepare the packaging extracts. Since both the chicken liver cDNA and chicken genomic DNA were not passaged through an hsdm$_k^+$ host to modify *Eco*K sites any sites present in the DNA could have been cleaved by *Eco*K activity. Evidence for the presence of one or more *Eco*K sites in the chicken pyruvate carboxylase gene in a region which would affect both cDNA and genomic clones such as the 3′ coding sequence must await the isolation of chicken pyruvate carboxylase genomic clones.

5.4.2 Analysis of mRNA Species in Chicken and Rat Liver

The mRNA for pyruvate carboxylase was shown by Freytag and Collier (1984) to be 4.2 kb long in RNA preparations from human liver, baboon liver and rat liver which represented all the species they examined. The length of the pyruvate carboxylase mRNA is therefore substantially longer than the 3.0–3.6 kb necessary to encode the pyruvate carboxylase protein and Freytag and Collier (1984) observed that the human pyruvate carboxylase mRNA contained ~400 bases of 3′ untranslated RNA. The 4.5 kb chicken mRNA species detected by the human pyruvate carboxylase probe was significantly longer than the 4.2 kb mRNAs which suggested that the chicken pyruvate carboxylase mRNA contained an even greater amount of untranslated RNA. The chicken enzyme appears by SDS-polyacrylamide gel electrophoresis

to be about M_r = 117,000 in its cytosolic protein precursor form (Srivastava *et al.*, 1983) which is significantly smaller than the processed rat pyruvate carboxylase subunit of M_r = 125,000 (Wallace and Easterbrook-Smith, 1985).

The hybridization of the rat liver pyruvate carboxylase cDNA clone to a 4.5 kb mRNA species confirmed the size of the chicken pyruvate carboxylase mRNA which was detected by the human pyruvate carboxylase probe. Interestingly the human pyruvate carboxylase probe also weakly detected three additional mRNA species of 3.3 kb, 2.2 kb and 1.7 kb which suggested that cross-hybridization was occurring to related mRNAs. The biotin carboxylases found in chicken liver include pyruvate carboxylase, propionyl-CoA carboxylase, acetyl-CoA carboxylase and 3-methylcrotonyl-CoA carboxylase (Moss and Lane, 1971). Could the bands detected in this Northern analysis have represented the mRNAs of one or more of these biotin carboxylase?

The mRNA for acetyl-CoA carboxylase has been shown to be ~10 kb in rat liver (Bai *et al.*, 1986) and since chicken acetyl-CoA carboxylase has a subunit size of M_r = 225,000 (Mackall *et al.*, 1977) its mRNA would similarly be expected to be ~10 kb. It could not therefore have been one of those detected. 3-Methylcrotonyl-CoA carboxylase has been found to have subunit sizes of M_r = 80,000 for the biotin containing α subunit and M_r = 62,000 for the β subunit in bovine kidney (Lau *et al.*, 1979). Although at present the size of the mRNAs for 3-methylcrotonyl-CoA carboxylase have not been determined for any organism. Chicken liver propionyl-CoA carboxylase has been found to be similar in subunit size and structure to propionyl-CoA carboxylase from other sources such as pig heart (Kaziro *et al.*, 1961), the nematode *Turbatrix aceti* [actually a more general acyl-CoA carboxylase, (Meyer *et al.*, 1978)], bovine kidney (Lau *et al.*, 1979), human liver (Kalousek *et al.*, 1980), *Mycobacterium smegmatis* (Haase *et al.*, 1982) and sheep liver (Goodall, *et al.*, 1985). The chicken liver enzyme has a biotin-containing α subunit of M_r = 77,000 and a β subunit of M_r = 56,000 (Whittle, 1986). The human propionyl-CoA carboxylase α subunit mRNA is 2.9 kb long and the α subunit cDNA clone which was characterized by Lamhonwah *et al.* (1986) appears to contain ~400 bp of 3′ untranslated sequence. The human propionyl-CoA carboxylase β subunit clone was

found to hybridize to two human fibroblast mRNA species one of 2.0 kb and weakly to an uncharacterized 4.5 kb species (Lamhonwah *et al.*, 1986).

The size of the 3.3 kb band was in keeping with the size required to encode the α subunits of either propionyl-CoA carboxylase or 3-methylcrotonyl-CoA carboxylase. However, the increased intensity of the hybridization signal from the human propionyl-CoA carboxylase α subunit probe would indicate that it was more likely to represent the chicken propionyl-CoA carboxylase α subunit mRNA. The 2.2 kb mRNA also appeared to be similar in size to the smaller human propionyl-CoA carboxylase β subunit mRNA. The human propionyl-CoA carboxylase α subunit clone was not observed to cross-hybridize to the 2.0 kb and 4.5 kb human fibroblast mRNA species which were detected by the human propionyl-CoA carboxylase β subunit clone (Lamhonwah *et al.*, 1986). However, in view of the low stringency washing conditions used in the present study it was more likely that the human propionyl-CoA carboxylase α subunit probe would be able to detect the chicken propionyl-CoA carboxylase β subunit mRNA if sequence homology existed. Again the human propionyl-CoA carboxylase α subunit probe yielded a stronger hybridization signal than the pyruvate carboxylase probe and therefore a tentative identification of the 2.2 kb as representing the propionyl-CoA carboxylase β subunit mRNA could be made. Experimental evidence for immunological cross-reactivity between antibodies directed against pyruvate carboxylase and propionyl-CoA carboxylase from both sheep and chicken liver has been described by Mottershead *et al.* (1984) and Oliver *et al.* (1984). Therefore the indication of cross-hybridization reported here provides further evidence for a family relationship between the biotin carboxylases.

The identity of 1.7 kb band remains speculative but it may represent an mRNA which fortuitously hybridized to the probe, a breakdown product from one of the biotin carboxylase mRNAs or the mRNA for an as yet unrecognized protein which is homologous to pyruvate carboxylase and propionyl-CoA carboxylase. The lack of any hybridization signal by the rat pyruvate carboxylase probe to smaller mRNAs was probably the result of the higher washing stringency employed with this probe.

The cross-hybridization of biotin carboxylase probes to other unexpected mR-

NAs has also been detected by Lamhonwah *et al.* (1986) as mentioned above, by Bai *et al.* (1986) using an acetyl-CoA carboxylase cDNA probe and by Kraus *et al.* (1986a). However, as yet there is no indication that these mRNAs were those of related enzymes. Experimental evidence for protein structural homology between the biotin carboxylases has already been provided by cross-reaction of both monoclonal and polyclonal antibodies directed to pyruvate carboxylase against propionyl-CoA carboxylase (Oliver *et al.*, 1984) and acetyl-CoA carboxylase (Mottershead *et al.*, 1984). Therefore Northern hybridization data presented here is the first experimental indication of homology at the RNA level.

5.4.3 Rat Liver Pyruvate Carboxylase cDNA Clones

The use of the human pyruvate carboxylase clone as a probe for rat liver pyruvate carboxylase cDNA clones was successful because of the very high degree of DNA sequence homology between the two species. In fact the level of homology is significantly greater than that in other coding sequences from these species. For example, there is 76.1% DNA sequence homology between the coding regions of rat and human serum albumins. This would suggest that pyruvate carboxylase is evolving at a slower rate than a non-catalytic protein such as albumin in these vertebrate species. The comparison of the rat pyruvate carboxylase cDNA sequence to the yeast pyruvate carboxylase gene sequence revealed a much lower level of homology than to the human enzyme. This appears to reflect the considerable period of time which has been estimated to be ~1000 million years since the separation of mammalian and yeast evolutionary lines (Dayhoff, 1976). Conservation of the DNA sequence has occurred to a much greater extent in the vicinity of the biotin attachment site codons. This appears to reflect a common evolutionary origin of this sequence as well as constraints on the structure of the protein product.

The chicken pyruvate carboxylase peptide sequence which was found to be homologous provides the first sequence comparison available between an avian and a mammalian pyruvate carboxylase which is not in the vicinity of the highly conserved biotin attachment site. Therefore this may serve as a guide to the more general level

of sequence homology in pyruvate carboxylase from these species. In yeast pyruvate carboxylase, the apparent insertion into the sequence of a block of ~50 amino acids provides evidence of structural evolution involving more substantial changes than individual point mutations in the gene. The possibility exists that this may represent a rearrangement of sequences as yet unidentified in rat pyruvate carboxylase or a deletion of rat pyruvate carboxylase sequence. The implications of gene rearrangement and further evidence for its occurrence in the biotin carboxylases is presented in Chapter 6.

Recent studies (Khew-Goodall, 1985; G. Booker and J.C. Wallace, personal communication) have shown that, in addition to the biotinyl prosthetic group, all the components of the enzyme's active site necessary to catalyse an acetyl-CoA-independent carboxylation of pyruvate (Keech and Attwood, 1985) are present in a C-terminal fragment of $M_r \simeq 75{,}000$. In electron microscopic images this represents the proximal 6.5 nm of each subunit in the tetramer (M. Rohde, personal communication). The lack of a consensus ATP binding site within the C-terminal 306 amino acids suggests that the ATP site should therefore be located within approximately 375 amino acids N-terminal to the sequence already determined. The acetyl-CoA binding site would be expected to be located within the cleaved $M_r \simeq 40{,}000$ fragment. At present it is difficult to reconcile this location for the acetyl-CoA binding site with either the electron microscopic studies which indicated that the acetyl-CoA binding site was in the vicinity of the biotin binding site near the inter-subunit junction (Chapman-Smith, 1981) or with the model proposed for acetyl-CoA binding in this region which was discussed in Chapter 1 (Keech, 1980). Further analysis of the enzyme using affinity labelling of the acetyl-CoA site accompanied by protein sequencing may solve this problem.

Predictions of the structure of proteins enable comparisons to be made which may reveal regions of similar structure in related biotin carboxylase which are not immediately apparent from an examination of sequence homology. Predictions of the secondary structure have been reported for the transcarboxylase 1.3S subunit (Wood and Kumar, 1985). Similarities may be observed between the predicted rat liver pyruvate carboxylase structure and the transcarboxylase predicted structure

especially in the biotin binding site region and also in the N-terminal region. It might be expected that there would be tertiary structure constraints on both the primary sequence and the secondary structure in order to produce the cleft environment in which the biotin has been reported to be positioned in electron microscopic studies of pyruvate carboxylase (Johannssen *et al.*, 1983). This might also be expected to influence the primary sequence in this region of other biotin carboxylases such as transcarboxylase.

The location of the lysyl residue representing the presumed site of biotinylation in rat liver pyruvate carboxylase, 35 amino acids from the C-terminus, was identical to the location found in several other biotin carboxylases (see Chapter 6). The expression and biotinylation *in vivo* of the *P. shermanii* transcarboxylase 1.3S biotinyl subunit by Murtif *et al.* (1985) provided an experimental system to study the sequence and structural requirements for biotinylation. In a recent study Murtif and Samols (1987) used site-directed mutagenesis of the 1.3S subunit clones to alter the sequence of the expressed protein. They determined that for biotinylation to occur, a hydrophobic amino acid such as isoleucine, valine or leucine was required to be located at the penultimate C-terminal position. These workers also established that the methionine residues adjacent to the biocytin were not required for biotinylation but were important for the biotinyl group to act as a carboxyl acceptor.

Chapter 6

Isolation and 3′ End Structure of the Rat Pyruvate Carboxylase Gene

6.1 Introduction

The biotin carboxylases catalyse essentially similar reactions involving the transfer of a carboxyl group either from bicarbonate or from a carboxylated donor molecule to an acceptor molecule (Wood and Barden, 1977). As was mentioned in Chapter 1, the subunit structures of these enzymes vary considerably both between enzymes and between the same enzyme in different species. Lynen (1975), Obermayer and Lynen (1976) and Lynen (1979) in reviewing the structure of the biotin enzymes proposed that the biotin enzymes could be divided into three groups which represent different stages in the evolution of an enzyme family (Fig. 1.4). The hypothesis also proposed that the generation of multifunctional enzyme subunits may have been the result of the fusion of the genes encoding individual monofunctional subunits.

The discovery approximately ten years ago of RNA splicing in adenovirus mRNA (Berget *et al.*, 1977; Chow *et al.*, 1977) was rapidly followed by a series of reports of the presence in eukaryotic genes of intervening non-coding DNA (Breathnach *et al.*, 1977; Glover and Hogness, 1977; Tonegawa *et al.*, 1978). Gilbert (1978) coined the term introns for intervening sequences and the term exons for expressed sequences and suggested that individual functional regions of proteins may be encoded by individual exons. This led to the development of a model of protein evolution wherein exonic regions encoded independently-folded domains in the protein and these exons could have been re-arranged and exchanged between different genes resulting in novel protein structures (Blake, 1978, 1979). In view of Lynen's hypothesis of biotin carboxylase evolution, this group of enzymes would appear to provide a model system for studying the evolution of an enzyme family. This is particularly pertinent because biotin carboxylases are present in both prokaryotes and eukaryotes and have been studied to a significant extent in both groups of organisms.

The sequence analyses which have been already performed on biotin carboxylase genes have been restricted to lower eukaryotes and prokaryotes. These include the recent sequencing of the yeast pyruvate carboxylase gene (F. Lim, personal communication; Morris *et al.*, 1987) and the sequencing of the transcarboxylase operon from *P.shermanii* (Murtif *et al.*, 1985; Thornton *et al.*, 1987). These studies have shown that neither the yeast pyruvate carboxylase gene nor the transcarboxy-

lase gene contain introns the avenue of structural analysis described above was not available in these genes. However, since introns have been found in most genes greater than 0.55 kb long in higher eukaryotes (Naora and Deacon, 1982) it could be expected that studies of eukaryotic biotin carboxylase genes would shed light on the structural relationship between the biotin carboxylases.

Investigations of higher eukaryotic biotin carboxylase genes have been made by three groups of workers. Freytag and Collier (1984) established that pyruvate carboxylase existed as a single copy gene on chromosome number 11 in the human genome without any pseudogene copies, which have been found for other genes (Proudfoot and Maniatis, 1980; Lacy and Maniatis, 1980). Lamhonwah *et al.* (1986) mapped the location of the human propionyl-CoA carboxylase α subunit gene *PCCA* and β subunit gene *PCCB* to different chromosomes, numbers 13 and 3 respectively. Kraus *et al.* (1986) also localized the *PCCB* gene to chromosome number 3 in the human genome using a rat propionyl-CoA carboxylase cDNA clone as the probe.

The absence of any previous sequence analyses of the gene structure of any of the biotin carboxylases in higher eukaryotes, particularly vertebrates, prompted an investigation of the rat pyruvate carboxylase gene. Both the human and rat pyruvate carboxylase cDNA clones, which were available for use as probes to screen a rat genomic library, represented the 3′ end of the cDNA. This ensured that the biotin attachment site region, which has great structural and functional significance in pyruvate carboxylase, could be expected to be present in any genomic clones selected with these probes.

The results presented in this chapter describe the isolation and characterization of two rat genomic clones. Sequence analysis of a restriction fragment common to both clones identified them as encoding the rat pyruvate carboxylase gene. The coding sequence was divided into two exons and the 5′ boundary of the most 3′ exon defined an 82 amino acid polypeptide. The homology present in protein sequence comparisons in this region of several biotin carboxylases suggests that the 3′ exon encodes a biotin-binding domain which has a common evolutionary origin in these and possibly all biotin carboxylases.

6.2 Specific Methods

6.2.1 Dale Deletion Sequencing System

M13mp18 and M13mp19 clones containing terminal insert deletions were prepared using the method described by Dale *et al.* (1985). Single-stranded DNA was prepared from recombinant M13 clones and was annealed with the 29mer oligonucleotide for M13mp18 clones or the 20mer oligonucleotide for M13mp19 clones. The double-stranded region of the oligonucleotide/M13 DNA hybrid was cleaved with *Eco*RI for M13mp18 clones or *Hin*dIII for M13mp19 clones. The linearized DNA was incubated with T4 DNA polymerase for 30 min, so that the 3′–5′ exonuclease activity associated with this enzyme could remove 5′ mononucleotides and aliquots were taken at 5 min intervals. After heat inactivation of the T4 DNA polymerase the time point samples were re-combined and the DNA was tailed with dATP for M13mp18 clones or dGTP for M13mp19 clones using terminal transferase. The tailed M13 DNA was then re-annealed with the 29mer for M13mp18 clones or the 20mer for M13mp19 clones and T4 DNA ligase was used to covalently close and re-circularize the M13 DNA. The ligated DNA was then transformed into *E. coli* JM101 and the resulting white plaques were picked. Single-stranded DNA was prepared and the size of the insert in the isolated clones was determined by electrophoresis of this DNA on a 0.8% (w/v) agarose, 1×TBE gel. Clones with inserts sequentially deleted by ~200 bp were chosen for sequencing using the dideoxy chain-termination method (Sections 2.2.22 and 2.2.23).

6.3 Results

6.3.1 Isolation of Rat Genomic Clones

Knowledge of the structure of the genes for biotin carboxylases especially those of higher eukaryotes containing introns might be expected to provide insights into structural and evolutionary relationships between these enzymes. A rat genomic library prepared from *Hae*III partially digested rat genomic DNA which was cloned

using DNA linkers into the *Eco*RI site of the λCharon 4A vector was kindly supplied by Dr. J. Bonner (Phytogen Corp., CA, USA). A total of 500,000 pfu from the library were plated and screened (Section 2.2.19). The isolated insert of the rat liver pyruvate carboxylase cDNA clone λRL1.1 was [^{32}P] labelled by nick-translation (Section 2.2.18.c) and used as the probe (Section 2.2.21.b). The filters were washed at moderate stringency (0.5×SSC, 0.1% SDS at 50°C) and ten positive signals were detected. After the second round of screening two strongly hybridizing clones were isolated which were termed λRG1.2 and λRG1.4.

6.3.2 Restriction Mapping of Rat Genomic Clones

DNA was prepared from the λRG1.2 and λRG1.4 clones as described in Section 2.2.14.f and the inserts of the clones were restriction mapped. The clone DNAs were digested with a series of restriction enzymes both singly and in combination and the digests were electrophoresed on 0.8% (w/v) agarose, 1×TBE gels. An example of a set of restriction digests of the λRG1.2 and λRG1.4 DNAs is given in Fig. 6.1a. The λRG1.2 and λRG1.4 clones were found to have inserts of 12.7 kb and 12.8 kb respectively and the resulting restriction maps indicated that these clones had several fragments in common. A comparison of the restriction maps of the two clones indicated that the inserts overlap each other by approximately 7 kb (Fig. 6.2).

6.3.3 Localization of Coding Region in Rat Genomic Clones

The coding region in the λRG1.2 and λRG1.4 clones was localized by probing a Southern transfer (Section 2.2.11.a) of the restriction digests shown in Fig. 6.1a. To determine if the clones contained the biotin attachment site, the 14 base B-14M oligonucleotide directed to the biotin attachment site was used to probe the Southern filter. The B-14M probe was 5′ [^{32}P] labelled (Section 2.2.18.a) and hybridized using the conditions described in Section 2.2.21.a. The filter was washed in 1×SSC, 0.1% SDS at 37°C and autoradiographed. The B-14M probe hybridized to an internal 1.4 kb *Bam*HI/*Hin*dIII fragment which was common to both clones

Figure 6.1

Restriction Digests of λRG1.2 and λRG1.4 DNA

a Restriction enzyme analysis of λRG1.2 and λRG1.4 DNA. Aliquots of 5.0 μg of λRG1.2 and λRG1.4 DNA were digested with 10 units of the indicated restriction enzyme in single digests and in double digests by redigestion with 10 units of the second enzyme. The fragments were recovered after phenol extraction and ethanol precipitation then electrophoresed on a 1.0% agarose, 1×TBE gel (Section 2.2.9a). The gel was stained with 0.3 μg/ml ethidium bromide and photographed under short-wave UV illumination. The sizes of the marker DNA bands are shown in kb.

Track M1 : λ_{wt}/*Hin*dIII marker,
Track 1 : λRG1.2/*Eco*RI,
Track 2 : λRG1.2/*Bam*HI,
Track 3 : λRG1.2/*Hin*dIII,
Track 4 : λRG1.4/*Eco*RI,
Track 5 : λRG1.4/*Bam*HI,
Track 6 : λRG1.4/*Hin*dIII,
Track 7 : λRG1.2/*Eco*RI/*Bam*HI,
Track 8 : λRG1.2/*Eco*RI/*Hin*dIII,
Track 9 : λRG1.2/*Bam*HI/*Hin*dIII,
Track 10 : λRG1.4/*Eco*RI/*Bam*HI,
Track 11 : λRG1.4/*Eco*RI/*Hin*dIII,
Track 12 : λRG1.4/*Bam*HI/*Hin*dIII,
Track M2 : SPP-1/*Eco*RI marker.

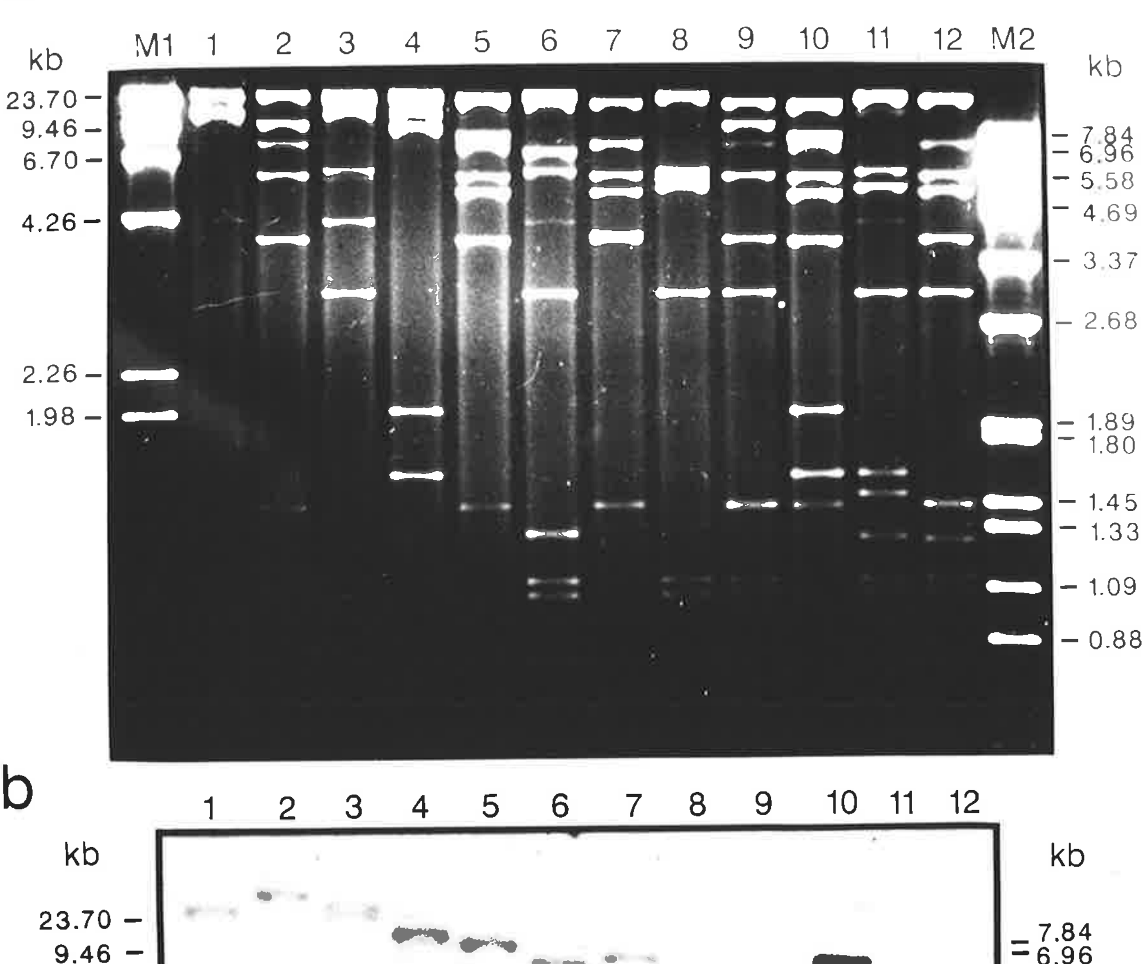
Figure 6.1 Continued

b Southern transfer of restriction digests of λRG1.2 and λRG1.4 DNA probed with the B-14M biotin site oligonucleotide. The DNA fragments in the gel shown in Fig. 6.1a were transferred to nitrocellulose using the Southern transfer method (Section 2.2.11a) and the filter was hybridized with the B-14M oligonucleotide which was 5′-[^{32}P] labelled (Section 2.2.21a). The filter was washed in 1×SSC, 0.1% SDS at 37°C and autoradiographed at -80°C with an intensifying screen. The size of the markers are indicated at the side of the figure in kb.

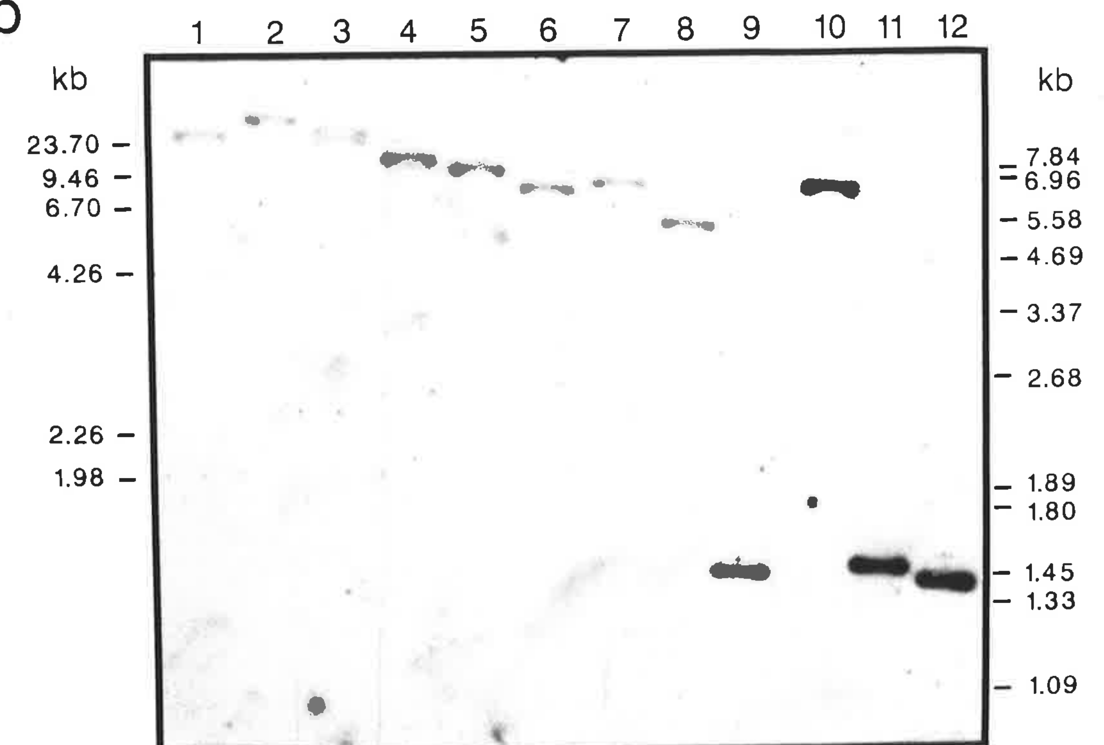
c Southern transfer of restriction digests of λRG1.2 and λRG1.4 DNA probed with the λRL1.1 insert. The filter probed in Fig. 6.1b was thoroughly washed in TE buffer at 100°C to remove the bound [^{32}P] labelled probe and was hybridized with the λRL1.1 insert (Section 2.2.21b) which was [^{32}P] labelled by nick-translation (Section 2.2.18c). The filter was washed in 0.5×SSC, 0.1% SDS at 65°C and autoradiographed at -80°C with an intensifying screen for 18 hr. The size of the markers are indicated at the side of the figure in kb.

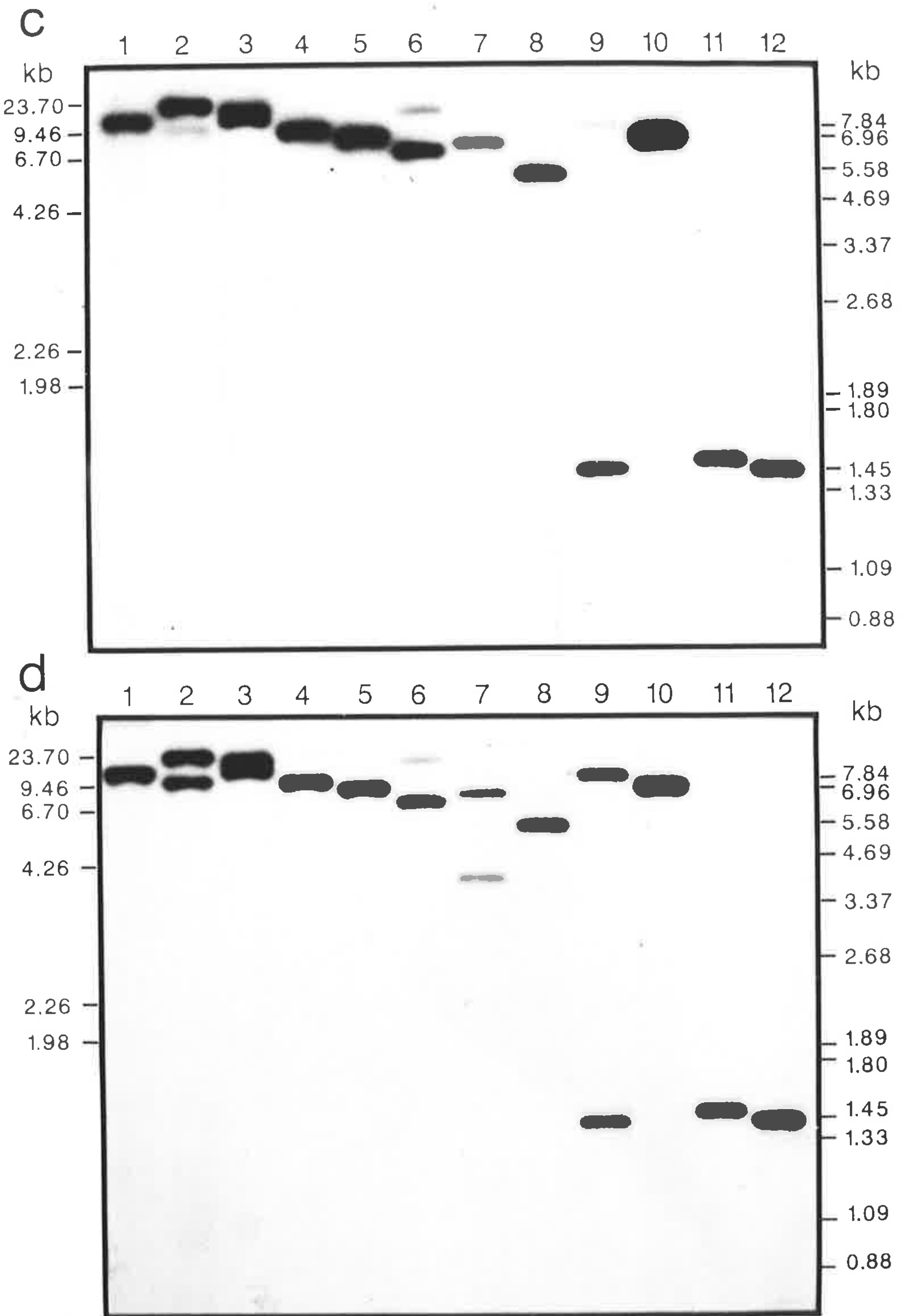
d Southern transfer of restriction digests of λRG1.2 and λRG1.4 DNA probed with the λRL2.35 insert. The filter probed in Fig. 6.1c was thoroughly washed in TE buffer at 100°C to remove the bound [^{32}P] labelled probe and was hybridized with the λRL2.35 insert (Section 2.2.21b) which was [^{32}P] labelled by nick-translation (Section 2.2.18c). The filter was washed in 0.5×SSC, 0.1% SDS at 65°C and autoradiographed at -80°C with an intensifying screen for 18 hr. The size of the markers are indicated at the side of the figure in kb.

a



b




c
1
2
3
4
5
6
7
8
9
10
11
12
kb
23.70
9.46
6.70
4.26
2.26
1.98
kb
7.84
6.96
5.58
4.69
3.37
2.68
1.89
1.80
1.45
1.33
1.09
0.88
d
1
2
3
4
5
6
7
8
9
10
11
12
kb
23.70
9.46
6.70
4.26
2.26
1.98
kb
7.84
6.96
5.58
4.69
3.37
2.68
1.89
1.80
1.45
1.33
1.09
0.88

(Fig. 6.1b). The Southern filter was subsequently reprobed with the λRL1.1 and λRL2.35 inserts after washing in TE buffer for 5 min at 90°C with a similar washing step between probings to remove [^{32}P] labelled DNA bound to the filter. Both of these probes were [^{32}P] labelled by nick-translation (Section 2.2.18.c) and were hybridized using the conditions described in Section 2.2.21.b. The filter was washed in 0.5×SSC, 0.1% SDS at 65°C for both probes.

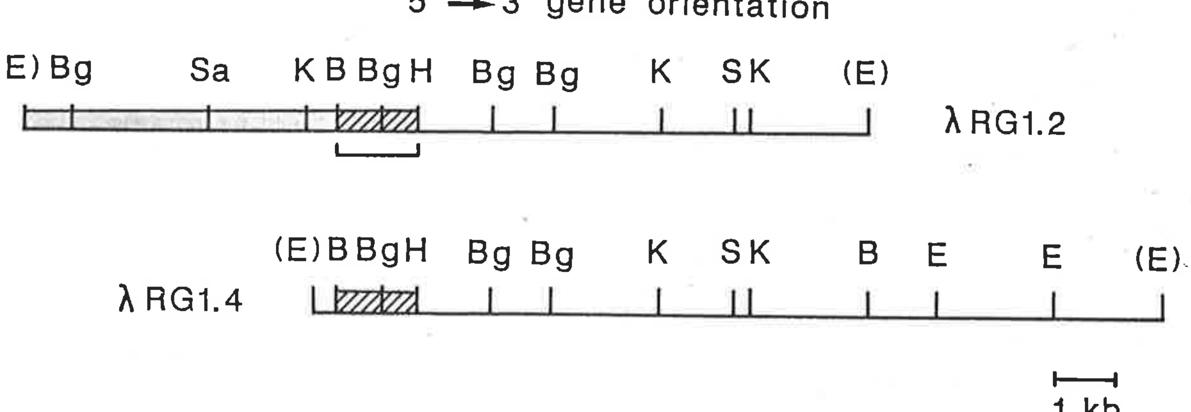
The probing of this filter with both the λRL1.1 and λRL2.35 inserts was designed to establish the extent of the coding sequence. Both of the cDNA clone inserts hybridized strongly to the 1.4 kb *Bam*HI/*Hin*dIII fragment which the B-14M probe detected in both clones. In addition the the λRL1.1 insert hybridized weakly to the terminal 4.3 kb *Eco*RI/*Bam*HI fragment in the λRG1.2 clone insert. The λRL2.35 insert hybridized to the terminal 4.3 kb *Eco*RI/*Bam*HI much more strongly (see Fig. 6.2 for a map of these fragments). This suggested that the coding region was split between the adjacent 1.4 kb *Bam*HI/*Hin*dIII and 4.3 kb *Eco*RI/*Bam*HI fragments and that the 1.4 kb fragment contained a majority of the λRL1.1 coding sequence representing the 3′ end of the gene. The difference between the λRL2.35 and λRL1.1 insert hybridization signals is particularly obvious in the *Bam*HI digestions of the λRG1.2 clone (tracks 2, 7 and 9 in Fig. 6.1). The additional coding sequence in the λRL2.35 clone did not hybridize to additional bands in the λRG1.4 clone because as indicated in the restriction map of the clones in Fig. 6.2 the 4.3 kb fragment of λRG1.2 extends beyond the end of the λRG1.4 clone by approximately 4 kb.

Hybridization by these probes to the same 1.4 kb *Bam*HI/*Hin*dIII fragment suggested that there was a high probability that these clones encoded a rat biotin carboxylase gene. The 1.4 kb *Bam*HI/*Hin*dIII fragment was therefore isolated from a restriction digest of λRG1.2 (Sections 2.2.9.a and 2.2.10) and subcloned into the plasmid vector pUC19. The resulting clone which was termed pRGBH.1 was used as the source of the 1.4 kb restriction fragment in subsequent studies.

Figure 6.2

Restriction Maps of λRG1.2 and λRG1.4 Rat Genomic Clones

The restriction maps of the λRG1.2 and λRG1.4 rat genomic clones are shown aligned with each other. The restriction sites are indicated as follows : B = *Bam*HI, Bg = *Bgl*II, E = *Eco*RI, H = *Hin*dIII, K = *Kpn*I, S = *Sma*I and Sa = *Sal*I. The restriction sites in brackets represent sites in the λCharon 4A vector. The λRG1.2 and λRG1.4 clones were found to have inserts of 12.7 kb and 12.8 kb respectively which overlap by approximately 7 kb. The 1.4 kb *Bam*HI/*Hin*dIII fragment in λRG1.2 which was subcloned into pUC19 to yield pRGBH.1 is marked. The fragments in these clones which hybridized to the B-14M biotin site probe is indicated by the box marked ( ). The fragment in λRG1.2 which hybridized strongly to the λRL2.35 insert is indicated by the box marked ( ). The λRL1.1 insert hybridized strongly to the fragment marked ( ) and weakly to the fragment marked ( ).


5' → 3' gene orientation
(E) Bg
Sa
K B Bg H
Bg Bg
K
S K
(E)
λ RG1.2
λ RG1.4
(E) B Bg H
Bg Bg
K
S K
B
E
E
(E)
1 kb

6.3.4 Northern Transfer of BRL-3A Cell RNA Probed with 1.4 kb *Bam*HI/*Hin*dIII Genomic Fragment

In view of the sequence homology which has been demonstrated between the biotin carboxylases (this report, Lamhonwah *et al.*, 1986) it was necessary to identify which of the biotin carboxylase genes the λRG1.2 and λRG1.4 clones represented. This could be achieved by identifying the size of the mRNA transcript from the gene encoded in these clones. Since the 1.4 kb *Bam*HI/*Hin*dIII fragment hybridized to the cDNA probes described above, it must have contained at least part of the coding region of the gene. The 1.4 kb fragment was therefore isolated from pRGBH.1 and used to probe a Northern transfer of the rat liver cell line BRL-3A RNA. The gene fragment hybridized to 3 poly(A)$^+$ RNA species of 4.2 kb, 10 kb and >20 kb (Fig. 6.3). Hybridization to a 4.2 kb species was consistent with the size of the mRNA species of rat liver pyruvate carboxylase however no larger molecular weight mRNA species had been observed previously when the λRL1.1 insert was used to probe a Northern transfer of this RNA, see Fig. 5.8. The possibility existed that the higher molecular weight bands resulted from hybridization of the non-coding regions of the 1.4 kb fragment to incompletely processed mRNA present in the poly(A)$^+$ RNA preparation.

6.3.5 Rat Genomic Southern Probed with 1.4 kb *Bam*HI/*Hin*dIII Genomic Fragment

In view of the equivocal nature of the Northern hybridization result described above an analysis of the hybridization of the λRL1.1 insert to rat genomic DNA was performed. The λRL1.1 insert was used as a probe to establish if; (i) a single copy of the gene present in λRG1.2 and λRG1.4 was present in the rat genome and (ii) a 1.4 kb fragment would be detected in a *Bam*HI/*Hin*dIII digest of rat genomic DNA.

Samples of 10 μg of rat genomic DNA (kindly supplied by L. Mattschoss, Dept. of Biochemistry, Univ. of Adelaide) were digested with the restriction enzymes

Figure 6.3

Northern Transfer of BRL-3A Cell RNA Probed with pRGBH.1 1.4 kb Insert

Poly(A)$^+$ RNA and poly(A)$^-$ RNA was prepared from BRL-3A rat liver cell line RNA (Sections 2.2.1 and 2.2.2). RNA samples containing 10 μg and the DNA markers were denatured by glyoxalation and electrophoresed on a 1.0% agarose, 10 mM sodium phosphate pH 6.5 gel as described in Section 2.2.9b. The RNA and DNA markers were transferred to nitrocellulose (Section 2.2.11b) and the filter was hybridized to the 1.4 kb *Bam*HI/*Hin*dIII rat genomic fragment (Section 2.2.21.b) prepared from the pRGBH.1 subclone (Sections 2.2.9a and 2.2.10). The probe was [^{32}P] labelled by nick-translation (Section 2.2.18c) and the filter was washed at 0.5×SSC, 0.1% SDS at 50°C. The filter was autoradiographed for 18 hr at -80°C with an intensifying screen. The sizes of the λ_{wt}/*Hin*dIII marker DNA bands are shown in kb.

Track 1 : 10 μg of BRL-3A cell poly(A)$^+$ RNA,

Track 2 : 10 μg of BRL-3A cell poly(A)$^-$ RNA,

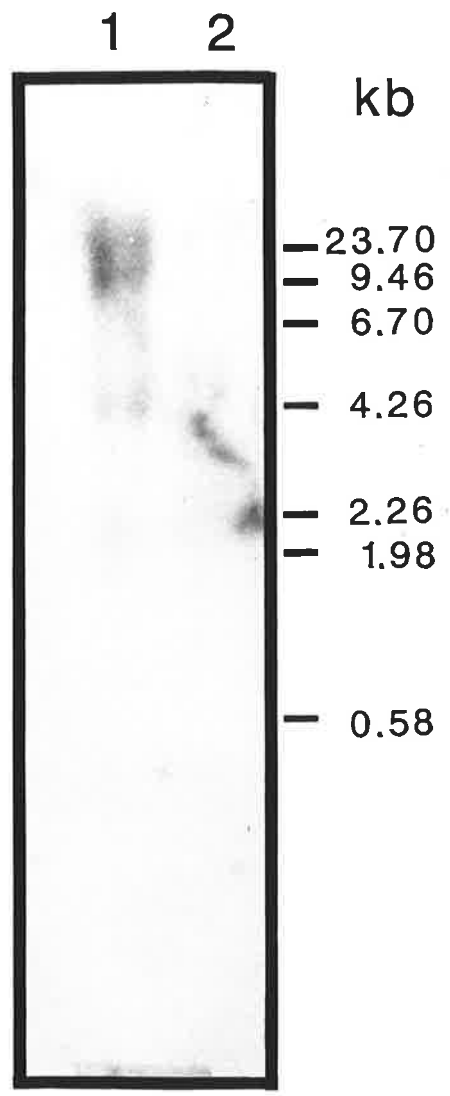

1
2
kb
23.70
9.46
6.70
4.26
2.26
1.98
0.58

Figure 6.4

Southern Transfers of Rat Genomic DNA Probed with λRL1.1 Insert

a Samples of 10 μg of rat genomic DNA were digested with the restriction enzymes *Eco*RI, *Bam*HI and *Hin*dIII both singly and in combination. The digested DNA was electrophoresed on a 0.8% (w/v) agarose, 1×TBE gel (Section 2.2.9a) then transferred to nitrocellulose using the Southern transfer method (Section 2.2.11a). The filter was probed with the rat pyruvate carboxylase cDNA clone λRL1.1 insert (Section 2.2.21b) which was [^{32}P] labelled by nick-translation (Section 2.2.18c). The filter was washed at 0.5×SSC, 0.1% SDS at 65°C and autoradiographed for 72 hr at -80°C with an intensifying screen. The sizes of the λ_{wt}/*Hin*dIII and SPP-1/*Eco*RI marker DNA bands are shown in kb.

Track 1 : *Bam*HI digestion of 10 μg of rat genomic DNA,

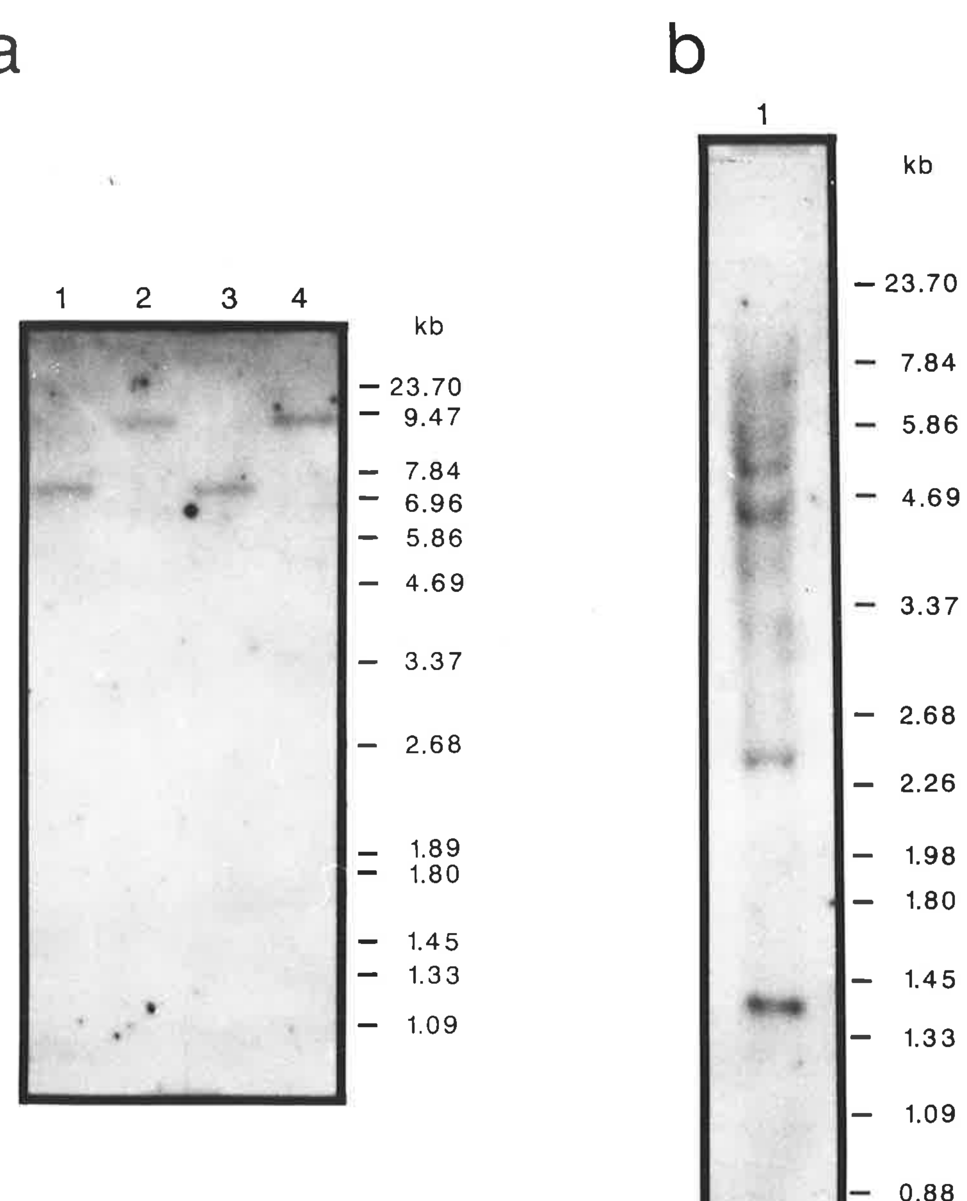
Track 2 : *Hin*dIII digestion of 10 μg of rat genomic DNA,

Track 3 : *Eco*RI/*Bam*HI digestion of 10 μg of rat genomic DNA,

Track 4 : *Eco*RI/ *Hin*dIII digestion of 10 μg of rat genomic DNA.

b A 50 μg sample of rat genomic DNA was digested with the restriction enzymes *Bam*HI and *Hin*dIII and the digested DNA was electrophoresed on a 0.8% (w/v) agarose, 1×TBE gel (Section 2.2.9a). The DNA was transferred to nitrocellulose using the Southern transfer method (Section 2.2.11a) and the filter was probed with the rat pyruvate carboxylase cDNA clone λRL1.1 insert (Section 2.2.21b) which was [^{32}P] labelled by nick-translation (Section 2.2.18c). The filter was washed at 0.5×SSC, 0.1% SDS at 65°C and autoradiographed for 72 hr at -80°C with an intensifying screen. The sizes of the λ_{wt}/*Hin*dIII and SPP-1/*Eco*RI marker DNA bands are shown in kb.

Track 1 : *Bam*HI/*Hin*dIII digestion of 50 μg of rat genomic DNA.


a
1
2
3
4
kb
23.70
9.47
7.84
6.96
5.86
4.69
3.37
2.68
1.89
1.80
1.45
1.33
1.09
b
1
kb
23.70
7.84
5.86
4.69
3.37
2.68
2.26
1.98
1.80
1.45
1.33
1.09
0.88

*Eco*RI, *Bam*HI and *Hin*dIII both singly and in combination. The digested DNA was electrophoresed on a 0.8% (w/v) agarose, 1×TBE gel and transferred to nitrocellulose (Sections 2.2.9.a and 2.2.11.a). The Southern transfer was probed with the λRL1.1 insert and washed at moderate stringency (0.5×SSC, 0.1% SDS at 50°C) and autoradiographed.

The probe hybridized to an 8.5 kb *Bam*HI fragment, a >10 kb *Hin*dIII fragment, an 8.5 kb *Eco*RI/*Bam*HI fragment and a >10 kb *Eco*RI/*Hin*dIII fragment (Fig. 6.4a). The hybridization to an 8.5 kb *Bam*HI fragment was consistent with the restriction mapping of the *Bam*HI sites in the λRG1.2 and λRG1.4 genomic clones (Fig. 6.2). The lack of any *Eco*RI sites in the 8.5 kb *Bam*HI fragment from λRG1.4 was confirmed by the failure of *Eco*RI re-digestion to cleave the 8.5 kb *Bam*HI fragment of rat genomic DNA. The hybridization of the probe to *Hin*dIII and *Eco*RI/*Hin*dIII fragments of >6 kb was also consistent with the restriction map of the λRG1.2 clone because there were no *Hin*dIII or *Eco*RI sites within 6 kb from the *Hin*dIII site bordering the 1.4 kb *Bam*HI/*Hin*dIII fragment to the end of the clone insert. Hybridization of the λRL1.1 probe to these single restriction fragments indicated that the gene was present as a single copy in the rat genome.

As a result of problems in detecting a hybridization signal from *Bam*HI/*Hin*dIII digested rat genomic DNA this experiment was repeated with this combination of restriction enzymes using 50 μg of rat genomic DNA instead of 10 μg of genomic DNA . In this experiment the λRL1.1 insert hybridized strongly to a band at 1.4 kb and weakly to several other bands at 2.4 kb, 4.8 kb and 6 kb (Fig. 6.4b). The restriction map data (Fig. 6.2) indicated that it was unlikely these bands represented partial restriction digest fragments. It was therefore concluded that the weaker hybridization, which was not apparent using smaller amounts of DNA (Fig. 6.4a), represented cross-hybridization to other genes possibly those of other biotin carboxylases. Hybridization of the λRL1.1 insert to a band at 1.4 kb in a Southern transfer of *Bam*HI/*Hin*dIII digested rat genomic DNA provided additional strong evidence that these clones encoded the pyruvate carboxylase gene. However, to identify the clones it was necessary to determine the sequence of the region hybridizing to the cDNA probes.

6.3.6 Sequencing of λRG1.2 1.4 kb *Bam*HI/*Hin*dIII Fragment

The 1.4 kb *Bam*HI/*Hin*dIII fragment was isolated from pRGBH.1 and was subcloned into M13mp18 and M13mp19 for sequence analysis. The deletion sequencing system described by Dale *et al.* (1985) (Section 6.2.1) was used to generate a set of clones with overlapping deletions in their inserts. Agarose gel electrophoresis was then used to estimate the size of the deleted clone inserts and the clones with inserts of the required size were selected such that the entire 1.4 kb fragment could be readily sequenced. The dideoxy chain termination system was used to determine the sequence (Sections 2.2.22 and 2.2.23) and the strategy employed for sequencing of the 1.4 kb fragment is shown in Fig. 6.5.

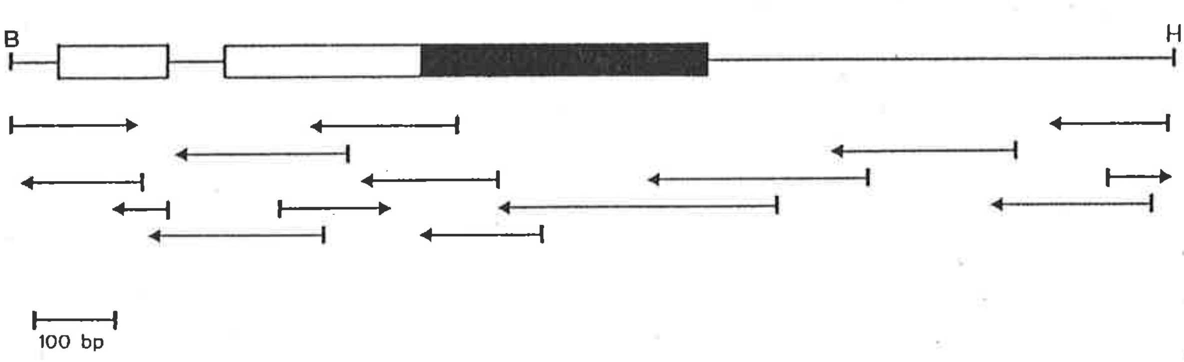
The sequence of the coding regions of the 1.4 kb *Bam*HI/*Hin*dIII fragment shown in Fig. 6.6 is identical with the previously determined sequence of the λRL1.1 rat liver pyruvate carboxylase cDNA clone. This established that λRG1.2 and λRG1.4 contained the 3′ end of the rat pyruvate carboxylase gene. An examination of the sequence revealed that the portion of the coding region present in this fragment was separated by introns into two exons. An intron/exon boundary was identified 56 bp from the terminal *Bam*HI site containing a consensus acceptor splice site as determined by Mount (1982) with a characteristic tract of >10 pyrimidine bases prior to the intron/exon boundary. The 5′ exon contained 141 bp and was terminated by a donor splice site (Mount, 1982). The intron which separated the two exons was remarkably short since it only contained 77 bp. The intron/exon boundary of the most 3′ exon was also very similar to the consensus acceptor splice site. The 3′ exon contained 610 bp which included 246 bp of the coding region and 364 bp of 3′ non-coding sequence.

The end of the gene was defined in the genomic clones using the site of polyadenylation in the cDNA clone sequence. This site can be assigned to either position 1282 or position 1283 in the cDNA clone, see Fig. 5.9b (positions 884 and 885 respectively in the genomic sequence) because the A at position 1283 was not distinguished from the poly (A) tail.

Figure 6.5

Sequencing Strategy for 1.4 kb *Bam*HI/*Hin*dIII Genomic Fragment

The 1.4 kb *Bam*HI/*Hin*dIII fragment was isolated from pRGBH.1 and subcloned into M13mp18 and M13mp19 (Section 2.2.16). A set of overlapping clones with deleted inserts was prepared using the Dale clone deletion system described in Section 6.2.1. The clones were sequenced using the dideoxy chain termination system (Sections 2.2.22 and 2.2.23) and the extent and orientation of each sequence is marked by an arrow. Exonic regions are indicated by boxes with an open box representing coding sequence and a solid box representing non-coding sequence. Intronic and intergenic regions are indicated by a line.



B
H
100 bp

Figure 6.6

Sequence of 1.4 kb *Bam*HI/*Hin*dIII Fragment from λRG1.2 Rat Genomic Clone

The 1482 bp of DNA sequence of the 1.4 kb *Bam*HI/*Hin*dIII fragment is shown together with the inferred amino acid sequence of the coding regions of the two exons. The 5′ exon is 141 bp long and the 3′ exon is 610 bp long including 246 bp of coding sequence and 364 bp of 3′ non-coding sequence. The intron which separates the two exons is 77 bp long. The donor and acceptor splice sites (Mount, 1982) are underlined as is the proposed polyadenylation signal, A-A-T-A-A-A, from position 859–864. The termination codon is marked by (*) and the site of biotinylation is boxed.

```
                                                           val  glu  leu  glu  arg  gly  lys  thr  leu  his  ile  lys
ggatccagggcccacggtcatgacatcctccccagctgtctcctgtcttctccagGTTGAGCTGGAACGGGGCAAGACCTTGCACATCAA
         10        20        30        40        50        60        70        80        90

 ala  leu  ala  val  ser  asp  leu  asn  arg  ala  ala  gln  arg  gln  val  phe  phe  glu  leu  asn  gly  gln  leu  arg  ser  ile  leu  val  lys  asp
AGCCCTGGCTGTAAGCGACCTGAACCGTGCTGCGCAGAGGCAGGTGTTCTTTGAACTCAATGGGCAGCTTCGATCCATTCTGGTTAAAGA
         100       110       120       130       140       150       160       170       180

 thr  gln  ala  met  lys
CACCCAGGCCATGAAGgtacagtgccatcagagccagtcagtggtagggatgggcctgacctgctgtctaatgtcatctctgtcctgtct
         190       200       210       220       230       240       250       260       270

     glu  met  his  phe  his  pro  lys  ala  leu  lys  asp  val  lys  gly  gln  ile  gly  ala  pro  met  pro  gly  lys  val  ile  asp  val  lys  val
gcagGAGATGCACTTCCATCCCAAGGCCTTGAAGGATGTGAAGGGCCAAATTGGGGCCCCTATGCCTGGAAGGTCATAGACGTCAAGGT
         280       290       300       310       320       330       340       350       360

 ala  ala  gly  ala  lys  val  val  lys  gly  gln  pro  leu  cys  val  leu  ser  ala  met  [lys]  met  glu  thr  val  val  thr  ser  pro  met  glu  gly
GGCAGCAGGAGCCAAGGTGGTTAAGGGCCAGCCCCTCTGTGTGCTCAGCGCCATGAAGATGGAGACTGTGGTGACTTCGCCCATGGAGGG
         370       380       390       400       410       420       430       440       450

 thr  ile  arg  lys  val  his  val  thr  lys  asp  met  thr  leu  glu  gly  asp  asp  leu  ile  leu  glu  ile  glu  *
CACTATCCGAAAGGTTCACGTGACCAAGGACATGACTCTGGAAGGCGATGACCTCATCCTAGAGATTGAGTGATCTTACTCCAGACTGGC
         460       470       480       490       500       510       520       530       540

AGCCTGGCCAACCCTACCCCAAGCCTCTCAACAGAAGCTGTGCAGCCAGGGCAGGCCCAGCGAGTACCTGAGGGCTAGGCTTGAGGTCCT
         550       560       570       580       590       600       610       620       630

GTCCCATGGGCACACACACACTACCTGCAATGGCCCTCCCATTGCCCTTCAGCTATTTGTCCTTGTCTTGCTGGCAGGCAGTTCTCACAT
         640       650       660       670       680       690       700       710       720

GTTCTCTCTTGCCAAATAAGGGTCTACTCCTGTGGGAGACCAAGGTGTACAGTAGGTGGCCTTGTACCTGGGAGAGGGGTTCTACCTCTG
         730       740       750       760       770       780       790       800       810

GGGGTAGAGGGAAGAAGACCTAATTCATAGGTCCTGGGAAATTTGCTCAATAAAGTGGCCTTCCCTTGCCCTCCacactaggtcatgtac
         820       830       840       850       860       870       880       890       900

agtctactccaccctagtggcagtgtgggtctctttctctcaaactctcatgggatctaggtcagatccgtcgcctttttttctttaaa
         910       920       930       940       950       960       970       980       990

ctctctagagctcggtaaagtagacacccacaagcctgtctgcctatgttctcaggagacttgttagagcaggccactgcgtaccggccc
         1000      1010      1020      1030      1040      1050      1060      1070      1080

ctcacctctcaaccagcttccacacaggtaggagacaagagttcccagggaggcagtgtatgactatgatactttgttataacctcgtga
         1090      1100      1110      1120      1130      1140      1150      1160      1170

gttatgtcacagctaaccttggggtatgtcactagtaaatgacacatgtccatttgctcactgaccacttcaggagcaggtttcagaac
         1180      1190      1200      1210      1220      1230      1240      1250      1260

agccttaaatgtagagacttagaggatagtcacatttttcaaaaggcaaaggaaaacaaaccgttctagaagtgtggtgagcaggtaaga
         1270      1280      1290      1300      1310      1320      1330      1340      1350

gtggagttacttggtagagtccttgctactttcaggctgtgacctttgtcctgggttcctccagggactgggtgaggcagagggtagca
         1360      1370      1380      1390      1400      1410      1420      1430      1440

gccatatagaaccagaagcctctacctctgagagagaagctt
         1450      1460      1470      1480
```

6.3.7 Sequence Homology Between Rat Liver Pyruvate Carboxylase and Other Biotin Carboxylases

The amino acid sequences of those biotin enzymes for which sequence data is available were aligned with the inferred amino acid sequence of the two exons of the rat pyruvate carboxylase gene using the ALIGNSIG program (Section 2.2.24), see Fig. 6.7. Very strong sequence homology was evident in the immediate vicinity of the biotin attachment site and at the C-termini of these proteins. This homology decreased in the direction of the N-termini of most proteins with the exception of human pyruvate carboxylase which is very highly conserved with respect to rat pyruvate carboxylase throughout its sequence. In the sequence available for comparison there are only four conservative amino acid substitutions which gives an overall level of amino acid homology in this region between the rat and human enzymes of 96%. The level of amino acid homology is much less between the rat and yeast enzymes (43%) as might be expected by the greater evolutionary separation between these species relative to rat and human. The time since the separation of the rat and human lines is estimated to be 75 million years whereas the time since the separation of the rat and yeast lines is estimated to be 1000 million years (Dayhoff, 1976). The sequence obtained recently (Hoffman *et al.*, 1987) for a tomato biotin-containing protein was included in the comparison and was homologous to rat pyruvate carboxylase. The level of homology tends to suggest that the clone probably represents a plant biotin carboxylase.

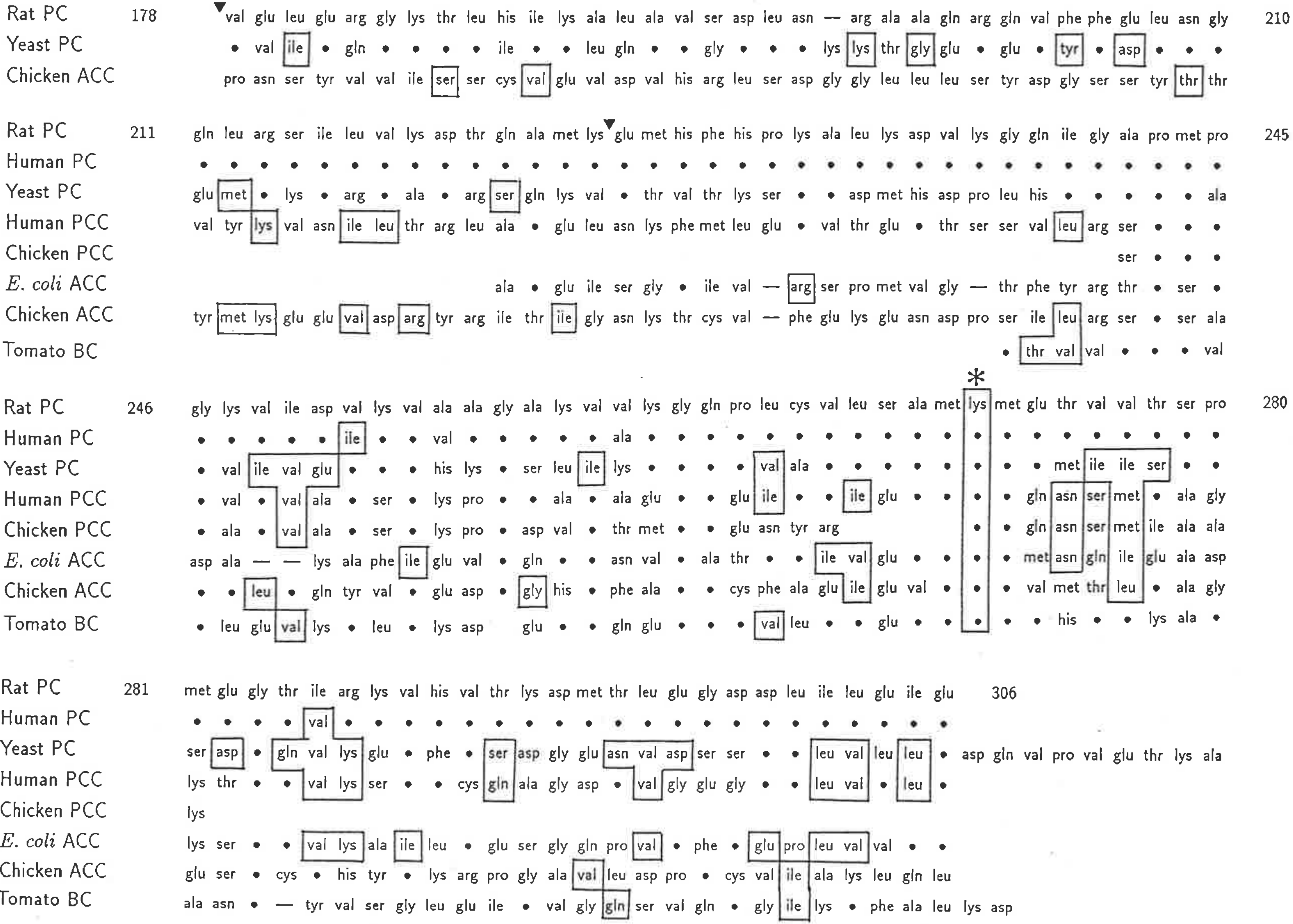
The position of the lysine residue which represents the site of biotinylation, 35 amino acids from the C-terminus, is identical in many of the biotin enzymes which have been currently sequenced including human pyruvate carboxylase (Freytag and Collier, 1984), human propionyl-CoA carboxylase (Lamhonwah *et al.*, 1987), *P. shermanii* transcarboxylase (Murtif *et al.*, 1985) and *E. coli* acetyl-CoA carboxylase (Sutton *et al.*, 1977). As discussed in Section 5.4.3, yeast pyruvate carboxylase has a 7 amino acid C-terminal extension (Morris *et al.*, 1987) and in chicken liver acetyl-CoA carboxylase the biocytin appears to be located much further from the C-terminus, see Fig. 6.7. Unfortunately experimental data supporting the proposed

Figure 6.7

Homology of C-Terminal of Rat Pyruvate Carboxylase to Other Biotin Carboxylases

The published sequences of several biotin carboxylases are shown aligned with the inferred protein sequence of rat liver pyruvate carboxylase using the ALIGN-SIG program (Section 2.2.24). Spaces were introduced into the sequences by the alignment program and are shown as dashes. The sequences compared with rat pyruvate carboxylase are marked by (•) where there is sequence identity and are boxed where there is a conservative difference. The lysine to which biotin is attached is boxed and marked by (*). The numbering of rat pyruvate carboxylase residues takes residue 1 as the most N-terminal inferred amino acid. The positions of introns in the rat pyruvate carboxylase gene is marked (▾). The Z scores which represent the number of standard deviations greater than random sequence homology between the sequences were calculated using 20 randomizations of the data. Where $Z<3$ then $p>0.001, <0.01$ and where $Z\geq3$ then $p<0.001$.

The sequence data was obtained from the following sources :- Yeast PC : yeast pyruvate carboxylase (F. Lim, personal communication, Morris *et al.*, 1987), Human PC : human liver pyruvate carboxylase (Freytag and Collier, 1984; this report; Lamhonwah *et al.*, 1987), Human PCC : human propionyl-CoA carboxylase (Lamhonwah *et al.*, 1987), Chicken PCC : chicken liver propionyl-CoA carboxylase (Whittle, 1986), *E. coli* ACC : *E. coli* acetyl-CoA carboxylase (Sutton *et al.*, 1977), Chicken ACC : chicken liver acetyl-CoA carboxylase (Takai *et al.*, 1987), Tomato BC : a putative biotin carboxylase from tomato (Hoffman *et al.*, 1987). The end of the sequences represent the C-termini of the proteins with the exception of chicken liver propionyl-CoA carboxylase which has not been determined and chicken liver acetyl-CoA carboxylase.



The Z scores for the alignment of the rat liver pyruvate carboxylase sequence with the following sequences are : Yeast PC = 31.42, Human PC = 75.36, Human PCC = 9.70, Chicken PCC = 6.58, *E. coli* ACC = 4.87, Chicken ACC = 2.72, Tomato BC = 12.50.

position of the biotin chicken liver acetyl-CoA carboxylase has not as yet been provided (Takai *et al.*, 1987).

The alignment of the rat pyruvate carboxylase sequence with the transcarboxylase 1.3S subunit sequence using the ALIGNSIG program (Section 2.2.24) introduced a large gap of 48 residues (Fig. 6.8a). This suggested that either an insertion had occurred into the rat pyruvate carboxylase sequence or sequence had been deleted from the transcarboxylase subunit. The possibility of this kind of rearrangement was given credence when the positions of the introns in the rat pyruvate carboxylase gene were found to be very close to the ends of this gap (Fig. 6.8a).

The sequences of the transcarboxylase 5S and 12S subunit gene sequences were recently determined (Thornton et al., 1987) and were kindly supplied by Dr. D. Samols (Case Western Reserve University, OH, USA). A comparison of the inferred amino acid sequences of these subunits with the amino acid sequence of rat liver pyruvate carboxylase revealed a significant level of homology between the central region of the 5S subunit and the N-terminal region of rat liver pyruvate carboxylase (Fig. 6.8b). This is of particular interest since the transcarboxylase 5S subunit contains the pyruvate binding site and the presence of homology suggests that the pyruvate carboxylase pyruvate binding site may be located in the homologous region. Areas of significant homology were not observed between the available rat liver pyruvate carboxylase sequence and the 12S subunit sequence which contains the CoA-ester binding site. The sequence of the transcarboxylase 5S and 12S subunit genes are shown in Appendices A and B respectively.

6.3.8 Intron/Exon Boundaries and Biotin Carboxylase Sequence Homology

The positions of the intron/exon boundaries in the rat pyruvate carboxylase gene were examined to determine if structural information about the biotin carboxylases could be obtained. The position of the 5′ boundary of the most 3′ exon established that an 82 amino acid segment of rat pyruvate carboxylase was encoded by this exon. This is identical to the number of amino acids determined for the

Figure 6.8

Homology Between Both the Transcarboxylase 1.3S and 5S Subunits and Rat Liver Pyruvate Carboxylase

a The sequence of the transcarboxylase 1.3S subunit (Maloy *et al.*, 1979) was aligned with the inferred amino acid sequence of rat liver pyruvate carboxylase using the ALIGNSIG program (Section 2.2.24). The homology present between the 1.3S subunit and the C-terminal region of the rat liver pyruvate carboxylase sequence is shown and is split into two regions separated by a large gap. The positions of introns in the rat pyruvate carboxylase gene are marked (▼). Where there is sequence identity the residues are boxed and where there is a conservative difference this is indicated by shaded boxes. The lysine to which biotin is attached is boxed and marked by (*). The Z score = 9.40, (p <0.001) which represents the number of standard deviations greater than random sequence homology between the sequences were calculated using 20 randomizations of the data.

b The sequence of the transcarboxylase 5S subunit (see Appendix A) was aligned with the inferred amino acid sequence of rat liver pyruvate carboxylase using the SEQHP program (Section 2.2.24). The homology present between the central region of the 5S subunit and the N-terminal region of the rat liver pyruvate carboxylase sequence is shown. Where there is sequence identity the residues are boxed and where there is a conservative difference this is indicated by shaded boxes.

a

PC 121 arg ser ser gly arg arg ala val leu pro pro leu cys gly gly val pro ala ala met tyr pro asp val phe ala gln phe lys asp 150
TC 1.3S 1 met lys leu lys val thr val asn gly thr ala tyr asp val asp 15

PC 151 phe thr ala thr phe gly pro leu asp ser leu asn thr arg leu phe leu gln gly pro lys ile ala glu glu phe glu ▼ val glu leu 180
TC 1.3S 16 val asp val asp lys ser his glu asn pro met gly thr ile leu phe gly gly gly thr gly gly ala pro ala pro arg ala ala gly 45

PC 181 glu arg gly lys thr leu his ile lys ala leu ala val ser asp leu asn arg ala ala gln arg gln val phe phe glu leu asn gly 210
TC 1.3S 46 gly ala gly —

PC 211 gln leu arg ser ile leu val lys asp thr gln ala met lys ▼ glu met his phe his pro lys ala leu lys asp val lys gly gln ile 240
TC 1.3S — ala gly lys ala gly glu gly glu ile 57

PC 241 gly ala pro met pro gly lys val ile asp val lys val ala ala gly ala lys val val lys gly gln pro leu cys val leu ser ala 270
TC 1.3S 58 pro ala pro leu ala gly thr val ser lys ile leu val lys glu gly asp thr val lys ala gly gln thr val leu val leu glu ala 87

PC 271 met lys* met glu thr val val thr ser pro met glu gly thr ile arg lys val his val thr lys asp met thr leu glu gly asp asp 300
TC 1.3S 88 met lys met glu thr glu ile asn ala pro thr asp gly lys val glu lys val leu val lys glu arg asp ala val gln gly gly gln 117

PC 301 leu ile leu glu ile glu 306
TC 1.3S 118 gly leu ile lys ile gly 123

b

TC 5S 255 ser val ala glu met leu glu gly thr gly tyr thr thr asn leu asp tyr asp arg leu his lys ile arg asp his phe lys ala ile 284
PC 1 ala leu val ala val pro lys gly leu leu trp thr thr glu val pro leu glu arg val phe asp tyr ser glu tyr trp glu gly ala 30

TC 5S 285 arg pro lys tyr lys lys phe glu ser lys — thr leu — val asp thr ser — ile phe lys ser gln ile pro gly gly met leu 311
PC 31 arg gly leu tyr ala ala phe asp cys thr ala thr met lys ser gly asn ser asp val tyr glu asn glu ile pro gly gly gln tyr 60

TC 5S 312 ser asn met glu ser glu leu arg ala gln gly ala glu asp lys met asp glu val met ala glu val pro arg val arg arg pro ala 341
PC 61 thr asn leu his phe gln ala his ser met gly leu gly ser lys phe lys glu val lys lys ala tyr val glu ala asn gln ser ala 90

TC 5S 342 pro val phe pro ala pro gly his pro val gln pro asp 354
PC 91 gly gly — pro his gln gly asp thr ile leu gln asp 102

M_r $\simeq$ 9100 C-terminal proteolytic fragment of the independent biotin-containing subunit (M_r $\simeq$ 22,500) of *E. coli* acetyl-CoA carboxylase (Sutton *et al.*, 1977). Such an agreement is unlikely to have occurred by chance and suggests that this exon may be encoding a single functional protein "domain" which could represent an ancestral biotin carrier protein gene.

The transcarboxylase 1.3S biotinyl subunit which contains 123 amino acids is reasonably similar in size to the polypeptide encoded in the 3′ exon as well as to the *E. coli* acetyl-CoA carboxylase biotin carrier protein. In fact, if the 47 amino acids encoded by the more 5′ exon in the rat pyruvate carboxylase gene are added to the 82 amino acid segment then the total of 129 amino acids resembles very closely the size of the 1.3S subunit in size. An examination of the level of homology between the transcarboxylase 1.3S subunit and the rat pyruvate carboxylase sequence reveals that homology is much greater within the 82 amino acid segment (34.2%) than within the 41 amino acid encoded by the more 5′ exon (2.4%).

6.3.9 Structural Homology Between Biotinyl Domains

The secondary structure of the proposed biotinyl domains of rat pyruvate carboxylase, transcarboxylase and *E. coli* acetyl-CoA carboxylase were determined using the CHOUDOT program (Section 2.2.24). These profiles revealed a quite similar structure with possibly more structural features in common than the comparison of the primary sequences would suggest. The turn propensity and α-helix or β-sheet predictions show that a similar pattern of secondary structural is present in all three proteins. The greatest structural homology in the comparison is present between rat pyruvate carboxylase and *E. coli* acetyl-CoA carboxylase. The structure predictions indicate that the N-termini of the regions used for this analysis has a generally lower level of secondary structure than the rest of the proteins indicating that this may be a domain boundary (Fig. 6.9).

Figure 6.9

Protein Structure Prediction for the C-Terminal Region of Three Biotin Carboxylases Using the CHOUDOT Program

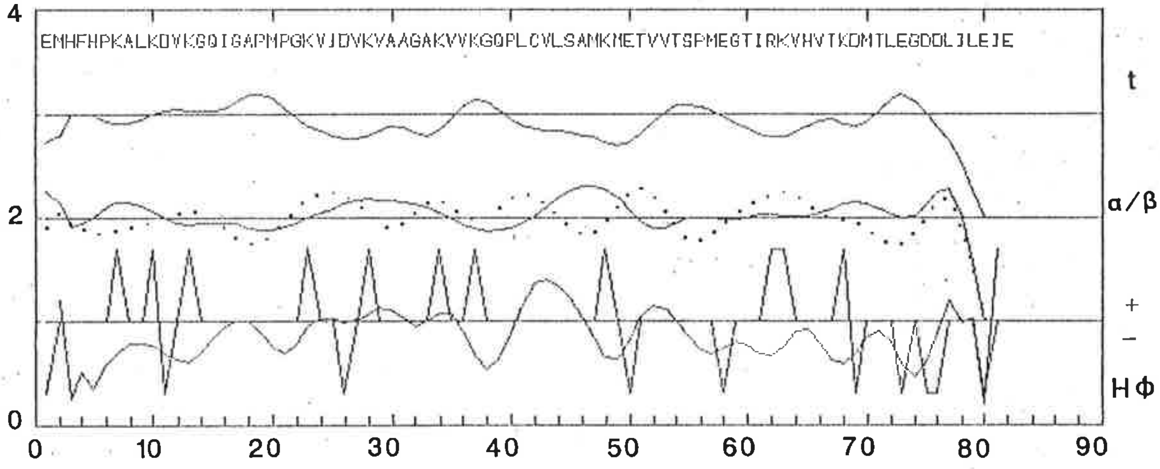
The C-terminal 82 amino acids of the sequence of rat liver pyruvate carboxylase, *P. shermanii* transcarboxylase 1.3S subunit and *E. coli* biotin carrier protein were analysed using the CHOUDOT program (Section 2.2.24) to predict their structure. In each profile from the top to the bottom the following is shown: the amino acid sequence, the reverse-turn propensity (t), the α-helix (solid line) and β-sheet (dotted line) propensities (α/β), the charge profile (+ or −) and hydrophobicity (Hϕ). The amino acids are numbered on the bottom line with number 1 representing the amino acid 82 residues from the C-terminus. The hydropathy plot was produced using the parameters of Kyte and Doolittle (1982). Three iterations of data smoothing were applied and the propensities were smoothed.

a Rat liver pyruvate carboxylase C-terminal 82 amino acids.

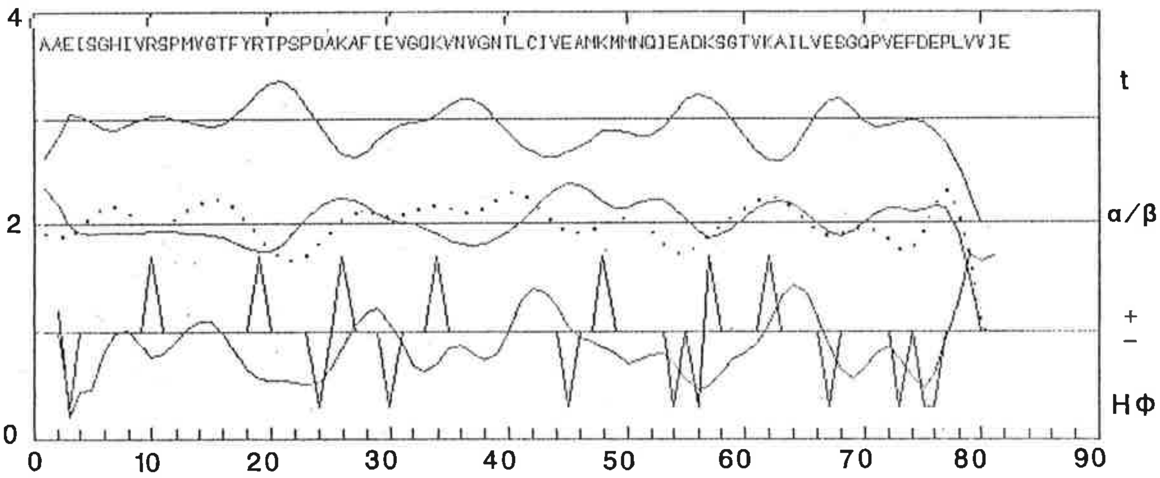
b *E. coli* acetyl-CoA carboxylase biotin carrier subunit C-terminal 82 amino acids (Sutton *et al.*, 1977).

c *P. shermanii* transcarboxylase 1.3S subunit C-terminal 82 amino acids (Maloy *et al.*, 1979).

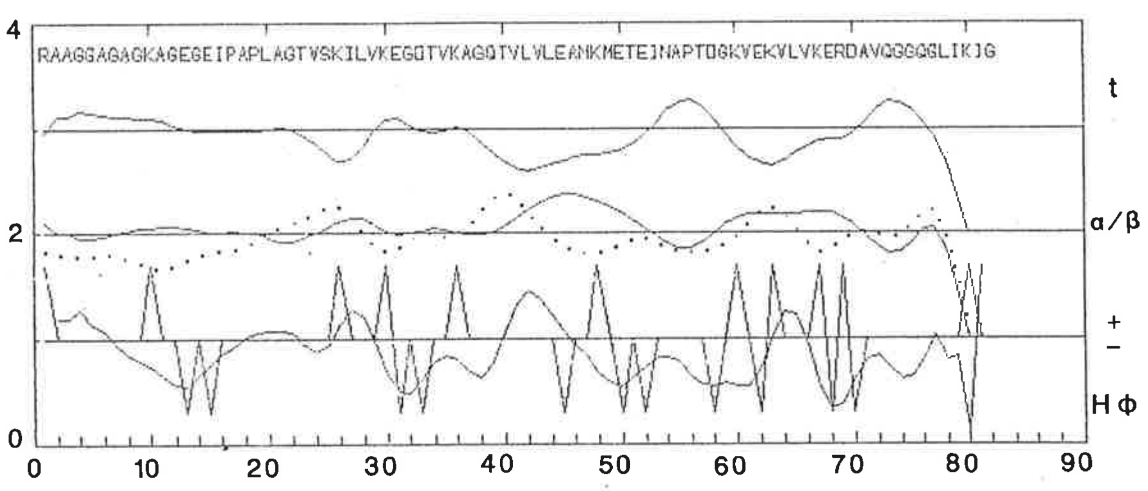
a



b



c



6.4 Discussion

6.4.1 Rat Pyruvate Carboxylase Gene Structure

The 3′ end of the rat pyruvate carboxylase gene contains similarities to the 3′ ends of other eukaryotic genes. The most 3′ exon contains the 3′ untranslated region intact and is quite large as a result. The most 3′ exons of other genes have been found in several cases to be the largest exons in the genes, for example in the ovalbumin (Heilig *et al.*, 1980), tissue-type plasminogen activator (Ny *et al.*, 1984), glyceraldehyde-3-phosphate dehydrogenase (Stone *et al.*, 1985) and type II procollagen (Upholt and Sandell, 1986) genes.

In a survey of the sizes of introns and exons in eukaryotic genes, Naora and Deacon (1982) found that the median size of exons was 140 bp and the median size of introns was 100–200 bp. They also determined that the size of an intron was unrelated to the size of adjacent exons although the total intron size of a gene was related to the total exon size. Therefore according to these criteria the more 5′ exon of the rat pyruvate carboxylase gene is of median size since it contains 141 bp however the 77 bp intron is significantly smaller than the median size. The very small size of this intron was rather unusual because in a study of the rabbit β-globin mRNA, Wieringa *et al.* (1984) determined that an intron size of less than approximately 80 bp resulted in incorrect splicing. Upholt and Sandell (1986) suggested that 78–80 bp was the smallest permissible size for introns which would result in consistently correct splicing. Therefore the rat pyruvate carboxylase gene intron is approximately the smallest size determined to be permissible by these workers and is also one of the smallest introns found in the genes of higher eukaryotes.

6.4.2 Biotin Carboxylase Biotin Domain

The analysis of the rat pyruvate carboxylase gene presented in this report has significant consequences for the current understanding of the biotin carboxylases. The precise matching of the size of the protein encoded by the rat pyruvate carboxylase

3′ exon and the size of the $M_r \simeq 9100$ proteolytic fragment of the *E. coli* acetyl-CoA carboxylase biotin carboxylase carrier protein suggests that this exon defines a functional protein domain. The fact that the region N-terminal to the $M_r \simeq 9100$ fragment is highly susceptible to proteolytic cleavage suggests that this may be an interdomain sequence in the *E. coli* acetyl-CoA carboxylase biotin carboxyl carrier protein. The sequence and structural homology between biotin carboxylases in this region suggests that this exon may represent the modern "relic" of an ancestral gene for a biotin-binding protein which in the vertebrate biotin carboxylases has become fused to other structural elements to produce a multifunctional enzyme subunit. The extent of the homology is such that the probability of this region representing convergent evolution of the biotin carboxylases as discussed by Rylatt *et al.* (1977) appears unlikely. In contrast to this situation the glycolytic enzymes have been found to contain a remarkably similar repeating α-helix/β-sheet protein structural motif which show very little sequence homology and provide a good example of convergent evolution (Fothergill-Gilmore, 1986).

The possibility that the protein encoded by the rat pyruvate carboxylase 3′ exon represents a structural domain is supported by observations of the structure of other genes. The proposal that exons may be correlated with structural regions of proteins (Blake 1978,1979) is now quite widely supported by experimental data from many genes in a variety of organisms. The immunoglobulin protein family shows indisputable correlation between exons and structural domains (Sakano, 1979; Tonegawa,1983). The analyses by Gō (1981) using X-ray crystallographic and protein sequence data show that structural units at a sub-domain level which he terms "modules" correlate strongly with exons in both human haemoglobin (Gō, 1981) and chicken lysozyme (Gō, 1983). Using his system of data analysis Gō (1981) predicted the position of an intron in the haemoglobin gene which he claimed had been lost and an intron was found in precisely the predicted position in the gene of the related protein, leghaemoglobin from soybean (Jensen *et al.*, 1981). Examples of a relationship between exons and structural units have also been observed in several glycolytic enzymes which are perhaps more relevant to pyruvate carboxylase such as chicken glyceraldehyde-3-phosphate dehydrogenase (Stone *et al.*, 1985),

chicken pyruvate kinase (Lonberg and Gilbert, 1985), maize triosephosphate isomerase (Marchionni and Gilbert, 1986) and rabbit phosphofructokinase (Lee *et al.*, 1987).

There is preliminary experimental evidence for the presence of an independently folded biotin-containing domain using partial proteolysis of pyruvate carboxylase to cleave the more exposed residues located between structural units so that the resulting fragments describe domains as described by Kirschner and Bisswanger (1976). Proteolytic digestion of both sheep liver and rat liver pyruvate carboxylase resulted in the removal of a biotin-containing protein fragment of $M_r \simeq 6{,}000$ (G. Booker, personal communication). In addition when proteinase K was used to partially proteolyse yeast pyruvate carboxylase, the enzyme was cleaved 110 residues from the C-terminus. The location of this cleavage point was confirmed by determining the sequence of the N-terminus of the small proteolytic fragment (F. Occhiodoro, personal communication).

The postulate that the biotin-binding region represents an independent structural region is in accordance with the prevailing view in the literature that structural units of either a domain-like size of 100–200 residues or a modular size of 20–40 residues (Gō 1981,1983) are very often encoded by individual exons (for recent reviews, see Doolittle, 1985; Gilbert *et al.*, 1986; Darnell and Doolittle, 1986).

6.4.3 Relationship Between Functional Protein Regions and Gene Structure

The evolution of multifunctional proteins was argued by Gilbert (1978) to be facilitated by the re-arrangement and fusion of the genes for monofunctional proteins. There is now little doubt that exon duplication and shuffling have occurred in the evolution of many proteins. Several examples of this phenomenon have already been described. The chicken calcium-dependent protease appears to have a composite structure with regions homologous to a cathepsin-like protease and a calmodulin-like domain (Ohno *et al.*, 1984). The tissue-type plasminogen activator has regions that are separately homologous to plasminogen, epidermal growth fac-

tor and fibronectin respectively (Ny *et al.*, 1984). However, functional regions are not necessarily confined to a single exon, for example the serine proteinase binding function in ovomucoid is encoded by two exons (Stein *et al.*, 1980) as is the haem binding function of leghaemoglobin (Jensen *et al.*, 1981). In considering the manner in which new functions may have evolved this is not as unreasonable as it appears at first. A new functional site may have evolved at the cleft created by the novel juxtaposition of two structural protein domains.

The biotin-containing subunits of *E. coli* acetyl-CoA carboxylase and *P. shermanii* transcarboxylase are functionally conserved and are small proteins although the transcarboxylase subunit is ~50% larger than the acetyl-CoA carboxylase subunit. Does the transcarboxylase 1.3S biotin carrier subunit represent a fusion of sequences of the two rat pyruvate carboxylase exons, the proposed biotin-binding structural unit with an additional module added to the N-terminus? The level of protein sequence homology between rat pyruvate carboxylase and the 1.3S transcarboxylase subunit decreases rapidly in the N-terminal direction from the biotin-binding region especially in the vicinity of the first intron/exon boundary. This suggests that if introns were originally present in the transcarboxylase 1.3S gene there may have been a removal of selective pressure to conserve the splice site sequence when the introns were lost. Within the 1.3S subunit N-terminal 41 residues there is no significant homology to the protein encoded by the more 5′ exon of the rat pyruvate carboxylase. The N-terminal 26 amino acids of the 1.3S subunit have been demonstrated to bind to both the 5S and 12S subunits and thereby to promote assembly of transcarboxylase (Kumar *et al.*, 1982; Wood and Kumar, 1985). Therefore, if the N-terminal 41 residues were derived from the same region as the more 5′ exon the evolution of a specialized intersubunit binding function may have removed traces of homology.

The homology of rat pyruvate carboxylase to the other biotin carboxylases is high within the region bordered by the 3′ exon but diminishes outside it. In chicken acetyl-CoA carboxylase, which also has a single multifunctional subunit, the lack of homology to the rat pyruvate carboxylase sequence away from the biotin-binding region may be explained by any one of three processes; (i) a rearrangement of

the fused gene elements, (ii) an insertion of alternative gene elements or (iii) a deletion of gene elements. Until further sequence becomes available for rat pyruvate carboxylase no assessment of the relative likelihood of these explanations can be offered.

Functionally significant areas of the biotin carboxylases might be expected to be revealed by sequence homology between the enzymes. Evolutionary pressure for sequence conservation would be expected to be strong in areas where, for example, there are structural constraints induced by binding sites and catalytic centres. Certainly the homology between the biotin site regions of these enzymes is very high, however an examination of the homology of the available rat pyruvate carboxylase sequence to the other biotin enzymes fails to reveal additional areas of significant conservation with the exception of the transcarboxylase 5S subunit sequence. This subunit contains the active sites for the transcarboxylase second partial reaction which transfers a carboxyl group from biotin to pyruvate thus forming oxaloacetate. This is precisely the reaction catalysed by the pyruvate carboxylase second partial reaction site and therefore the homology between the N-terminal region of the rat pyruvate carboxylase sequence and the central portion of the transcarboxylase 5S subunit would appear to position the pyruvate binding site in this region. This is in contrast to the suggestion of Lamhonwah *et al.* (1987) that the pyruvate binding site is located near the N-terminus of human pyruvate carboxylase. The recent finding that the three transcarboxylase subunit genes are linked in the *P. shermanii* genome in the order 12S, 5S and 1.3S is consistent with the proposed location of the pyruvate site adjacent to the biotin-binding site in rat pyruvate carboxylase.

The C-terminal domain in rat pyruvate carboxylase may be derived from a primordial gene encoding a biotin binding protein which early in the evolution of the biotin carboxylases became fused to other elements of the pyruvate carboxylase gene. This primordial gene may have similarly given rise to the gene encoding the 1.3S subunit of transcarboxylase and perhaps the 3′ ends of the other biotin carboxylases as proposed by Obermayer and Lynen (1976). A hypothesis to account for the origin of introns in the coding sections of genes (Place *et al.*, 1986) proposed

that the 3′-most exon is most likely to be complete and to code for an intact domain. Whilst this hypothesis has not been tested, the structure of the 3′ exon of rat pyruvate carboxylase fulfils these criteria and thereby supports it.

Chapter 7

Conclusions

7.1 The Search for Chicken Liver Pyruvate Carboxylase Clones

The primary aims of this investigation have been the identification of recombinant DNA clones encoding pyruvate carboxylase and the characterization of such clones by sequence analysis to provide sequence information about pyruvate carboxylase. The studies which had been carried out in this research group in the past had centered on the characterization of the enzymic activity and the structural features of pyruvate carboxylase from two main sources - chicken liver and sheep liver. The choice therefore tended to be biased towards either of these species because of the wealth of experimental data available for these enzymes. Chicken liver was chosen because the reported levels of pyruvate carboxylase activity in the late embryonic chicken liver were the highest from any vertebrate source.

The study began with an investigation of the embryonic chicken liver as a source of mRNA for use principally in development of cDNA probes for pyruvate carboxylase and secondly as an enriched source of pyruvate carboxylase mRNA which might be suitable for use in constructing a cDNA library. The level of pyruvate carboxylase protein and the rate of synthesis of pyruvate carboxylase was examined during embryonic liver development for the first time and was confirmed to be high during late embryonic development. To detect the pyruvate carboxylase mRNA an oligonucleotide probe was synthesized which was complementary to the inferred biotin attachment site RNA sequence. This oligonucleotide was demonstrated to bind specifically to embryonic chicken liver poly(A)$^+$ RNA and the size class of the RNA to which it was binding was consistent with the size of the mRNA expected at this time to encode pyruvate carboxylase.

The use of the human pyruvate carboxylase cDNA clone as a probe was successful in establishing the size of the chicken liver pyruvate carboxylase mRNA as 4.5 kb which was significantly larger than initially expected. Cross-hybridization of the human pyruvate carboxylase probe to lower molecular weight mRNA bands and the correspondence of these to bands detected by the human propionyl-CoA car-

boxylase α subunit cDNA probe provides the first evidence for homology between chicken biotin carboxylases at the RNA level.

In view of the homology between the human pyruvate carboxylase probe and the chicken pyruvate carboxylase probe it was surprising that no cDNA or genomic clones were detected in any of the libraries which were screened. The features of cDNA and genomic library construction which could influence whether pyruvate carboxylase clones were present provide several possible explanations for the failure to detect pyruvate carboxylase clones. The use of rec^- strains of *E. coli* in future screenings of chicken genomic libraries should obviate any problem with plating of pyruvate carboxylase genomic clones if homopolymeric tracts are present in the pyruvate carboxylase gene. Recently λ packaging extracts have become commercially available that are prepared from *E. coli* $hsdr_k^-$ strains and therefore do not contain *Eco*K activity. The use of cDNA libraries prepared with these extracts would eliminate the possibility of chicken pyruvate carboxylase clones being lost during library construction owing to the presence of *Eco*K sites in the cDNA.

The protein sequence data obtained for chicken liver pyruvate carboxylase provides the largest amount of directly determined sequence information for pyruvate carboxylase from any species. It is hoped that this information will aid the identification of the sequence from chicken pyruvate carboxylase cDNA or genomic clones in the future.

Chicken liver pyruvate carboxylase has catalytic properties which differentiate it from the enzymes already partially or fully sequenced. These include complete dependence on acetyl-CoA for activity. A continuation of the effort to isolate clones encoding chicken pyruvate carboxylase is undoubtedly warranted for the valuable information which could be revealed by comparisons of the sequence data obtained, with the sequence data for pyruvate carboxlases from other species. This would begin to provide a structural basis for the catalytic differences noted between these enzymes.

7.2 The Isolation of Chicken Serum Albumin Clones

The chicken serum albumin clones which were inadvertently isolated in trying to isolate chicken pyruvate carboxylase genomic and cDNA clones provided the first complete pre-proalbumin sequence from this species. While this was obviously not the aim at the time, the sequence information did provide a probable explanation for the selection of these clones. The strategy employed in the isolation of these clones used a specific oligonucleotide to prime the synthesis of extended cDNA probes and had been used successfully in other situations by several workers. In the case of pyruvate carboxylase it appeared to offer a means of circumventing the cross-hybridization which was expected between the oligonucleotide biotin site probe and clones present in cDNA and genomic libraries for other biotin carboxylases. This seemed to be a valid proposition since the pyruvate carboxylase mRNA was present in the embryonic chicken liver in much greater amounts than the mRNAs of other biotin carboxylases. The failure of the strategy was due to the unfortunate and unforeseeable cross-hybridization which seems to have occurred between the oligonucleotide and the highly abundant chicken serum albumin mRNA. The conclusion which must be drawn from this result is that this strategy runs the risk of this type of non-selective priming and therefore should be used cautiously.

7.3 Rat Liver Pyruvate Carboxylase cDNA Clones and Future Experiments

The isolation of cDNA clones encoding rat liver pyruvate carboxylase provides the first isolation of pyruvate carboxylase clones from this species. Since the inferred sequence represents approximately 30% of the total sequence it is the largest amount of amino acid sequence for any mammalian pyruvate carboxylase. The sequence was analysed with respect to its content both in terms of the DNA sequence and inferred protein sequence. Strong sequence homology even beyond the immediate

vicinity of the biotin attachment site was apparent in comparisons with both the DNA and protein sequences of pyruvate carboxylase from other species. However, with respect to yeast pyruvate carboxylase, the homology was very low in some areas suggesting that rearrangements by insertion or deletion of sequence had occurred since the separation of the rat and yeast evolutionary lines. Further analysis of rat pyruvate carboxylase as an example of a mammalian biotin carboxylase will require additional sequence information especially if functionally important regions can be identified by comparison with the complete sequence of yeast pyruvate carboxylase. Therefore an immediate future aim would be the isolation of cDNA clones representing cDNA from more 5′ regions of the mRNA. This would probably necessitate the construction of a new rat liver cDNA library since the library which was screened in this investigation appears unlikely to contain such clones.

7.4 Rat Pyruvate Carboxylase Gene Structure and Future Experiments

The rat pyruvate carboxylase genomic clones described in this thesis represent the first sequence analysis of the gene structure for any vertebrate biotin carboxylase. The coding region of the rat pyruvate carboxylase gene was found to be separated by introns as is characteristic of genes from higher eukaryotes. The position of the intron/exon boundaries was established by sequence analysis and the most 3′ exon revealed important structural information about the biotin carboxylases. The most 3′ exon is very large and the coding region is homologous to the biotinyl domains of other biotin carboxylases. The location of the 5′ boundary of this exon and the sequence homology which extends beyond the vicinity of the biotin attachment site between the biotin enzymes suggests that the evolution of the biotin carboxylases has been divergent in character rather than convergent. It is possible that a biotin-binding domain encoded by an ancestral gene was the common progenitor of this region in at least several and possibly all biotin carboxylases.

The homology apparent between the *P. shermanii* transcarboxylase 5S subunit

and the N-terminal end of the rat pyruvate carboxylase inferred amino acid sequence localizes the probable region of the pyruvate binding site in both pyruvate carboxylase and transcarboxylase. It also suggests that this region in pyruvate carboxylase also arose as the result of divergence from a common ancestor and perhaps most importantly provides the first experimental evidence for the Lynen hypothesis of biotin carboxylase evolution through gene fusion (Lynen, 1975; Obermayer and Lynen, 1976).

This study provides the starting point for the characterization of the complete intron/exon structure of the rat pyruvate carboxylase gene. The λRG1.2 clone has been demonstrated to contain further coding sequences in the *Eco*RI/*Hin*dIII fragment immediately 5′ of the 1.4 kb *Bam*HI/*Hin*dIII fragment which was sequenced. The exonic regions in the *Eco*RI/*Hin*dIII fragment could be readily mapped and sequenced. Additional clones representing the more 5′ regions of the gene could be obtained by re-screening the rat genomic library using a restriction fragment from the 5′ end of λRG1.2 as a probe. In this way by "walking" along the gene, clones encoding the complete gene could be obtained. Of particular interest would be the location of the intron/exon boundaries for those exons encoding either important functional regions or regions which are homologous to other biotin carboxylases.

An extension of the observed structural homology between the independent biotinyl subunits of transcarboxylase and *E. coli* acetyl-CoA carboxylase would be provided by subcloning the DNA representing the 3′ exon into an expression vector and expressing the encoded protein in *E. coli*. The transcarboxylase 1.3S biotinyl subunit has already been expressed in *E. coli* and was biotinylated by the *E. coli* holocarboxylase synthetase. If the region encoded by the 3′ exon represents an ancestral biotinyl domain as has been postulated here, then it could be predicted that it also would be biotinylated and could be at least partly functionally active if used to replace the biotinyl subunits of transcarboxylase and acetyl-CoA carboxylase.

The structure of the gene control regions at the 5′ end of the gene would also be of interest since the control of pyruvate carboxylase synthesis has been demonstrated to lie at the level of transcription (Angus *et al.*, 1981). Pyruvate carboxylase activity has been shown to be regulated during cell differentiation in mouse 3T3 fibroblasts

(Angus *et al.*, 1981) and during development of the liver in the rat (Chang, 1977) as well as other animals (for a review, see Barritt, 1985). The level of rat liver pyruvate carboxylase protein has also been demonstrated to be increased by an increase in either glucagon or glucocorticoids (Seitz *et al.*, 1976) and in experimentally induced diabetes (Weinberg and Utter, 1980). For any study of the transcriptional control of the rat pyruvate carboxylase gene to proceed knowledge of the DNA sequence at the 5′ region of the gene will be essential.

The isolation and characterization of rat pyruvate carboxylase cDNA and genomic clones in this study provides the groundwork for the investigations described above to proceed.

Bibliography

Alberts, A.W. and Vagelos, P.R. (1972) Acyl-CoA carboxylases. *The Enzymes* **6**, 37–82.

Aldred, A.R., Howlett, G.J. and Schreiber, G. (1984) Synthesis of rat transferrin in *Escherchia coli* containing a recombinant bacteriophage. *Biochem. Biophys. Res. Comm.* **122**, 960–965.

Ambrose, M.C. and Perham, R.N. (1976) Spin-label study of the mobility of enzyme-bound lipoic acid in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli. Biochem. J.* **155**, 429–432.

Angus, C.W., Lane, M.D., Rosenfeld, P.J. and Kelly, T.J. (1981) Increase in translatable mRNA for mitochondrial pyruvate carboxylase during differentiation of 3T3 preadipocytes. *Biochem. Biophys. Res. Comm.* **103**, 1216–1222.

Arinze, J.C. and Mistry, S.P. (1970) Hepatic acetyl CoA carboxylase, propionyl CoA carboxylase and pyruvate carboxylase activities during embryonic development and growth in chickens. *Proc. Soc. Exp. Biol. Med.* **135**, 553–556.

Ashman, L.K., Keech, D.B., Wallace, J.C. and Nielsen, J. (1972) Sheep kidney pyruvate carboxylase : Studies on its activation by acetyl coenzyme A and characteristics of its acetyl coenzyme A independent activity. *J. Biol. Chem.* **247**, 5818–5824.

Ashman, L.K., Wallace, J.C. and Keech, D.B. (1973) Desensitisation of pyruvate carboxylase against acetyl-CoA stimulation by chemical modification. *Biochem. Biophys. Res. Comm.* **51**, 924–931.

Attwood, P.V., Mayer, F. and Wallace, J.C. (1986) Avidin as a probe of the conformational changes induced in pyruvate carboxylase by acetyl-CoA and pyruvate. *FEBS Lett.* **203**, 191–196.

Aviv, H. and Leder, P. (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* **69**, 1408–1412.

Bai, D.H., Pape, M.E., Lopez-Casillas, F., Luo, X.C., Dixon, J.E. and Kim, K-H. (1986) Molecular cloning of cDNA for acetyl-coenzyme A carboxylase. *J. Biol. Chem.* **261**, 12395–12399.

Bais, R. and Keech, D.B. (1972) The magnesium ion (Mg^{2+}) activation of sheep kidney pyruvate carboxylase. *J. Biol. Chem.* **247**, 3255–3261.

Ballard, F.J. and Oliver, I.T. (1963) Glycogen metabolism in embryonic chicken and neonatal rat liver. *Biochim. Biophys. Acta* **71**, 578–588.

Barden, R.E., Fung, C-H., Utter, M.F. and Scrutton, M.C. (1972) Pyruvate carboxylase from chicken liver; steady state kinetic studies indicate a “two-site” ping-pong mechanism. *J. Biol. Chem.* **247**, 1323–1333.

Barden, R.E., Taylor, B.L., Isohashi, F., Frey, W.H., Zander, G.L., Lee, J.C. and Utter, M.F. (1975) Structural properties of pyruvate carboxylase from chicken liver and other sources. *Proc. Natl. Acad. Sci. USA.* **72**, 4308–4312.

Barritt, G.J. (1985). Regulation of enzymatic activity. In Keech, D.B. and Wallace, J.C., editors, *Pyruvate Carboxylase.* pages 141–177, CRC Series in Enzyme Biology, CRC Press, Boca Raton, USA.

Barritt, G.J., Zander, G.L. and Utter, M.F. (1976). The regulation of pyruvate carboxylase activity in gluconeogenic tissues. In Hanson, R.W. and Mehlman, M.A., editors, *Gluconeogenesis, Its Regulation in Mammalian Species.* pages 3–46, John Wiley and Sons, New York, USA.

Beaty, N.B. and Lane, M.D. (1982) Acetyl coenzyme A. Rapid purification of the chick liver enzyme and steady state kinetic analysis of the carboxylase-catalysed reaction. *J. Biol. Chem.* **257**, 924–929.

Benton, W.D. and Davis, R.W. (1977) Screening λgt recombinant clones by hybridization to single plaques *in situ*. *Science* **196**, 180–182.

Berget, S.M., Moore, C. and Sharp, P.A. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA.* **74**, 3171–3175.

Bickle, T.A. (1982). The ATP-dependent restriction enzymes. In Linn, S.M. and Roberts, R.J., editors, *Nucleases.* pages 85–108, Cold Spring Harbor Laboratory, New York, USA.

Blake, C.C.F. (1978) Do genes-in-pieces imply proteins in pieces? *Nature* **273**, 267.

Blake, C.C.F. (1979) Exons encode protein functional units. *Nature* **277**, 598.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. and Boyer, H.W. (1977) Construction and characterization of new cloning vehicles.II. A multipurpose cloning system. *Gene* **2**, 95–113.

Bonner, W.M. and Laskey, R.A. (1974) A film detection method for tritium-labelled proteins and nucleic acids in polyacrylamide gels. *Eur. J. Biochem.* **46**, 83–88.

Breathnach, R., Mandel, J.L. and Chambon, P. (1977) Ovalbumin gene is split in chicken DNA. *Nature* **270**, 314–319.

Bridgeland, E.S. and Jones, K.M. (1967) Formation of dicarboxylic acids and phosphoenol pyruvate in *Arthrobacter globiformis*. *Biochem. J.* **104**, 9P–10P.

Brown, J.R. (1975) Structure of bovine serum albumin. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **34**, 591.

Brown, J.R. (1976) Structural origins of mammalian albumin *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **35**, 2141–2144.

Brown, D.D. and Suzuki, Y. (1974) The purification of the messenger RNA for silk fibroin. *Methods Enzymol.* **30**, 648–654.

Casadaban, M.J. and Cohen, S.N. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli. J. Mol. Biol.* **138**, 179–207.

Cathala, G., Savouret, J-F., Mendez, B., West, B.L., Karin, M., Martial, J.A. and Baxter, J.D. (1983) A method for isolation of intact, translationally active ribonucleic acid. *DNA* **2**, 329–335.

Cazzulo, J.J. and Stoppani, A.O.M. (1967) Purification and properties of pyruvate carboxylase from bakers yeast. *Arch. Biochem. Biophys.* **121**, 596–608.

Cazzulo, J.J., Sundaram, T.K. and Kornberg, H.L. (1970) Properties and regulation of pyruvate carboxylase from *Bacillus stearothermophilus. Proc. R. Soc. London Ser. B* **176**, 1–19.

Chaconas, G. and Van de Sande, J.H. (1980) 5′-^{32}P Labeling of RNA and DNA restriction fragments. *Methods Enzymol.* **65**, 75–85.

Chan, S.J., Noyes, B.E., Agarwal, K.L. and Steiner, D.F. (1979) Construction and selection of recombinant plasmids containing full-length complementary DNAs corresponding to rat insulins I and II. *Proc. Natl. Acad. Sci. USA.* **76**, 5036–5040.

Chang, L.O. (1977) The development of pyruvate carboxylase in rat liver mitochondria. *Pediatr. Res.* **11**, 6–8.

Chapman-Smith, A. *Studies on the Interaction of Sheep Liver Pyruvate Carboxylase with the Allosteric Effector, Acetyl-CoA.* PhD thesis,University of Adelaide, Adelaide, South Australia (1981).

Chirgwin, J.M., Przybyla, A.E., Macdonald, R.J. and Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299.

Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.T. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8.

Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Non-chromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R factor DNA. *Proc. Natl. Acad. Sci. USA.* **69**, 2110–2114.

Cohen, N.D., Duc, J.A., Beegen, H. and Utter, M.F. (1979) Quaternary structure of pyruvate carboxylase from *Pseudomonas citronellolis. J. Biol. Chem.* **254**, 9262–9269.

Cohen, N.D., Utter, M.F., Wrigley, N.G. and Barrett, A.N. (1979a) Quaternary structure of yeast pyruvate carboxylase: biochemical and electron microscope studies. *Biochemistry* **18**, 2197–2203.

Cohen, N.D., Beegan, H., Utter, M.F. and Wrigley, N.G. (1979b) A re-examination of the electron microscopic appearance of pyruvate carboxylase from chicken liver. *J. Biol. Chem.* **254**, 1740–1747.

Collins, J. and Hohn, B. (1978) Cosmids: A type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. *Proc. Natl. Acad. Sci. USA.* **75**, 4242–4246.

Compere, S.J., Lively, M.O. and MacGillivray, R.T.A. (1981) Amino-terminal sequence of chicken preproalbumin. *Eur. J. Biochem.* **116**, 437–440.

Cornish-Bowden, A. (1980) The random character of protein evolution and its effect on the reliability of phylogenetic information deduced from amino acid sequences and compositions. *Biochem. J.* **191**, 349–354.

Dale, R.M.K., McClure, B.A. and Houchins, J.P. (1985) A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: Application to sequencing the corn mitochondrial 18S rDNA. *Plasmid* **13**, 31–40.

Dalton, S., Coleman, J.R. and Wells, J.R. (1986) Transcription of the histone H5 gene is not S-phase regulated. *Mol. Cell Biology* **6**, 601–606.

Darnell, J.E. and Doolittle, W.F. (1986) Speculations on the early course of evolution. *Proc. Natl. Acad. Sci. USA.* **83**, 1271–1275.

Das, H.K., Biro, P.A., Cohen, S.N., Erlich, H.A.and Von Gabain, A., Lawrance, S.K., Lemaux, P.G., McDevitt, H.O., Peterlin, B.N., Schull, M-F., Sood, A.K. and Weissman, S.M. (1983) Use of synthetic oligonucleotide probes complementary to genes for human HLA-DR α and β as extension primers for the isolation of 5′-specific genomic clones. *Proc. Natl. Acad. Sci. USA.* **80**, 1531–1535.

Dayhoff, M.O. (1976). *Atlas of Protein Sequence and Structure.* National Biomedical Research Foundation Washington, D.C., USA, 5th edition.

Deeley, R.G., Gordon, J.I., Burns, A.T.H., Mullinix, K.P., Binastein, M. and Goldberger, R.F. (1977) Primary activation of the vitellogenin gene in rooster. *J. Biol. Chem.* **252**, 8310–8319.

De Vivo, D.C., Haymond, M.P., Leckie, Y.L., Bursmann, D.B., McDougae, D.B. and Pagliara, A.S. (1977) The clinical and biochemical implications of pyruvate carboxylase deficiency. *J. Clin. Endocrinol. Metab.* **45**, 1281–1296.

Dodgson, J.B., Strommer, J. and Engel, J.D. (1979) Isolation of the chicken β-globin gene and a linked embryonic β-like globin gene from a chicken DNA recombinant library. *Cell* **17**, 879–887.

Doolittle, R.F. (1985) The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* **10**, 233–237.

Dugaiczyk, A., Lawn, S.W. and Dennison, O.E. (1982) Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. *Proc. Natl. Acad. Sci. USA.* **79**, 71–75.

Duggleby, R.G., Kinns, H. and Rood, J.I. (1981) A computer program for determining the size of DNA restriction fragments. *Anal. Biochem.* **110**, 49–55.

Dworkin, M.D. and Dawid, I.B. (1980) Use of a cloned library for the study of abundant polyA$^+$ RNA during *Xenopus laevis* development. *Dev. Biol.* **76**, 449–464.

Easterbrook-Smith, S.B., Hudson, P.J., Goss, N.H., Keech, D.B. and Wallace, J.C. (1976) Pyruvate carboxylase: Mechanisms of the second partial reaction. *Arch. Biochem. Biophys.* **176**, 709–720.

Efstratiadis, A. and Villa-Komaroff, L. (1979). Cloning of double-stranded cDNA. In Setlow, J.K. and A., Hollaender, editors, *Genetic Engineering. Principles and Methods.* pages 15–36, Plenum Press, New York, USA, Volume 1.

Fall, R.R., Glaser, M. and Vagelos, P.R. (1976) Acetyl coenzyme A carboxylase: Circular dichroism studies of *Escherichia coli* biotin carboxyl carrier protein. *J. Biol. Chem.* **251**, 2063–2069.

Feir, H.A. and Suzuki, I. (1969) Pyruvate carboxylase of *Aspergillus niger*: kinetic study of a biotin-containing enzyme. *Can. J. Biochem.* **47**, 697–710.

Feldhoff, D.J. and Ledden, D.J. (1983) Evidence for the spontaneous formation of interspecies hybrid molecules of human, rat and bovine serum albumins. *Biochem. Biophys. Res. Comm.* **114**, 20–27.

Felicioli, R.A., Gabrielli, F. and Rossi, C.A. (1967) The synthesis of phosphoenolpyruvate in the gluconeogenesis. Enzyme levels of chick embryo livers. *Eur. J. Biochem.* **3**, 19–24.

Flick, P.K., Chen, J., Alberts, A.W. and Vagelos, P.R. (1978) Translation of rat liver fatty acid synthetase mRNA in a cell-free system derived from wheat germ.

Proc. Natl. Acad. Sci. USA. **75**, 730–734.

Fothergill-Gilmore, L.A. (1986). Domains of glycolytic enzymes. In Hardie, D.G. and Coggins, J.R., editors, *Multidomain Proteins - Structure and Evolution.* pages 85–174, Elsevier, Amsterdam.

Fowler, A.V. and Zabin, I. (1978) Amino acid sequence of β-galactosidase. XI. Peptide ordering procedures and the complete sequence. *J. Biol. Chem.* **253**, 5521–5525.

Frey, W.H. and Utter, M.F. (1977) Binding of acetyl CoA to chicken liver pyruvate carboxylase. *J. Biol. Chem.* **252**, 51–56.

Freytag, S.O. and Collier, K.J. (1984) Molecular cloning of a cDNA for human pyruvate carboxylase. Structural relationship to other biotin-containing carboxylases and regulation of mRNA content in differentiating preadipocytes. *J. Biol. Chem.* **259**, 12831–12837.

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97–120.

Gilbert, W. (1978) Why genes in pieces? *Nature* **271**, 501.

Gilbert, W., Marchionni, M. and McKnight, G. (1986) On the antiquity of introns. *Cell* **46**, 151–154.

Glover, D.M. and Hogness, D.S. (1977) A novel arrangement of the 18S and 28S sequences in a repeating unit of Drosophila melanogaster rDNA. *Cell* **10**, 167–176.

Gō, M. (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature* **291**, 90–92.

Gō, M. (1983) Modular structural units, exons and function in chicken lysozyme. *Proc. Natl. Acad. Sci. USA.* **80**, 1964–1968.

Goodall, G.J. *Biotin Carboxylases.* PhD thesis,University of Adelaide, Adelaide, South Australia (1981).

Goodall, G.J., Johannssen, W., Wallace, J.C. and Keech, D.B. (1985). Sheep liver propionyl-CoA carboxylase: Purification and some molecular properties. In Dakshinamurti, K. and Bhagavan, H.N., editors, *Biotin.* pages 396–397, Annals of the New York Academy of Sciences, New York, USA, Volume 447.

Gordon, J.I., Burns, A.T.H., Christmann, J.L. and Deeley, R.G. (1978) Cloning of a double-stranded cDNA that codes for a portion of chicken preproalbumin: A general method for isolating a specific DNA sequence from partially purified mRNA. *J. Biol. Chem.* **253**, 8629–8639.

Goss, N.H., Dyer, P.Y., Keech, D.B. and Wallace, J.C. (1979) An electron microscopic study of pyruvate carboxylase. *J. Biol. Chem.* **254**, 1734–1739.

Green, N.M. (1963) Avidin. The use of [^{14}C] biotin for kinetic studies and for assay. *Biochem. J.* **89**, 585–591.

Grunstein, M. and Hogness, D.S. (1975) Colony hybridization: A method for the isolation of cloned DNA that contain a specific gene. *Proc. Natl. Acad. Sci. USA.* **72**, 3961–3965.

Guchhait, R.B., Polakis, S.E., Dimroth, P., Stoll, E., Moss, J. and Lane, M.D. (1974) Acetyl coenzyme A carboxylase system of *Escherichia coli*: Purification and properties of the biotin carboxylase, carboxyltransferase and carboxyl carrier protein components. *J. Biol. Chem.* **249**, 6633–6645.

Haarasilta, S. and Taskinen, L. (1977) Location of three key enzymes of gluconeogenesis in baker's yeast. *Arch. Microbiol.* **113**, 159–161.

Haase, F.C., Henrikson, K.P., Treble, D.H. and Allen, S.H.G. (1982) The subunit structure and function of the propionyl coenzyme A carboxylase of *Mycobacterium smegmatis. J. Biol. Chem.* **257**, 11994–11999.

Haché, R.J.G., Wiskocil, R., Vasa, M., Roy, R.N., Lau, P.C.K. and Deeley, R.G. (1983) The 5′ non-coding and flanking regions of the avian very low density apolipoprotein II and serum albumin genes: Homologies with the egg white protein genes. *J. Biol. Chem.* **258**, 4556–4564.

Harvey, R.P., Robins, A.J. and Wells, J.R.E. (1982) Independently evolving chicken histone H2B genes: Identification of a ubiquitous H2B-specific 5′ element. *Nucleic Acids Res.* **10**, 7851–7863.

Heilig, R., Perrin, F., Gannon, F., Mandel, J.L. and Chambon, P. (1980) The ovalbumin gene family: Structures of the X gene and evolution of duplicated split genes. *Cell* **20**, 625–637.

Hendrick, D. and Moller, F. (1973) Pyruvate carboxylase in the developing chick embryo. *Comp. Biochem. Physiol.* **45B**, 197–221.

Higgins, C.F., Hiles, I.D., Salmond, G.P.C., Gill, D.R., Downie, J.A., Evans, I.J., Holland, I.B., Gray, L., Buckel, S.D., Bell, A.W. and Hermondson, M.A. (1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* **323**, 448–450.

Hobbs, A.A., Mattschoss, L.A., May, B.K., Williams, K.E. and Elliott, W.H. (1986) The cDNA and protein sequence of a phenobarbital-induced chicken cytochrome P-450. *J. Biol. Chem.* **261**, 9444–9449.

Hoffman, N.E., Pichersky, E. and Cashmore, A.R. (1987) A tomato cDNA encoding a biotin-binding protein. *Nucleic Acids Res.* **15**, 3928.

Hollemans, M., Runswick, M.J., Fearnley, I.M. and Walker, J.E. (1983) The sites of labeling of the β-subunit of bovine mitochondrial F_1-ATPase with 8-azido-ATP. *J. Biol. Chem.* **258**, 9307–9313.

Houghton, M., Stewart, A.G., Doel, S.M., Emtage, J.S., Eaton, M.A.W., Smith, J.C., Patel, T.P., Lewis, H.M., Porter, A.G., Birch, J.R., Cartwright, T. and

Carey, N.H. (1980) The amino-terminal sequence of human fibroblast interferon as deduced from reverse transcripts obtained using synthetic oligonucleotide primers. *Nucleic Acids Res.* **8**, 1913–1931.

Hudson, P.J., Keech, D.B. and Wallace, J.C. (1975) Pyruvate carboxylase : Affinity labelling of the pyruvate binding site. *Biochem. Biophys. Res. Comm.* **65**, 213–219.

Hudson, P., Haley, J., Cronk, M., Shine, J. and Niall, H. (1981) Molecular cloning and characterization of cDNA sequences coding for rat relaxin. *Nature* **291**, 127–131.

Hunkapillar, M.W., Hewick, R.M., Dreyer, W.J. and Hood, L.E. (1983) High-sensitivity sequencing with a gas-phase sequenator. *Methods Enzymol.* **91**, 399–413.

Huynh, T.V., Young, R.A. and Davis, R.W. (1985). Constructing and screening cDNA libraries in λgt10 and λgt11. In Glover, D.M., editor, *DNA Cloning.* pages 49–78, IRL Press, Oxford, England.

Ish-Horowicz, D. and Burke, J.F. (1981) Rapid and efficient cosmid vector cloning. *Nucleic Acids Res.* **9**, 2989–2998.

Jekel, P.A., Weijer, W.J. and Beintema, J.J. (1983) Use of endoproteinase lys-C from *Lysobacter enzymogenes* in protein sequence analysis. *Anal. Biochem.* **134**, 347–354.

Jensen, E.Ø., Paludan, K., Hyldig-Nielsen, J.J., Jørgensen, P. and Marcker, K.A. (1981) The structure of a chromosomal leghaemoglobin gene from soybean. *Nature* **291**, 677–679.

Johannssen, W., Attwood, P.V., Wallace, J.C. and Keech, D.B. (1983) Localisation of the active site of pyruvate carboxylase by electron microscopic examination of avidin-enzyme complexes. *Eur. J. Biochem.* **133**, 201–206.

Kalousek, F., Darigo, M.D. and Rosenberg, L.E. (1980) Isolation and characterization of propionyl-CoA carboxylase from normal human liver: Evidence for a protomeric tetramer of nonidentical subunits. *J. Biol. Chem.* **255**, 60–65.

Kanehisa, M. *Ideas: Integrated Database and Extended Analysis System for Nucleic Acids and Proteins.* Users Manual - National Institutes of Health, Bethesda, USA (1985).

Kaziro, Y., Ochoa, S., Warner, R.C. and Chen, J-Y. (1961) Metabolism of propionic acid in animal tissues. VIII. Crystalline propionyl carboxylase. *J. Biol. Chem.* **236**, 1917–1923.

Keech, D.B. (1980) Pyruvate carboxylase and acetyl-CoA. *Proc. Aust. Biochem. Soc.* **13**, P1–P2.

Keech, D.B. and Attwood, P.V. (1985). The reaction mechanism. In Keech, D.B. and Wallace, J.C., editors, *Pyruvate Carboxylase.* pages 109–140, CRC Series in Enzyme Biology, CRC Press, Boca Raton, USA.

Keech, D.B. and Barritt, G.J. (1967) Allosteric activation of sheep kidney pyruvate carboxylase by magnesium ions (Mg^{2+}) and the magnesium adenosine triphosphate ion ($MgATP^{2-}$). *J. Biol. Chem.* **242**, 1983–1987.

Keech, D.B. and Utter, M.F. (1963) Pyruvate carboxylase: Properties. *J. Biol. Chem.* **238**, 2609–2614.

Keil, B. (1977). Enzymatic degradation of proteins yielding large fragments. In Previero, A. and M.A., Coletti-Previero, editors, *Solid Phase Methods in Protein Sequence Analysis* pages 287–292, INSERM Symposium No. 5, Elsevier/North Holland Biomedical Press, Amsterdam.

Kessler, S.W. (1981) Use of protein A-bearing staphylococci for the immunoprecipitation and isolation of antigens from cells. *Methods Enzymol.* **73B**, 442–459.

Khew-Goodall, Y-S. *Pyruvate Carboxylase: Its Interactions with Acetyl CoA.* PhD thesis,University of Adelaide, Adelaide, South Australia (1985).

Kirschner, K. and Bisswanger, H. (1976) Multifunctional proteins. *Annu. Rev. Biochem.* **45**, 143–166.

Kobayashi, I. and Ikeda, H. (1977) Formation of recombinant DNA of bacteriophage λ by *recA* function of *Escherichia coli* without duplication, translation and maturation. *Mol. Gen. Genet.* **153**, 237–245.

Kornberg, H.L. (1966). Anaplerotic sequences and their role in metabolism. In Campbell, P.N. and Greville, G.D., editors, *Essays in Biochemistry,* pages 1–31, Academic Press, London.

Koshland, D.E., Némethy, G. and Filmer, D. (1966) Comparison of experimental binding data and theoretical modes in proteins containing subunits. *Biochemistry* **5**, 365–385.

Kosow, D.P., Huang, S.G. and Lane, M.D. (1962) Propionyl holocarboxylase synthesis. I. Preparation and properties of the enzyme system. *J. Biol. Chem.* **237**, 3633–3639.

Kraus, J.P. and Rosenberg, L.E. (1982) Purification of low-abundance messenger RNAs from rat liver by polysome immunoadsorption. *Proc. Natl. Acad. Sci. USA.* **79**, 4015–4019.

Kraus, J.P., Williamson, C.L., Firgaira, F.A., Yang-Feng, T.L., Münke, M., Franke, U. and Rosenberg, L.E. (1986) Cloning and screening with nanogram amounts of immunopurified mRNAs: cDNA cloning and chromosomal mapping of cystathione β-synthase and the β subunit of propionyl-CoA carboxylase. *Proc. Natl. Acad. Sci. USA.* **83**, 2047–2051.

Krieg, P.A., Robins, A.J., Colman, A. and Wells, J.R.E. (1982) Chicken histone H5 mRNA: the polyadenylated RNA lacks the conserved histone 5$'$ terminator sequence. *Nucleic Acids Res.* **10**, 6777–6785.

Kumar, G.K., Bahler, C.R., Wood, H.G. and Merrifield, R.B. (1982) The amino acid sequences of the biotinyl subunit essential for the association of transcarboxylase. *J. Biol. Chem.* **257**, 13828–13834.

Kyte, K. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.

Lacy, E. and Maniatis, T. (1980) The nucleotide sequence of a rabbit β-globin pseudogene. *Cell* **21**, 545–553.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Lamhonwah, A-M., Barankiewicz, T.J., Willard, H.F., Mahuran, D.J., Quan, F. and Gravel, R.A. (1986) Isolation of cDNA clones coding for the α and β chains of human propionyl-CoA carboxylase: Chromosomal assignments and DNA polymorphisms associated with *PCCA* and *PCCB* genes. *Proc. Natl. Acad. Sci. USA.* **83**, 4864–4868.

Lamhonwah, A-M., Quan, F. and Gravel, R.A. (1987) Sequence homology around the biotin-binding site of human propionyl-CoA carboxylase and human pyruvate carboxylase. *Arch. Biochem. Biophys.* **254**, 631–636.

Land, H., Grez, M., Hauser, H., Lindmayer, W. and Schütz, G. (1981) 5′-Terminal sequences of eukaryotic mRNA can be cloned with high efficiency. *Nucleic Acids Res.* **9**, 2251–2266.

Lane, M.D., Moss, J. and Polakis, S.E. (1974) Acetyl coenzyme A carboxylase. *Curr. Top. Cell Regul.* **8**, 139–195.

Lau, E.P., Cochran, B.C., Munson, L. and Fall, R.R. (1979) Bovine kidney 3-methylcrotonyl-CoA and propionyl-CoA carboxylases: Each enzyme contains nonidentical subunits. *Proc. Natl. Acad. Sci. USA.* **76**, 214–218.

Lawn, R.M., Adelman, J., Bock, S.C., Franke, A.E., Houck, C.M., Najarian, R.C., Seeburg, P.H. and Wion, K.L. (1981) The sequence of human serum albumin cDNA and its expression in *E. coli. Nucleic Acids Res.* **9**, 6103–6114.

Ledden, D.J. and Feldhoff, R.C. (1982) Characterization of fragments of human albumin purified by Cibacron blue F3GA affinity chromatography. *Biochem. J.*

205, 331–337.

Ledden, D.J. and Feldhoff, R.C. (1983) Purification and characterization of peptic fragments derived from the carboxyl-terminal half of human albumin. *J. Prot. Chem.* **2**, 303–319.

Lee, C-P., Kao, M-C., French, B.A., Putney, S.D. and Chang, S.H. (1987) The rabbit muscle phosphofructokinase gene : Implications for protein structure, function and tissue specificity. *J. Biol. Chem.* **262**, 4195–4199.

Lonberg, N. and Gilbert, W. (1985) Intron/exon structure of the chicken pyruvate kinase gene. *Cell* **40**, 81–90.

Lynen, F. (1975). Structures of biotin carboxylases. In Richter, D., editor, *Energy regulation and biosynthesis in molecular biology.* pages 671–698, W. de Gruyter, Berlin, West Gremany.

Lynen, F. (1979) New experiments of biotin enzymes. *Crit. Rev. Biochem.* **7**, 103–119.

Lynen, F., Knappe, J., Jütting, G. and Ringelmann, E. (1959) Die biochemische Funktion des Biotins. *Angew. Chem.* **71**, 481–486.

Mackall, J.C. and Lane, M.D. (1977) Changes in mammary-gland acetyl-coenzyme A carboxylase associated with lactogenic differentiation. *Biochem. J.* **162**, 635–642.

Maloy, W.L., Bowien, B.U., Zwolinski, G.K., Kumar, K.G., Wood, H.G., Ericsson, L.H. and Walsh, K.A. (1979) Amino acid sequence of the biotinyl subunit from transcarboxylase. *J. Biol. Chem.* **254**, 11615–11622.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) The isolation of structural genes from libraries of eucaryotic DNA. *Cell* **15**, 687–701.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular Cloning. (A Laboratory Manual).* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

Marchionni, M. and Gilbert, W. (1986) The triosephosphate isomerase gene from Maize: Introns antedate the plant-animal divergence. *Cell* **46**, 133–141.

Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res.* **14,Supplement**, r151–r197.

Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* **74**, 560–564.

Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560.

Mayer, F., Wallace, J.C. and Keech, D.B. (1980) Further electron microscopic studies on pyruvate carboxylase. *Eur. J. Biochem.* **112**, 265–272.

McAllister, H.C. and Coon, M.J. (1966) Further studies on the properties of liver propionyl coenzyme A holocarboxylase synthetase and the specificity of holocarboxylase formation. *J. Biol. Chem.* **241**, 2855–2861.

McClure, W.R. and Lardy, H.A.and Cleland, W.W. (1971) Rat liver pyruvate carboxylase III. Isotopic exchange studies of the first partial reaction. *J. Biol. Chem.* **246**, 3584–3590.

McDonell, M.W., Simon, M.N. and Studier, F.W. (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J. Mol. Biol.* **110**, 119–146.

McMaster, G.K. and Carmichael, G.C. (1977) Analysis of single-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA.* **74**, 4835–4838.

Messing, J. (1979). A multipurpose cloning system based on the single-stranded DNA bacteriophage M13. In *Recombinant DNA Technical Bulletin,* pages 43–48, NIH Publications, Bethesda, USA, Volume 2.

Messing, J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**, 269–276.

Messing, J., Crea, R. and Seeburg, P.H. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**, 309–321.

Mevarech, M., Noyes, B.E. and Agarwal, K.L. (1979) Detection of gastrin-specific mRNA using oligodeoxynucleotide probes of defined sequence. *J. Biol. Chem.* **254**, 7472–7475.

Meyer, H., Nevaldine, B. and Meyer, F. (1978) Acyl-coenzyme A carboxylase of the free-living nematode *Turbatrix aceti.* 1. Its isolation and molecular characteristics. *Biochemistry* **17**, 1822–1827.

Monod, J., Wyman, J. and Changeaux, J-P. (1965) On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118.

Montgomery, D.L., Hall, B.D., Gillam, S. and Smith, M. (1978) Identification and isolation of the yeast cytochrome c gene. *Cell* **14**, 673–680.

Morris, S.M., Winberry, L.K., Fisch, J.E., Back, D.W. and Goodridge, A.G. (1984) Development and nutritional regulation of the messenger RNAs for fatty acid synthase, malic enzyme and albumin in the livers of embryonic and newly-hatched chicks. *Mol. Cell. Biochem.* **64**, 63–68.

Morris, C.P., Lim, F. and Wallace, J.C. (1987) Yeast pyruvate carboxylase: Gene isolation. *Biochem. Biophys. Res. Comm.* **145**, 390–396.

Moss, J. and Lane, M.D. (1971) The biotin-dependent enzymes. *Adv. Enzymol.* **35**, 321–442.

Mottershead, D.G., Oliver, K.J., Brazier, J.A. and Wallace, J.C. (1984) Immunological cross-reaction of some biotin carboxylases. *Proc. Aust. Biochem. Soc.* **16**, 69.

Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459–472.

Murray, N.E., Brammar, W.J. and Murray, K. (1977) Lambdoid phages that simplify the recovery of *in vitro* recombinants. *Mol. Gen. Genet.* **150**, 53–61.

Murtif, V.L. and Samols, D. (1987) Mutagenesis affecting the carboxyl terminus of the biotinyl subunit of transcarboxylase: Effects on biotination. *J. Biol. Chem.* **262**, 11813–11816.

Murtif, V.L., Bahler, C.R. and Samols, D. (1985) Cloning and expression of the 1.3s biotin-containing subunit of transcarboxylase. *Proc. Natl. Acad. Sci. USA.* **82**, 5617–5621.

Nader, W.F., Isenberg, G. and Sauer, H.W. (1986) Structure of *Physarum* actin gene locus *ardA*: a nonpalindromic sequence causes inviability of phage lambda and *recA*-independent deletions. *Gene* **48**, 133–144.

Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsödi, J., Boll, W., Cantell, K. and Weissmann, C. (1980) Synthesis in *E. coli* of a polypeptide with human leukocyte interferon activity. *Nature* **284**, 316–320.

Naora, H. and Deacon, N.J. (1982) Relationship between the total size of exons and introns in protein-coding genes of higher eukaryotes. *Proc. Natl. Acad. Sci. USA.* **79**, 6196–6200.

Noll, H. (1967) Characterization of macromolecules by constant velocity centrifugation. *Nature* **215**, 360–363.

Novotný, J. and Auffray, C. (1984) A program for the prediction of protein secondary structure from nucleotide sequence data : application for histocompatibility antigens. *Nucleic Acids Res.* **12**, 243–255.

Noyes, B.E., Mevarech, M., Stein, R. and Agarwal, K.L. (1979) Detection and partial sequence analysis of gastrin mRNA by using an oligodeoxynucleotide probe. *Proc. Natl. Acad. Sci. USA.* **76**, 1770–1774.

Ny, T., Elgh, F. and Lund, B. (1984) The structure of the human tissue-type plasminogen activator gene: Correlation of the intron and exon structures to functional and structural domains. *Proc. Natl. Acad. Sci. USA.* **81**, 5355–5359.

Obermayer, M. and Lynen, F. (1976) Structure of biotin enzymes. *Trends Biochem. Sci.* **1**, 169–171.

Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. and Suzuki, K. (1984) Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein? *Nature* **312**, 566–570.

Oizumi, J., Shaw, K.N.F., Giudici, T.A., Carter, M., Donnell, G.N. and Ng, W.G. (1983) Pyruvate carboxylase deficiency with renal tubular acidosis and cystinuria. *J. Inherited Metab. Dis.* **6**, 89–93.

Okayama, H. and Berg, P. (1982) High-efficiency cloning of full-length cDNA. *Mol. Cell Biology* **2**, 161–170.

Oliver, K.J., Rylatt, D.B., Bunderson, P. and Wallace, J.C. (1984) Study of the common structural determinants of the biotin carboxylases using monoclonal antibodies. *Proc. Aust. Biochem. Soc.* **16**, 29.

Osmani, S.A. and Scrutton, M.C. (1983) The sub-cellular localisation of pyruvate carboxylase and of some other enzymes in *Aspergillus nidulans*. *Eur. J. Biochem.* **133**, 551–560.

Osmani, S.A., Mayer, F., Marston, F.A.O., Selmes, I.P. and Scrutton, M.C. (1984) Pyruvate carboxylase from *Aspergillus nidulans* : Effects of regulatory modifiers on the structure of the enzyme. *Eur. J. Biochem.* **139**, 509–518.

Ou, W., Ito, A., Morohashi, K., Fujii-Kuriyama, Y. and Omura, T. (1986) Processing-independent *in vitro* translation of cytochrome P-450(SCC) precursor

across mitochondrial membranes. *J. Biochem.* **100**, 1287–1296.

Parnes, J.R., Velan, B., Felsenfeld, A., Ramanathan, L., Ferrini, U., Appella, E. and Seidman, J.G. (1981) Mouse β_2-microglobulin cDNA clones: A screening procedure for cDNA clones corresponding to rare mRNAs. *Proc. Natl. Acad. Sci. USA.* **78**, 2253–2257.

Pelham, H.R.B. and Jackson, R.J. (1976) An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* **67**, 247–256.

Peters, Jr., T. (1970) Serum Albumin. *Adv. Clin. Chem.* **13**, 37–111.

Peters, Jr., T. (1985) Serum Albumin. *Adv. Prot. Chem.* **37**, 161–245.

Place, A.R., Anderson, S.M. and Sofer, W. (1986). Introns and domain-coding regions in the dehydrogenase genes. In Hardie, D.G. and Coggins, J.R., editors, *Multidomain Proteins - Structure and Evolution.* pages 175–194, Elsevier, Amsterdam.

Proudfoot, N.J. and Brownlee, G.G. (1976) 3′ Non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211–214.

Proudfoot, N.J. and Maniatis, T. (1980) The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication. *Cell* **21**, 537–544.

Reed, R.G., Feldhoff, R.C., Clute, O.L. and Peters, Jr., T. (1975) Fragments of bovine serum albumin produced by limited proteolysis. Confirmation and ligand binding. *Biochemistry* **14**, 4578–4583.

Richards, R.I. and Wells, J.R.E. (1980) Chicken globin genes: Nucleotide sequence of cDNA clones coding for the α-globin expressed during hemolytic anemia. *J. Biol. Chem.* **255**, 9306–9311.

Richards, R.I., Shine, J., Ullrich, A., Wells, J.R.E. and Goodman, H.M. (1979) Molecular cloning and sequence analysis of adult chicken β globin cDNA. *Nucleic Acids Res.* **7**, 1137–1146.

Rigby, P.W.J., Dieckmann, C., Rhodes, C. and Berg, P. (1977) Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**, 237–251.

Robinson, B.H., Oei, J., Sherwood, W.G., Applegarth, D., Wong, L., Haworth, J., Goodyer, P., Casey, R. and Zaleski, L.A. (1984) The molecular basis for the two different clinical presentations of classical pyruvate carboxylase deficiency. *Am. J. Hum. Genet.* **36**, 283–294.

Rohde, M., Lim, F. and Wallace, J.C. (1986) Pyruvate carboxylase from *Saccharomyces cerevisiae.* Quaternary structure, effects of allosteric ligands and binding of avidin. *Eur. J. Biochem.* **156**, 15–22.

Rosen, A.M. and Geller, D.M. (1977) Chicken microsomal albumin: Amino terminal sequence of chicken proalbumin. *Biochem. Biophys. Res. Comm.* **78**, 1060–1066.

Rosen, J.M., Liarakos, C.D. and O'Malley, B.W. (1973) Effect of estrogen on gene expression in the chick oviduct. I. Deoxyribonucleic acid-deoxyribonucleic acid renaturation studies. *Biochemistry* **12**, 2803–2809.

Rosenberg, S.M. (1985) Eco K restriction during *in vitro* packaging of coliphage lambda DNA. *Gene* **39**, 313–315.

Rosenberg, S.M., Stahl, M.M., Kobayashi, I. and Stahl, F.W. (1985) Improved in vitro packaging of coliphage lambda DNA: A one-strain system free from endogenous phage. *Gene* **38**, 165–175.

Rubin, C.M. and Schmid, C.W. (1980) Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucleic Acids Res.* **8**, 4613–4619.

Rylatt, D.B., Keech, D.B. and Wallace, J.C. (1977) Pyruvate carboxylase: Isolation of the biotin-containing tryptic peptide and the determination of its primary structure. *Arch. Biochem. Biophys.* **183**, 113–122.

Sakano, H., Rogers, J.H., Hüppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S. (1979) Domains and the hinge region of an imminoglobulin heavy chain are encoded in separate DNA segments. *Nature* **277**, 627–633.

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

Sargent, T.D., Yang, M. and Bonner, J. (1981) Nucleotide sequence of cloned rat serum albumin messenger RNA. *Proc. Natl. Acad. Sci. USA.* **78**, 243–246.

Schiele, U., Niedermeier, R., Stürzer, M. and Lynen, F. (1975) Investigations of the structure of the 3-methylcrotonyl-CoA carboxylase from *Achromobacter*. *Eur. J. Biochem.* **60**, 259–266.

Schmid, B., Wachter, E., Sebald, W. and Neupert, W. (1984) Processing peptidase of *Neurospora* mitochondria. Two-step cleavage of imported ATPase subunit 9. *Eur. J. Biochem.* **144**, 581–588.

Schwinghamer, M.W. and Symons, R.H. (1977) Translation of the four major RNA species of cucumber mosaic virus in plant and animal cell-free systems and in toad oocytes. *Virology* **79**, 88–108.

Scrutton, M.C. (1974) Pyruvate carboxylase. Studies of activator-independent catalysis and of the specificity of activation by acyl derivatives of coenzyme A from rat liver. *J. Biol. Chem.* **249**, 7057–7067.

Scrutton, M.C. and Utter, M.F. (1965) Pyruvate carboxylase: Interaction of the enzyme with adenosine triphosphate. *J. Biol. Chem.* **240**, 3714–3723.

Scrutton, M.C. and White, M.D. (1973) Pyruvate carboxylase: Specific inactivation of acetyl coenzyme A–dependent oxaloacetate synthesis during modification of the enzyme by trinitrobenzene sulphonate. *J. Biol. Chem.* **248**, 5541–5544.

Scrutton, M.C., Keech, D.B. and Utter, M.F. (1965) Pyruvate carboxylase. IV. partial reactions and the locus of activation of acetyl CoA. *J. Biol. Chem.* **240**, 574–581.

Scrutton, M.C., Pearce, P.H. and Fatebene, F. (1977) Pyruvate carboxylase: Effect of reaction components and anologues of acetyl–coenzyme A on the rate of inactivation in the presence or absence of trinitrobenzene sulphonate. *Eur. J. Biochem.* **76**, 219–231.

Seitz, H.J., Kaiser, M., Krone, W. and Tarnowski, W. (1976) Physiologic significance of glucocorticoids and insulin in the regulation of hepatic gluconeogenesis during starvation in rats. *Metabolism* **25**, 1545–1555.

Seubert, W. and Remberger, U. (1961) Reinigung und Wirkungsweise der Pyruvatcarboxylase aus *Pseudomonas citronellosis. Biochem. Z.* **334**, 401–414.

Sinsheimer, R.L. (1977) Recombinant DNA. *Annu. Rev. Biochem.* **46**, 415–438.

Smithies, O., Blechl, A.E, Dennison-Thompson, K., Newell, N., Richards, J.E., Slightom, J.L., Tucker, B.W. and Blattner, F.R. (1978) Cloning human γ-globin and mouse α-type globin DNA: Characterization and partial sequencing. *Science* **202**, 1284–1288.

Spiro, R.G. (1973) Glycosylation. *Adv. Prot. Chem.* **27**, 349–467.

Srivastava, G., Borthwick, I.A., Brooker, J.D., Wallace, J.C., May, B.K. and Elliott, W.H. (1983) Hemin inhibits transfer of pre-δ-aminolevulinate synthase into chick embryo liver mitochondria. *Biochem. Biophys. Res. Comm.* **117**, 344–349.

Staden, R. (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10**, 2951–2961.

Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. *Cell* **21**, 681–687.

Stone, E.M., Rothblum, K.N. and Schwartz, R.J. (1985) Intron-dependent evolution of chicken glyceraldehyde phosphate dehydrogenase gene. *Nature* **313**, 498–500.

Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H. and Itakura, K. (1981) Use of synthetic oligonucleotides as hybridization probes: Isolation of cloned cDNA sequences for human β_2-microglobulin. *Proc. Natl. Acad. Sci. USA.* **78**, 6613–6617.

Sutton, M.R., Fall, R.R., Nervi, A.M., Alberts, A.W., Vagelos, P.R. and Bradshaw, R.A. (1977) Amino acid sequence of *Escherichia coli* biotin carboxyl carrier protein. *J. Biol. Chem.* **252**, 3934–3940.

Szostak, J.W., Stiles, J.I., Tye, B-K., Chiu, P., Sherman, F. and Wu, R. (1979) Hybridization with synthetic oligonucleotides. *Methods Enzymol.* **68**, 419–428.

Takai, T., Wada, K. and Tanabe, T. (1987) Primary structure of the biotin-binding site of chicken liver acetyl-CoA carboxylase. *FEBS Lett.* **212**, 98–102.

Taylor, B.L, Frey, W.H., Barden, R.E., Scrutton, M.C. and Utter, M.F. (1978) The use of the ultracentrifuge to determine the catalytically competent forms of enzymes with more than one oligomeric structure. *J. Biol. Chem.* **253**, 3062–3069.

Thomas, P.S. (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA.* **77**, 5201–5205.

Thornton, C., Samols, D., Kumar, G.K., Haase, C. and Wood, H.G. (1987) Sequence and expression of the *Propionibacterium shermanii* transcarboxylase operon. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **46**, 2144.

Tonegawa, S. (1983) Somatic generation of antibody diversity. *Nature* **302**, 575–581.

Tonegawa, S., Maxam, A.M., Tizard, R., Bernhard, O. and Gilbert, W. (1978) Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. USA.* **75**, 1485–1489.

Upholt, W.B. and Sandell, L.J. (1986) Exon/intron organization of the chicken type II procollagen gene: Intron size distribution suggests a minimal intron size. *Proc. Natl. Acad. Sci. USA.* **83**, 2325–2329.

Utter, M.F. and Keech, D.B. (1960) Formation of oxaloacetate from pyruvate and CO_2. *J. Biol. Chem.* **235**, PC17–PC18.

Utter, M.F. and Keech, D.B. (1963) Pyruvate carboxylase. I. Nature of the reaction. *J. Biol. Chem.* **238**, 2603–2608.

Utter, M.F., Barden, R.E. and Taylor, B.L. (1975) Pyruvate carboxylase: an evaluation of the relationship between structure and mechanism and between structure and catalytic activity. *Adv. Enzymol.* **42**, 1–72.

Valentine, R.C. (1968). Shapes and symmetries of biological molecules. In *Proc. 4th Reg. Conf. (Europe) Electron Microscopy.* pages 3, Tipografia Poliglotta Vaticana, Rome.

Valentine, R.C., Wrigley, N.G., Scrutton, M.C., Irias, J.J. and Utter, M.F. (1966) Pyruvate carboxylase. VIII. The subunit structure as examined by electron microscopy. *Biochemistry* **5**, 3111–3116.

Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. (1978) A bacterial clone synthesizing proinsulin. *Proc. Natl. Acad. Sci. USA.* **75**, 3727–3731.

Wahl, G.M., Strern, M. and Stark, G.R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulphate. *Proc. Natl. Acad. Sci. USA.* **76**, 3683–3687.

Wallace, J.C. (1985). Distribution and biological functions of pyruvate carboxylase in nature. In Keech, D.B. and Wallace, J.C., editors, *Pyruvate Carboxylase.* pages 5–64, CRC Series in Enzyme Biology, CRC Press, Boca Raton, USA.

Wallace, J.C. and Easterbrook-Smith, S.B. (1985). The structure of pyruvate carboxylase. In Keech, D.B. and Wallace, J.C., editors, *Pyruvate Carboxylase.* pages 65–108, CRC Series in Enzyme Biology, CRC Press, Boca Raton, USA.

Wallace, R.B., Johnson, M.J., Suggs, S.V., Miyoshi, K., Bhatt, R. and Itakura, K. (1981) A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing

in the plasmid vector pBR322. *Gene* **16**, 21–26.

Weeks, D.P., Beerman, N. and Griffith, O.M. (1986) A small-scale five-hour procedure for isolating multiple samples of CsCl-purified DNA: Applications to isolations from mammalian, insect, higher plant, algal, yeast and bacterial sources. *Anal. Biochem.* **152**, 376–385.

Weinberg, M.B. and Utter, M.F. (1980) Effect of streptomycin-induced diabetes mellitus on the turnoverof rat liver pyruvate carboxylase and pyruvate dehydrogenase. *Biochem. J.* **188**, 601–608.

Whittle, A. *Propionyl-CoA Carboxylase - A Structural Analysis.* Honours thesis, University of Adelaide Adelaide, South Australia (1986).

Wieringa, B., Hofer, E. and Weissmann, C. (1984) A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. *Cell* **37**, 915–925.

Wilbur, W.J. and Lipman, D.J. (1983) Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA.* **80**, 726–730.

Wood, H.G. and Barden, R.E. (1977) Biotin enzymes. *Annu. Rev. Biochem.* **46**, 385–413.

Wood, H.G. and Kumar, G.K. (1985). Transcarboxylase: Its quaternary structure and the role of the biotinyl subunit in the assembly of the enzyme and in catalysis. In Dakshinamurti, K. and Bhagavan, H.N., editors, *Biotin.* pages 1–22, Annals of the New York Academy of Sciences, New York, USA, Volume 447.

Wood, H.G. and Zwolinski, G.K. (1976) Transcarboxylase: Role of biotin, metals and subunits in the reaction and its quaternary structure. *Crit. Rev. Biochem.* **4**, 47–122.

Wyman, A.R., Wolfe, L.B. and D., Borstein. (1985) Propagation of some human DNA sequences in bacteriophage λ vectors requires mutant *Escherichia coli* hosts. *Proc. Natl. Acad. Sci. USA.* **82**, 2880–2884.

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

Appendix A

The amino acid sequence inferred from the DNA sequence of the *P. shermanii* transcarboxylase 5S subunit gene is shown below. This data was kindly supplied by Dr. D. Samols (Case Western Reserve University, Cleveland, Ohio, USA). The amino acids are numbered with the N-terminal Met taken as number 1.

met ser pro arg glu ile glu val ser glu pro arg glu val gly ile thr glu leu val leu arg asp ala his gln ser leu met ala thr arg met ala met glu asp met val gly
10 20 30 40

ala cys ala asp ile asp ala ala gly tyr trp ser val glu cys trp gly gly ala thr tyr asp ser cys ile arg phe leu asn glu asp pro trp glu arg leu arg thr phe arg
50 60 70 80

lys leu met pro asn ser arg leu gln met leu leu arg gly gln asn leu leu gly tyr arg his tyr asn asp glu val val asp arg phe val asp lys ser ala glu asn gly met
90 100 110 120

asp val phe arg val phe asp ala met asn asp pro arg asn met ala his ala met ala ala val lys lys ala gly lys his ala gln gly thr ile cys tyr thr ile ser pro val
130 140 150 160

his thr val glu gly tyr val lys leu ala gly gln leu leu asp met gly ala asp ser ile ala leu lys asp met ala ala leu leu lys pro gln pro ala tyr asp ile ile lys
170 180 190 200

ala ile lys asp ile arg pro glu asp ala asp gln pro ala leu his ser thr thr gly val thr glu val ser leu met lys ala ile glu ala gly val asp val val asp thr ala
210 220 230 240

ile ser ser met ser leu gly pro gly his asn pro thr glu ser val ala glu met leu glu gly thr gly tyr thr thr asn leu asp tyr asp arg leu his lys ile arg asp his
250 260 270 280

phe lys ala ile arg pro lys tyr lys lys phe glu ser lys thr leu val asp thr ser ile phe lys ser gln ile pro gly gly met leu ser asn met glu ser glu leu arg ala
290 300 310 320

gln gly ala glu asp lys met asp glu val met ala glu val pro arg val arg arg pro ala pro val phe pro ala pro gly his pro val gln pro asp arg arg his ala gly leu
330 340 350 360

phe asn val met met gly glu tyr lys arg met thr gly glu phe ala asp ile met leu gly tyr tyr gly ala thr pro ala asp arg asp pro lys trp ser val gly glu glu his
370 380 390 400

arg arg ala ile thr gln arg pro ala asp his asp pro lys val val lys leu ala glu glu gln ser gly lys lys pro ile thr gln arg pro ala asp leu leu pro pro glu trp
410 420 430 440

glu glu gln ser lys glu pro arg pro lys gly phe asn gly thr asp glu asp val leu thr tyr ala leu phe pro gln val ala pro val phe phe glu ser arg pro arg ala ala
450 460 470 480

glu arg gly ser his arg cys pro ala glu gly arg gly glu gly asp glu lys cys arg arg gly arg ser arg his leu gln arg glu arg gly gly thr val arg lys ser pro phe
490 500 510 520

ser arg arg glu asp asp cys gln ser trp leu lys thr thr ile
530

Appendix B

The DNA sequence and the inferred amino acid sequence of the *P. shermanii* transcarboxylase 12S subunit gene is shown below. This data was kindly supplied by Dr. D. Samols (Case Western Reserve University, Cleveland, Ohio, USA). The bases are numbered with the 5′ base of the Met codon taken as number 1.

```
met ala glu asn asn asn leu lys leu ala ser thr met glu gly arg val glu gln leu ala glu gln arg gln val ile glu ala gly gly gly glu arg leu val glu lys gln his
ATGGCTGAAAACAACAATTTGAAGCTCGCCAGCACCATGGAAGGTCGCGTGGAGCAGCTCGCAGAGCAGCGCCAGGTGATCGAAGCCGGTGGCGGCGAACGTCTCGTCGAGAAGCAACAT 120

ser gln gly lys gln thr ala arg glu arg leu asn asn leu leu asp pro his ser phe asp glu val gly ala phe arg asn his arg thr thr leu phe gly met asp lys ala val
TCCCAGGGTAAGCAGACCGCTCGTGAGCGCCTGAACAACCTGCTCGATCCCCATTCGTTCGACGAGGTCGGCGCTTTCCGCAACCACCGCACCACGTTGTTCGGCATGGACAAGGCCGTC 240

val pro ala asp gly val val thr gly arg gly thr ile leu gly arg pro val his ala ala ser gln asp phe thr val met gly gly ser ala trp arg asp ala val his glu gly
GTCCCGGCAGATGGCGTGGTCACCGGCCGTGGCACCATCCTTGGTCGTCCCGTGCACGCCGCGTCCCAGGACTTCACGGTCATGGGTGGTTCGGCTTGGCGAGACGCAGTCCACGAAGGT 360

arg arg asp asp gly thr ala leu leu thr gly thr pro phe leu phe phe tyr asp ser gly gly arg ile gln glu gly ile asp ser leu ser gly tyr gly lys met phe phe ala
CGTCGAGACGATGGAACAGCGCTGCTCACCGGCACGCCCTTCCTGTTCTTCTACGATTCGGGCGGCCGGATCCAGGAGGGCATCGACTCGCTGAGCGGTTACGGCAAGATGTTCTTCGCC 480

asn val lys leu ser gly val val pro gln ile ala ile ile ala gly pro cys ala cys ala ser tyr ser pro ala leu thr asp phe ile ile met thr lys lys ala his met phe
AACGTGAAGCTGTCGGGCGTCGTGCCGCAGATCGCCATCATTGCCGGCCCCTGTGCGTGCGCCTCGTATTCGCCGGCACTGACTGACTTCATCATCATGACCAAGAAGGCCCATATGTTC 600

ile thr gly pro gln val ile lys ser val thr gly glu asp val thr ala asp glu leu gly gly ala glu pro ile trp pro ser arg ala ile tyr phe val ala glu asp asp asp
ATCACGGGCCCCCAGGTCATCAAGTCGGTCACCGGCGAGGATGTCACCGCTGACGAACTCGGTGGCGCTGAGCCCATATGGCCATCTCGGGCAATATACTTCGTGGCCGAGGACGACGAC 720

ala ala glu leu ile ala lys lys leu leu ser phe leu pro gln asn asn thr glu glu ala ser phe val asn pro asn asn asp val ser pro asn thr glu leu arg asp leu val
GCCGCGGAGCTCATTGCCAAGAAGCTGCTGAGCTTCCTTCCGCAGAACAACACTGAGGAAGCATCCTTCGTGAACCCGAACAATGACGTCAGCCCCAATACCGAGCTGCGCGACCTCGTT 840

pro ile asp gly lys lys gly tyr asp val arg asp val ile ala lys ile val asp trp gly asp tyr leu glu val lys ala gly tyr ala thr asn leu val thr ala phe ala arg
CCGATTGACGGCAAGAAGGGCTATGACGTGCGCGATGTCATTGCCAAGATCGTCGACTGGGGTGACTACCTCGAGGTCAAGGCCGGCTATGCCACCAACCTCGTGACCGCCTTCGCCCGG 960

val asn gly arg ser val gly ile val ala asn gln pro ser val met ser gly cys leu asp ile asn ala ser asp lys ala ala glu phe val asn phe cys asp ser phe asn ile
GTCAATGGTCGTTCGGTGGGCATCGTGGCCAATCAGCCTTCGGTGATGTCGGGTTGCCTCGACATCAACGCCTCTGACAAGGCCGCCGAATTCGTGAATTTCTGCGATTCGTTCAACATC 1080

pro leu val gln leu val asp val pro gly phe leu pro val gln gln glu tyr gly gly ile ile arg his gly arg lys met leu tyr ala tyr ser glu ala thr val pro lys ile
CCGCTGGTGCAGCTGGTCGACGTGCCGGGCTTCCTGCCCGTGCAGCAGGAGTACGGCGGCATCATTCGCCATGGGCGCAAGATGCTGTACGCCTACTCCGAGGCCACCGTGCCGAAGATC 1200

thr cys leu ala thr pro thr ala ala pro thr trp pro cys ala thr val thr leu val pro thr pro cys thr pro val pro ser ala glu ile ala val met gly ala glu gly ala
ACGTGTCTCGCAACGCCTACGGCGGCTCCTACCTGGCCATGTGCAACCGTGACCTTGGTGCCGACGCCGTGTACGCCTGTGCCCAGCGCCGAGATTGCGGTGATGGGCGCCGAGGGTGCG 1320

ala asn val ile phe arg lys glu ile lys ala ala asp asp pro asp ala met arg ala glu lys ile glu glu tyr gln asn gly ser thr arg arg thr trp arg ala arg gly gln
GCAAATGTGATCTTCCGCAAGGAGATCAAGGCTGCCGACGATCCCGACGCCATGCGCGCCGAGAAGATCGAGGAGTACCAGAACGGTTCAACACGCCGTACGTGGCGCGCCCGCGGTCAG 1440

val asp asp val leu asp pro ala asp thr arg arg lys ile ala ser ala leu glu met tyr ala thr lys arg gln thr arg pro ala lys lys pro trp lys leu pro leu leu ser
GTCGACGACGTGCTTGACCCGGCTGATACCCGTCGAAAGATTGCTTCCGCCCTGGAGATGTACGCCACCAAGCGTCAGACCCGCCCGGCGAAGAAGCCATGGAAACTTCCCCTGCTGAGC 1560

glu glu glu ile met ala asp glu glu glu lys asp leu met ile ala thr leu asn lys arg val ala ser leu glu ser glu leu gly ser leu gln ser asp thr gln gly val thr
GAGGAGGAAATTATGGCTGATGAGGAAGAGAAGGACCTGATGATCGCCACGCTCAACAAGCGCGTCGCGTCATTGGAGTCTGAGTTGGGTTCACTCCAGAGCGATACCCAGGGTGTCACC 1680

glu asp val leu thr ala ile ser ala val ala ala tyr leu gly asn asp gly ser ala glu val val his phe ala pro ser pro asn trp val arg glu gly arg arg ala leu gln
GAGGACGTACTGACGGCCATTTCGGCCGTTGCGGCCTATCTCGGCAACGATGGATCGGCTGAGGTCGTCCATTTCGCCCCGAGCCCGAACTGGGTCCGCGAGGGTCGTCGGGCTCTGCAG 1800

asn his ser ile arg *
AACCATTCCATTCGTTGA 1810
```

Appendix C

Publications

1. Cassady, A.I., Wells, J.R.E. and Wallace, J.C. (1983) Isolation and analysis of chicken genomal pyruvate carboxylase clones. *Proc. Aust. Biochem. Soc.* **15**, 65. (Abstract)

2. Cassady, A.I., Morris, C.P. and Wallace, J.C. (1987) Sequencing and characterization of clones encoding rat liver pyruvate carboxylase. *Proc. Aust. Biochem. Soc.* **19**, 20. (Abstract)

3. Cassady, A.I., Morris, C.P., Lim, F. and Wallace, J.C. (1987) Computer-assisted structural analysis of the biotin carboxylases. *Proc. Aust. Biochem. Soc.* **19**, 26. (Abstract)

4. Cassady, A.I., Morris, C.P. and Wallace, J.C. Isolation and characterization of rat liver pyruvate carboxylase cDNA and genomic clones: Relationship between the 3′ exon and the biotin carboxyl carrier domain. Submitted to *Proc. Natl. Acad. Sci. USA* (1987)

5. Cassady, A.I., Salkilld, K. and Wallace, J.C. The complete sequence of chicken pre-proalbumin. Manuscript in preparation.