CHARACTERISATION OF THE PYRUVATE CARBOXYLASE GENE AND STUDIES ON THE REGULATION OF ITS EXPRESSION IN RAT

by

Sarawut Jitrapakdee M.Sc. (Biochemistry)

A thesis submitted to the University of Adelaide

for the degree of Doctor of Philosophy

Department of Biochemistry

University of Adelaide

Adelaide, South Australia

March, 1999

TABLE OF CONTENTS

SUMMARY

Pyruvate carboxylase (PC) [EC 6.4.1.1.] is a multi-functional enzyme belonging to the biotin carboxylase family, enzymes that use the biotin moiety as a mobile carboxy-group carrier between the two catalytic domains. Mammalian PC is located in the mitochondrial matrix where it plays a vital role in gluconeogenesis, lipogenesis and other biosynthetic pathways. Although cDNA or genomic sequences encoding PC from a number of organisms have been reported they have not improved our understanding of the molecular mechanisms that control the PC expression. This study aimed to unravel the molecular basis of what controls PC expression at the gene level.

Full length rat PC cDNA has been isolated by performing 5' RACE-PCR. Full length cDNA contains an open reading frame of 3537 bp encoding a polypeptide with a calculated M_r of 129,848. The inferred protein sequence shows a high degree of sequence similarity to PC from other species. Three functional domains of rat PC i.e. the biotin carboxylation domain, the transcarboxylation domain and the biotinyl domain have been determined by a combination of limited proteolysis, sequence comparisons with PC from other organisms and with other biotin enzymes. Comparison with the known structure of the biotin carboxylation subunit of *Escherichia coli* acetyl-CoA carboxylase, suggested the functional importance of 11 highly conserved residues including Cys265, His271, Lys273, Glu311, Glu324, Asn326, Arg328, Gln330, Val331, Glu332 and Arg377.

The rat PC gene spans over 40 kb comprising nineteen coding exons and four 5'-untranslated region exons. Organisation of exons in the gene is consistent with the domain structure of the enzyme, suggesting a close relationship between exons and protein domains. Fluorescence *in situ* hybridisation of rat metaphase chromosomes localised PC to the long arm of chromosome 1, band 1q43. Five alternate forms of rat PC transcripts encoding the same enzyme have been identified and they exhibit tissue-specific expression. These mRNA isoforms contain the same coding sequence but differ in their 5'-untranslated regions (UTRs). Direct comparison of the nucleotide sequences of 5'-UTR exons to 5'-UTR sequences of PC mRNA isoforms clearly showed that alternative splicing of 5'-UTR exons of primary transcripts alternatively transcribed from two promoters, generated the multiple transcripts with 5'-end heterogeneity. Class I mRNAs (rUTR A, rUTR B and

rUTRC) transcribed from the proximal promoter are expressed in gluconeogenic tissues (liver and kidney) and lipogenic tissues (liver, adipose tissue and lactating mammary gland). In contrast, class II mRNAs (rUTR D and rUTR E) transcribed from the distal promoter are constitutively expressed in most tissues.

The proximal promoter lacks TATA and CAAT boxes but includes a sequence that is typical of an initiator HIP-1 box surrounding the transcription initiation site while the distal promoter contains three copies of CAAT boxes and multiple Sp1 binding sites. Several potential transcription factor binding sites were also identified in both promoters. A series of 5'-nested deletion constructs of both promoters were fused to a firefly luciferase reporter gene and transiently expressed in COS-1 cells. The results showed that the 153- and 187-bp, preceding the transcription initiation sites of the proximal and the distal promoters, respectively were required for basal transcription. The distal promoter was found to drive the expression of the luciferase reporter gene at a level 4-8-fold higher than the proximal promoter when transiently expressed in COS-1, CHO-K and HepG2 cell lines. Insulin was found to selectively inhibit the expression of the proximal promoter-luciferase reporter gene by 50% but not the distal promoter, suggesting the presence of an insulin-responsive element. A half maximum effect was found at 1 nM insulin.

Regulation of PC expression is developmentally regulated. The abundance of PC mRNAs was low in fetal liver but increased by 2-4-fold within 7 days after birth concomitant with an 8-fold increase in the amount of immunoreactive PC and its activity, and then decreased during the weaning period. Reverse-transcriptase polymerase chain reaction (RT-PCR) analysis indicated that the proximal promoter was activated during the suckling period and reduced in activity at weaning.

In genetically obese Zucker rats, adipose PC was 4-5 fold increased concomitant with a 5-6 fold increase in mRNA level, whereas a 2-3 fold increase in hepatic PC was observed in parallel with a 4 fold increase in its mRNA. RT-PCR analysis also showed that the proximal promoter was activated in the hyperlipogenic condition.

Conversely, transcription of the proximal promoter was not detectable in various liver cell lines suggesting that this promoter was not functional under cell culture conditions. In rat pancreatic islets and insulinoma cells only rUTR D and rUTR E transcripts, generated

from the distal promoter of the PC gene, were expressed. Glucose increased PC transcripts from the distal promoter when the insulinoma cells were maintained in 10 mM glucose.

From these data it can be concluded that the proximal promoter of the rat PC gene plays a major role in gluconeogenesis and lipogenesis while the distal promoter is necessary for anaplerosis.

In vitro translation and *in vivo* polysome profile analysis indicated that rUTR C and rUTR E transcripts were translated with similar translational efficiencies which are substantially greater than that of the rUTR D transcript, suggesting that 5'-UTRs play a role in translational control.

Full length human PC cDNA clones were assembled and cloned upstream of the puromycin resistance gene driven by the very strong promoter of human polypeptide chain elongation factor-1α, in pEFIRES-Puro plasmid. The resulting construct, pEF-PC was stably transfected into human kidney cells (293T). Because puromycin exerts selective pressure on the whole expression cassette, a high dose of puromycin was added to the culture medium to select only highly expressing clones. Stably pooled clones expressed a very high level of human PC, up to 4-6% of soluble protein or 20-30 times more than non-transfected cells. After the selective precipitation of the mitochondrial extract with 40% ammonium sulfate, the recombinant human PC can be purified by a single step monomeric avidin affinity chromatography with an overall yield of 20%. PC purified by this method yielded a single band on SDS-PAGE and exhibited a specific activity of 20 U/mg.

STATEMENT

This thesis contains no material that has been accepted for the award of any degree or diploma by any other University. To the best of my knowledge it contains no material that has previously been published by any other person, except where due reference has been made in the text. I consent to this thesis, when deposited in the University library, being available for photocopying and loan.

Sarawut Jitrapakdee

ACKNOWLEDGEMENTS

ABBREVIATIONS

Abbreviations used throughout this thesis are in accordance with those described in The *Journal of Biological Chemistry*. Additional abbreviations are shown below.

ACC	acetyl-CoA carboxylase
BC	biotin carboxylase
bp	base pair
BSA	bovine serum albumin
cpm	counts per minute
CPS	carbamoylphosphate synthetase
CREB	cAMP responsive element binding protein
DMSO	dimethylsulfoxide
DMEM	Dulbecco's modified Eagle's medium
DTT	dithioththreitol
FSE1	fat-specific element 1
HIP-1	housekeeping initiator protein 1
IRE	insulin responsive element
kb	kilobase(s)
LD-PCR	long distance PCR
MCC	methylcrotonyl-CoA carboxylase
min	minute
NF-1	nuclear factor 1
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PC	pyruvate carboxylase
PCC	propionyl-CoA carboxylase
PCR	polymerase chain reaction
PEPCK	phosphoenolpyruvate carboxykinase
PMSF	phenylmethylsulfonyl fluoride
RACE-PCR	rapid amplification cDNA ends
rpm	revolutions per minute
RT-PCR	reverse transcriptase PCR
TC	transcarboxylase
TCA	tricarboxylic acid cycle
U	unit
UTR	untranslated region

PUBLICATIONS ARISING FROM THIS THESIS

Refereed journals

1. **Jitrapakdee, S.,** Booker, G.W., Cassady, A.I. & Wallace, J.C. (1996)
    Cloning, sequencing and expression of rat liver pyruvate carboxylase.
    Biochem. J. **316**, 631-637.

2. **Jitrapakdee, S.,** Walker, M.E. & Wallace, J.C. (1996) Identification of novel
    alternatively spliced pyruvate carboxylase mRNAs with a divergent 5'-untranslated
    regions which are expressed in a tissue-specific manner. *Biochem. Biophys. Res.*
    Commun. **223**, 695-700.

3. **Jitrapakdee, S.,** Booker, G.W., Cassady, A.I. & Wallace, J.C. (1997)
    The rat pyruvate carboxylase gene structure : Alternate promoters generate multiple
    transcripts with 5'-end heterogeneity. *J. Biol. Chem.* **272**, 20520-20528.

4. Webb, G.C., **Jitrapakdee, S.,** Bottema, C.D.K & Wallace, J.C. (1997)
    Assignment of the pyruvate carboxylase gene to rat chromosome band 1q43 by
    in situ hybridization. *Cytogenet. Cell Genet.* **79**, 151-152.

5. Wallace, J.C., **Jitrapakdee, S.** & Chapman-Smith, A. (1998) Pyruvate
    Carboxylase (Review). *Int. J. Biochem. Cell Biol.* **30**, 1-5.

6. Hobbs, S., **Jitrapakdee, S.** & Wallace, J.C. (1998) Development of a bicistronic
    vector driven by polypeptide chain elongation factor 1α promoter for the creation of
    stable mammalian cell lines that express very high levels of recombinant proteins.
    Biochem. Biophys. Res. Commun. **252**, 368-372.

7. **Jitrapakdee, S.,** Gong, Q., MacDonald, M.J. & Wallace J.C. (1998) Regulation
    of rat pyruvate carboxylase gene expression by alternate promoters during
    development, in genetically obese rats and in insulin secreting cells: Multiple
    transcripts with 5'-end heterogeneity modulate translation. *J. Biol. Chem.* **273**,
    34422-34428.

8. **Jitrapakdee, S.** & Wallace, J.C. (1999) Structure, function and regulation of
    pyruvate carboxylase (review). *Biochem. J.* (Accepted).

9. **Jitrapakdee, S.** , Walker, M.E. & Wallace, J.C. (1999) Overexpression and
    purification of recombinant human pyruvate carboxylase from mammalian cells
    (*manuscript in preparation*).

Conference abstracts

1. **Jitrapakdee, S.,** Booker, G.W., Cassady, A.I & Wallace, J.C. (1995) Rat Pyruvate Carboxylase: Heterogeneity in the 5'-end of mRNA and Primary Structure Deduced from cDNA. *Proceedings of the Australian Society of Biochemistry and Molecular Biology* **28**, POS 1-66.

2. **Jitrapakdee, S.,** Booker, G.W., Cassady, A.I. & Wallace, J.C. (1996) Alternative splicing of the 5'-UTR exons of rat pyruvate carboxylase gene generates multiple transcripts with 5'-end heterogeneity which are differentially expressed. *Proceedings of the Australian Society of Biochemistry and Molecular Biology* **29**, POS 297-03.

3. **Jitrapakdee, S.** & Wallace, J.C. (1997) Regulation of rat pyruvate carboxylase gene expression by alternate promoters. 10th International Congress of Biochemistry and Molecular Biology (San Francisco, USA) *FASEB J.* **11** (9), (Supp) 1275.

4. **Jitrapakdee, S.** & Wallace, J.C. (1997) Understanding regulation of the rat pyruvate carboxylase expression through analysis of the gene structure and the transcriptional unit. *Proceedings of the Australian Society of Biochemistry and Molecular Biology* **30**, B1-56.

5*. **Jitrapakdee, S.,** Booker, G. W., Webb, G.C., Cassady, A. I. & Wallace, J. C. (1998) Molecular characterisation of pyruvate carboxylase gene and studies on the regulation of its expression in rat. *Proceedings of the 20th Lorne Genome Conference* SYM 7-3.

6. **Jitrapakdee, S.,** Hobbs, S., Walker, M.E., Mulhausler, B. & Wallace, J.C. (1998) Expression of functionally active human pyruvate carboxylase using bicistronic cassette driven by polypeptide chain elongation factor-1α promoter. *Proceedings of the Australian Society of Biochemistry and Molecular Biology* **31** SYM 30-04.

7¶. **Jitrapakdee, S.,** MacDonald, M.J. & Wallace, J.C. (1998) Translation regulation of rat pyruvate carboxylase expression is mediated through the 5'-untranslated regions. *Proceedings of the Australian Society of Biochemistry andMolecular Biology* **31,** POS-094.

*Presented on acceptance of an award for research excellence by a young investigator
¶Awarded an ASBMB poster prize

CHAPTER 1

INTRODUCTION

AND

LITERATURE REVIEW

1.1 GLUCONEOGENESIS AND GLYCOLYSIS

Glycolysis is the metabolic pathway by which glucose is converted via fructose-1,6-biphosphate to pyruvate with the generation of energy in the form of ATP. This sequence of enzymatic reactions (Fig. 1.1) occurs in the cytosol and plays a key role in energy metabolism by providing a significant portion of the energy utilised by most living organisms. Glucose occupies a central role in metabolism, both as a fuel in certain tissues (e.g. brain, white blood cell and renal medulla) and as a precursor of essential structural carbohydrate and other biomolecules in most tissues. During fasting, cells must therefore obtain glucose from either the breakdown of glycogen in liver or by gluconeogenesis in liver and kidney. The non-carbohydrate precursors that can be converted to glucose include the glycolysis products lactate and pyruvate, citric acid cycle intermediates, and some amino acids. However these substances must be converted to oxaloacetate the starting material for gluconeogenesis. Gluconeogenesis, in part a reverse pathway of glycolysis occurs exclusively in liver and kidney cortex utilising most glycolytic enzymes except three steps including the conversions of pyruvate to oxaloacetate, fructose 1,6-bisphosphate to fructose-6-phosphate and glucose-6-phosphate to glucose which can not be simply reversed by pyruvate kinase, phosphofructokinase and hexokinase respectively, due to the large negative free energy changes in the direction of glycolysis (Krebs, 1954). However cells overcome this situation by having another four enzymes including pyruvate carboxylase acting in concert with phosphoenolpyruvate carboxykinase (PEPCK) and fructose-1,6-bisphosphatase and glucose-6-phosphatase which bypass the energy barriers as shown in Fig 1.1.

1.2 COORDINATE REGULATION OF GLYCOLYTIC AND GLUCONEOGENIC ENZYME GENES IN LIVER

Although many other enzymes are involved in the glycolytic and gluconeogeneic pathways, they catalyse equilibrium reactions which are not rate-limiting step. The control of gluconeogenesis and glycolysis has been comprehensively reviewed (Granner and Pilkis, 1990; Pilkis and Granner, 1992; Lemaigre and Rousseau, 1994). Three substrate cycles, glucose/glucose-6-phosphate, fructose-6-phosphate/fructose-1,6-bisphosphate and

1
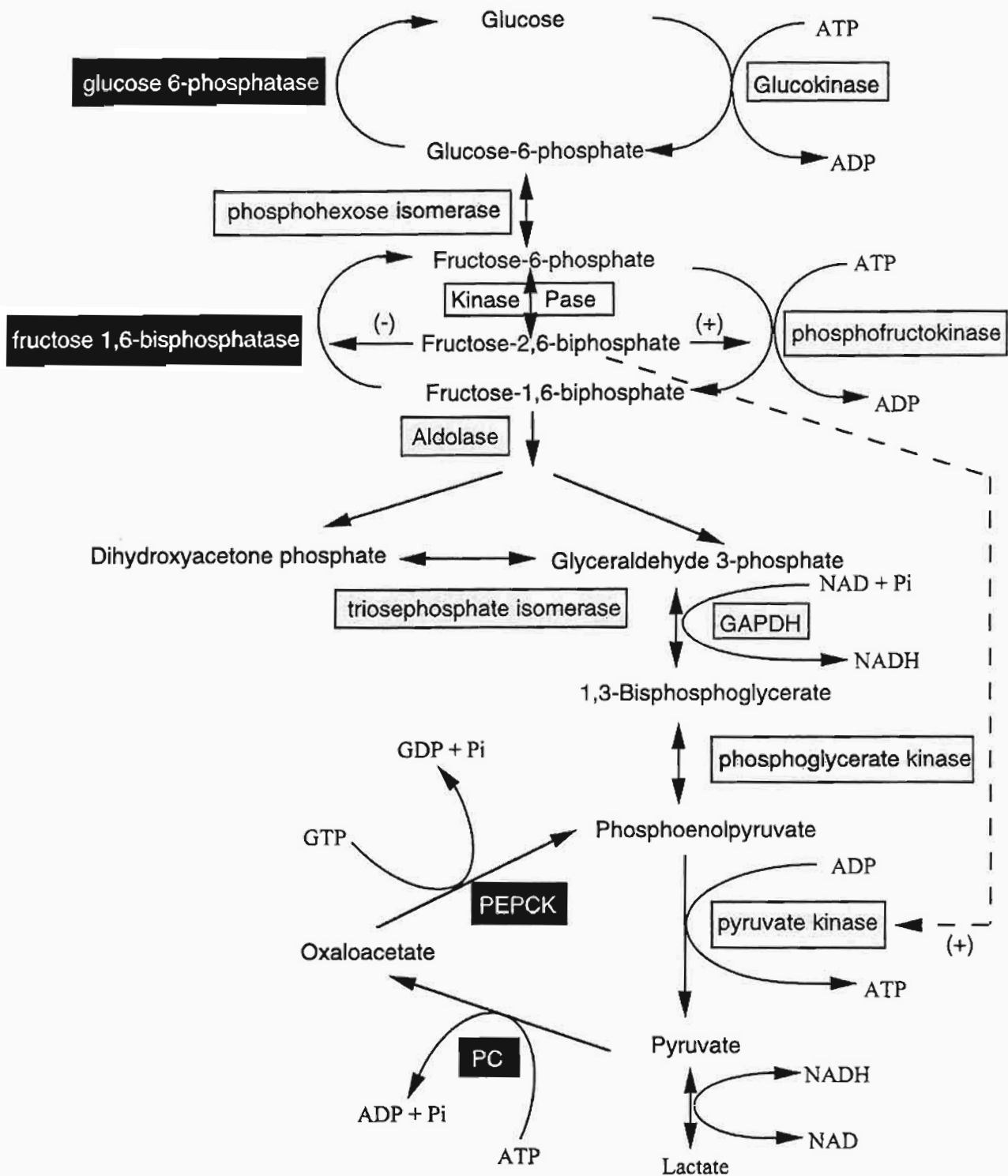
Figure 1.1 Schematic diagram of glycolysis and gluconeogenic pathways.
Different enzymes that catalyse different steps in these pathways are shown in
boxes. Fructose-2,6-bisphosphate is an activator of phosphofructokinase and
pyruvate kinase but acts as inhibitor of fructose-1,6-bisphosphatase. PC, pyruvate
carboxylase; PEPCK, phosphoenolpyruvate carboxykinase; GAPDH,
glyceraldehyde-3-phosphate dehydrogenase; Kinase/ase, 6-phosphofructo-2-
kinase/fructose-2,6-bisphosphatase.

phosphoenolpyruvate/pyruvate are irreversibly catalysed by different enzymes (Fig. 1.1) and are known to be targets for short-term or long-term regulation. Short-term regulation (seconds to minutes) of hepatic glucose can be achieved via hormonal changes which in turn alter the enzymatic activity of three enzymes namely, 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (bifunctional enzyme) and pyruvate kinase. High concentrations of plasma glucagon and catecholamines (and a low concentration of plasma insulin) enhance adenylate cyclase activity which results in a rise in the level of cAMP, and thus increased cAMP-dependent protein kinase activity. This signalling cascade results in the phosphorylation of pyruvate kinase and 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. This phosphorylated state of both enzymes reduces the conversion of phosphoenolpyruvate to pyruvate and fructose-1,6-bisphosphate to fructose-6-phosphate, thus driving glucose metabolism to the gluconeogenic direction (Pilkis and Granner, 1992). On the other hand, a high level of insulin (low levels of glucagon and catecholamine), favours glucose metabolism to the glycolytic direction i.e. binding of insulin to its receptor which exhibits tyrosine kinase activity and activating phosphodiesterase activity. This results in a reduced cAMP level and its downstream effector, which reduces phosphorylation of pyruvate kinase and 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. An increase in the level of fructose-1,6-bisphosphate is also known to act as a potent activator of pyruvate kinase. This can also act to amplify the pyruvate kinase reaction (Pilkis and Granner, 1992).

Long term regulation (minutes to hours) of glucose metabolism can also be achieved via hormonal changes. These occur during starvation, low carbohydrate diet intake, or prolonged exercise which lead to high levels of plasma glucagon, catecholamine and glucocorticoids. This results in an increased activity of PEPCK, fructose-1,6-bisphosphatase and glucose-6-phosphatase and a coordinate decrease in pyruvate kinase, phosphofructokinase and glucokinase activities (Granner and Pilkis, 1990; Pilkis and Granner 1992). These changes drive glucose metabolism in the direction of gluconeogenesis. However, the opposite effect occurs when animals are fed a carbohydrate-rich diet which results in an increased levels of plasma insulin and a low level of other hormones. Insulin stimulates transcription of the genes encoding key glycolytic enzymes and inhibits transcription of the genes encoding key gluconeogenic enzymes. The

2

mechanism of how insulin can exert its effect on the transcription of these genes is not well defined. It has been suggested that the binding of insulin to its receptors, triggers a signalling cascade resulting in phosphorylation of insulin receptor substrate 1 (IRS-1), which in turn activates downstream effectors including transcription factors (White and Hahn, 1994). Glucagon and catecholamine act via cAMP which exerts effects opposing insulin by inhibiting transcription of glycolytic enzyme genes while stimulating transcription of gluconeogenic enzyme genes (Granner and Pilkis, 1990; Pilkis and Granner, 1992). The mechanism of how cAMP can modulate the genes encoding glycolytic and gluconeogenic enzymes remains unclear, but it may be mediated through a cAMP-responsive element(s) located in the promoters of these genes. Whereas insulin and cAMP affect all the genes encoding key glycolytic and gluconeogenic enzymes, glucocorticoids have a more restricted action. They play a role in increasing transcription of PEPCK but have a more permissive effect on glucokinase. This coordinate regulation between genes for key glycolytic and gluconeogenic enzymes allows an efficient mechanism to control blood glucose under different physiological conditions. Although most studies have focused on understanding the regulation of genes encoding key glycolytic and gluconeogenic enzymes, many attempts have also been made to isolate and characterise the transcriptional units from other genes whose products are involved in glycolytic and gluconeogenic pathways in order to fully understand the molecular basics underlying the control of glucose metabolism (see Table 1 and Table 2).

Table I: Genes encoding glycolytic enzymes

Enzyme	Gene structure	organism	reference
glucokinase	single copy gene, isozymes produced by alternate promoters	rat human	Magnuson *et al.*, 1989 Magnuson and Shelton 1989 Stoffel *et al.*, 1992 Tanizawa *et al.*, 1992
phosphofructokinase	isoenzymes produced by separate genes	mouse human	Rongnoparut *et al.*, 1991 Elson *et al.*, 1990 Yamasaki *et al.*, 1991
6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase	isozymes produced by separate genes/alternate promoters	rat	Darville *et al.*, 1989 Darville *et al.*, 1991
aldolase	isoenzymes produced by separate genes	rat human	Burgess and Penhoet, 1985 Tsutsumi *et al.*, 1985
glyceraldehyde-3-phosphate dehydrogenase	single copy gene with pseudogenes	rat human	Piechaczyk *et al.*, 1984 Ercolani *et al.*, 1988
phosphoglycerate kinase	isoenzymes produced by separate genes	human	Michelson *et al.*, 1985
phosphoglycerate mutase	isoenzymes produced by separate genes	rat human	Ruiz-Lazano *et al.*, 1994 Tsujino *et al.*, 1989 Castella-Escola *et al.*, 1990
enolase	isoenzymes produced by separate genes	rat human	Sakimura *et al.*, 1987 Giallongo *et al.*, 1990 Oliva *et al.*, 1991 Peshavaria and Day, 1991 Giallongo *et al.*, 1993
pyruvate kinase	isoenzymes produced by separate genes	rat human	Noguchi *et al.*, 1987 Kanno *et al.*, 1992

Table 2: Genes encoding gluconeogenic enzymes

Enzyme	Gene structure	organism	reference
pyruvate carboxylase	single gene/alternate promoters	rat	chapters 4, 5 (Jitapakdee *et al.*, 1997)
		human	Carbone *et al.*, 1998
phosphoenol pyruvate carboxykinase	single gene/single promoter	rat	Beale *et al.*, 1985
fructose bisphosphatase	single gene/single promoter	rat	el-Maghrabi *et al.*, 1991
		human	el-Maghrabi *et al.*, 1995
glucose-6-phosphatase	single gene/single promoter	mouse	Shelly *et al.*, 1993
		rat	Argaud *et al.*, 1996
		human	Lei *et al.*, 1993

1.3 TRANSCRIPTIONAL REGULATION OF GLUCONEOGENIC ENZYME GENE EXPRESSION

1.3.1 PEPCK

PEPCK catalyses the conversion of oxaloacetate to phosphoenolpyruvate. PEPCK is regulated by nutrients and hormones. In some mammals there are two isoforms of PEPCK, encoded by separate genes. PEPCK-C is located in the cytosol whereas PEPCK-M is located in the mitochondrial matrix. The distribution of PEPCK isozymes has been found to vary from species to species (Hanson and Patel, 1994). Only the gene encoding a PEPCK-C isoform in rat has been extensively studied. PEPCK-C is encoded by a single copy gene which contains 10 exons and is controlled by a single promoter (Beale *et al.*, 1985). The minimal promoter of the gene contains a TATA box, a CCAAT box, a combined basal enhancer element and a cAMP-responsive element (CRE) all of which are required for basal transcription (Quinn *et al.*, 1988). The regulatory region of PEPCK-C confers a multihormonal response to insulin, cAMP and glucocorticoids, has initially been shown to be located within the first 600 bp upstream from the transcription initiation site (Magnuson *et al.*, 1987). Subsequent functional analysis of this promoter revealed the presence of a number of transcription factor binding sites including a CRE, a glucocorticoid

regulatory unit (GRU) consisting of glucocorticoid receptor binding sites (GRE) and accessory factor binding sites (AF1, AF2), an insulin-regulatory element (IRE), a thyroid hormone regulatory element (TRE), as well as other transcription factor binding sites located within this region (Hanson and Reshef, 1997) [see Fig. 1.2].
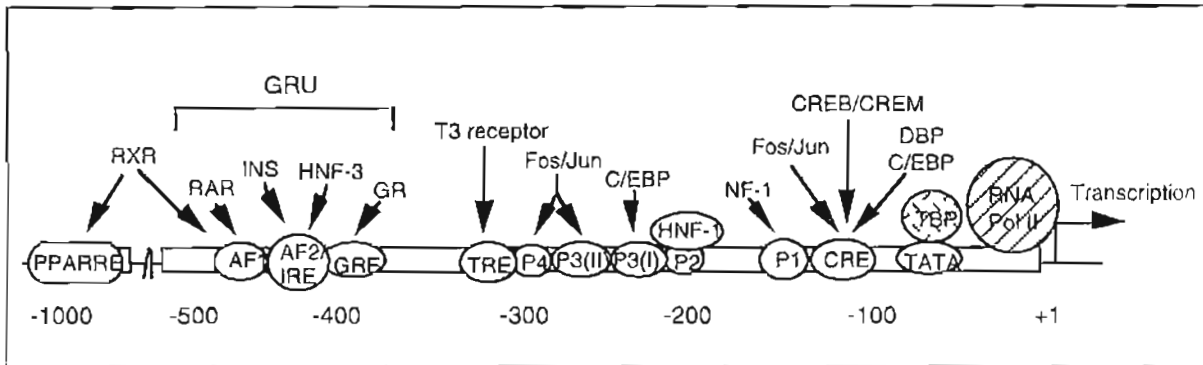
Figure 1.2 Transcription unit of the promoter of PEPCK-C gene (adapted from Hanson and Reshef, 1997). Different *cis*-acting elements are shown by ovals, and the different regulatory proteins that interact with these elements are also shown. CRE, cAMP regulatory element; TRE, thyroid-hormone regulatory element; GRE, glucocorticoid regulatory element; TBP, TATA binding protein; IRE, insulin regulatory element; PPARRE, peroxisome proliferatory-activated receptor regulatory element; AF-1, accessory factor-1; C/EBP, CCAAT/enhancer binding protein; NF-1, nuclear factor-1; CREBP, cAMP regulatory element binding protein; DBP, D-binding protein; HNF-1 and HNF-3, hepatic nuclear factor 1 and 3, respectively; GRU, glucocorticoid regulatory unit; GR, glucocorticoid receptor; RAR, retinoic acid receptor; RXR, retinoic X receptor; ins, insulin.

The increased plasma insulin following carbohydrate intake decreases the rate of PC synthesis (Granner *et al.*, 1983) due to the inhibition of PEPCK transcription (Granner *et al.*, 1983; Sasaki *et al.*, 1984). In contrast, the elevated plasma glucagon (via cAMP) during fasting induces PEPCK expression (O'Brien and Granner, 1990) via transcriptional activation (Granner *et al.*, 1983; Sasaki *et al.*, 1984). Glucocorticoids which act as permissive agents to promote gluconeogenesis, also promote transcription of the PEPCK gene (Petersen *et al.*, 1989). The actions of these hormones are mediated through hormone responsive elements located in the promoter of the PEPCK gene. cAMP and glucocortioids not only stimulate transcription of the PEPCK gene but also stabilise PEPCK mRNA against

degradation (Hod and Hanson, 1988; Petersen *et al.*, 1989). The transcriptional regulation of PEPCK has recently been reviewed by Hanson and Reshef (1997).

1.3.2 Fructose-1,6-bisphosphatase

Fructose-1,6-bisphosphatase catalyses the conversion of fructose-1,6-bisphosphate to fructose-6-phosphate. Starvation and diabetes induce fructose-1,6-bisphosphatase activity (Kitajima and Uyeda, 1983) and mRNA (el-Maghrabi *et al.*, 1988). The gene encoding this enzyme is comprised of 7 exons spanning over 23 kb. The promoter region of the gene for fructose-1,6-bisphosphatase contains a consensus TATA box and a Sp1 binding site (el-Maghrabi *et al.*, 1991). cAMP has been shown to increase level of fructose-1,6-bisphosphatase mRNA in rat hepatocytes and expression of the fructose-1,6-bisphosphatase promoter-driven luciferase gene, consistent with the presence of a CRE (el-Maghrabi *et al.* 1991). In contrast, insulin opposes the action of cAMP by decreasing mRNA encoding fructose-1,6-bisphosphatase in primary hepatocytes (el-Maghrabi *et al.*, 1991).

1.3.3 Glucose-6-phosphatase

Glucose-6-phosphatase is a multi-subunit microsomal enzyme which catalyses the terminal step in the hydrolysis of glucose-6-phosphate to glucose in the pathways of gluconeogenesis and glycogen breakdown in the liver (Nordlie and Sukalski, 1985). The enzyme activity is highly detectable in liver and kidney. Hepatic glucose-6-phosphatase is increased during starvation and diabetes (Nordlie *et al.*, 1968; Barzilai and Rossetti, 1993; Argaud *et al.*, 1996; Massillon *et al.*, 1996). The cDNA and gene (Shelly *et al.*, 1993; Lei *et al.*, 1993; Lange *et al.*, 1994; Argaud *et al.*, 1996; Scholl *et al.*, 1996) encoding the catalytic subunit of glucose-6-phosphatase have been cloned. Glucocorticoids and cAMP increase the level of glucose-6-phosphatase mRNA while insulin exerts the opposite effect and also has a dominant negative effect on the transcription of the glucose-6-phosphatase gene when insulin is present with glucocorticoids and cAMP (Lange *et al.*, 1994; Argaud *et al.*, 1996). The rat glucose-6-phosphatase gene is comprised of 5 exons spanning 13 kb and the promoter contains canonical TATA and CCAAT boxes (Argaud *et al.*, 1996). Functional analysis of promoters of human (Schmoll *et al.*, 1996; Lin *et al.*, 1997; Lin *et al.*, 1998) and mouse

(Streeper *et al.*, 1997) glucose-6-phosphatase genes revealed the presence of multihormonal-responsive units including those for insulin, cAMP and glucocorticoids.

1.4   PYUVATE CARBOXYLASE (PC)

Pyruvate carboxylase (PC) was first described by Utter and Keech (1960) in the course of defining the gluconeogenic pathway in chicken liver. The reaction PC catalyses is:

$$\text{Pyruvate} + HCO_3^- + ATP \xleftrightarrow{\quad \text{acetyl-CoA, } Mg^{2+} \quad} \text{oxaloacetate} + ADP + Pi$$
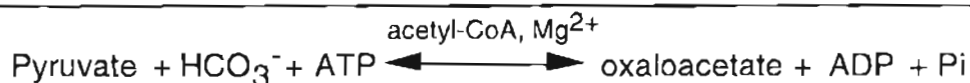
A few years later, PC from chicken was purified and its general properties and reactions were described (Utter and Keech, 1963; Keech and Utter, 1963).

1.5   BIOLOGICAL ROLES OF PC

PC has been found widely distributed among both vertebrates and invertebrates, as well as in many microorganisms (reviewed by Wallace, 1985). In prokaryotes, PC has been identified in *Pseudomonas spp, Rhodobacter spp, Bacillus spp, Rhizobium spp* etc. but not in Enterobactcriaceae (Wallace, 1985). The last group of bacteria can synthesise oxaloacetate directly from phosphoenolpyruvate by phosphoenolpyruvate carboxylase (Ashworth and Kornberg, 1966). A genetic study has indicated that expression of *E. coli* phosphoenolpyruvate carboxylase can complement the phenotypic effects of PC-deficiency in yeast (Flores and Gancedo, 1997). The anaplerotic role of PC has been shown to be essential for normal growth in *Rhizobium etli* (Dunn *et al.*, 1996). In *Rhodobacter capsulatus*, changes in PC specific activity under different growth conditions are mediated at the level of enzyme synthesis (Yakunin and Hallenbeck, 1997) while PC activity in *Bacillus stearothermophilus* (Cazzulo *et al.*, 1971) and *Rhizobium etli* (Dunn *et al.*, 1997) is regulated by the availability of biotin in the media. Conversion of an inactive form to an active form of PC in these two organisms has been shown to occur post-translationally through the biotinylation reaction not the transcriptional step (Dunn *et al.*, 1997).

In *Saccharomyces cerevisiae*, the two principal pathways for the replenishment of oxaloacetate levels are via the carboxylation of pyruvate by PC and from the glyoxylate cycle

(Barnett and Kornberg, 1960). During growth in the presence of glucose, the enzymes of the glyoxylate cycle are repressed (Gancedo, 1992) and PC catalyses the only known reaction to replenish the TCA cycle in this condition. The highest activity of PC was found in glucose-grown cells under anaerobic conditions (Haarasilta and Oura, 1975).

A recent discovery has shown that plants also contain PC which has been suggested to be an alternative gluconeogenic pathway in addition to the photosynthetic process during germination (Wurtele and Nikolau, 1990).

In mammals, PC is expressed in a tissue-specific manner with its activity found to be highest in the liver and kidney (gluconeogenic tissues), in adipose tissue and lactating mammary gland (lipogenic tissues) and in pancreatic islets, moderate in brain, heart, adrenal gland and least in white blood cells and skin fibroblasts (Wallace, 1985; Wexler *et al.*, 1994; Jitrapakdee *et al.*, 1996). The roles of PC in these different tissues are described below.

1.5.1 Gluconeogenesis

In fasting, gluconeogenesis accounts for up to 96% of total glucose production (Rothman *et al.*, 1991). The presence of very high PC activity together with other gluconeogenic enzymes including PEPCK, fructose 1,6-bisphosphatase and glucose 6-phosphatase in liver and kidney cortex suggests that the primary role of PC is to participate in gluconeogenesis in these tissues. During fasting or starvation in which endogenous glucose is required for certain tissues (brain, white blood cells and kidney medulla), PC expression together with other gluconeogenic enzymes have been shown to be elevated (Owen *et al.*, 1976). The reaction catalysed by PC is also known to be a rate-limiting point in the gluconeogenic pathway, since this step has been shown to be regulated by a number of metabolites, diets and some hormones (reviewed by Barritt, 1985).

1.5.2 Lipogenesis

PC was first considered to be only a gluconeogenic enzyme but was soon recognised to be expressed at a very high level during the differentiation of adipocytes (Mackall and Lane, 1977). PC appears to be particularly important in adipose tissue where it contributes to the generation of a substantial proportion of the NADPH required for lipogenesis (Ballard

Figure 1.3 Schematic representation of the anaplerotic functions of PC in mammalian tissues in relation to the biosynthetic pathways in which oxaloacetate is utilised. Four major biosynthetic pathways including gluconeogenesis in liver and kidney, fatty acid synthesis in adipose tissue and lactating mammary gland, synthesis of neurotransmitter precursors in astrocytes and the generation of metabolic coupling factor NADPH in pancreatic islets. PC, PC; MDH, malate dehydrogenase; ACC, acetyl-CoA carboxylase; PEP, phosphoenolpyruvate.
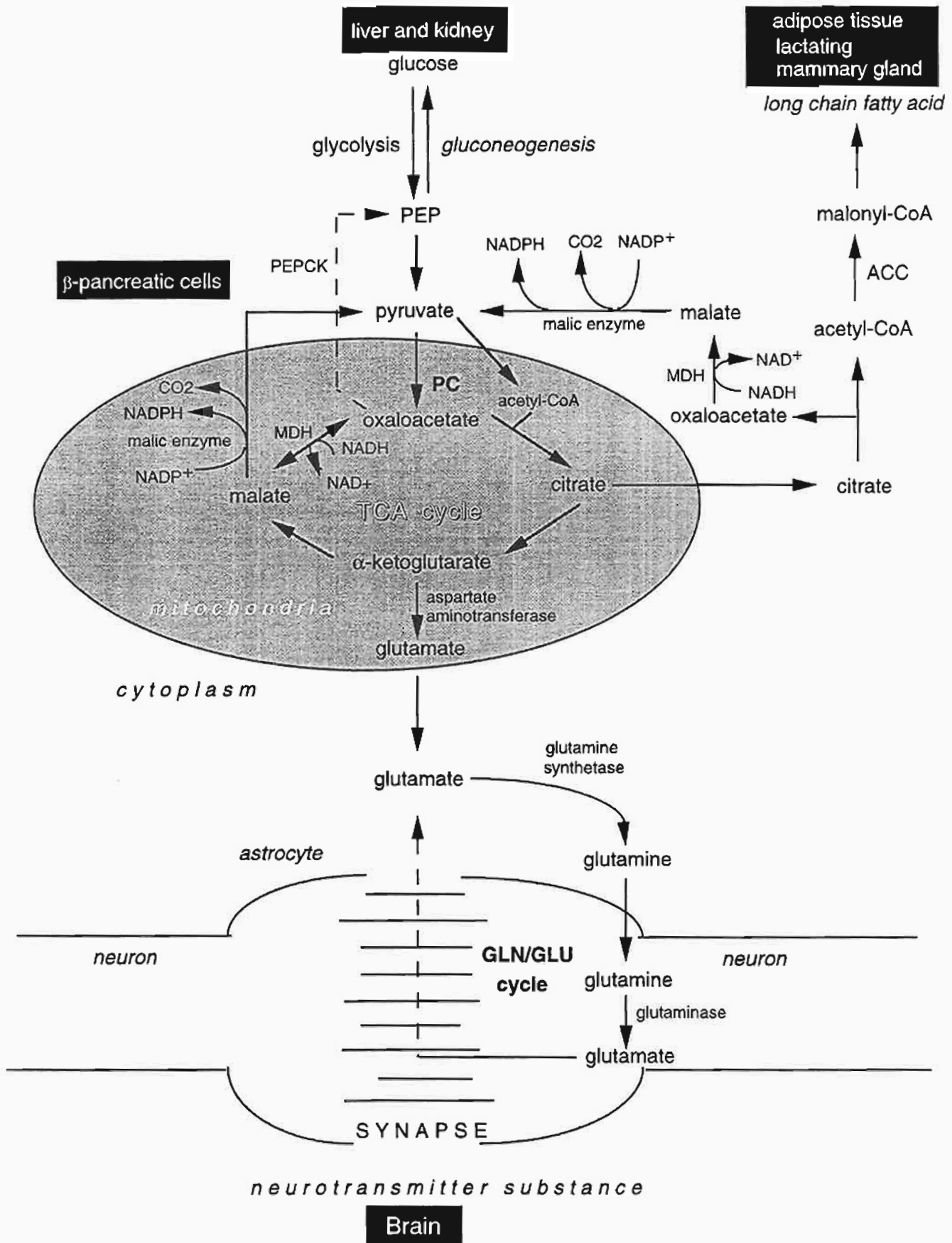
and Hanson, 1970). As shown in Fig. 1.3, the generation of NADPH is coupled to the transport of mitochondrial acetyl groups into the cytosol for fatty acid synthesis. Acetyl-CoA is generated in the mitochondria by the oxidative decarboxylation of pyruvate, and acetyl-groups are transported to the cytoplasm as citrate, which undergoes ATP-dependent cleavage to acetyl-CoA and oxaloacetate. This pathway requires a continual supply of oxaloacetate, which is produced by the activity of PC. Acetyl-CoA, a building block for long chain fatty acid synthesis, is then converted to malonyl-CoA by acetyl-CoA carboxylase. Meanwhile the oxaloacetate generated in the cytosol from citrate is reduced with NADH to malate which in turn is oxidatively decarboxylated in a reaction catalysed by the NADP-dependent malic enzyme (EC 1.1.1.40). The pyruvate thereby produced is taken up by the mitochondria to be carboxylated to oxaloacetate, while the NADPH produced is used in the pathway of fatty acid synthesis.

1.5.3 Role of PC in insulin signalling in pancreatic islets

Glucose is a potent stimulator of insulin secretion from β-pancreatic cells when extracellular levels are greater than 3 mM. Secretion of insulin in response to a high concentration of glucose results in the rapid uptake of glucose by β-pancreatic cells more than by other cell types (Heimberg *et al.*, 1993; Sekine *et al.*, 1994). This is a feature of β-pancreatic cells but not α-pancreatic cells (Schuit *et al.*, 1997). Signalling for glucose-induced insulin release is believed to require aerobic glycolysis plus the TCA cycle activity (MacDonald, 1981; MacDonald, 1990). Two mitochondrial enzymes, pyruvate dehydrogenase and PC have been shown to be elevated when pancreatic islets are grown in higher than physiological concentrations of glucose, suggesting that both enzymes are involved in the regulation of glucose-induced insulin release (MacDonald *et al.*, 1991). It is known that pancreatic islets contain a concentration of PC equivalent to that in gluconeogenic tissues, but lack PEPCK activity and its mRNA. This suggests that PC is not present for the purpose of gluconeogenesis (MacDonald and Chang, 1985; MacDonald *et al.*, 1992). The rapid uptake of glucose is thought to be mediated through the glucose-sensing enzyme, glucokinase which is rate-limiting for overall glucose utilization in β-pancreatic cells (Liang *et al.*, 1992; De Vos *et al.*, 1995). Glucose then undergoes oxidation to pyruvate (Schuit *et al.*,

1997) which is subsequently carboxylated by PC (MacDonald, 1995b; Schuit *et al.*, 1997). Higher concentrations of glucose also up-regulate the levels of PC protein (MacDonald, 1995a) and its mRNA (MacDonald *et al.*, 1991; MacDonald, 1995a).

It has been demonstrated that there is a pyruvate/malate shuttle operating across mitochondrial membrane as shown in Fig. 1.3. The high level of PC permits the rapid formation of oxaloacetate which is subsequently converted to malate by malic dehydrogenase (MDH) and exits the mitochondrial membrane to the cytosol where it is decarboxylated to pyruvate by malic enzyme to produce a putative coupling factor, NADPH (MacDonald, 1995b). Since this pathway occurs as a cycle, this shuttle can generate far more NADPH than the pentose phosphate pathway. Although the above pathways have been shown to be linked to insulin secretion, the metabolic signaling mechanism leading to insulin release remains unknown. However, it has been shown that an increase in glucose metabolism, which causes a rise in the ratio of ATP/ADP in islets, results in the closure of K_{ATP} channels (Ashcroft *et al.*, 1984), membrane depolarization, and influx of Ca^{2+}, which is required for exocytosis.

1.5.4 Role of PC in astrocytes

Although four gluconeogenic enzymes including glucose-6-phosphatase (Middleditch *et al.*, 1998), fructose-1,6-bisphosphatase (Liu and Fromm, 1988), PEPCK (Zimmer and Magnuson, 1990) and PC (Faff-Michalak and Albrecht, 1991) have been reported in brain, the specific activities of these enzymes are too low to ensure gluconeogenesis. However there are a few reports which have demonstrated that lactate, alanine, aspartate or glutamine could be converted to glycogen in astrocyte cultures (Wiesinger *et al.*, 1997; Schmoll *et al.*, 1995) but not in neurons due to the absence of PC (Shank *et al.*, 1985).

The anaplerotic role of PC has been proposed to be necessary for the production of glutamine, the precursor of excitatory amino acid neurotransmitters via the operation of the glutamate/glutamine cycle (Benjamin and Quastel, 1974; Cooper and Plum, 1987). As indicated in Fig. 1.3, when glutamate, a neurotransmitter substance is released from the nerve endings of neurons, it is taken up by astrocytes. Subsequently, glutamate is converted to glutamine by astrocytic glutamine synthetase and secreted into the extracellular fluid.

Glutamine is then taken up by neurons for conversion to glutamate, aspartate and γ-aminobutyric acid (Cooper and Plum, 1987; Shank *et al.*, 1993; Gamberino *et al.*, 1997). Oxaloacetate, produced by PC, can also participate in this glutamate/glutamine cycle and a recent study has shown that PC can alter the rate of *de novo* astrocytic synthesis of glutamate by increasing the amount of citric acid cycle intermediates (Gamberino *et al.*, 1997).

1.6   SYNTHESIS AND DEGRADATION

In *S. cerevisiae*, there are two isozymes of PC encoded by separate genes (PC1 and PC2) [Walker *et al.*,1991; Stucka *et al.*, 1991] while in mammals, no tissue-specific isozymes have been reported. The newly synthesised enzyme undergoes a post-translational modification in which one biotin moiety is covalently attached to a side chain of a specific lysine located near the C-terminal of each protomer. This reaction is known as the biotinylation reaction and is catalysed by biotin ligase [E.C. 6.3.4.1] (also known as holoenzyme synthetase). A two-step reaction has been shown to convert apoenzyme to holoenzyme, as indicated in the following equations (Lane *et al.*, 1964).

$$\text{Biotin + ATP} \longleftrightarrow \text{Biotinyl 5'-AMP + PPi} \qquad (1)$$

$$\text{Biotinyl 5'-AMP apocarboxylase} \longleftrightarrow \text{holocarboxylase + AMP} \qquad (2)$$

Holocarboxylase synthetase from mammals can biotinylate bacterial apocarboxylases (McAllister and Coon, 1966) and *vice versa* (Lane *et al.*, 1964). Recently, the genes encoding biotin ligase from *E. coli* (Howard *et al.*, 1985), *S. cerevisiae* (Cronan and Wallace, 1995) and human (Leon-Del-Rio *et al.*, 1995) have been cloned. Recent work has demonstrated that the biotin carboxyl carrier domain of *E. coli* acetyl-CoA carboxylase (ACC) undergoes a conformational change accompanying biotinylation (Chapman-Smith *et al.*, 1997). The presence of a prosthetic group, biotin on PC allows investigators to detect the biotinylated enzyme using either avidin or streptavidin conjugated with reporter enzymes. Deficiency of biotin ligase causes serious developmental and neonatal defects in humans (Suzuki *et al.*, 1994). The intracellular site of the biotinylation reaction by biotin ligase is not clear. In 3T3-L1 mouse adipocytes which contain very high levels of PC and ACC, most holocarboxylase synthetase activity was detected in the cytosol and only 30 percent of

activity was detected in the mitochondrial fraction (Chang and Cohen, 1983). A study with rat liver yielded similar results (Cohen *et al.*, 1985). In lower organisms ie. *Saccharomyces cerevisiae* (Sundaram *et al.*, 1971), *Methanobacterium thermoautotrophicum* (Mulkhopadhyay *et al.*, 1998), *Bacillus stearothermophilus* (Cazzulo *et al.*, 1971) and *Rhizobium etli* (Dunn *et al.*, 1996; Dunn *et al.*, 1997), the availability of biotin in the media has been shown to greatly enhance PC activity and this is thought to be mediated through biotinylation of the apoenzyme to the holoenzyme. L-aspartate is known to inhibit PC activity and biotinylation in yeast through an allosteric effect (Sundaram *et al.*, 1971).

In vertebrates, newly synthesised PC contains a leader sequence at the N-terminus comprising several positively charged and several hydroxylated amino acids (Jitrapakdee *et al.*, 1996). Upon translocation into the mitochondria, this targetting sequence undergoes cleavage resulting in a reduction of subunit molecular weight (Srivastava *et al.*, 1983). No further post-translational modifications including phosphorylation (Leiter *et al.*, 1978) have been reported. In confirmation of earlier studies on the intracellular localisation of PC (Böttger *et al.*, 1969; Taylor *et al.*, 1971), Rohde *et al.* (1991) employed an immunoelectron microscopic approach to show that rat PC is located exclusively in the mitochondrial matrix and close to the inner mitochondrial membrane (Rohde *et al.*, 1991). Conversely, PC in *S. cerevisiae* was shown to be exclusively cytoplasmic (Rohde *et al.*, 1991). Recent studies have indicated that the glycolytic and gluconeogenic enzymes present in the cell are associated in the form of multienzyme complexes (Srere, 1987; Berry *et al.*, 1993). PC is also found to be specifically associated in a binary complex with mitochondrial aspartate transferase or malate dehydrogenase, in a ternary complex with aspartate aminotransferase plus glutamate dehydrogenase and in a quaternary complex with aspartate aminotransferase, glutamate dehydrogenase and malate dehydrogenase (Fahein *et al.*, 1993). These interactions among PC and other mitochondrial enzymes are likely to profoundly influence their characteristics and kinetic properties (Srere, 1987) but are yet to be defined for PC.

The difference in location of PC between vertebrate mitochondria and yeast cytosol implies distinct features which control the flow of metabolites for gluconeogenesis, lipogenesis and the anaplerotic roles in these organisms. The vertebrate enzyme is known to be activated by a short-chain derivative of coenzyme A, preferably acetyl-CoA. In contrast,

the yeast enzyme is most effectively activated by long-chain acyl-CoA derivatives such as palmitoyl CoA and is inhibited by aspartate and 2-oxoglutarate, whereas the mitochondrial enzyme is not (Osmani *et al.*, 1985).

The half-life of PC in rat liver is about 4.6 days (Weinberg and Utter, 1979; Weinberg and Utter 1980) which is slightly longer than the value of 3.8 days for the average turnover time of mitochondrial protein. However in a human cell line HE(39)L (Chandler and Ballard, 1985) and in 3T3-L1 mouse adipocytes (Chandler and Ballard, 1986), PC has been shown to have shorter half lives. The degradation of PC in these cell lines has been suggested to be mediated via the mitochondrial autophagic/lysosomal degradative pathway (Chandler and Ballard, 1985).

1.7 BIOTIN CARBOXYLASE FAMILY

The biotin-dependent enzymes comprise a diverse group of enzymes, which includes carboxylases, viz. PC, acetyl-CoA carboxylase (ACC, EC 6.4.1.2), propionyl-CoA carboxylase (PCC, EC 6.4.1.3), 3-methylcrotonyl-CoA carboxylase (MCC, EC 6.4.1.4), urea carboxylase (EC 6.3.4.4), geranoyl-CoA carboxylase (EC 6.4.1.5); transcarboxylase (TC, EC 2.1.3.1); and decarboxylases e.g. oxaloacetate decarboxylase (ODC, EC 4.1.1.3), glutaconyl-CoA decarboxylase (EC 4.1.1.70), methylmalonyl-CoA decarboxylase (EC 4.1.1.41). Each of these enzymes contains the prosthetic group, biotin which is covalently bound to the ε-amino group of a specific lysine residue. The reactions which these biotin-dependent enzymes catalyse in different metabolic pathways (Wood and Barden, 1977; Chandler and Ballard, 1985) are shown in Fig. 1.4.

In mammals only four biotin-dependent carboxylases i.e. ACC, PC, PCC and MCC have been identified so far (Moss and Lane, 1972). The biotin carboxylase enzymes comprise three functional components i.e. biotin carboxylase (BC), biotin carboxyl carrier protein (BCCP) and carboxyl transferase (CT) which have been found in a wide variety of organisms. Although these enzymes are involved in diverse metabolic pathways such as gluconeogenesis, lipogenesis and breakdown of some amino acids, they share a common reaction mechanism. The reactions involve the ATP-dependent carboxylation of biotin

Figure 1.4 The reactions catalysed by the biotin-dependent enzymes.
The reactions catalysed by the biotin dependent enzymes are shown (Wood and Barden, 1977) along with known cofactors. Me^{2+} represents a divalent metal cation, acetyl-CoA is an allosteric activator of PC and citrate is an allosteric activator of acetyl-CoA carboxylase (Lane *et al.*, 1974)

CARBOXYLASES

pyruvate carboxylase [EC 6.4.1.1]
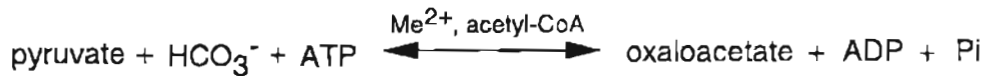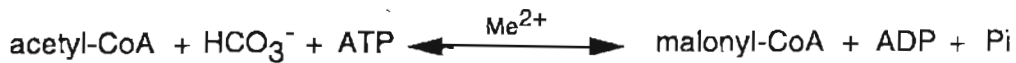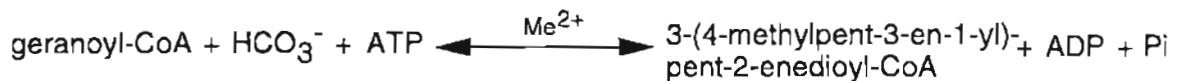
$$\text{pyruvate} + HCO_3^- + ATP \xrightarrow{\text{Me}^{2+},\ \text{acetyl-CoA}} \text{oxaloacetate} + ADP + Pi$$

acetyl-CoA carboxylase [EC 6.4.1.2]

$$\text{acetyl-CoA} + HCO_3^- + ATP \xrightarrow{\text{Me}^{2+}} \text{malonyl-CoA} + ADP + Pi$$

geranoyl-CoA carboxylase [EC 6.4.1.5]

$$\text{geranoyl-CoA} + HCO_3^- + ATP \xrightarrow{\text{Me}^{2+}} \text{3-(4-methylpent-3-en-1-yl)-pent-2-enedioyl-CoA} + ADP + Pi$$

3-methylcrotonyl-CoA carboxylase [EC 6.4.1.4]

$$\text{3-methylcrotonyl-CoA} + HCO_3^- + ATP \xrightarrow{\text{Me}^{2+}} \text{3-methylglutaconyl-CoA} + ADP + Pi$$

propionyl-CoA carboxylase [EC 6.4.1.3]

$$\text{propionyl-CoA} + HCO_3^- + ATP \xrightarrow{\text{Me}^{2+}} \text{methylmalonyl-CoA} + ADP + Pi$$

urea carboxylase [EC 6.3.4.3]

$$\text{urea} + HCO_3^- + ATP \xrightarrow{\text{Me}^{2+}} \text{allophanate} + ADP + Pi$$
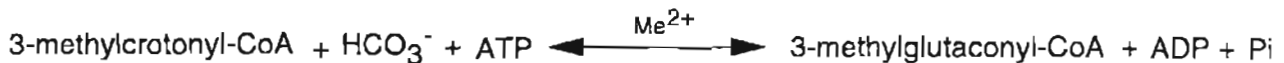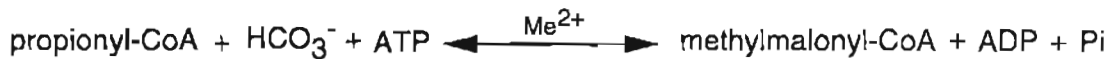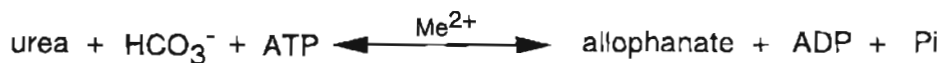
DECARBOXYLASES

glutaconoyl-CoA decarboxylase [EC 4.1.1.70]

$$\text{glutaconyl-CoA} + H^+ + 2\,Na^+ \longleftrightarrow \text{crotonyl-CoA} + CO_2 + 2\,Na^+$$

methylmalonyl-CoA decarboxylase [EC 4.1.141]

$$\text{methylmalonyl-CoA} + H^+ + 2\,Na^+ \longleftrightarrow \text{propionyl-CoA} + CO_2 + 2\,Na^+$$

oxaloacetate decarboxylase [6.1.1.3]

$$\text{oxaloacetate} + H^+ + 2\,Na^+ \longleftrightarrow \text{pyruvate} + CO_2 + 2\,Na^+$$

TRANSCARBOXYLASE

transcarboxylase [2.1.3.1]

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

which serves as a "swinging arm" in transferring CO_2 to different acceptor molecules (Knowles, 1989).

Given the similar reaction mechanisms which they catalyse, it has been postulated that biotin carboxylases represent various stages in evolution of these enzymes, starting from unifunctional subunits leading to a multifunctional enzyme through successive gene fusions (Obermayer and Lynen 1976). Recent studies which have focused on the molecular cloning and determination of the primary structures deduced from cDNA and gene sequences of the biotin-dependent enzymes, have thus made it possible to identify common amino acid residues that might play a critical role in the catalytic reaction. Comparisons of amino acid sequences among the three functional domains of the biotin carboxylases have strongly supported the evolutionary hypothesis described above (Samols *et al.*, 1988). Recently, the molecular evolution of biotin carboxylases has been evaluated by constructing a phylogenetic tree based on amino acid sequence similarities of different biotin enzymes (Toh *et al.*, 1993; Tu and Hagedorn, 1997). However, whether this evolution has occurred through the process of exon shuffling will remain unknown until the structures of various biotin carboxylase genes in higher organisms have been determined. Chapter 5 reports the first genomic structure of a mammalian biotin carboxylase gene, i.e. the rat PC gene.

1.8 REACTION MECHANISM

The reaction mechanism of PC has been extensively reviewed by Attwood (1995). The overall reaction catalysed by PC occurs in two spatially separated subsites and can be summarised into two partial reactions (equations 1 and 2) as follows.

$$\text{Enz-biotin} + \text{ATP} + \text{HCO}_3^- \xrightarrow[\text{acetyl-CoA, Mg}^{2+}]{} \text{Enz-biotin-CO}_2 + \text{ADP} + \text{Pi} \qquad (1)$$

$$\text{Enz-biotin-CO}_2 + \text{pyruvate} \rightleftharpoons \text{oxaloacetate} + \text{Enz-biotin} \qquad (2)$$
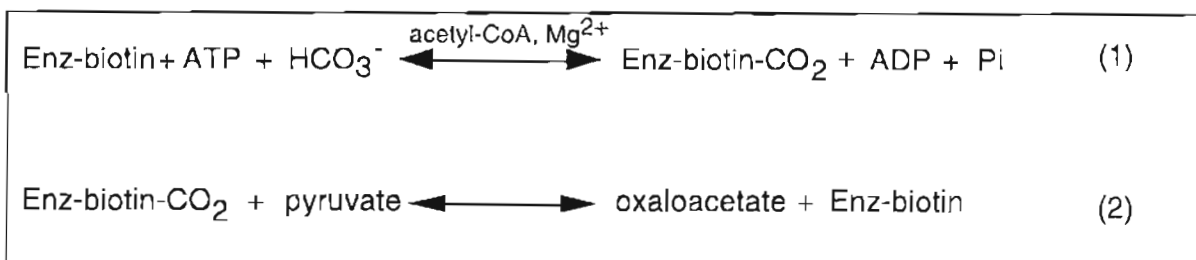
The first partial reaction involves the formation of the enzyme-carboxybiotin complex most probably via the formation of a very labile carboxyphosphate intermediate (Knowles,

1989). The evidence comes from the studies demonstrating that the biotin carboxylase subunit from *E. coli* ACC can phosphorylate ADP from carbamyl phosphate, which is an analogue of carboxyphosphate, to form ATP (Polakis *et al.*,1974). The observation that sheep kidney and chicken liver PCs are also capable of catalysing the phosphorylation of ADP by carbamoyl phosphate, provides further support for such an intermediate (Ashman and Keech, 1975; Attwood and Graneri, 1991). Support for this proposed mechanism was provided by the isolation of a putative enzyme-bound carboxyphosphate intermediate (Phillips *et al.*, 1992). In the presence of acetyl-CoA, the carboxyl group is subsequently transferred to biotin to form carboxybiotin which is the product of the first partial reaction (Phillips *et al.*, 1992; Attwood, 1993). Further evidence favouring the existence of a carboxyphosphate intermediate is to be found in the similarity of the primary structure of biotin carboxylase to that region of carbamoylphosphate synthetase (CPS) shown by site-directed mutagenesis to be involved in the reaction mechanism for the synthesis of carboxyphosphate (Stapleton *et al.*, 1996; Javid-Majd *et al.*, 1996).

PC from most species is allosterically activated by acetyl-CoA. PC isolated from chicken is absolutely dependent on acetyl-CoA (Utter and Keech, 1963) while PC isolated from rat liver (McClure *et al.*, 1971), sheep (Ashman *et al.*, 1972) and *Bacillus stearothermophilus* (Libor *et al.*, 1978) is highly dependent on acetyl-CoA as well. In contrast, yeast PC is less dependent on acetyl-CoA (Cazzulo and Stoppani, 1968). Unlike other sources, PCs from *Pseudomonas citronellolis* (Scubert and Remberger 1961), *Aspergillus niger* (Bloom and Johnson, 1962) and *M. thermoautotrophicum* (Mukhopadhyay *et al.*, 1998) are acetyl-CoA independent.

PCs from many sources possesse a reactive lysine residue whose integrity is essential for full enzymatic activity. Modification of the enzyme with amino group-selective reagents causes the preferential loss of the catalytic activity that is stimulated by the allosteric activator by acetyl-CoA (Scrutton and White 1973; Ashman *et al.*, 1973). The loss of acetyl-CoA-dependent activity is due to the modification of a single lysine per active site (Scrutton and White, 1973). Acetyl-CoA has also been shown to protect the loss of enzyme activity against this modification (Scrutton and White 1973; Ashman *et al.*, 1973; Chapman-smith *et al.*, 1991). This essential lysine has been suggested to form part of the acetyl-CoA binding

16

site rather than being protected by a conformational change following activator binding (Chapman-Smith *et al.*, 1991). Sequence comparison of yeast PC to other biotin-dependent enzymes which bind acetyl-CoA, i.e. ACC and the β-subunit of PCC or propionyl-CoA, i.e. 12S subunit of transcarboxylase (TC), as well as with other acyl-CoA binding enzymes have so far been unable to identify any putative acetyl-CoA binding site on PC. This may reflect the fact that ACC and PCC bind acetyl-CoA as a substrate whereas PC binds acetyl-CoA as an allosteric ligand (Lim *et al.*, 1988).

A number of ionisable groups of the amino acid side chains of the enzyme were proposed to form part of an active site of the enzyme to enolise biotin (Attwood and Cleland, 1986; Tipton and Cleland, 1988). Modification of the side-chains of lysine ($-NH_3$ group) and cysteine (-SH group) by *o*-phthalaldehyde results in an inactivation of the first partial reaction (Werneberg and Ash, 1993) supporting their crucial role as an ion-pair (Jitrapakdee *et al.*, 1996).

In addition to its role in complexing with ATP, Mg^{2+} is an essential cofactor of the reaction as revealed by kinetic studies of PC (Keech and Barritt, 1967; McClure *et al.*, 1971; Bais and Keech, 1972; Attwood and Granneri, 1992). Evidence in favour of there being two extrinsic divalent cation binding sites, in addition to the intrinsic divalent cation (see below) was obtained using electron paramagnetic resonance (EPR) spectroscopy to show that extrinsic Mn^{2+} binds to PC in the presence of CrATP, a potent competitive inhibitor of MgATP (Reed and Scrutton, 1974). Recently, EPR has again been used to demonstrate that two equivalents of the divalent oxyvanadyl cation, VO^{2+}, bind at the first subsite of PC: one is involved in nucleotide substrate binding, while the other interacts strongly with bicarbonate (Werneberg and Ash, 1997). These authors have suggested the roles of this second extrinsic cation could include orientation of the bicarbonate for attack on the γ-phosphoryl group of ATP as well as minimising charge repulsion between these anionic substrate species.

Kinetic studies (Ashman *et al.*, 1972; Barden and Scrutton, 1974) have previously shown that certain monovalent cations (K^+, NH_4^+, Rb^+, Cs^+ and Tl^+) are effective (3-7 fold) activators of PC, with an apparent equilibrium-ordered binding interaction with HCO_3^- which binds first. Recently, direct evidence supporting this relationship was obtained when

binding constants for K^+ and Tl^+, measured by their quenching of intrinsic protein fluorescence, were shown to agree well with the activator constants, and HCO_3^- was shown to enhance the affinity of chicken PC for Tl^+ by 2-fold (Werneburg and Ash, 1997). Together these data suggest that the monovalent cation binds in the vicinity of bicarbonate in the first subsite. However, it was concluded from a study of the superhyperfine coupling between the electron spin of VO^{2-} and the nuclear spin of Tl^+ that this activating monovalent cation is unlikely to share a ligand with either the enzymic or nucleotide VO^{2+} cation (Werneburg and Ash, 1997).

The second partial reaction involves the transfer of the carboxy group from carboxybiotin to pyruvate to form oxaloacetate. It was proposed that binding of pyruvate induced the carboxybiotin to move into the second subsite where the carboxybiotin is destabilised (Easterbrook-Smith *et al.*, 1976). Goodall *et al.* (1981) confirmed this proposal and also showed that a number of pyruvate analogues can induce the translocation of carboxybiotin to the second subsite. This process requires the removal of a proton from pyruvate, carboxyl group transfer and reprotonation of biotin. There is evidence of an involvement of a cysteine-lysine ion pair in the second subsite (Attwood *et al.*, 1986a). Werneberg and Ash (1993) reported the presence of a second cysteine-lysine pair, as revealed by the modification of such an ion pair by *o*-phthalaldehyde resulting in the loss of the second partial reaction. Taken together, Attwood (1995) proposed the detail of the second partial reaction by which the cysteine-lysine pair stabilises the enol form of biotin and participates in the proton transfer between biotin and pyruvate via the -SH group of cysteine.

1.9   STRUCTURE OF PYRUVATE CARBOXYLASE

1.9.1 Primary and domain structure

In the early period of studies, many attempts have been made to obtain primary structure of PC by protein sequencing. The biotin attachment site of PC from avian and sheep was first reported to be AMKM with biotin covalently attached to the ε-amino group of lysine via an amide bond (Rylatt *et al.*, 1977). Similar results were reported at about the same time for biotinyl subunit of transcarboxylase and *E. coli* ACC (Wood and Barden 1977; Sutton *et al.*, 1977). Clearly members of the family of biotin carboxylases were too

large to determine their primary structure efficiently and precisely by protein sequencing. With the development of recombinant DNA technology, more sequences for PC and other biotin carboxylases were obtained by inferring their primary structures from cDNA and genomic clones. Freytag and Collier (1984) first determined a partial sequence of human PC from cDNA cloning. Several years later Lamhonwah *et al.* (1987) determined the C-terminal 83 amino acid residues of human PC and were able to show that there is a significant homology around the biotin-attachment site of human PC with PCC, ACC and transcarboxylase. Yeast (*S. cerevisiae*) PC was the first biotin carboxylase whose complete amino acid sequence was determined (Lim *et al.*, 1988). These authors also showed that there is a significant sequence homology of yeast PC to different regions of other biotin carboxylases as well as to parts of carbamyl phosphate synthetase and lipoamide transferases, which share some common substrate binding or functional properties. Taken together with limited proteolysis studies, these data led to the identification of three functional domains i.e. ATP/HCO$_3$binding domain, pyruvate-binding domain and the biotinyl domain (Lim *et al.*, 1988). Some years later, primary structures of PC from different organisms including bacteria (Dunn *et al.*, 1996; Kondo *et al.*, 1997; Kurnst *et al.*, 1997; Mukhopadhyay *et al.*, 1998; Koffas *et al.*, 1998), other yeast (*Pichia pastoris*) (Menendez *et al.*, 1998) mosquito (Tu and Hagedorn, 1997) mouse (Zhang *et al.*, 1993), rat (Jitrapakdee *et al.*, 1996; Chapter 3, Lehn *et al.*, 1995), human (MacKay *et al.*, 1994; Wexler *et al.*, 1994; Walker *et al.*, 1995b), have also been shown to contain the same three functional domains as yeast PC.

1.9.2 Three-dimensional structure

Native PCs from a number of sources including bacteria, yeast, insect and mammals consist of four identical subunits (α_4) of approximately 120-130 kDa (Wallace and Easterbrook-Smith, 1985) except for *Pseudomonas citronellolis* (Goss *et al.*, 1981), *Azobacter vinelandii* (Scrutton and Taylor, 1974) and *Methanobacterium thermoautotrophicum* (Mukhopadhyay *et al.*, 1998) in which each protomer consists of 2 polypeptide chains of 75 kDa (α) and 52 kDa (β) subunits arranged as an ($\alpha\beta$)$_4$ structure. Although the primary structure of PC from a number of organisms have been reported in recent years, a high resolution three-dimensional structure of this enzyme has yet to be

established. This is most likely due to the failure to obtain a sufficient amount of homogeneous enzyme to be crystallised and studied by X-ray diffraction. The very large size of PC precludes its structure being determined by NMR. The quaternary structure of PC has so far been obtained by electron microscopic studies. Electron microscopy has revealed that the PC from chicken, rat and sheep are indistinguishable tetrahedron-like structures, composed of two pairs of subunits in different planes orthogonal to each other (Fig. 1.5A). The opposite pairs contact each other on their convex surfaces with a midline cleft separating two distinct domains and running along the longitudinal axis of each monomer. Since this midline cleft becomes less visible in the presence of acetyl-CoA, it has been suggested that this cleft area might be the active site of the enzyme (Mayer *et al.*, 1980). Using avidin as a structural probe, Johannssen *et al.* (1983) localised the biotin moieties in the midline cleft on the external surface of each subunit, within approximately 3 nm of the

A   B
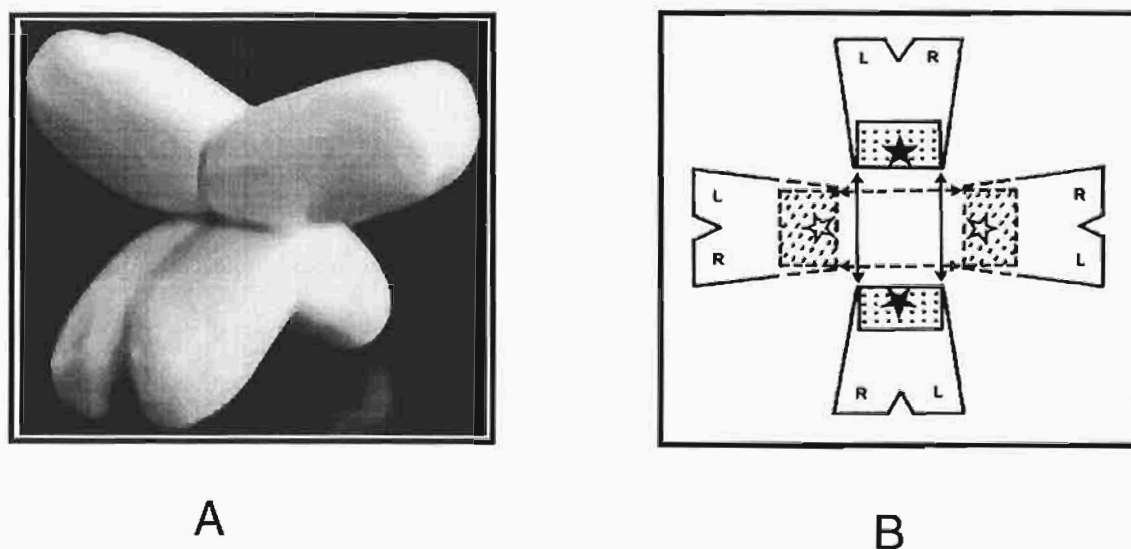
Figure 1.5   Quaternary structure of vertebrate PC derived from electron microscopic studies. A, PC consists of four identical subunits arranged as tetrahedron-like structure with a midline cleft separating two distinct domains and running along the longitudinal axis of each monomer [reproduced from Mayer *et al.* (1980)]. B, Exploded face-view of the enzyme tetramer with an indication of the bound avidin molecule (shaded) and the sites of the biotin-binding areas indicated by ★ for the avidin bound to upper pair of PC subunit, and by ☆ on the avidin bound to the lower pair of PC subunits [reproduced from Johannssen *et al.* (1983)].

intersubunit junction (Fig. 1.5B). PC from *Aspergillus nidulans* (Osmani *et al.*, 1984), *S. cerevisiae* (Rohde *et al.*, 1986) and even *Pseudomonas citronellolis* (Fuchs *et al.*, 1988) in which PC is arranged as an $(\alpha\beta)_4$ configuration, all appeared to have a tetrahedron-like structure with a midline cleft similar to that in the vertebrate enzyme.

Dilution of sheep (Ashman *et al.* 1972; Khew-Goodall *et al.*, 1991) and chicken (Attwood *et al.*, 1993) PC resulted in inactivation of PC activity accompanied by dissociationof the active tetramers into inactive dimers and monomers as revealed by electron microscopic and high resolution gel filtration studies. Acetyl-CoA has also been shown to prevent both the dissociation of the tetrameric enzyme and the associated loss of activity. Addition of acetyl-CoA to partially dilution-inactivated enzyme prevented further loss of enzymatic activity and of tetrameric structure (Khew-Goodall *et al.*, 1991; Attwood *et al.*, 1993). This ligand was similarly effective in preventing the cold-induced loss of both activity and tetrameric structure of chicken PC (Irias *et al.*, 1969). Apart from stabilising the quaternary structure of PC, addition of acetyl-CoA was also shown to cause conformational changes in PC, as revealed by spectrophotometric (Frey and Utter, 1977), ultracentrifugal (Taylor *et al.*, 1978) and electron microscopic (Attwood *et al.*, 1986b) studies.

1.10 PHYSIOLOGICAL STATES THAT ALTER PC EXPRESSION

PC is one of a number of important metabolic enzymes whose expression is regulated in a differential manner between particular tissues to achieve an appropriate overall response to various physiological and pathological stimuli. Long term regulation involves changes in the total amount of PC through alterations in the rate of enzyme synthesis in liver, kidney and adipose tissue (Barritt, 1985). Different physiological conditions have been shown to alter the level of PC expression: these include nutritional alterations, diabetes, hormonal changes, neonatal development, adipogenesis and lactation.

1.10.1 Nutrition and xenobiotics

Fasting in rats has been shown to induce two- to three-fold increases in hepatic PC activity (Wimhurst and Manchester, 1970). Similar results have also been reported in other animals e.g. cow (Ballard *et al.*, 1968), guinea pig (Söling *et al.*, 1970) and sheep (Taylor *et al.*, 1971; Lemons *et al.*, 1986). Little information is available on the

effects of refeeding starved rats. However, a small increase in total PC activity has been detected in kidney. The hormonal mechanisms which regulate the total amount of PC activity during fasting and refeeding still remain unclear. Increases in the level of PC activity during the starvation have been correlated with increases in the plasma concentration of glucagon and glucocorticoid (Seitz *et al.*, 1976). PC activity has also been shown to be reduced by 50% in the diabetes-prone BHE/cdb rats fed a diet containing 6% menhaden oil, rich in long-chain highly unsaturated fatty acids (Wickwire and Berdanier, 1997).

A recent study in rats has also shown that in both carbon tetrachloride administration-induced liver degeneration and in alloxan-induced diabetic there is an increase in PC protein and its activity (Salto *et al.*, 1996). Chronic administration to rats of lipoic acid, a chemical that has structural similarity to biotin, lowered the activities of biotin-dependent carboxylases including PC and MCC to 28-36% of those of control animals. The decrease in these carboxylases was thought to be due to the competition for biotin transport into cells by binding to the biotin transporters in the cell membrane or displacing biotin from holocarboxylase synthetase, and was reversed by dietary biotin supplementation (Zempleni *et al.*, 1997).

Cadmium has long been known to increase the activities of gluconeogenic enzymes including hepatic and renal PC in rats (Chapatwala *et al.*, 1980; Chapatwa *et al.*, 1982). A recent study demonstrated that PC transcript was up-regulated 2-4 fold in response to cadmium treatment *Caenorhabditis elegans*. The mechanism by which cadmium induces PC expression has not yet been elucidated, although it has been suggested that cadmium could act via either a cAMP- or a calcium-mediated pathway (Liao and Freedman, 1998).

1.10.2 Diabetes

The rate of hepatic gluconeogenesis is increased dramatically in the diabetic state, concomitant with an increase in the activities of other key gluconeogenic enzymes i.e. PEPCK, fructose-1,6-bisphosphate, glucose-6-phosphatase (Filsell *et al.*, 1969; Wimhurst and Manchester, 1970) and PC (Weinberg and Utter, 1979). In streptozotocin-induced diabetic rats, the hepatic PC activity was increased 2-fold over that of control rats. This increase in enzymatic activity which resulted from an increased amount of protein due to an enhanced rate of synthesis, is thought to be mediated by a high plasma glucagon to insulin

ratio (Weinberg and Utter, 1979). Administration of insulin to diabetic rats brought the amount of PC and its activity back to the control values.

Apart from being both a gluconeogenic and a lipogenic enzyme, PC also plays an important role in glucose-induced insulin secretion in pancreatic islets as described earlier. In the Goto-Kakizaki (GK) rat, a genetic model of non-insulin dependent diabetic mellitus (NIDDM) in which glucose-induced insulin secretion in pancreatic β-cells is impaired, it has been found that PC activity was 45% of that in the normal rat islets due to a decrease in the amount of PC protein. However, administration of insulin to GK rats resulted in the recovery of PC activity to that of normal rat (MacDonald *et al.*, 1996). A low level of pancreatic PC activity was also concomitant with a decrease in glucose transporter, GLUT2 (Ohneda *et al.*, 1993) and mitochondrial glycerol phosphate dehydrogenase, another enzyme believed to play a role in glucose-induced insulin secretion (MacDonald *et al.*, 1996). Down-regulation of these three proteins in the pancreatic β-cell in NIDDM is proposed to be an adaptive response by the cell to protect itself from a high glucose concentration by modulating glucose metabolism (MacDonald *et al.*, 1996).

1.10.3 Hormonal alterations

It has long been known that thyroid hormone affects the hepatic gluconeogenic rate in rats by increasing the activity of gluconeogenic enzymes, including PC (Böttger *et al.*, 1970). Experiments carried out by Weinberg and Utter (1979) showed that hepatic PC activity was increased 2-fold in hyperthyroid rats, whereas in hypothyroid rats PC decreased 2-fold. Inhibition of *de novo* protein synthesis with actinomycin D reduced PC activity in thyroxin-treated thyroidectomized rat suggesting that thyroid hormone increases the rate of PC synthesis (Böttger, 1970). The mechanism of action of thyroid hormone on PC expression remains to be defined at the molecular level.

Glucocorticoids have been shown to acutely stimulate gluconeogenesis (Friedmann *et al.*, 1967; Rinard *et al.*, 1969) and result in an increased glucose output by rat hepatocytes (Jones *et al.*, 1993). It was suggested that glucocorticoids would induce the gluconeogenic enzymes, PC and PEPCK. Short term treatment of rats with dexamethasone, an analogue of glucocorticoids caused an increase in PC activity. It has been suggested (Krause-Friedmann and Feng, 1996) that glucocorticoids act in relieving the restraint on PC by altering the

substrate supply and the intramitochondrial concentrations of effectors (Martin *et al.*, 1984) perhaps via a Ca^{2+} influx-mediated mechanism. It has also been found that glucocorticoids are not necessary to maintain basal metabolic gluconeogenic rates in adrenalectomized rats (Martin *et al.*, 1984; Ciprés *et al.*, 1994).

Glucagon has long been demonstrated to increase the rate of pyruvate carboxylation in mitochondria isolated from rat hepatocytes without changing the level of PC (Adam and Haynes, 1969; Garrison and Haynes, 1975). The effect of glucagon could be detected within 6 minutes and reached a maximum within 10 minutes after the liver or hepatocytes were exposed to hormone (Garrison and Haynes, 1975). The precise mechanism by which glucagon acts on PC is not well understood. Initially it was suggested that glucagon causes an increase in the transmembrane pH gradient, which in turn stimulates the rate of pyruvate transport into the mitochondria (Halestrap, 1978; Thomas and Halestrap, 1981). Subsequent experiments using a more potent inhibitor of pyruvate transporter, have shown that in fact glucagon did not exert its effect by this mechanism but rather it stimulates the respiratory chain leading to an activation of pyruvate carboxylation (Halestrap and Armston, 1984). Further evidence to support the later hypothesis were obtained from the studies using an inhibitor of the respiratory chain in mitochondria isolated from rat hepatocytes. This led to the conclusion that glucagon stimulates respiratory-chain activity via a Ca^{2+} influx mechanism (McCormack *et al.*, 1990). The increase in respiratory chain activity (O_2 uptake) stimulates gluconeogenesis by generating ATP and by providing reducing equivalents to the cytosol. The increase in O_2 uptake therefore indirectly stimulates pyruvate uptake into the mitochondria (Pryor *et al.*, 1987; Owen and Halestrap, 1993). The mechanism(s) by which glucagon affects pyruvate metabolism have recently been reviewed (Krause-Friedmann and Feng, 1996).

Adrenalin has also been known to stimulate pyruvate carboxylation by isolated liver mitochondria (Garrison and Borland, 1979). Little is known about the mechanism of how adrenalin acts on pyruvate metabolism although it has been shown that adrenalin also acts via Ca^{2+} mediated pathway as those of glucagon.

1.10.4 Postnatal gluconeogenesis

As the maternal circulation provides glucose for the developing fetus, gluconeogenesis does not occur in fetal liver but is triggered rapidly soon after birth (Ballard and Oliver, 1965). The increase in PC activity is accompanied by increases in the activities of other gluconeogenic enzymes including phosphoenolpyruvate carboxykinase, fructose-1,6-bisphosphatase and glucose-6-phosphatase suggesting the gluconeogenic pathway begins to function (Ballard and Hanson, 1967; Yeung et al., 1967). In rats, PC activity is highest in day 7 after birth (suckling period) and begins to decline in the weaned rat to the adult level (Yeung et al., 1967). Alteration of plasma glucagon and insulin during weaning period has been proposed to involve up-regulation of hepatic PC expression (Girard et al., 1973).

1.10.5 Genetic obesity

In Zucker fatty rats (fa/fa), PC expression has been shown to be elevated 2-5 fold upon the onset of obesity (Lynch et al., 1992). This increase in PC level is also accompanied by an increase in the levels of other lipogenic enzymes including ACC, fatty acid synthetase and ATP-citrate lyase. Given the lipogenic role of PC, as mentioned above, it has been proposed that oxaloacetate produced by PC is drastically consumed in obesity, thus contributing to the hypertrophy of adipose tissue during the development of obesity (Lynch et al., 1992).

It has also been shown that during the *in vitro* differentiation of mouse 3T3-L1 preadipocytes to mature adipocytes, this conversion is accompanied by increases in lipogenic enzymes including fatty acid synthetase (Ahmad et al., 1979; Student et al., 1980) and PC (Mackall et al., 1977; Freytag and Utter, 1980; Zhang et al., 1995). The increase in PC activity is concomitant with an increase in the rate of enzyme synthesis (Freytag and Utter, 1980; Freytag and Utter, 1983; Zhang et al., 1995) and its mRNA (Angus and Lane, 1981; Freytag and Collier, 1984; Zhang et al., 1995). The turnover of PC in differentiated cells was similar to that of undifferentiated cells with the enzyme having a half-life of 28-35 hours (Freytag and Utter, 1983). The induction of PC in this cell line is consistent with its role in lipogenesis. Recently, it has been shown that cAMP down-regulated the PC mRNA and PC activity by decreasing the transcription rate of the PC gene and/or stability of mRNA (Zhang

et al., 1995). Despite these decreases in its mRNA and enzyme activity, the level of PC protein was not affected. The inactivation of PC activity did not involve the loss of biotin from the holoenzyme but it was suggested that it involved the loss of the active tetrameric form of the enzyme. The mechanism by which this is effected has not been elucidated. Thus cAMP not only exerts its effects by inactivating the protein but also affects the transcription rate of PC or the stability of its mRNA (Zhang *et al.*, 1995).

1.11 PC DEFICIENCY

Given the diverse functions of PC described above, it is apparent that PC plays very significant role in metabolism. This conclusion is supported by the effects of PC deficiency whether it occurs in yeast or human. In *S. cerevisiae*, a number of mutants have been reported and shown to affect growth phenotype (Walker *et al.*, 1991; Walker and Wallace, 1991; Brewster *et al.*, 1994). Furthermore, defects in both PC1 and PC2 genes resulted in a failure to grow on glucose minimum media (Stucka *et al.*, 1991; Brewster *et al.*, 1994). The same effect has been reported when the single PC gene locus was disrupted in yeast *Pichia pastoris* (Menéndez *et al.*, 1998).

In humans, PC deficiency is an autosomal, recessively-inherited disease. Patients who suffer from the disease have less than 5% of normal PC activity when assayed in skin-fibroblast cultures (Atkin *et al.*, 1979). The main clinical features associated with this deficiency are congenital lactic acidosis (Atkin *et al.*, 1979) and deterioration of the central nervous system (Robinson, 1982). Lactic acidosis is associated with the decline in both gluconeogenesis and citric acid cycle activity leading to an accumulation of alanine, lactate and pyruvate and a decrease in oxaloacetate and glucose (Robinson *et al.*, 1983). Two groups of patients have been reported. The first group of patients suffers from a mild to moderate lactic acidaemia, delayed development and psychomotor retardation and may survive many years. This group of patients have some residual immunoreactive PC and its mRNA detected by Northern blot analysis (known as CRM^{+ve} phenotype, patients exhibit cross reactive material with anti-PC antibodies) (Robinson *et al.*, 1987). The second group of patients represent a more serious disease, with a severe lactic acidaemia accompanied by hyperammonaemia, citrullinaemia and hyperlysinaemia, and rarely survive longer than 3

months after birth. In contrast to the first group, these patients lack immunoreactive PC and its mRNA (CRM^{-ve} phenotype) (Robinson *et al.*, 1987; Robinson *et al.*, 1996). The two forms of the disease have distinct ethnic groups in which they occur (Robinson *et al.*, 1987; Robinson *et al.*, 1996). The CRM^{+ve} phenotype has been reported among North American native peoples whereas the CRM^{-ve} pheotype has been reported in United Kingdom and France (Robinson *et al.*, 1987).

To date four patients carrying different single point mutations including substitutions of Val 145 -> Ala, Arg 451 -> Cys, Ala 610 -> Thr and Met 743 -> Ile in both alleles respectively have been shown to be responsible for some forms of the disease. The first two mutations were found within the exons encoding the biotin carboxylation domain (Wexler *et al.*, 1998) whereas the last two mutations (Carbone *et al.*, 1998) were identified within the exons encoding the transcarboxylation domain of the enzyme (see Fig. 3). The first case (Val145Ala) showed barely detectable levels of immunoreactive PC and its activity, suggesting this mutation affects protein stability (Wexler *et al.*, 1998). In contrast, the other cases contained a normal level of immunoreactive PC but lower PC activity suggesting these mutations affect the catalytic activity of the enzyme (Wexler *et al.*, 1998; Carbone *et al.*, 1998). The carriers who contain heterozygous alleles of these mutations (Val 145 -> Ala, Arg 451 -> Cys) are able to survive but PC activity detected in skin fibroblasts was about 50% of normal (Wexler *et al.*, 1998). However the genotype(s) of the CRM^{-ve} phenotype patients have not been identified. It has been suggested that the mutation responsible for this form of disease may be splicing mutation which results in the absence of PC mRNA (Carbone *et al.*, 1998). The severity of PC deficiency may also be influenced by environmental factors such as stress and fasting (Robinson *et al.*, 1996). Another group of patients, who also show PC deficiency, are those suffering multiple carboxylase deficiency due to a defect in the biotin metabolism. This group of patients shows elevated levels of organic acids, which are metabolites of acetyl-CoA, propionyl-CoA and 3-methylcrotonyl-CoA, as well as lacticacidemia.

1.12 GENE STRUCTURE OF PC

A gene encoding the yeast PC1 isoenzyme was first cloned by Lim *et al.* (1988). Walker *et al.* (1991) and Stucka *et al.* (1991) independently discovered that in fact yeast contains two genes encoding two isoenzymes (PC1 and PC2). The PC 1 gene is located on chromosome VII while the PC 2 gene is located on chromosome II (Walker *et al.*, 1991; Stucka *et al.*, 1991). Neither PC1 nor PC2 gene contains an intron. Walker *et al.* (1991) found that disruption of the PC1 gene, reduced PC activity of DBY 746 yeast to 10-20%. In contrast, Stucka *et al.* (1991) found that disruption of either PC1 or PC2 gene in the W303 strain retained 50% of total PC activity. However, disruption of both genes resulted in complete loss of enzyme activity (Stucka *et al.*, 1991; Brewster *et al.*, 1994). It was found later that there is a polymorphism of the PC2 gene in these different yeast strains used by both groups of investigators, as indicated by amino acid differences in the PC2 gene (Val *et al.*, 1995). The most significant difference is a single base substitution near the 3'-end of the gene which altered the reading frame encoding the biotin domain of the enzyme. This C-terminal variant has been shown to affect *in vitro* biotinylation of the enzyme (Val *et al.*, 1995).

In mosquito, there is evidence for the presence of two PC isoforms of similar size i.e. 133 and 128 kDa. These two isoenzymes exhibit tissue-specific expression (Tu and Hagedorn, 1997). However, it is uncertain whether these PC isoforms are the products of two separate genes or a single gene with allelic polymorphism in the genome (Tu and Hagedorn, 1997).

Unlike yeast, only one gene has been identified in humans and has been mapped to the long arm of chromosome 11 by the somatic cell hybrid technique (Freytag and Collier, 1984). Recently, the human PC gene has been sublocalised to chromosome 11q13.4 by fluorescence *in situ* hybridisation (Walker *et al.*, 1995a). However, the gene structures for PC in higher organisms hade not been reported except for the rat gene whose partial sequence had been determined (Cassady, 1987; Booker, 1990). Chapters 4 and 5 present the complete gene structure of rat PC, and represents the first report for a mammalian biotin-dependent enzyme. The human PC gene structure had recently been reported (Carbone *et al.*, 1998) follow that of the rat PC gene (Jitrapakdee *et al.*, 1997). The genomic

organisation of the PC genes in these two mammalian species is discussed in Chapter 5 of this thesis.

1.13 TRANSCRIPTIONAL REGULATION OF YEAST PC1 AND PC2 GENES

Despite the central importance of PC in intermediary metabolism, little information is known about the transcriptional regulation of PC in eukaryotes. This is mainly due to a lack of the gene structure, particularly at the 5'-end which contains regulatory elements. In yeast, although PC1 and PC2 isozymes exhibit high sequence similarity at both amino acid level and nucleotide level (Stucka *et al.*, 1991; Val *et al.*, 1995), their 5'-non coding regions are markedly different, suggesting that both genes are regulated differently (Brewster *et al.*, 1994). Two copies of TATA boxes located at positions -117 and -110 relative to the initiation codon were identified in the PC1 promoter (Lim *et al.*, 1988) but only the downstream TATA box is functional and responsible for transcribing PC1 mRNA with distinct transcription initiation sites (Brewster, 1994). The basal promoter of PC1 is located within the first 149 bp of the 5'-non coding region which consisted of a TATA element and a UAS1 transcription factor binding site (Brewster, 1994). In contrast, Menéndez and Gancedo (1998) have localised the basal promoter of PC1 within the first 330 bp of the 5'-non coding region which consisted of a TATA element and an UAS1 transcription factor binding site whereas the basal promoter of PC2 is located within the first 291 bp of the 5'-non coding region.

PC1 and PC2 appear to carry out different metabolic functions. PC1 expression has been shown to be relatively constant throughout the main growth phase during growth on glucose minimal media, while the PC2 expression is characterised by a high level of transcript production in the early growth phase. Both genes are repressed throughout the latter stages. During growth on ethanol minimal media, PC1 and PC 2 expression exhibit a similar pattern i.e. decline from early to mid log phase. However, the level of PC1 expression is 10-fold above the level of PC2 during this fermentative growth (Brewster *et al.* 1994). Further studies with either PC1 or PC2 null mutants also indicated that the lack of either PC gene had little effect on the level and pattern of expression of the other PC gene,

suggesting that both genes are regulated differently. However, the PC1 null mutant or double null mutant of DBY746 showed a strong requirement for L-aspartate in ethanol minimal media (Brewster *et al.*, 1994). This strongly suggested that the PC1 isoenzyme plays a crucial role in maintaning growth on ethanol media and more specifically for the establishment of glucose-dependent growth on glucose minimal media. In contrast, the role of PC2 isoenzyme remains unclear but it is believed to support growth on a glycolytic carbon source (Brewster *et al.*, 1994). Different carbon sources have also been shown to affect PC1 and PC2 expression differently (Brewster *et al.*, 1994; Menéndez and Gancedo, 1998).

1.14 ALTERNATIVE SPLICING IN THE CONTROL OF GENE EXPRESSION

Eukaryotic pre-mRNAs undergo a series of nuclear processing events. In addition to 5' capping with 7-methyl guanosine (Salditt-Georgieff *et al.*, 1980) and the attachment of a poly(A) tail to the 3' end of nascent transcripts (Birnstiel *et al.*, 1985), the intron sequences must be removed precisely and the exons ligated correctly to avoid disrupting the open reading frame. A process in which all exons in the genes are incorporated into mature mRNA through the invariant ligation of individual exons is know as 'constitutive splicing'. This type of splicing yields a single gene product from each transcriptional unit even when its coding sequence is split into many exons. However, there are many genes for which individual exons on the transcriptional unit are ligated alternatively to produce different forms of mature transcripts. This alternative splicing can exclude individual exon sequences from the mature mRNA in some transcripts but include them in others (Breitbart *et al.*, 1987) [see Fig. 1.6]. Alternative splicing in most cases generates protein isoforms sharing extensive regions of identity and differing only in some parts of the proteins, thus allowing for the fine modulation of protein function. Alternative splicing occurring in the coding regions of the proteins can affect almost all aspects of the protein function including localisation, or modulation of protein function (Smith *et al.*, 1989). However, some genes produce different mature mRNAs which diverge only at the 5'-untranslated region resulting from an alternative splicing of the 5'-noncoding exons. The 5'-untranslated regions are known to be
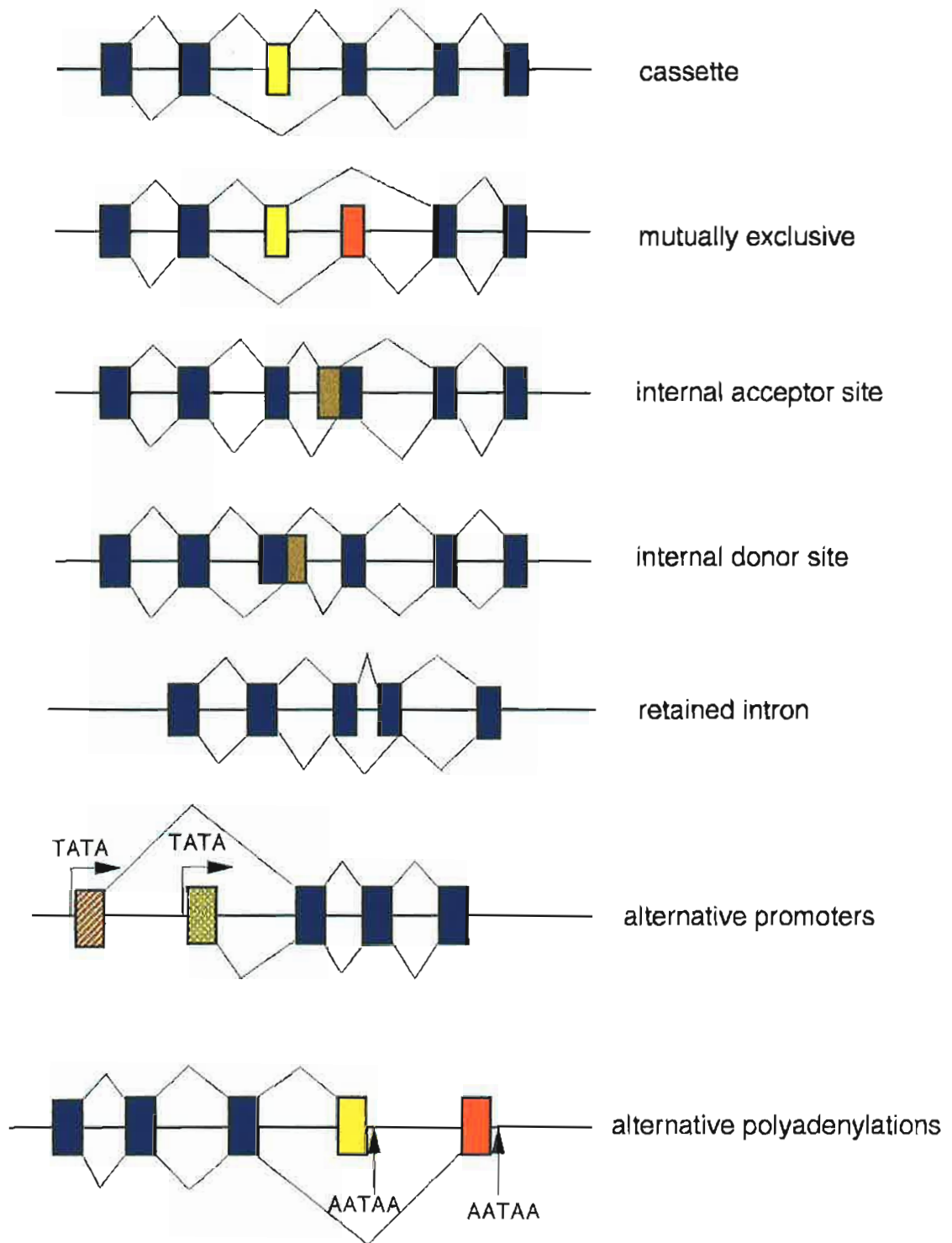
Figure 1.6 Modes of alternative splicing (Smith *et al.*, 1989). Exons are shown in boxes. Splicing pathways are shown by the diagonal lines. Promoters and poly (A) sites are denoted TATA and AATAA, respectively.

involved in regulating translational efficiency (Kozak 1987) and RNA stability (Brawerman, 1987; Raghow 1987). Different modes of alternative splicing are shown in Fig. 1.6 below.

As a mechanism to regulate gene expression and generate protein diversity, alternative splicing has advantages over gene rearrangement and extensive multigene families. It allows for switches in protein isoforms or expression pattern without changes in the cell's genetic content (Smith *et al.*, 1989).

1.15 PROMOTER SWITCHING IN BIOTIN CARBOXYLASE GENE

So far there is only one biotin carboxylase gene i.e. rat acetyl-CoA carboxylase (ACC) for which regulation at the gene level has been extensively studied. ACC catalyses the rate-limiting step in the biosynthesis of long chain fatty acids. Long term control mechanisms regulate cellular levels of ACC by modulating expression of the ACC gene. Studies on the molecular basis of long-term regulation of rat ACC have revealed that ACC is expressed from two promoters of a single-copy gene, whose transcripts are differentially spliced to produce mature mRNA with 5'-end heterogeneity (Luo *et al.*, 1989). Rat ACC mRNA can be grouped into two classes. Class I mRNA are transcriptional products of the promoter I which is inducible and they contain exon 1 as their leading exon. Class II mRNA are transcribed from the constitutively expressed promoter II, and their leading exon is exon 2.

At least five species of ACC mRNA (Lopez-Cassillas *et al.*, 1989; Lopez-Cassillas and Kim, 1989) have been characterised thus far. All five major species contain exon 5, which has the initiation codon AUG, but differ in their 5' untranslated regions resulting from different combinations of exons 1 through 4 (Luo *et al.*, 1989). In spite of this extensive heterogeneity in their 5' untranslated region, both class I and class II ACC mRNA appear to have the same open reading frame which consists of 7035 nucleotides. The starting AUG codon of the open reading frame in exon 5 is common to all the species of ACC mRNA (Lopez-Casillas *et al.*, 1989; Lopez-Cassillas and Kim, 1989) (Fig. 1.7).

Studies that have employed chimeric constructs by placing a reporter gene under the control of either one of the ACC gene promoters have revealed that ACC gene promoter diversity permits a very precise and selective regulation of ACC gene expression at the level

31
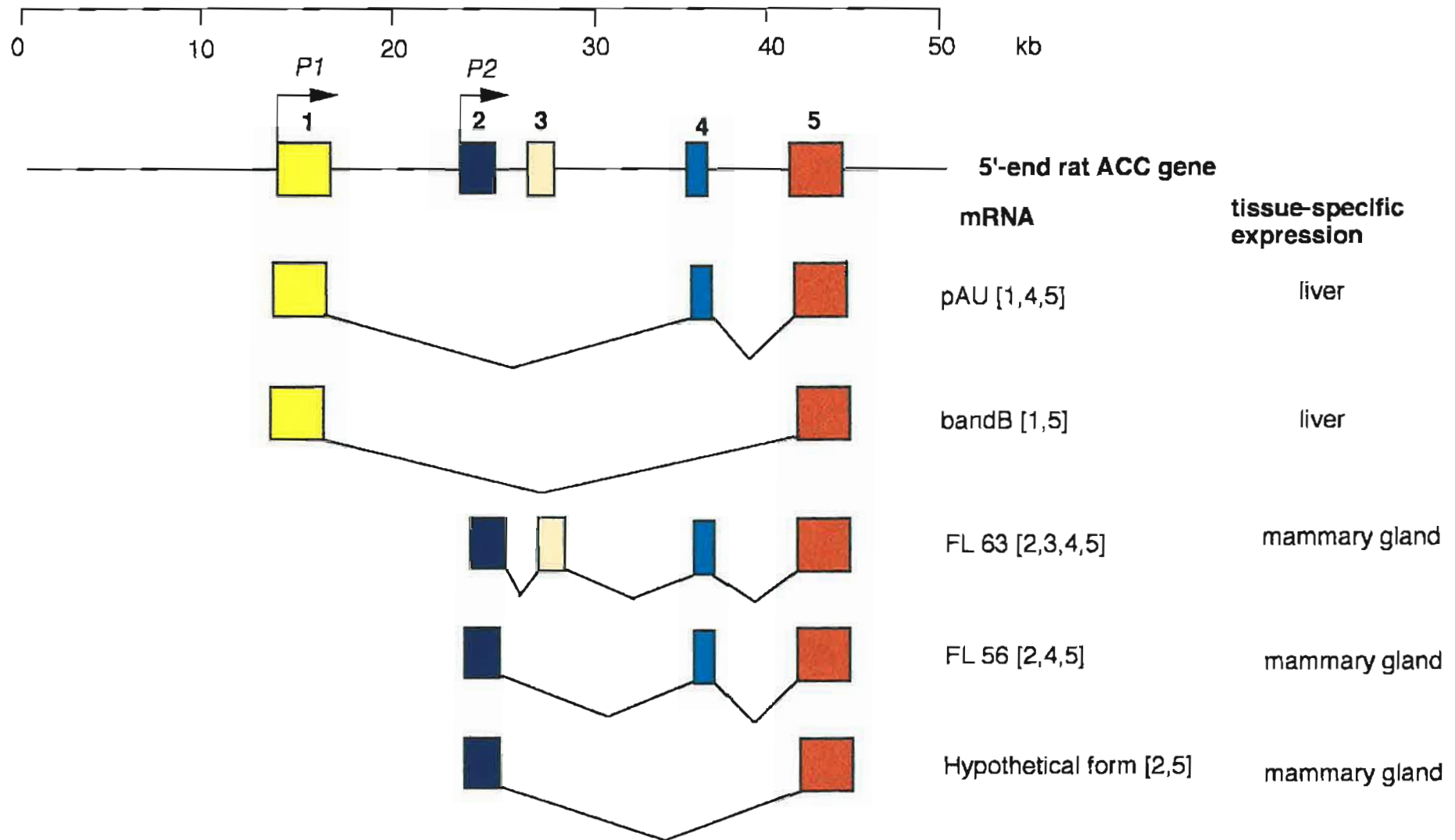
Figure 1.7 Organisation of the 5'-end of rat ACC gene and the alternatively spliced transcripts. Exons are shown in boxes and numbered. Alternate promoters are indicated as P1 and P2. Different ACC mRNA isoforms and their tissue-specific expression are also shown.

of transcription (Luo and Kim 1990). In addition, the *in vivo* activities of two classes of ACC mRNA are also highly regulated in a tissue-specific manner (Lopez-Casillas *et al.*, 1991). It was also found that this 5' untranslated region heterogeneity in the ACC mRNA affected the translation efficiency (Lopez-Casillas and Kim, 1991). No evidence of alternate promoters is found in the chicken ACC gene (El Khadir-Mounier *et al.*, 1996). In human, two ACC genes encoding 2 isozymes have been reported (Ha *et al.*, 1994; Ha *et al.*, 1996; Abu-Elheiga *et al.*, 1995).

1.16 PROJECT RATIONALE

The main objective in our laboratory over the years has focused on understanding the structure and function of PC including the determination of the primary structure of PC from different organisms including yeast, chicken, rat and human. Information of the primary structures of PC obtained from different organisms would be invaluable for identification of structurally important residues which might play a common role in the catalytic reaction. At the time when this project commenced, partial cDNA clones encoding human and rat pyruvate carboxylases had been ongoing in our laboratory. Partial cDNA clones representing half of the full length of rat PC had previously been isolated in by two former Ph.D. students (A.I. Cassady, 1987 and G.W. Booker, 1990). Meantime, Dr. M. E.Walker had found several human PC cDNA clones having the same coding sequence but differing at their 5'-untranslated regions. This raised the possibility that human PC mRNA could have been alternatively spliced from the 5'-end of the gene. The isolation of full length rat PC cDNA as well as the identification of multiple transcripts both in human and rat genes were therefore undertaken as a part of this study.

Determination of PC gene structure would be invaluable for this study. Clearly, the functional domains of the various biotin-dependent enzymes are related in structure and most probably resulted from shuffling of exons for individual components of the complex multifunctional polypeptide by gene fusion during evolution. Comparison of exon/intron organisation of PC to other biotin carboxylase genes would test such a hypothesis. In addition, the isolation of the 5'-end of the PC gene would also provide very important information on the regulatory sequences that modulate PC expression.

An ultimate goal of this study is to unravel the molecular basis that controls PC expression under different physiological and pathological circumstances including postnatal development and in genetic obesity.

The specific aims of this thesis were to

1. Complete isolation of the cDNA sequence encoding rat PC; determine the domain structure and predict the active sites of this enzyme by sequence comparisons with other biotin enzymes.

2. Identify and characterise multiple mRNA species generated either by the use of alternative promoters or by differential splicing of the primary transcripts from the rat and human PC genes.

3. Complete the isolation of the gene structure for rat PC and determine its chromosomal localisation.

4. Characterise alternate promoters of rat PC gene by reporter gene analysis.

5. To establish a ribonuclease protection assay and an RT-PCR assay to measure the level of each PC mRNA isoform in various tissues.

6. To investigate the transcriptional and translational regulation of PC expression *in vivo* during postnatal development and in genetically obese rats and in different cell lines.

7. To overexpress recombinant human PC in mammalian cells and purify this recombinant material with the ultimate aim to examine the effect of the identified point mutations in PC on PC activity.

CHAPTER 2

MATERIALS AND GENERAL MOLECULAR

BIOLOGY METHODS

MATERIALS

2.1 **CHEMICALS** The following chemicals were obtained from Sigma Chemical Co., St Louis, MO., USA: agarose (type 1), ampicillin, ATP (disodium, grade I), β-mercaptoethanol, bovine serum albumin (BSA), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (BCIG), D-biotin, Coomassie Brilliant Blue (G_{250}, R_{250}), cycloheximide, dithiothreitol (DTT), ethidium bromide, ethylenediaminetetraacetic acid (EDTA), guanidium thiocyanate, bovine pancreatic insulin, lithium chloride, 3-[N-Morpholino]propane-sulfonic acid (MOPS), nicotinamide adenine dinucleotide (reduced form) [NADH], nitroblue tetrazolium (NBT), phenylmethylsulphonyl fluoride (PMSF), o-nitrophenyl-β-D-galactopyranoside (ONPG), polyoxyethylene-sorbitan monolaurate (Tween-20), puromycin, rubidium chloride, salmon sperm DNA, sodium dodecyl sulfate (SDS), sodium pyruvate, glycyl glycine, N,N,N',N'-tetramethylethylenediamene (TEMED), Tris-(hydroxymethyl)methylamine (Tris) and Triton X-100. Glycogen was purchased from Boehringer Mannheim. Beetle luciferin was purchased from Promega, Madison, WI, U.S.A. Acrylamide and bisacrylamide were purchased from BioRad Laboratories Inc., Hcrates, CA, U.S.A. 5-bromo-4-chloro-3-indolyl-β-D-bromogalactopyranoside (X-gal) was obtained from Progen Industries Ltd., Brisbane, Qld. Sephadex G25, Sepharose CL-6B and Ficoll 400 were purchased from Pharmacia, Uppsala, Sweden. Other fine chemicals were purchased from Merck Pty. Ltd., Kilsyth, Vic., Australia.

2.2 RADIOCHEMICALS

[α-^{32}P] dATP, [α-^{32}P] dCTP, [α-^{32}P] UTP and [γ-^{32}P] ATP were purchased from Brcsatec Ltd., Adelaide, South Australia. NaH^{14}CO$_3$ was purchased from Amersham Australia, North Ryde, NSW, Australia.

2.3 OLIGONUCLEOTIDES

All oligonucleotides used in this study were purchased from Brcsatec Ltd., Adelaide, South Australia. The oligonucleotides used for RACE-PCR, RT-PCR and LD-PCR were described in Chapters 3, 4 and 5 while the ones that were use to sequence PC cDNA and gene are listed below.

Oligonucleotide name	*Sequence*
PC4	5'-CCGCACAGCTCTGGCAAAGCCCT-3'
PC5	5'-GCAGGAAGGGGATGTTGGTCT-3'
PC6	5'-GGCACAACCATTGATTCGGAT-3'
PC7A	5'-GGTGGGTAGCAGGGGCAATCTC-3'
PC8	5'-GGGACCACTGGAACGCCTGC-3'
PC9	5'-ACAGCCCCCGTTGCCTCCCCA-3'
PC12	5'-GCAGGCGTTCCAGTGGTCCC-3'
PC17	5'-GAGATTGCCCCTGCTACCCACC-3'
PC18	5'-CTTTACACCTCGGACACGGAACTCC-3'
PC19	5'-GCCAGGGCTCTGCTCATCTTGGTG-3'
PC20	5'-GACCTGGTCCATGCTCAGATCCAT-3'
PC22	5'-GGAGTTCCGTGTCCGAGGTGTAAAG-3'
EMBL/L (lambda left arm)	5'-GGCCTAATACGACTCACTATAGGG-3'
EMBL/R (lambda right arm)	5'-GGCCATTTAGGTGACACTATAGAAG-3'
forward sequencing primer	5'-GTAAAACGACGGCCAGT-3'
reverse sequencing primer	5'-CACACAGGAAACAGCTATGACCATC-3'

2.4   MOLECULAR BIOLOGY KITS

The 5'-AmpliFinder RACE kit and Rat GenomeWalker Kit were purchased from CLONTECH, Palo Alto, CA, U.S.A.   The *fmol*^{TM} DNA sequencing system, TA cloning system and luciferase assay system were all purchased from Promega, Madison, WI, USA. The ribonuclease protection assay kit (RPA II), MaxiScriptTM kit and RNA isolation kit were purchased from Ambion, Houston, TX, USA.   The GIGAprime labeling kit was purchased from Bresatec Ltd., Adelaide, South Australia.   The Geneclean II kit was obtained from BIO 101, CA, U.S.A.

2.5   ENZYMES AND PROTEIN MARKERS

All restriction enzymes were purchased from Pharmacia, Uppsala, Sweden and Bresatec Ltd., Adelaide, South Australia.   T4 DNA ligase, T4 DNA polymerase, T4

polynucleotide kinase, *Taq* polymerase and Moloney murine leukemia virus reverse transcriptase (MMLV-RT) were purchased from Promega, Madison, WI, U.S.A. *Pwo* and avian myeloblastosis virus reverse transcriptase (AMV-RT) were purchased from Boehringer Mannheim. Advantage™, *Tth* polymerase mix was purchased from CLONTECH, Palo Alto, CA, U.S.A. Malic dehydrogenase, RNase A, DNaseI, anti-rabbit IgG-antibodies, and avidin-conjugated with alkaline phosphatase were purchased from Sigma. Biotinylated protein markers was obtained from BioRad. ^{14}C-labelled molecular weight markers were purchased from Amersham.

2.6   BACTERIAL STRAINS

E. coli DH5α: supE44ΔlacU169(p80lacZΔM15) *hsdR17* recA1 endAA1 gyrA96 thi-1 relA1 host for recombinant plasmids (New England, Biolabs, CA)

E. coli LE 392: F⁻ c14⁻(McrA⁻) *hsdR514*(r$_k^-$ m$_k^+$) *supE44 supF58 lacY1* or Δ(*lacIZY*)6 *galK2 galT22 metB1 trpR55*. (New England, Biolabs, CA)

2.7   BACTERIAL GROWTH MEDIA

LB broth: 1% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) yeast extract (Difco), 1% (w/v) NaCl, adjusted to pH 7.0 with NaOH.

LB agar plates were made by adding 1.5% (w/v) Bacto-agar (Difco) to the LB broth

LMN agar plate: 0.1% tryptone, 0.5% (w/v) NaCl, 0.4% (w/v) maltose and 0.2% (w/v) MgSO$_4$, 1.5%(w/v) Bacto-agar.

NZCYM agar plate: 1% (w/v) Nzamine, 0.5% (w/v) NaCl, 1% (w/v) casamino acid, 0.5 % (w/v) yeast extract, 0.25% MgSO4 and 0.7% (w/v) Bacto-agar.

Long term storage of plasmids in bacteria was prepared by adding 40% glycerol to an overnight culture of *E. coli* in LB-broth and then stored at -80°C.

2.8   MAMMALIAN CELL LINES

African green monkey kidney cell line (COS-1; ATCC: CRL 1750); human hepatoma cell line (HepG2; ATCC:HB 8065); human embryonic kidney cell line (293T; ATCC: CRL

1573); human embryonic kidney (293T Tet-On: Clontech); Chinese hamster ovarian line (CHO-K; ATCC: CCL 61); fetal rat liver cell line (FRL4.1; kind gift from Dr George Yeoh, Department of Biochemistry, University of Western Australia), rat liver cell lines (Reuber H35; ATCC: CRL 1548, BRL 3A; ATCC: CRL 1442, McA-RH 777; ATCC: CRL 1601and FTO3; obtained from the Institute of Cell and Tumor Biology, Heidelberg, Germany); rat mammary gland carcinoma (Institute of Medical and Veterinary Science, Adelaide); rat myoblast (L6; ATCC: CRL 1458) and rat insulinoma cell line (obtained from Dr Michael MacDonald, Children Diabetes Center, University of Wisconsin Medical School, Madison, Wisconsin, U.S.A.) were used in this thesis.

For long term storage, mammalian cell lines were prepared by freezing cells in 10% DMSO in fetal calf serum at -80°C or in liquid nitrogen.

2.9 CELL CULTURE MEDIA

Dulbecco's Modified Eagle's Medium (DMEM) (Gibco), 28 mM $NaHCO_3$, 19 mM glucose and 20 mM Hepes, pH 7.3.

RPMI 1840 (Gibco) supplemented with 50 mM β-mercaptoethanol, 1 mM sodium pyruvate, 10 mM Hepes, pH 7.35.

Ham's F12 with L-glutamine (Gibco) supplemented with 28 mM $NaHCO_3$, pH 7.4 All media was supplemented with 50,000 units/l of gentamycin (Gibco) and filter sterilised prior to use.

Fetal calf serum : Gibco

2.10 GENOMIC LIBRARIES

1.   Rat liver genomic library constructed in EMBL3 Sp6/T7 (CLONTECH, Palo Alto, CA, U.S.A.).

2.   Rat cosmid genomic library constructed in pWE 15 (CLONTECH, Palo Alto, CA, U.S.A.).

2.11 PLASMIDS

1. pBluescript II (SK); Stratagene, La Jolla, CA, U.S.A.

2. pGL-3 basic vector, pGL-3 promoter vector, pGL-2 control vector; pGEM3Z (+), Promega, Madison, WI, U.S.A.

3. pRSV-βGal; kindly provided by Dr. K. Surinya, Department of Biochemistry, University of Adelaide, Adelaide.

4. pTZ-hPC50 encoding most of the coding region of the human PC cDNA was given by Dr. M. Walker, University of Adelaide, Adelaide.

5. pIRES1*Neo* and pBI, CLONTECH, Palo, Alto, CA, U.S.A.

6. pEFIRES-puro obtained from Dr. Steven Hobbs, CRC for Cancer Theruapeutics, Institute of Cancer Research, London, U.K.

7. prGAPDH, kind gift from Dr Tim Sadlon, Department of Biochemistry, University of Adelaide, Adelaide.

2.12 DNA AND RNA TECHNIQUES

2.12.1 Small scale preparation of plasmid DNA for restriction enzyme digestion, sequencing and in vitro transcription

A single colony of *E. coli* was grown in 2 ml of LB broth containing 100 μg/ml of ampicillin and incubated at 37°C with shaking overnight. One and a half millilitres of overnight culture were transferred to a 1.5 ml microtube and centrifuged at 13,000 rpm for 2 min. The supernatant was discarded and the pellet was suspended in 100 μl of TES solution [25 mM Tris-HCl, pH 8.0, 10 mM EDTA pH 8.0, 15% (w/v) sucrose] and left on ice for 5 min. Two hundred microlitres of lysis solution [1% (w/v) SDS in 0.2 M NaOH] were then added and left on ice for 5 min. The lysates were then neutralised with 150 μl of 3 M potassium acetate pH 5.2 and left on ice for 5 min. Cell debris and chromosomal DNA were removed by centrifugation at 13,000 rpm for 15 min. The supernatant was transferred to a new tube and RNA removed by digesting with 4 μl of 10 mg/ml of RNAse A at 37°C for 30 min before being extracted twice with equal volumes of phenol and chloroform [1:1 (v/v)] The aqueous phase was separated from the organic phase by centrifugation at 13,000 rpm for 2 min. The aqueous phase was then transferred to a new tube and extracted once with an

equal volume of chloroform. The plasmid DNA was precipitated by adding 2 volumes of cold absolute ethanol and kept at -20°C for 1 h. The plasmid DNA was then recovered by centrifugation at 13,000 for 15 min and washed once with 70% ethanol. The pellet was dried under vacuum and dissolved in 30 μl sterile distilled water.

2.12.2 Large scale preparation of plasmid DNA by phenol/chloroform/PEG precipitation for transfection

A single colony of *E. coli* was grown in 10 ml of LB broth containing 100 μg/ml ampicillin with vigorous shaking at 37°C for 6 h. Five millilitres of culture were then added to 250 ml of the same culture medium and grown at 37°C with vigorous shaking at 37°C overnight. The cells were harvested by centrifugation at 5000 g at 4°C for 10 min. Cells were suspended in 6 ml of TES buffer supplemented with 2 mg/ml lysozyme and left on ice for 30 min. Twelve millilitres of freshly prepared 0.2 M NaOH-1% (w/v) SDS were then added, gently mixed and left on ice for 10 min. The lysates were neutralised by adding 7.5 ml of 3 M potassium acetate pH 4.8, gently mixed and incubated on ice for 10 min. Cells debris and chromosomal DNA were removed by centrifugation at 10,000 x g for 15 min at 4°C. The supernatant was extracted with an equal volume of phenol/chloroform [1:1 (v/v)] twice, and then an equal volume of chloroform once. The supernatant containing plasmid DNA was precipitated by adding 2 volumes of cold absolute ethanol and kept at -20°C for 1 h. The plasmid was recovered by centrifugation at 10,000 x g for 15 min at 4°C. The pellet was washed once with 70% ethanol and dried under vacuum before being dissolved in 1.6 ml sterile distilled water. The plasmids were further purified by adding 400 μl of 4 M NaCl and 800 μl of 33.3% polyethylene glycol (PEG) 6000, left on ice for 1 h and centrifuged at 10,000 x g for 15 min. The pellet was washed once with 70% ethanol and dried under vacuum. The dried pellet was then dissolved in 400 μl of sterile distilled water and extracted twice with an equal volume of phenol/chloroform [1:1 (v/v)] and extracted once with equal volume of chloroform. The plasmid was precipitated by adding 2 volumes of cold absolute ethanol and kept at -20°C for 1 h and centrifuged at 10,000 x g for 15 min. The pellet was then washed once with 70% ethanol, dried and dissolved in 0.5 ml sterile distilled water.

2.12.3 Purification of DNA fragments from agarose gel using Geneclean kit

A slice of agarose containing a DNA fragment was excised under UV-transilluminator and put in 1.5 ml microtube. A half volume of TBE modifier and 4.5 volumes of NaI solution were added to a slice of gel and incubated at 65°C for 5 min. Five microlitres of Glassmilk were then added and the DNA allowed to bind at 4°C for 5 min. Glassmilk-bound DNA was pelleted by centrifugation at 13,000 rpm for 10 sec and washed 3 times with New Wash solution. DNA was eluted from Glassmilk with 20 μl of water by incubation at 65°C for 5 min. The Glassmilk was pelleted, as above, and the supernatant was kept.

2.12.4 Ligation

Plasmid DNA and insert DNA were digested with restriction enzymes that generated compatible ends. The ligation was routinely carried out in a total volume of 20 μl containing DNA insert : vector = 2:1 (mass ratio), 1X ligation buffer (66 mM Tris-HCl, pH 7.5, 5 mM MgCl$_2$, 1.0 mM DTT, 1.0 mM ATP) and 2 U of T4 DNA ligase at 14°C overnight.

2.12.5 Bacterial transformation

A single colony of *E. coli* DH 5α was grown in 2 ml of LB broth at 37°C overnight. Overnight culture (330 μl) was then subcultured into 10 ml of LB broth and grown to A$_{600}$ of 0.5-0.6. Five millilitres of the culture was transferred to a new flask containing 100 ml of LB broth and grown to A$_{600}$ of 0.5-0.6. Cells were pelleted by centrifugation at 5000 x g for 10 min and suspended in 40 ml of transformation buffer-1 (30mM potassium acetate, 100 mM rubidium chloride, 10 mM CaCl$_2$, 50 mM MnCl$_2$ and 15% glycerol, pH 5.8) and incubated on ice for 10 min. Cells were pelleted by centrifugation at 5000 x g for 10 min and resuspended in 4 ml of transformation buffer-2 (10 mM MOPS acid, 10 mM rubidium chloride, 75 mM CaCl$_2$ and 15% glycerol, pH 6.5) and incubated on ice for 15 min. Competent cells were taken in aliquots (200 μl) and kept at -80°C until required.

Typically, 200 μl of frozen competent cells were thawed out, mixed with 10 μl of ligation reaction and left on ice for 30 min. DNA and cells mixture were heat-shocked at 42°C for 2 min. The transformed cells were recovered by incubation in 2 ml of LB broth at 37°C for 45 min before plating. The recombinant clones were distinguished from non-recombinant clones on the basis of the α-complementation of β-galactosidase (lac Z) gene expression, using LB plates containing IPTG as the inducer and X-gal as substrate of β-galactosidase. Blue colonies indicate the non-recombinant clones while white colonies indicate the recombinant clones as a result of disruption of the β-galactosidase gene on the plasmid.

2.12.6 Isolation of high molecular weight DNA from tissues

One gram of frozen tissue was ground to a powder in liquid nitrogen, suspended in 10 ml of extraction buffer (0.1 M EDTA, pH 8.0, 10 mM Tris-HCl, pH 8.0 and 0.5% SDS) and incubated at 37°C with gentle shaking. Large pieces of tissues were removed by centrifugation at 10,000 x g for 10 min. The proteins were digested by adding 100 μl of 10 mg/ml of proteinase K at 50°C for 1.5 h. Another 100 μl of the same concentration of proteinase K was added and further incubated at the same temperature for 1.5 h. The proteins were then extracted with an equal volume of phenol/chloroform [1:1 (v/v)] twice, and then an equal volume of chloroform once. The chromosomal DNA was precipitated by adding 0.2 volume of 10 M ammonium acetate and 2 volumes of cold absolute ethanol. The DNAs were spooled from solution with a hooked glass rod, washed once with 70% ethanol and dissolved in 1 ml TE (10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA, pH 8.0). The chromosomal DNA prepared was used for PCR and Southern blot analysis.

2.12.7 Labelling of DNA with ^{32}P

2.12.7.1 End filling of DNA markers

The reaction was carried out in a total volume of 20 μl containing 1 μg of DNA marker, 2 μl of 10x T4 DNA polymerase buffer (33 mM Tris-acetate, pH 7.9, 10 mM magnesium acetate and 66 mM potassium acetate), 2 μl of 10 μg/ml BSA, 1 μl of 10 mM DTT, 1 μl each of 2 mM dTTP, dCTP and dGTP, 25 μCi of [α-^{32}P] dCTP, 8 U of T4 DNA

polymerase and water. The reaction was performed at 37°C for 10 min and terminated by adding 2 μl of 0.5 M EDTA.

2.12.7.2 Labelling 5'-end of oligo by kinasing

The reaction was carried out in a total volume of 50 μl containing 500 ng of oligonucleotide, 5 μl of 10x kinase buffer (500 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 50 mM DTT and 1.0 mM spermidine), 50 μCi of [γ-^{32}P]ATP, 10 U of T4 polynucleotide kinase and water. The reaction was performed at 37°C for 30 min and terminated by adding 2 μl of 0.5 M EDTA. The reaction volume was brought up to 100 μl with water and then extracted with equal volume of phenol/chloroform [1:1 (v/v)] twice followed by an equal volume of chloroform once. The phosphorylated oligonucleotides were precipitated by adding 0.1 volume of 3 M sodium acetate pH 5.2 and 2 volumes of cold absolute ethanol and centrifuged at 13,000 for 15 min. The pellet was washed once with 70% ethanol, dried under vacuum and dissolved in 50 μl of water. Labelled oligonucleotide can be added directly to the prehybridised filter without heat denaturation of probe.

2.12.7.3 Random priming

DNA fragments were labelled with ^{32}P using the Giga Prime labelling kit. Briefly, the reaction was carried out in 20 μl containing 0.5-1.0 μg of heat denatured DNA template and random decanucleotide primers. Single strand probes were synthesised as short fragments by *E. coli* DNA polymerase I in the presence of [α-^{32}P] dATP at 37°C for 1 h. Labelled DNA fragments were ethanol precipitated in the presence of 0.1 volume of 7.5 M ammonium acetate and 25 μg yeast tRNA. The probes were recovered by centrifugation at 13,000 rpm for 15 min and suspended in 100 μl of water. Two microlitres were counted in a Liquid Scintillation Counter. The specific activity of the probe was approximately 1-5 x 10^8 cpm/μg DNA obtained. Labelled probes were heat-denatured at 100°C for 10 min and incubated on ice for 5 min before adding in hybridisation buffer.

2.12.8 Southern blot hybridisation

Fractionated DNA in agarose gel was denatured by soaking the gel in 0.5 M NaOH-1.5 M NaCl 2 x 15 min followed by neutralisation with 0.5 M Tris-HCl pH 7.4-1.5 M NaCl 2 x 15 min. A stack of Whatman papers were cut with an area slightly bigger than the gel,

soaked in 10x SSC (1.5M NaCl, 0.15 M sodium citrate) and placed in a tray containing 10x SSC as transfer buffer. The gel was then placed over the Whatman paper, followed by a sheet of nylon membrane which had been soaked in 10x SSC. A stack of paper towels was then placed over the nylon membrane. The DNA was allowed to transfer to the membrane overnight. The membrane was then washed briefly in 2x SSC and was ready for hybridisation.

For detecting specific DNA sequences with oligonucleotide probes (kinase reaction), the membrane was prehybridized in a solution containing 5x SSC, 5x Denhardt's solution (0.1% polyvinylpyrrolidone, 0.1% BSA, 0.1% Ficoll 400) and 100 µg/ml heat denatured salmon sperm DNA. For detecting specific DNA sequences with polynucleotide probes (random priming), the membrane was prehybridized in a solution containing 5x Denhardt's solution, 1% SDS, 50 mM Tris-HCl, pH 7.4, 1M NaCl, 10% PEG6000, 40% deionized formamide and 100 µg/ml heat denatured salmon sperm DNA. After incubation of the membrane in this solution for at least 2 h, 10 ml of fresh solution including probes were then added and hybridisation was performed for 12 h. The membrane was then washed 2 x15 min in 2x SSC-0.1% SDS at 42°C followed by another 2 x 15 min washing in 0.1x SSC-0.1% SDS at 42°C.

2.12.9 DNA sequencing

2.12.9.1 fmol^{TM} DNA sequencing

The 5'-end of the sequencing primer was labelled in a 10 µl reaction containing 100 ng of primer, 1x T4 polynucleotide kinase buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl$_2$, 5.0 mM DTT, 0.1 mM spermidine) and 50 µCi of [γ-^{32}P]ATP. The reaction was carried out at 37°C for 30 min before inactivating the enzyme by heating at 95-100°C for 2 min. A master mix was then prepared containing 1-2 µg of plasmid DNA, 5 µl of 5X sequencing buffer (250 mM Tris-HCl, pH 9.0, 10 mM MgCl$_2$), 1.5 µl of labelled primer and 5U of *Taq* polymerase (sequencing grade). The reaction volume was brought up to a final of 17 µl with water. Four microlitres of master mix were then transferred to 0.5 ml microtubes containing 2 µl each of ddATP, ddTTP, ddCTP and ddGTP separately and overlaid with 100µl of mineral oil. The reaction was subjected to 30 cycles of amplification. Each cycle

consists of denaturation at 95°C for 30 sec (2 min for the first cycle), annealing at 50°C for 30 sec and extension at 70°C for 1 min. In the cases where the length of primer was more than 24 nucleotides and or with GC content greater than 50%, annealing and extension were performed at 70°C for 30 sec. Other parameters were essentially the same as described above. The reaction was terminated by the addition of 4 µl of stop solution (10 mM NaOH, 95% formamide, 0.05% bromphenol blue, 0.05% xylene cyanole) and heated at 95-100°C before electrophoresis.

2.12.9.2 Dye Primer sequencing

Plasmid DNA isolated according to the method described in 2.12.1 was further purified by adding 1 µl of 10 mg/ml RNase A to 20 µl of plasmid sample and incubating at 37°C for 30 min. Twenty five microlitres of 7.5 M ammonium acetate and 40 µl of isopropanol were then added, mixed and left at -20°C for 1 h. The tube was then centrifuged at 13,000 rpm for 15 min and the pellet was washed once with 300 µl of 75% ethanol in 50 mM sodium acetate and once with 100% ethanol. The pellet then was dried and dissolved in 20 µl of water. The sequencing reaction was performed in a single 0.5 ml tube containing 1 µg of plasmid template, 8 µl of dye terminator mix and 100 ng of primer. The reaction was adjusted to a final volume of 20 µl and overlaid with 100 µl of mineral oil before being subjected to PCR. The PCR profile consisted of 25 cycles of denaturation at 96°C for 30 sec, annealing at 50°C for 15 sec and extension at 60°C for 4 min. After the cycles was completed, the mineral oil was removed and the sequencing products were precipitated by adding 0.1 volume of 3 M sodium acetate and 2 volumes of absolute ethanol. The DNA was precipitated by centrifuging the tube at 13,000 rpm for 15 min. The pellet was dried under vacuum and submitted to the sequencing service at the Institute of Medical and Veterinary Science, Adelaide.

2.12.10   Isolation of RNA

2.12.10.1   Isolation of RNA using the guanidine method

Total RNA was isolated as described by Chomczynski and Sacchi (1987), with some modifications. Approximately 1 g of frozen tissue was homogenised in 10 ml of denaturation solution (4 M guanidinium thiocyanate, 25 mM sodium citrate pH 7.0, 0.5%

(v/v) sarkosyl and 0.1 M β-mercaptoethanol) using a Polytron™ at full speed for 1 min. One millilitre of 2 M sodium acetate pH 4.0, 10 ml of phenol and 2 ml of chloroform were added with vigorous mixing and left on ice for 15 min. The supernatant was removed after centrifugation at 10,000 x g for 30 min at 4°C. The RNA was isolated from DNA by precipitation with 2 volumes of isopropanol and centrifugation at 10,000 x g for 30 min at 4°C. The pellet was dissolved in 5 ml TE/SDS, reprecipitated by adding 5 ml of isopropanol and kept at -20°C for 1 h. The RNA was pelleted by centrifugation at 10,000 x g for 30 min at 4°C and redissolved in 2.5 ml TE/SDS. Small RNAs were precipitated overnight at 0°C by adding 10 ml of 4 M LiCl. RNA was recovered by centrifugation at 10,000 x g for 30 min at 4°C, dissolved in 1 ml TE/SDS and precipitated with 0.1 volume of 3 M sodium acetate, pH 5.2 and 2 volumes of cold absolute ethanol. Finally, RNA was recovered by centrifugation as above, washed with 70% ethanol, vacuum dried and dissolved in 0.5 ml of RNAse-free water.

2.12.10.2.   *Isolation of RNA using RNAqueous™ kit*

Small scale preparations of RNA were carried out using RNAqueous™ kit. Briefly, 50-100 mg of frozen tissue was ground in liquid nitrogen to a powder. Twelve volumes of Lysis/Binding solution was then added to the tissue powder and homogenised with a Polytron™ homogeniser. An equal volume of 64% ethanol was added to the lysate and mixed gently. Lysate/ethanol mixture was then applied to the RNAqueous cartridge filter which was placed in a collection tube and centrifuged at 13,000 rpm for 30 sec. The supernatant was discarded and the cartridge was washed once with 700 µl of Wash Solution #1 and twice with 500 µl of Wash Solution #2/3. RNA was eluted from the cartridge filter twice with 60 µl of Elution solution by incubating the tube at 65-70°C for 10 min followed by centrifugation at 13,000 rpm for 1 min. One-half volume of LiCl precipitation solution was then added, incubated at -20°C for at least 30 min and centrifuged at 13,000 rpm for 15 min. The pellet was washed with 70% ethanol, vacuum dried and dissolved in 50 µl of RNAse free water.

2.13 Quantitaion of DNA and RNA by spectrophotometry

DNA and RNA were quantitated by measuring their absorbance at 260 nm. Samples were routinely measured in a 1 ml volume.

concentration of DNA (μg/μl) $=$ A_{260} x 50 x dilution factor

concentration of RNA (μg/μl) $=$ A_{260} x 40 x dilution factor

CHAPTER 3

ISOLATION OF FULL LENGTH cDNA AND

DOMAIN STRUCTURE OF THE RAT PC

3.1 INTRODUCTION

Clearly, amino acid sequences of eukaryotic proteins can be obtained from their mRNAs. Isolation of cDNA would therefore be an effective means to gain insight into the primary structure of proteins by inferring amino acid sequences from codons in the cDNAs. Also the relatively large size of PC limits the possibility of determining its primary structure by protein sequencing effectively. Comparisons of primary structures of PC from different sources and to known structures of other biotin-dependent carboxylases would allow the identification of structurally important residues as well as other conserve residues which might play a common role in the catalytic reaction.

During their Ph.D. candidature, A.I. Cassady and G.W. Booker had previously isolated partial cDNA clones encoding rat PC by screening a rat liver cDNA library constructed in lambda phage. Attempts to isolate from this library other clones representing the remaining rat PC cDNA were unsuccessful. This is most likely due to the relatively large size of the rat PC mRNA (approximately ~4.2 kb) and the inherent limitations of oligo-dT-primed cDNA libraries. As an alternative approach, PCR was employed to obtain full length rat PC cDNA.

PCR is a highly sensitive and specific method to amplify a fragment of DNA *in vitro* using a thermostable DNA polymerase. This technique was first introduced by Mullis *et al.* (1985). Modified PCRs were then applied to different areas of research. As stated above, obtaining full length transcripts from a cDNA library is rarely achieved if the size of transcript is very large but methods have been developed which utilise PCR to selectively amplify the 5'-end of cDNA. This PCR is known as RACE (rapid amplifiation of cDNA ends). RACE-PCR was first described by Frohman *et al.* (1988) and Belyavsky *et al.* (1989). This technique combines three features: reverse transcription of cDNA from mRNA, homopolymeric tailing with terminal transferase at the 3'-end of cDNA followed by PCR with the gene-specific primer and a primer complementary to a homopolymeric tail. However, low efficiency of the terminal transferase reaction reduces the power of this technique. Consequently, a modified RACE-PCR was developed involving the ligation of a single-stranded oligonucleotide anchor to the 3'-end of the first-strand cDNA using T4 RNA ligase (Edwards *et al.*, 1991). A Clontech 5'-AmpliFinder RACE kit (Apte and Siebert,

47

1993) employing this principle was used as a tool to obtain a full length rat PC cDNA clone as illustrated in Fig. 3.1.

Briefly, first strand cDNA is synthesized with the first antisense gene-specific primer (GSP1) by AMV reverse transcriptase. The RNA template is then hydrolysed with NaOH. Single strand cDNA is purified and ligated to a single-stranded anchor oligonucleotide at the 3'-end. This anchor oligonucleotide contains the phosphate group necessary for ligation at the 5'-end and an *Eco* RI restriction site to facilitate cloning, while the 3'-end is blocked with an amine group to prevent concatamer formation. Semi-nested PCR using an anchor primer (complementary to anchor adaptor) and a second gene-specific primer (GSP2) internal to GSP1 is then performed to increase the specificity of the reaction. Amplified products can then be cloned and sequenced.

Figure 3.1 A, Amplification of the 5'-end of a gene using the 5'-AmpliFINDER RACE method. P1 and P2 are nested gene-specific primers. Arrows indicate PCR primers, the hatched box indicates the AmpliFINDER anchor. **B,** 5'-AmpliFINDER anchor oligonucleotide sequence, at its 5'-end was modifed with an amino blocking group to prevent concatamer formation; a phosphate group was added to the 5'-end of the anchor primer which also contains an *Eco*RI site (underlined). The complementary region of anchor sequence and anchor primer are bolded.

A
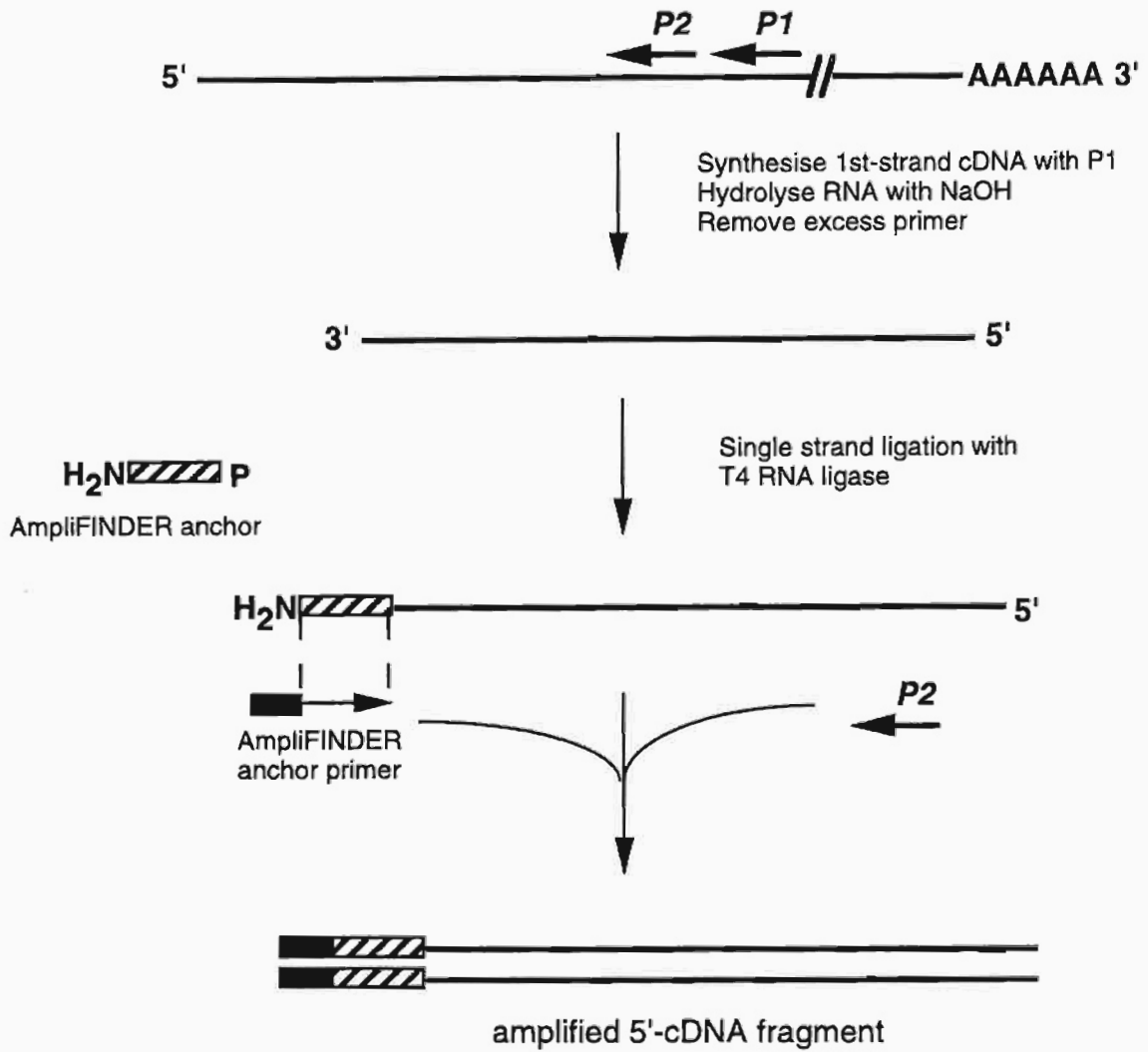
B
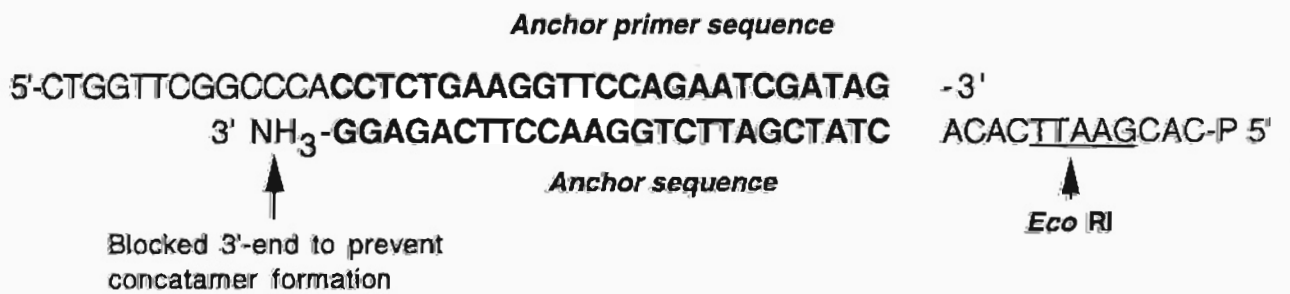
3.2 SPECIFIC METHODS

3.2.1 RACE-PCR

3.2.1.1 Reverse transcription

Briefly 10 μg of total rat liver RNA were hybridised to 100 ng of cDNA synthesis primer in a total volume of 10 μl. The reaction mixture was incubated at 65°C to disrupt the secondary structure of RNA for 10 min and slowly cooled down to 52°C. The 25 μl master mix containing 9.2 μl of water, 9 μl of 4X reverse transcriptase buffer (1X=42 mM Tris-HCl, pH 8.3, 8mM $MgCl_2$ and 8 mM DTT), 1.6 μl of RNase inhibitor (40 U/μl), 3.7 μl of 10 mM each of dNTP and 0.5 μl of AMV reverse transcriptase (25 U/μl) was made and pre-warmed at 52°C for 5 min. Twenty microlitres of master mix were then added to the RNA annealed primers and incubated at 52°C for 1 h. The reaction was terminated by adding 1 μl of 0.5 M EDTA pH 7.0. The RNA was hydrolysed by adding 2 μl of 6M NaOH and incubated at 65°C for 30 min followed by neutralisation with 2 μl of 6M acetic acid.

3.2.1.2 cDNA purification and Anchor ligation

cDNA was purified by adding 8 μl of GENO-BIND™ suspension, mixed and incubated on ice for 10 min with occasional vortexing. The GENO-BIND™-bound cDNA was pellet by centrifugation at 13,000 rpm at 4°C for 10 sec. The supernatant was removed and the pellet was washed with 80% ethanol and centrifuged as described above. The pellet was air-dried and then dissolved in 50 μl of DEPC-treated water. cDNA was eluted from GENO-BIND™ by incubating the tube at 65°C for 5 min followed by centrifugation at 13,000 rpm for 5 min. The supernatant containing cDNA was transferred to a new tube and the cDNA was precipitated by adding 15 μg of glycogen, 5 μl of 2 M sodium acetate and 100μl of 95% ethanol. After 2 h of incubation at -20°C, the cDNA was pelleted by centrifugation at 13,000 rpm at 4°C for 15 min, washed once with 80% ethanol, air-dried and dissolved in 6 μl of DEPC-treated water. The ligation of adaptor at the 5'-end of single-stranded cDNA was performed in a 0.5 ml microtube containing 2.5 μl of purified cDNA, 2μl of AmpliFINDER Anchor (4 pmol), 5 μl of 2X T4 RNA ligase buffer (100 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 1 mM hexamine cobalt chloride, 10 μM ATP and 50% PEG6000) and 0.5 μl of T4 RNA ligase (20 U/μl). The ligation reaction was incubated at room temperature (~20-25°C) for 16-20 h.

3.2.1.3 PCR amplification

Hot start PCR was carried out to enhance the specificity of PCR amplification. In brief, 48 µl of the reaction mixture containing 10 mM Tris, pH 8.3, 50 mM KCl, 1.5 mM MgCl$_2$, 0.2 mM each of dNTP, 0.01% gelatin, 2.5 U of *Taq* polymerase and 1.0 µl of anchor-ligated cDNA were assembled in a 0.5 ml microtube and overlaid with 2 drops of mineral oil. The reaction was heated at 82°C for 1 min before adding 1 µl of each primer (nested gene-specific and AmpliFINDER Anchor primer, 100 ng/µl of each) and proceeding with the PCR. The thermal profile consisted of 35 cycles of denaturation at 94°C 45 sec, annealing at 60°C for 45 sec and extension at 72°C for 2 min (7 min for the final extension). The PCR products were analysed by running on 2% agarose gel.

3.2.2 Sequence analyses by computer

BLAST searches (Altschul *et al.*, 1997) were made courtesy of the NCBI (Bethesda, MD, U.S.A.) of the non-redundant database comprising the Protein Data Base (PDB) plus GenBank (release no. 91) plus GenBankupdate plus EMBL (release no. 43) plus EMBL update by TBLASTN, and on the non-redundant database comprising PDP plus SwissProt plus SPupdate plus PIR plus GenPept plus GPupdate by BLASTP, with an overlapping series of 11 fragments, each of approximately, 200 residues, from the rat PC primary structure as the 'query' sequences.

3.3 RESULTS

3.3.1 Isolation of cDNA by RACE-PCR

Two antisense primers were used to synthesize two regions of PC cDNA from total rat liver RNA. The first region was synthesised with a primer designed from a conserved sequence of mouse (Zhang *et al.*, 1993) and human (Wexler *et al.*, 1994; MacKay *et al.*, 1994) PC cDNA; MRACE I (GSP 1) (nucleotide positions +358 to +380) followed by PCR with another internally conserved antisense primer; MRACE II (GSP2) (Fig.3.2) corresponding to nucleotide positions +124 to +150 of mouse (Zhang *et al.*, 1993) and human (Wexler *et al.*, 1994 and MacKay *et al.*, 1994) PC cDNAs and adaptor primer. A smear pattern was observed using an annealing temperature of 60°C (data not shown). However, increasing the annealing temperature to 65°C improved specific amplification, i.e. discrete bands ~ 0.2 kb were seen upon agarose gel electrophoresis (Fig. 3.3). These individuals bands were then purified from the gel, digested with *Eco*RI and cloned into *Eco*RI-digested pBluescript. Sequencing of one of many clones [designated as RACE (Anchor/MRACE II)], showed high sequence homology to both human and rat PC cDNA within the coding region. However, sequences at the 5'-UTR were different from those of mouse (Zhang *et al.*, 1993) and human PC cDNAs (Wexler *et al.*, 1994; MacKay *et al.*, 1994). Indeed, sequence divergence at the 5'-UTR was observed in different clones, suggesting that they were alternatively spliced at the 5'-non-coding region. Characterisation of these multiple transcripts in full detail is described in Chapter 4 of this thesis.

The second region of cDNA was synthesised with the PC1 primer (GSP1) designed from the 5'-end of λRL 2.53 (Booker, 1990), followed by PCR with a nested antisense primer, PC2 (GSP2), external to the PC1 binding region of λRL 2.53 clone (Booker, 1990) (Fig. 3.2) and an adaptor primer. PCR reactions were carried out under different conditions, including various annealing temperatures (55°C, 60°C and 65°C) and different amounts of *Taq* DNA polymerase (2U and 4U), to get the longest PCR products which might overlap with the 5'-end of the RACE (Anchor/MRACE II) clone. Multiple bands of different sizes were observed after agarose gel electrophoresis (Fig. 3.4A). Southern blot hybridisation with an oligonucleotide (PC3), internal to PC2 was carried out to verify the specificity of

Figure 3.2 Schematic diagram of overlapping lambda phage clones representing the C-terminal region of the rat PC isolated by Cassady (1987) and Booker (1990) relative to the primary structure of the enzyme based on mouse (Zhang *et al.*, 1993) and human (Wexler *et al.*, 1994; MacKay *et al.*, 1994) PCs. Nucleotide sequences at the 5'-end of λRL 2.53 and those of the conserved sequence at the 5'-end of the ATP/HCO$_3^-$ domain of mouse and human PCs are shown. PC1, PC2, PC3, MRACE I, MRACE II and HUM 19 primer binding sites used for RACE-PCR are indicated by arrows. Different lambda phage clones are boxed. The ATP/HCO$_3^-$, pyruvate and biotinyl domains are also indicated.

```
            G       C AT               AA      A    AC      CC                          C               C
ATGCTGAAGTTCCAAACAGTTCGAGGGGGGCCTGAGGCTCCTGGGTGTCCGCCGATCCTCCTCGGCCCCTGTTGCCTCCCCAAATGTCCGGGGTCTGGAGTATAAGCCTATC
TACGACTTCAAGGTTTGTCAAGCTCCCCCGGACTCCGAGGACCCACAGGCGGCTAGGAGGAGCCGGGGACAACGGAGGGGTTTACAGGCCCCAGACCTCATATTCGGATAG
 1  M L K F Q T V R G G L R L L G V R R S S A P V A S P N V R G L E Y K P I   37
                                                    ◄──────────────────────
                                                           HUM19
```

```
          C                         T   C G C   G      C      C A C      T          G C
AAGAAAGTAATGGTGGCCAACAGAGGTGAGATTGCCATCCGAGTGTTTCGTGCGTGCACAGAGCTGGGTATCCGCACAGTGGCTGTCTACTCGGAGCAGGACACAGGGCAG
TTCTTTCATTACCACCGGTTGTCTCCACTCTAACGGTAGGCTCACAAAGCACGCACGTGTCTCGACCCATAGGCGTGTCACCGACAGATGAGCCTCGTCCTGTGTCCCGTC
 38  K K V M V A N R G E I A I R V F R A C T E L G I R T V A V Y S E Q N T G N   74
               ◄─────────────────
                    MRACEII
```

```
      C       A      T  C C C      C       C C        C                    C                 G CAAC
ATGCACAGGCAGAAAGCTGATGAAGCCTACCTTATTGGCCGTGGCCTGGCACCTGTGCAGGCCTACCTGCACATTCCAGACATCATCAAGGTGGCCAAGGAAAATGGTGTA
TACGTGTCCGTCTTTCGACTACTTCGGATGGAATAACCGGCACCGGACCGTGGACACGTCCGGATGGACGTGTAAGGTCTGTAGTAGTTCCACCGGTTCCTTTTACCACAT
 75  M H R Q K A N E A Y L I G R G L A P V Q A Y L H I P N I I K V A K E N G V   111
                                                                  ◄──────────────────
                                                                        MRACEI
```



```
         ○
H2N ─[ ][ ATP/HCO3⁻ ][ ][ PYRUVATE ][ ][ BIO ][ ]─ COOH
```

λRL 1.1
λRL 2.35
λRL 2.53

```
5'-GAGCTCATCCCCAACATCCCATTCCAGATGCTACTGAGGGGGGGCCAATGCTGTGGGCTACACCAACTACCCTGACAACGTGGTCTTCAAGTTCTGTGAGGTGGCCAAAGAGAATGGCATGGAC-3'
3'-CTCGAGTAGGGGGTTGTAGGGTAAGGTCTACGATGACTCCCCCCCGGTTACGACACCCGATGTGGTTGATGGGACTGTTGCACCAGAAGTTCAAGACACTCCACCGGTTTCTCTTACCGTACCTG-5'
  ◄──────────                          ◄───────                ◄──────────
      PC3                                 PC2                       PC1
                                              EcoRI
```

Figure 3.3 RACE-PCR products amplified with anchor/MRACE II primers using different dilutions of cDNA (UD, undiluted; 1:10, 1 in 10 dilution), were analysed by 2% agarose gel electrophoresis. M, markers.
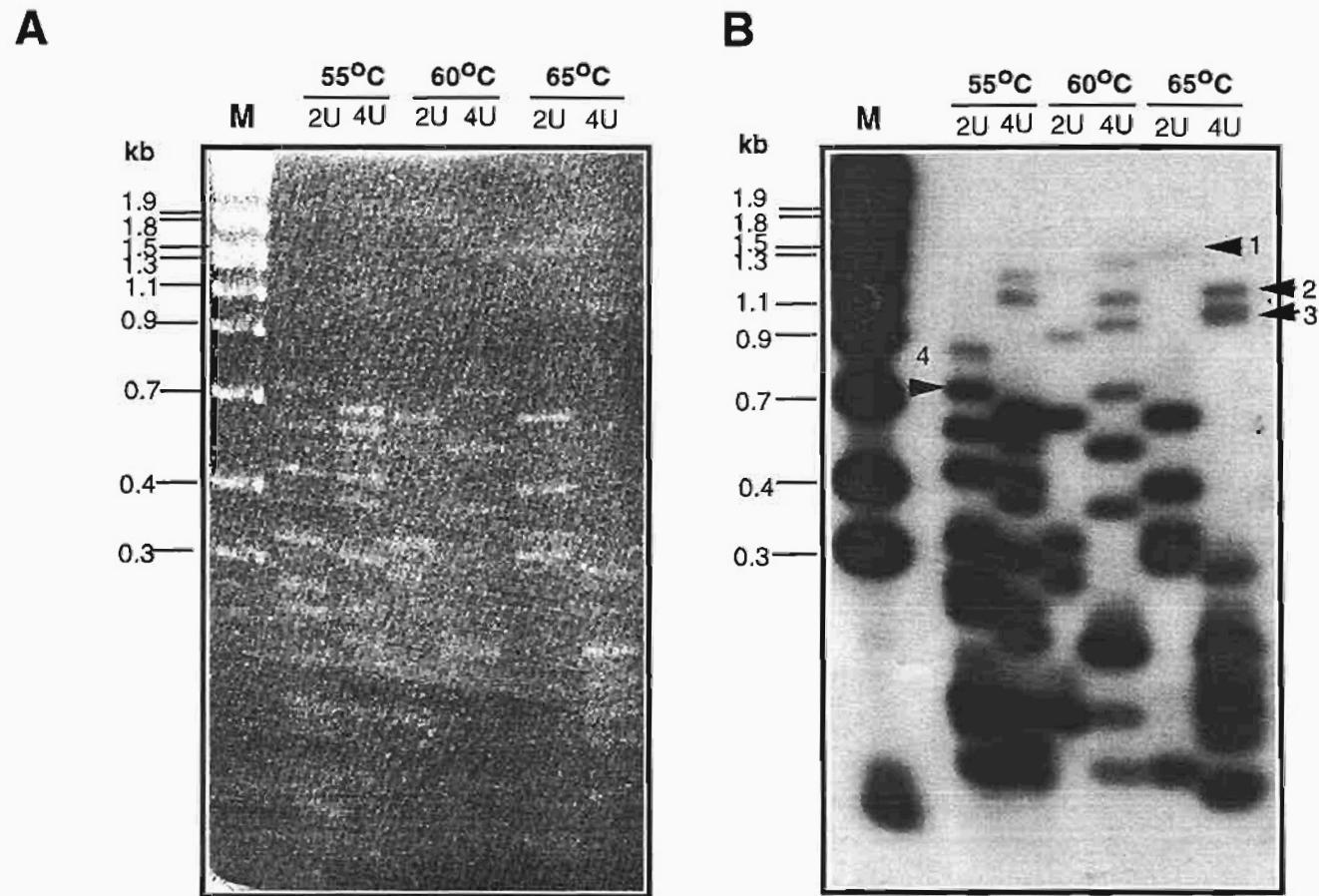
Figure 3.4 A, RACE-PCR products obtained from Anchor/PC1/PC2 primers amplified with different annealing temperatures (55°C, 60°C and 65°C) and different units of *Taq* polymerase, and subsequently analysed by 0.7% agarose gel electrophoresis.

B, Autoradiogram of Southern blot analysis of RACE-PCR products in A, probed with ^{32}P-labelled internal oligonucleotide probe, PC3. Arrows (1 to 4) indicate the fragments that were reamplified and sequenced.

PCR. As indicated in Fig. 3.4B, all fragments seen on the gel were hybridised to the probe, suggesting that they all represented rat PC cDNA.

Direct sequencing of these bands, indicated that they were truncated transcripts possibly due to premature termination during cDNA synthesis. The longest products of size 0.7 kb, 0.9 kb, 1.1 kb and 1.5 kb were individually purified from the gel, reamplified (Fig. 3.5A) and sequenced in both directions as shown in Fig. 3.5B. However, sequencing at the 5'-end of the 1.5 kb product, designated RACE (Anchor/PC2) and the 3'-end of RACE (Anchor/MRACE II) fragments showed that they had not overlapped. Approximately 200 nucleotides were missing from the junction between them, based on the cDNA sequences of mouse (Zhang *et al.*, 1993) and human PC (Wexler *et al.*, 1994; MacKay *et al.*, 1994). This gap was finally filled by performing another RT-PCR. The cDNA synthesis primer, PC7A was designed from the 5'-end of RACE (Anchor/PC2) fragment and used for reverse transcription followed by PCR with an antisense primer (PC8), external to PC7A binding site of RACE (Anchor/PC2) fragment and sense primer (PC9) designed from the 3'-end of RACE (Anchor/MRACE II) fragment. A 0.4 kb product [RT-PCR (PC8/PC9)], which overlapped the above clones, was generated (Fig. 3.6). A summary of the overlapping lambda clones and the PCR fragments spanning an entire rat PC cDNA are shown in Fig. 3.7.

3.3.2 cDNA sequence and inferred amino acid sequence of rat PC

The nucleotide sequences of the RACE (Anchor/MRACE II), RT-PCR (PC8/PC9) and RACE (Anchor/PC2) were determined both directions. Together with nucleotide sequences of λRL 2.53, λRL 2.35 and λRL 1.1 which had previously been sequenced (Cassady, 1987; Booker, 1990), the complete nucleotide sequence spanning the entire coding region of rat PC was obtained. Fig. 3.8 shows the complete nucleotide sequence of the rat PC cDNA encompassing 4024 nucleotides and including a 5'-UTR, coding region, stop codon and 3'-untranslated region. The largest open reading frame was 3537 nucleotides including the stop codon, encoding a polypeptide of 1178 amino acid residues with a calculated M_r of 129,848. The 3'-untranslated region of rat PC is 421 nucleotides
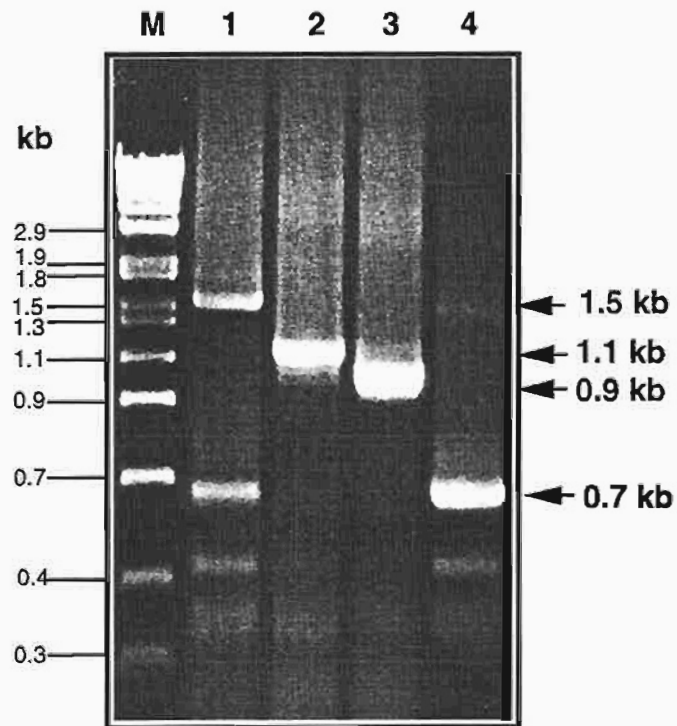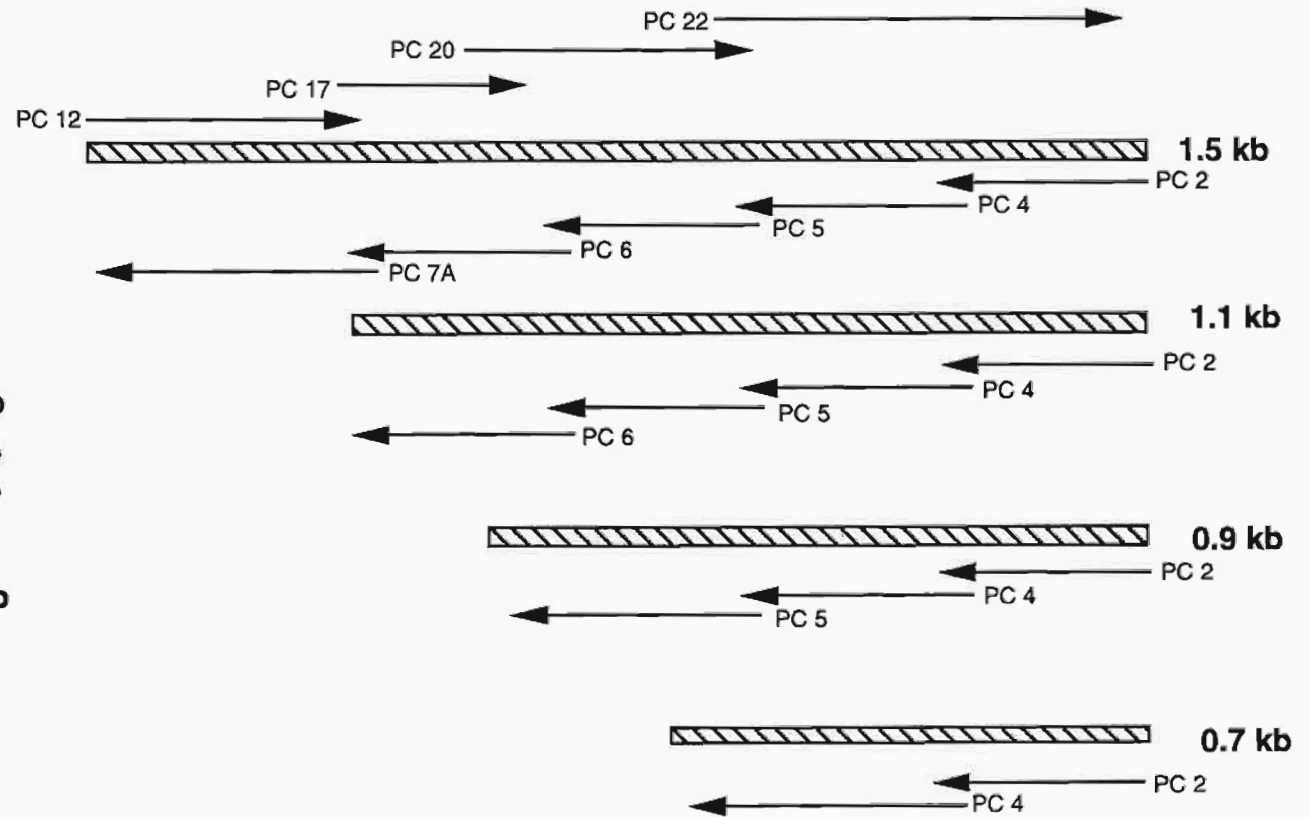
Figure 3.5 A, The 1.5 kb (lane 1), 1.1 kb (lane 2), 0.9 kb (lane 3) and 0.7 kb (lane 4) PCR products shown in Fig. 3.4 were reamplified using the same sets of primers, and analysed by 1.5% agarose gel electrophoresis; M, markers.

B, Sequencing strategy of these 4 PCR products with different primers. Arrows indicate the direction of sequencing.
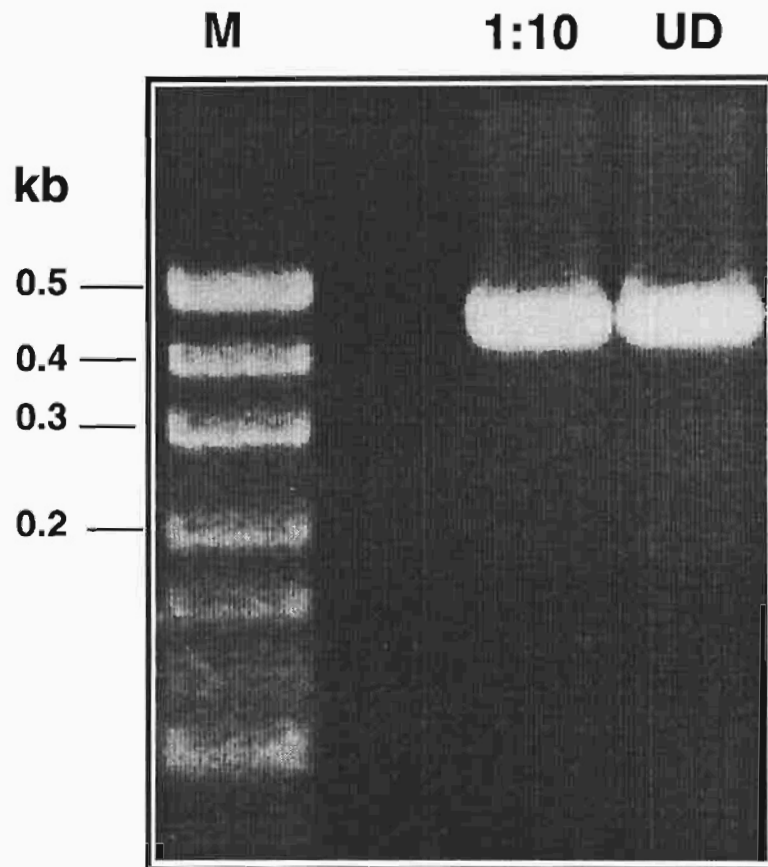
Figure 3.6 RT-PCR products, amplified with PC 8/PC 9 primers using different dilutions of cDNA (UD, undiluted; 1:10, 1 in 10 dilution) were analysed by 2% agarose gel electrophoresis. M, markers.
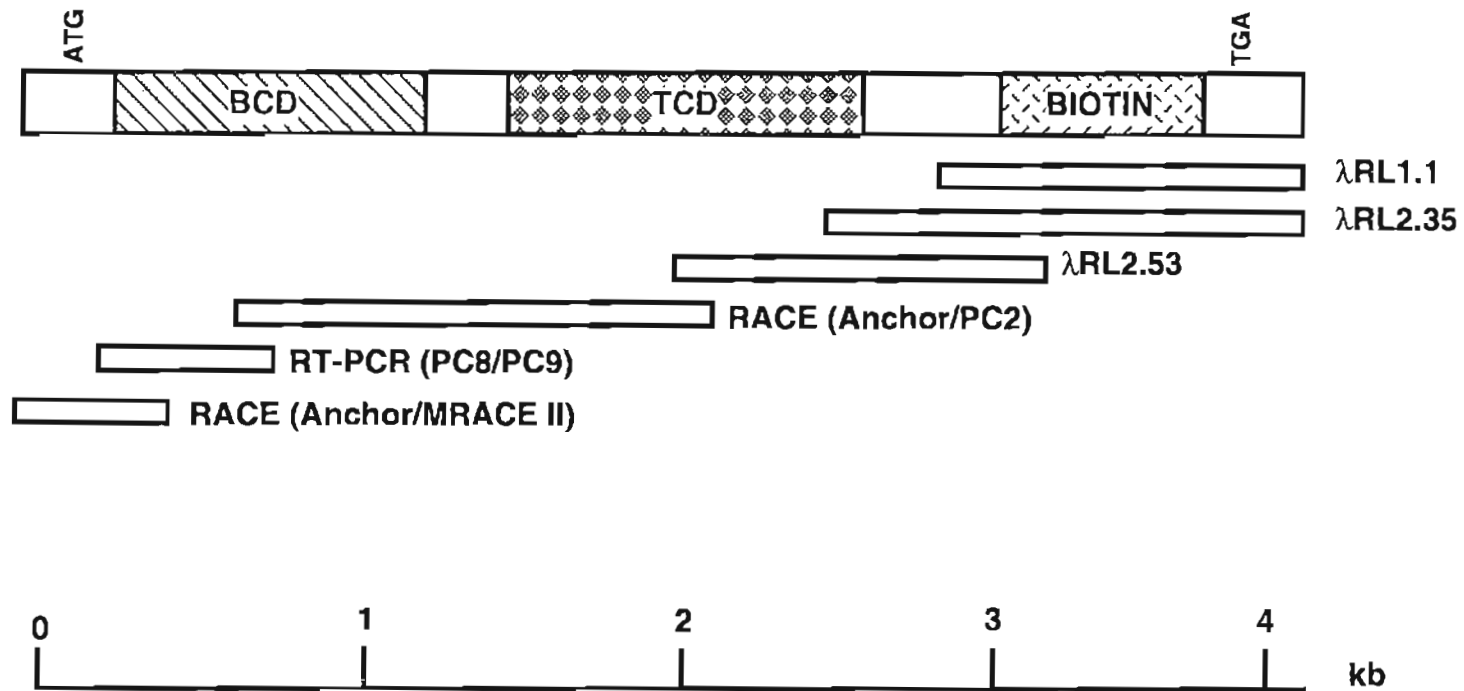
Figure 3.7   Schematic diagram of the isolation of full length rat PC cDNA which was deduced from three overlapping clones, λRL 1.1, λRL2.35 and λRL 2.53 and from three overlapping PCR fragments, RACE (Anchor/PC2), RT-PCR (PC8/PC9) and RACE (Anchor/MRACEII), which extend further to the N-terminal part of PC. The PCR products were named after the primers used in PCR amplification. Three functional domains within the coding region [Biotin carboxylation domain (BCD), transcarboxylation domain (TCD) and the biotinyl domain (BIOTIN)], the ATG start codon, the TGA stop codon, poly(A) tail and the relative length of each clone are shown.

Figure 3.8 Complete nucleotide sequence and inferred amino acid sequence of rat PC. Amino acid residues are numbered at the left; nucleotide sequence positions are numbered at the right. The predicted mitochondrial targeting sequence is shown in the black box. The domain boundaries, identified by highly significant degrees of similarity with other biotin enzymes, are indicated by square brackets. The conserved putative metal-binding motif (HXHXH) at residues 769-773 is identified by asterisks. The amino acid residues, which were identified by Edman degradation sequence analysis of the Mr approximately 75,000 chymotryptic fragment (Booker, 1990), are shown in a grey box. The putative carboxyl binding site (EXWGGATFDVAMRFLYECPWXRL) (Samols *et al.*, 1988) within the transcarboxylation domain is underlined. The biotin attachment site (AMKM) is in bold type.

```
                              GTTCCCTGTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG

      ATGCTGAAGTTCCAAACAGTTCGAGGGGGCCTGAGGCTCCTGGGTGTCCGCCGATCCTCCACAGCCCCCGTTGCCTCCCCAAATGTCCGG   90
  1    M  I  K  F  Q  I  V  P  G  G  L  R  L  L  G  V  R  P  S  S  T  A  P  V  A  S  P  N  V  R

      CGTCTGGAGTACAAGGCCATCAAGAAAGTAATGGTGGCCAACAGAGGTGAGATTGCCATCCGAGTGTTTCGTGCCTGCACAGAGCTGGGT  180
 31    R  L  E  Y  K [P  I  K  K  V  M  V  A  N  R  G  E  I  A  I  R  V  F  R  A  C  T  E  L  G

      ATCCGCACAGTGGCTGTCTACTCGGAGCAGGACACAGGCCAGATGCACCGGCAGAAAGCTGATGAAGCCTACCTTATTGGCCGTGGGCTG  270
 61    I  R  T  V  A  V  Y  S  E  Q  D  T  G  Q  M  H  R  Q  K  A  D  E  A  Y  L  I  G  R  G  L

      GCTCCTGTGCAAGCCTACCTGCACATTCCAGACATCATTAAGGTGGCCAAGGAGAATGGTGTAGATGCTGTGCACCCTGGCTATGGGTTC  360
 91    A  P  V  Q  A  Y  L  H  I  P  D  I  I  K  V  A  K  E  N  G  V  D  A  V  H  P  G  Y  G  F

      CTCTCAGAGAGAGCAGACTTGCCCAGGCCTGCCAAGATGCTGGAGTCCGATTCATTGGTCCAAGCCCAGAGGTGGTCCGCAAGATGGGA  450
121    L  S  E  R  A  D  F  A  Q  A  C  Q  D  A  G  V  R  F  I  G  P  S  P  E  V  V  R  K  M  G

      GACAAGGTGGAAGCCCGGGCCATTGCCATTGCTGCAGGAGTTCCAGTGGTCCCTGGCACTAATTCCCCCATCAATTCCCTGCATGAGGCA  540
151    D  K  V  E  A  R  A  I  A  I  A  A  G  V  P  V  V  P  G  T  N  S  P  I  N  S  L  H  E  A

      CACGAGTTCTCTAACACCTATGGTTTCCCTATTATCTTCAAGGCTGCCATGGAGGTGGGGGGCCGTGGCATGAGGGTTGTGCATAGCTAC  630
181    H  E  F  S  N  T  Y  G  F  P  I  I  F  K  A  A  Y  G  G  G  G  R  G  M  R  V  V  H  S  Y

      GAGGAGCTGGAAGAGAATTACACCCGGGCCTACTCTGAGGCCTTGGCAGCCTTGGGAATGGGGCATTGTTTGTGGAGAAATTCATTGAG  720
211    E  E  L  E  E  N  Y  T  R  A  Y  S  E  A  L  A  A  F  G  N  G  A  L  C  V  E  K  F  E  E

      AAGCCAAGACACATTGAGGTGCAGATCCTAGGGGACCAATATGGGAACATCTTGCACTTGTATGAGCGGGACTGCTCCATCCAGCGGCG  810
241    K  P  R  H  I  E  V  Q  I  L  G  D  Q  Y  G  N  I  L  H  L  Y  E  R  D  C  S  I  Q  R  R

      CACCAGAAGGTCGTCGAGATTGCCCCTGCTACCCACTTGGACCCCCAACTTCGGTCACGCCTCACCAGTGACTCTGTCAAACTTGCCAAG  900
271    H  Q  K  V  V  E  I  A  P  A  T  H  L  D  P  Q  L  R  S  R  L  T  S  D  S  V  K  L  A  K

      CAGGTTGGCTATGAGAATGGCAGGCACTGTCGAGTTCCTGGTGGACAAGCATGGCAAGCACTACTTCATCGAGGTCAATTCCCGCCTGCAG  990
301    Q  V  G  Y  E  N  A  G  T  V  E  F  L  V  D  K  H  K  H  Y  F  I  E  V  N  S  R  L  Q

      GTGGAGCACACGGTCACTGAGGAGATTACAGATGTGGACCTGGTCCATGCTCAGATCCATGTGTCTCCGAAGGCCGGAGCCTGCCTGACCTG 1080
331    V  E  H  T  V  T  E  E  I  T  D  V  D  L  V  H  A  Q  I  H  V  S  E  G  R  S  L  P  D  L

      GGGCTGCGGCAGGAAAACATCCGAATCAATGGTTGTGCCATTCAGTGTCGGGTCACCACTGAGGACCCTGCACGCAGCTTCCAGCCAGAC 1170
361    G  L  R  Q  E  N  I  R  I  N  G  C  A  I  Q  C  R  V  T  T  E  D  P  A  R  S  F  Q  P  D

      ACTGGCCGCATTGAGGTTTTCCGGAGTGGTGAGGGCATGGGCATCCGCCTGGACAATGCCTCAGCATTCCAGGGAGCTGTCATATCCCCC 1260
391    T  G  R  I  E  V  F  R  S  G  E  G  M  G  I  R  L  D  N  A  S  A  F  Q  G  A  V  I  S  P

      CACTATGACTCCCTGCTCGTCAAAGTCATTGCCCATGGCAAAGACCACCCTACAGCTGCCACCAAGATGAGCAGAGCCCTGGCGGAGTTC 1350
421    H  Y  D  S  L  L  V  K  V  I  A  H  G  K  D  H  P  T  A  A  T  K  M  S  R  A  L  A  E  F

      CGTGTCCGGAGGTGTAAAGACCAACATCCCCTTCCTGCAGAATGTTCTCAACAACCAGCAGTTCCTAGCGGGCATTGTGGACACCCAGTTC 1440
451    R  V  R  G  V  K  T  N  I  P  F  L  Q  N  V  L  N  N  Q  Q  F  L  A  G  I  V  D  T  Q  F

      ATCGATGAGAACCCCGAGCTGTTCCAGCTGCGGCCTGCACAGAACCGGGCCCAGAAGTTGCTACATTACCTTGGACACGTCATGGTCAAT 1530
481    I  D  E  N  P  E  L  F] Q  L  R  P  A  Q  N  R  A  Q  K  L  L  H  Y  L  G  H  V  M  V  N

      GGCCCTACCACTCCAATCCCCGTCAAGGTCAGTCCCAGCCCTGTGGACCCCATTGTTCCTGTGGTGCCCATAGGCCCACCCCCAGCTGGT 1620
511    G  P  T  T  P  I  P  V  K  V  S  P  S  P  V  D  P  I  V  P  V  V  P  I  G  P  P  P  A  G

      TTCAGAGACATCCTTCTGCGAGAGGGGCCAGAGGGCTTTGCCAGAGCTGTGCGGAATCACCAGGGGCTGCTGCTAATGGACACAACTTC 1710
541    F  R  D  I  L  L  R  E  G  P  E  G  F  A  R  A  V  R  N  H  Q  G  L  L  L  M  D  T  T  F

      CGGGATGCCCACCAGTCACTACTTGCCACTAGAGTGCGCACACACGATCTCAAAAAGATTGCACCCTACGTTGCCCACAACTTCAACAAC 1800
571    R  D  A  H  Q  S  L  L  A  T  R  V  R  T  H  D  L  K  K  I  A  P  Y  V  A  H  N  F  N  N

      CTCTTCAGCATAGAGAACTGGGGAGGAGCCACATTTGACGTGGCCATGCGCCTTCTTGTATGAGTGCCCCTGGCGGCGGCTCCAGGAGCTC 1890
600    L [F  S  I  E  N  W  G  G  A  T  F  D  V  A  M  R  F  L  Y  E  C  P  W  R  R  L  Q  E  L

      CGGGAGCTCATCCCCAACATCCCATTCCAGATGCTACTGAGGGGGCAATGCTGTGGGCTACACCAACTACCCTGACAACGTGGTCTTC 1980
631    R  E  L  I  P  N  I  P  F  Q  M  L  L  R  G  A  N  A  V  G  Y  T  N  Y  P  D  N  V  V  F

      AAGTTCTGTGAGGTGGCCAAAGAGAATGGCATGGACGTCTTCCGGATCTTTGACTCCCTTAACTACCTGCCAAACATGCTGCTGGGCATG 2070
661    K  F  C  E  V  A  K  E  N  G  M  D  V  F  R  I  F  D  S  L  N  Y  L  P  N  M  L  L  G  M

      GAAGCAGCTGGCAGTGCTGGGGGTGTGGTGGAAGCTGCCATCTCCTACACGGGTGACGTGGCTGACCCCAGTCGCACTAAATACTCACTG 2160
691    E  A  A  G  S  A  G  V  V  E  A  A  I  S  Y  T  G  D  V  A  D  P  S  R  T  K  Y  S  L

      GAGTACTACATGGGCTTAGCTGAAGAACTGGTGCGAGCCGGCACTCACATCCTCTGCATTAAGGACATGGCAGGCCTGCTGAAGCCTGCA 2250
721    E  Y  Y  M  G  L  A  E  E  L  V  R  A  G  T  H  I  L  C  I  K  D  M  A  G  L  L  K  P  A

      GCATGCACCATGCTGGTCAGCTCCCTCCGGGACCGGTTCCCCGACCTCCCACTGCACATCCATACCCATGACACATCAGGGTCAGGTGTG 2340
751    A  C  T  M  L  V  S  S  L  R  D  R  F  P  D  L  P  L  H  *   *  H  T  H  *  T  S  G  S  G  V

      GCAGCCATGTTGGCCTGTGCACAAGCTGGGGCTGATGTTGGATGTGGCAGTCGACTCTATGTCTGGATGACCTCACAGCCCAGCATG 2430
781    A  A  M  L  A  C  A  Q  A  G  A  D  V  V  D  V  A  V  D  S  M  S  G  M  T  S  Q  P  S  M

      GGGGCCCTGGTGGCCTGTACCAAAGGGACTCCTCTGGACACAGAGGTACCCCTGGAGCGTGTGTTTGACTACAGTGAGTATTGGGAAGGC 2520
811    G  A  L  V  A  C  T  K  G  T  P  L  D  T  E  V  P  L  E  R  V  F  D  Y  S  E  Y  W  E  G

      GCTCGGGGGCTGTATGCAGCCTTTGATTGCACGGCTACCATGAAGTCTGGCAACTCAGACGTGTATGAGAATGAGATCCCAGGGGGCCAG 2610
841    A  R  G  L  Y  A  A  F  D  C  T  A  T  M  K  S  G  N  S  D  V  Y  E  N  E  I  P  G  G  Q

      TACACCAACCTACACTTCCAGGCCCACAGCATGGGACTTGGCTCCAAGTTCAAGGAGGTCAAGAAGGCCTATGTGGAGGCTAACCAGATG 2700
871    Y  T  N  L  H  F  Q  A  H  S  M  G  L  G  S  K  F  K  E  V  K  K  A  Y  V  E  A  N  Q  M

      CTGGGGGACCTCATCAAGGTGACACCATCCTCCAAGATTGTGGGGGACTGGCCCAGTTCATGGTGCAGAACGGGTTGAGCCGGGCAGAG 2790
901    L  G  D  L  I  K  V  T  P  S  S  K  I  V  G  D  L  A] Q  F  M  V  Q  N  G  L  S  R  A  E

      GCAGAAGCTCAGGCAGAAGAGCTGTCCTTCCCCCGCTCTGTGGTGGAGTTCCTGCAGGGCTACATTGGCATTCCCCATGGGGGTTTCCCT 2880
931    A  E  A  Q  A  E  E  L  S  F  P  R  S  V  V  E  F  L  Q  G  Y  I  G  I  P  H  G  G  F  P

      GAACCCTTCCGGTTCTAAGGTGCTAAAGGACCTGCCAAGAATAGAAGGGAGGCCTGGAGCCTCCCTCCCTCCCTTGAACCTGAAGGAGCTG 2970
961    E  P  F  R  S  K  V  L  K  D  L  P  R  I  E  G  R  P  G  A  S  L  P  P  L  N  L  K  E  L

      GAGAAGGACCTGATTGATAGGCATGGAGAGGAGGTGACCCCAGAGGACGTTCTCTCTGCAGCCATGTACCCTGATGTCTTTGCTCAGTTC 3060
991    E  K  D  L  I  D  R  H  G  E  E  V  T  P  E  D  V  L  S  A  A  M  Y  P  D  V  F  A  Q  F

      AAAGACTTCACGGCTACCTTTGGCCCCCTGGATAGCCTGAATACTCGTCTCTTTCTTCAAGGACCCAAAATTGCAGAGGAGTTTGAGGTT 3150
1021   K  D  F  T  A  T  F  G  P  L  D  S  L  N  T  R  L  F  L  Q  G  P  K  I  A  E  E  F  E  V

      GAGCTGGAACGGGGCAAGACCTTGCACATCAAAGCCCTGGCTGTAAGCGACCTGAACCGTGCTGGCCAGAGGCAGGTGTTCTTTGAACTC 3240
1051   E  L  E  R  G  K  T  L  H  I  K  A  L  A  V  S  D  L  N  R  A  G  Q  R  Q  V  F  F  E  L

      AATGGGCAGCTTCGATCCATTCTGGTTAAAGACACCCAGGCCATGAAGGAGATGCACTTCCATCCCAAGGCCTTGAAGGATGTGAAGGGC 3330
1081   N  G  Q  L  R  S  I  L  V  K  D  T  Q  A  M  K  E  M  H  F  H  P  K  A  L  K  D  V  K  G

      CAAATTGGGGCCCCTATGCCTGGGAAGGTCATAGACGTCAAGGTGGCAGCAGGAGCCAAGGTGGTTAAGGGCCAGCCCCTCTGTGTGCTC 3420
1111   Q  I  G  A  P  M  P  G  K  V  I  D  V  K  V  A  A  G  A  K  V  V  K  G  Q  P  L  C  V  L

      AGCGCCATGAAGATGGAGACTGTGGTGACTTCGCCCATGGAGGGCACTATCCGAAAGGTTCACGTGACCAAGGACATGACTCTGGAAGGC 3510
1141   S  A  M  K  M  E  T  V  V  T  S  P  M  E  G  T  I  R  K  V  H  V  T  K  D  M  T  L  E  G

      GATGACCTCATCCTAGAGATTGATGATCTTACTCCAGACTGGCAGCCTGGCCAACCCTACCCCAAGCCTCTCAACAGAAGCTGTGCAGCC 3600
1171   D  D  L  I  L  E  I  E]

      AGGGCAGGCCCAGGCAGTACCTGAGCTAGCCCTTGAGGTCCTGTCCCATGGGACAGCACACACACTACCTGCAATGGCCCTCCCATTCCC 3690
      TTCAGCTATTTGTCCTTGTCTTGCTGGCAAGGCAGTTCTCACATATTCATCTCTTGCCCAAATAAGGGTCTGCTCCTCGTGGGAGACCACA 3780
      GGTGTACAGTAGGTGGCCTTGTACCTGGGAGAGGGGTTCTACCTCTGGGGGTAGAGGGAAGAAGACCTAATTCATAGGTCCTGGGAAATT 3870
      TGCTCAATAAAAGTGGCCTTCCCTTGCCCTCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4024
```

long with a consensus polyadenylation signal (AATAAA) located 27 nucleotides 5' to a long poly (A) tail.

As a nuclear-encoded protein (Freytag *et al.*, 1984; Walker *et al.*, 1995a; Webb *et al.*, 1997), PC is synthesised as a precursor protein in the cytosol and translocated to the mitochondrial matrix with a consequent reduction of molecular weight (Srivastava *et al.*, 1983). Many imported mitochondrial proteins are synthesized with an N-terminal leader sequence of 20-80 amino acid residues which typically contain several positively charged and several hydroxylated amino acids and no acidic amino acids (Hendrick *et al.*, 1989), as is also the case with rat PC. The mature N-terminal sequence of the rat PC was previously reported from peptide sequencing to be Ser-Gly-Pro-Val-Ala-Pro-Leu-Asn-Val-Leu-Leu-Leu-Glu-Tyr-Pro (Thampy *et al.*, 1988). This report seems to conflict with previous studies which have shown that the N-terminus of mature PC of other species were N-terminally blocked (Rylatt *et al.*, 1977; Lim *et al.*, 1988). In addition, this sequence was not found in the amino acid sequence inferred from the cDNA and nor does it exhibit any significant sequence identity with any biotin-dependent enzyme in the non-redundant databases. The proteolytic cleavage at the N-terminus of premature PC during the translocation to mitochondria is likely to occur between Ser 20 and Thr 21. Cleavage at this site is a common feature of most mitochondrially imported proteins with their targeting sequence always ending two residues after an arginine (Hendrick *et al.*, 1989). The mitochondrial targeting sequence of rat PC showed high sequence similarity with mouse (Zhang *et al.*, 1993) and human (Wexler *et al.*, 1994; MacKay *et al.*, 1994), namely 95% and 80% respectively. The inferred protein sequence of rat PC is 96.6% identical with mouse (Zhang *et al.*, 1993) and 96.3% identical with human PCs (Wexler *et al.*, 1994; MacKay *et al.*, 1994), 68.4% identical with mosquito PC (Tu and Hagedorn, 1997) and 53.3% identical with yeast PC isoenzyme PC1 (Lim *et al.*, 1988) and PC2 (Stucka *et al.*, 1991; Val *et al.*, 1995).

Comparison of the entire primary structure of rat PC to PC from other species, revealed that rat PC also contained three functional domains i.e. the N-terminal region (residues 1-488), central region (residues 601-918) and C-terminal region (residues 1103-

1178) corresponding to the ATP/HCO$_3^-$ binding domain, pyruvate-binding domain and the biotin carrier domain, respectively.

While this part of the study was submitted for publication in the *Biochemical Journal* in early November, 1995, Lehn and coworkers independently published rat PC cDNA sequence in late November in the same year (Lehn *et al.*, 1995). However, three residue differences were observed in the above report including Pro^{222} (Ser in this thesis), Asp^{866} (Ile) and Gly^{977} (Arg). These residues reported in this thesis however are highly conserved across mammalian species (Jitrapakdee and Wallace, 1999).

3.4 DISCUSSION

3.4.1 Domain structure of rat PC

Although biotin-dependent carboxylases participate in different arrays of metabolic pathways, they share many common feature of their reaction mechanisms, which involve two partial reactions: the carboxylation of the covalently attached biotin moiety and transcarboxylation from the carboxy-biotin to an acceptor molecule. Comparison of primary structures of PC to other biotin-dependent enzymes would be important to identify some common motifs that might be catalytically or structurally significant. Furthermore comparison of the primary structure of PC to the known 3-dimensional structure of any biotin carboxylase or other homologue proteins would enable prediction of the general overall structure of PC.

3.4.1.1 Biotin carboxylase (BC) domain

Multiple sequence alignment of the N-terminal region of PC from rat, human (Wexler *et al.*, 1994; MacKay *et al.*, 1994; Walker *et al.*, 1995b), mosquito (Tu and Hagendorn, 1997), bacterial (Dunn *et al.*, 1996, Kunst *et al.*, 1997), α-subunit of propionyl-CoA carboxylase (Browner *et al.*, 1989), yeast ACC (Al-Feel et al., 1992), the biotin carboxylase (BC) subunit of *E. coli* ACC (Li and Cronan, 1992a) and carbamoyl phosphate synthetase from *E. coli* (Nyunoya and Lusty, 1983) and yeast (Lusty *et al.*, 1983) were made and is shown in Fig. 3.9. Also shown are the secondary structure elements identified from the X-ray crystal structure of the *E. coli* BC subunit of ACC (Waldrop *et al.*, 1994). It is clear from this sequence alignment that these sequences are highly conserved and align well with the secondary structure of BC, implying that overall folding of these proteins is likely to be conserved. This region of yeast PC had originally been shown to have significant homology to other ATP-binding proteins, the carbamyl phosphate synthetase and ACC (Lim *et al.*, 1988). At the time no other complete cDNA sequences of other biotin carboxylases were available and the 3-D structure of BC subunit of *E. coli* ACC and carbamoyl phosphate synthetase had not been determined. The alignment made in this chapter extends the boundary of previously described homologous regions (Lim *et al.*, 1988) to the N-terminus of the *E. coli* BC subunit, which corresponds to residue 33 of the inferred sequence of rat PC. The proposed boundary between the BC and transcarboxylase (TC) domains is

consistent with the results of limited proteolysis experiments. Pure native rat PC, when digested with chymotrypsin, is cleaved into two fragments (Booker, 1990). The N-terminal sequence of the large fragment (Mr~75,000 Da) containing the biotin moiety was determined as Gln-Leu-Arg-Pro-Ala-Gln-Asn-Arg-Ala-Gln-Lys-Leu-Leu-His-Tyr-Leu-Gly (Booker, 1990). This N-terminal sequence matches exactly to the residues inferred from cDNA-derived sequeces for residues Gln 489 to Gly 505, and places the chymotrypsin-labile peptide bond between Phe 488 and Gln 489 (Fig. 3.8). This corresponds to the C-terminal extent of the sequence conservation and the structure of the BC subunit of *E. coli* ACC. It seems not unreasonable to call this conserved structural and functional unit within PC as the 'BC domain' since it probably forms the first partial reaction site.

The residues highlighted as white text in Fig. 3.9 are those that are likely to play an important role in catalysis. The positions of these residues within the structure of the BC subunit of the *E. coli* ACC are shown in Fig. 3.10. Waldrop *et al.* (1994) identified the active site pocket containing His^{256} (equivalent to His^{271} in rat PC), Lys^{238} (Lys^{273}), Glu^{276} (Glu^{311}), Glu^{288} (Glu^{324}), Asn^{290} (Asn^{326}), Arg^{292} (Arg^{328}), Gln^{294} (Gln^{330}), Val^{295} (Val^{331}), Glu^{196} (Glu^{332}) and Arg^{338} (Arg^{377}) as these residues are appropriately placed to interact with an enzyme-bound biotin or phosphate molecule. Previous studies proposed a number of ionisable groups on the amino acid side chains of PC which would form part of the active site to enolise biotin and thus render it capable of fixing CO_2 to a biotin molecule (Attwood and Cleland, 1986; Tipton and Cleland, 1988; Attwood, 1995). Recent chemical modification studies have implicated the presence of an essential cysteine-lysine ion pair of chicken PC (Werneburg and Ash, 1993). Modification with *o*-pthalaldehyde resulted in inactivation of the first partial reaction (Wernberg and Ash, 1993). Although the cDNA sequence for chicken PC is not available, the sulphydryl group of Cys^{230} (Cys^{265} in rat) is 4.2 angstroms from the ε-amino group of Lys^{238} (Lys^{273} in rat), which is sufficiently close to allow for cross-linking of these two residues by *o*-pthalaldehyde. Indeed, these Cys and Lys residues are invariant within all other biotin carboxylases shown in Fig. 3.9 and are likely to play a role in the enolisation of biotin in the first partial reaction. Mutation of these Cys and Lys individually to Ala in yeast PC1 isozyme dramatically diminished its activity, confirming the crucial role of these two residues in the catalytic reaction (Nezic and Wallace,

Figure 3.9 Multiple sequence alignment of the biotin carboxylation domain of PC to other biotin-dependent enzymes. The amino acid sequences from a representative selection of eukaryotic and prokaryotic examples of PC and other biotin-dependent enzymes and other enzymes shown by ψ-BLAST search (Altschul *et al.*, 1997) to be related, were compared using Clustal W. (Thompson *et al.*, 1994). The highly conserved amino acid residues found in different groups of enzymes are showed by shaded boxes. The open boxes represent the residues of ACC and CPS of *E. coli* superimposed using the Homology/Insight program [Molecular Simulations Inc., San Diego, CA] (see Fig. 3.11) The Cys and Lys pair is indicated by asterisks. Also shown are the α-helices (α) and β-strands (β) observed in the X-ray crystal structure of the BC subunit of *E. coli* ACC (Waldrop *et al.*, 1994) Sources: PC *H. sapiens* (human) [Wexler *et al.*, 1994; MacKay *et al.*, 1994; Walker *et al.*, 1995b], PC *Aedes aegypti* (mosquito) [Tu and Hagedorn, 1997], *S. cerevisiae* PC1 [Lim *et al.*, 1988], PC *Bacillus subtilis* [Kunst *et al.*, 1997], PC *Rhizobium etli* [Dunn *et al.*, 1996], PCC (Propionyl-CoA carboxylase α–subunit) rat [Browner *et al.*, 1989]; ACC *S. cerevisiae* [Al-Feel *et al.*, 1992]; ACC *E. coli* [Li and Cronan, 1992a]; CPS *E. coli* [Nyunoya and Lusty, 1983], CPS *S. cerevisiae* [Lusty *et al.*, 1983].

```
PC   H.sapiens       33  -EYKP-IKKVMVANR----------GEIAIRVFRAC---------TELGIRTVAIYSEQDTGQ--MHRQK--ADEAYLIGRG-------LAPVQAYLHIPDIIKVAKENNVDAVHPGYG--- 119
PC   R.norvegicus    33  -EYKP-IKKVMVANR----------GEIAIRVFRAC---------TELGIRTVAVYSEQDTGQ--MHRQK--ADEAYLIGRG-------LAPVQAYLHIPDIIKVAKENGVDAVHPGYG--- 119
PC   A.aegypti       34  -EYKP-IRSVLVANR----------GEIAIRVFRAC---------NELGIRSVAVYSEQDKH--MHRQK--ADESYMVGKG-------LAPVEAYLSIPEIIRVCKENDVDAVHPGYG--~ 120
PC   S.cerevisiae    21  --KILVANR----------GEIPIRIFRTA---------HELSMQTVAIYSHEDRLS--THKQK--ADEAYVIGEV-----GQYTPVGAYLAIDEIISIAQKHQVDFIHPGYG--- 103
PC   B.subtilis       1  MSQQS-IQKVLVANR----------GEIAIRVFRAC---------TELNIRTVAVYSKEDSGS--YHRYK--ADEAYLVGEG-------KPIDAYLDIEGIIDIAKRNKVDAIHPGYG--- 88
PC   R.etli           1  MP-ISKILVANR----------SEIAIRVFRAA---------NELGIKTVAIWAEEDKLA--LHRFK--ADESYQVGRGPHLAR-DLGPIESYLSIDEVIRVAKLSGADAIHPGYG--- 91
PCC  R. norvegicus   38  T-FDKILIANR----------GEIACRVIKTC---------RKMGIRTVAIHSDVDASS--VHVKM--ADEAVCVGP-------APTSKSYLNMDAIMEAIKKTGAQAVHPGYG--- 120
ACC  S.cerevisiae    58  -V-ISKILIANN----------GIAAVKEIRSVRKWAYETFGDDRTVQFVAMATPEDLEANAEYIRM--ADQYIEVPGG-------TNNNNYANVDLIVDIAERADVDAVWAGWG--- 152
ACC  E.coli           1  M-LDKIVIANR----------GEIALRILRAC---------KELGIKTVAVHSSADRDL--KHVLL--ADETVCIGPA-------PSVKSYLNIPATISAAEITGAVAIHPGYG--- 83
CPS  E.coli           1  MPKRTD-IKSILIUGAGPIVIGQACEFDYSGAQACKAL---------REEGYRVILVN-SNFAT--IMTDPEMADATYIE---------------PIHWEVVRKIIEKERPDAVLPTMGGQT 94
CPS  S.cerevisiae    21  -KPQLVEGVNSVLVIGSGGLSIGQAGEFDYSGSQAIKAL---------KEDNKFTILVN-PNIAT--NQTSHSLADKIYYL--------------PVTPEYITYIIELERPDAILLTFGGQT 115
                                        ββββ              αααααααααααα              βββββββ            αααααα      βββββββ                αααααααα        βββ

PC   H.sapiens      120  --FLSERADFAQACQDAG--VRFIGPSPEVVRKMGDKVEARAIAIAAGVPVVPGTEGP-----------------------------V-ERYEEVKAFTDTYGFPIIKAAFGGGGRGMRVVHSYEELEEN 216
PC   R. norvegicus  120  --FLSERADFAQACQDAG--VRFIGPSPEVVRKMGDKVEARAIAIAAGVPVVPGTNSP-----------------------------I-NSLHEAHEFSNTYGFPIIFKAAYGGGGRGMRVVHSYEELEEN 216
PC   A.aegypti      121  --FLSERSDFAQAVIDAG--LRFIGPSPKVVQQMGDKVAARKAAIEAGVPIVPGTDGP-----------------------------V-TTKEEALDFCKKHGLPVIFKAAYGGGGRGMRVVRKMEEVEDN 217
PC   S.cerevisiae   104  --FLSEADKVKAG--ITWIGPSPAEVIDSVGDKVSARNLAAKANVPTVPGTGP------------------------------I-ETVEEALDFVNEYGYPVIIKAAFGGGGRGMRVVREGDDVADA 200
PC   B.subtilis      89  --FLSENIHFARRCEEEG--IVFIGPKSEHLDMFGDKVKAREQAEKAGIPVIPGSDG------------------------------A-ETLEAVEQFGQANGYPIIIKASLGGGGRGMRIVRSESEVKEA 185
PC   R.etli          92  --LLSESPEFVDACNKAG--IIFIGPKADTMRQLGNKVAARNLAISVGVPVVPATEPL------------------------------P-DDMAEVAKMAAAIGYPVMLKASWGGGGRGMRVIRSEADLAKE 188
PCCα R. norvegicus 121  --FLSENKEFAKCLAAED--VTFIGPDTHAIQAMGDKIESKLLAKRAKVNTIPGFDGV-----------------------------L-KDADEAVRIAREIGYPVMIKASAGGGGKGMRIPWDDEETRDG 217
ACC  S.cerevisiae  153  --HASENPLLPEKLSQSKRKVIFIGPPGNAMRSLGDKISSTIVAQSAKVPCIPWSGTGVDTVHVDEKTGLVSVDDDIYQKGCCTSPEDGLQKAKRIGFPVMIKASEGGGGKGIRQVEREEDFIAL 275
ACC  E.coli         84  --FLSENANFAEQVERSG--FIFIGPKAETIRLMGDKVSAIAAMKRAGVPCIPGSDGP------------------------------LGDDMDKNRAIAKRIGYPVIIKASGGGGGRGMRVVRGDAELAQS 181
CPS  E. coli         95  ALNCALELERQGVLEEFG--VTMIGATADAIDKAEDRRRFDVAMKKIGLETARSG------------------------------IAHTMEEALAVAADVGFPCIIRPSFTMGGSGGGIAYNREEFEEI 191
CPS  S.cerevisiae   116  GLNCGVALDESGVLAKYN--VKVLGTPIKTLITSEDRDLFASALKDINIPIAESF------------------------------ACETVDEALEAAERVKYPVIVRSAYALGGLGSGFANNASEMKEL 212
                                       ααααααααααααα      ββββ        ααααααααααααααααααααααα                             αααααααααααα             ββββββ    αααααα

PC   H.sapiens      217  YTRAYSEAWPAFGNGALFVEKFIEKPRHIEVQILGDQYGNILHLYERDCSIQRRHQKV--VEIAPAAHLDPQLRTRLTSDSVKLAKQVGY-ENAGTVEFL--VDRH-GKHYFIEVNSRLQVEHTV 335
PC   R. norvegicus  217  YTRAYSEAWPAFGNGALFVEKFIEKPRHIEVQILGDQYGNILHLYERDCSIQRRHQKV--VEIAPATHLDPQLRSRLTSDSVKLAKQVGY-ENAGTVEFL--VDKH-GKHYFIEVNSRLQVEHSV 335
PC   A.aegypti      218  FQRASSEAKAAFGNGAMFIEKFIERPRHIEVQLLGDKAGNVVHLYERDCSVQRRHQKV--VEIAPAPRLPREVRDKMTEYAVKLAKHVGY-ENAGTVEFL--CDES-GNFYFIEVNARLQVEHTV 336
PC   S.cerevisiae   201  FQRATSEARTAFGNGTCFVERFLDKPKHIEVQLLADNHGNVVHLFERDCSVQRRHQKV--VEVAPAKTLPREVRDAILTDAVKLAKECGY-RNAGTAEFL--VDNQ-NRHYFIEINPRIQVEHTI 319
PC   B.subtilis     186  YERAKSEAMAAFGNDEVYVEKLIENPKHIEVQVIGDKYGNVVHLYERDCSVQRRHQKV--IEVAPSVSLSPELRDQICEAAVALAKNVMY-INAGTVEFL--VAN--NEFYFIEVNPRVQVEHTI 303
PC   R.etli         189  VTEAKREAMAAFGKDEVYLEKVERARHVESQILGDTHGNVVHLFERDCSVQRRNQKV--VERAPAPYLSEAQRQELAAYSLKIAGATNY-IGAGTVEYL--MDADTGKFYFIEVNPRIQVEHTV 308
PCC  R. norvegicus 218  FRFSSQEAAASSFGDDRLLIEKFIDNPRHIEQVLGDKHGNALWLNERECSIQRRNQKV--VEEAPSIFLDPETRRAMGEQAVAWPKAVKY-SSAGTVEFL--VDSQ-KNFYFLEMNTRLQVEHPV 336
ACC  S.cerevisiae  276  YHQAANEIPGSP----IFIMKLAGRARHLEVQLLADQYGTNISLFGRDCSVQRRHQKI--IEEAPVTIAKAETFHEMEKAAVRLGKLVGY-VSAGTVEYL--YSHDD-GKFYFLELNPRLQVEHPT 392
ACC  E.coli        182  ISMTRAEAKAAFSNDMVYMEKYLENPRHVEIQVLADGQGNAIYLAERDCSMQRRHQKV--VEEAPAPGITPELRRYIGERCAKACVDIGY-RGAGTFEFLFF-EN--GEFYFIEMNTRIQVEHPV 298
CPS  E. coli       192  CARGLDLSPTK----ELLIDESLIGWKEVEMEVVRDKNDNCIIVCSIENFDAMGIHTGDSITVAPQTLTDKEYQIMRNASMAVLREIGVETGGSNVQFAVNPKN--GRLIVIEMNPRVSRSSAL 310
CPS  S.cerevisiae  213  AAQSLSLAP------QILVEKSLKGWKEVEYEVVRDRVGNCITVCNMENFDPLGVHTGDSMVFAPSQTLSDEEFHMLRSAAIKIIRHLGVIG-ECNVQYALQPDG--LDYRVIEVNARLSRSSAL 328
                                         αααααααα          ββββββββ      ββββββββββ            βββββββββ            βββββ        ααααααααααααα          ββββββββββ          ββββββββββ      ααα

PC   H.sapiens      336  TEEITDVDLVHAQIHVAEGRSL--PDLGLRQ---EN------IR-INGCAIQCRVTTEDPARTFQPDTGRIEVFRSGEGMGIRLDNASAFQGAVISPHYDSLLVKVIAHGKDHPTAATKMSRAL 447
PC   R. norvegicus  336  TEEITDVDLVHAQIHVSEGRSL--PDLGLRQ---EN------IR-INGCAIQCRVTTEDPARTFQPDTGRIEVFRSGEGMGIRLDNASAFQGAVISPHYDSLLVKVIAHGKDHPTAATKMSRAL 447
PC   A.aegypti      337  TEEITGIDLVQSQIRVAEGMTL--PELGYNQ---EN------IK-TQGYAIQCRVTTEDPANDFQPSTGRLEVFRSGEGMGIRLDSASAYAGAIISPYYDSLLVKVISHASDLQSSAAKMNRAL 448
PC   S.cerevisiae   320  TEEITGIDIVAAQIQIAAGASL--PQLGLFQ---DK------IT-TRGFAIQCRITTEDPAKNFQPDTGRIEVYRSAGGNGVRLDGGNAYAGTIISPHYDSMLVKCSCSGSTYEIVRRKMIRAL 431
PC   B.subtilis     304  TEMITGVDTVQGGHSL--HSKKVNIPEQKD------IR-TIGYAIQCRVTTEDPQNDFMPDTGKIMAYRSGGGFGVRLDTGNSFQGAVITYSNGFGVRLDGGTSYSGAIITRYYDPLLVKVTAWAPNPLEAISRMVRNL 423
PC   R.etli         309  TEVVTGIDIVKAQIHILDGAAIGTPQSGVPN---QED------IR-LNGHALQCRVTTEDPEHNFIPDYGRITAYRSASGFGIRLDGGTSYSGAIITRYYDPLLVKVTYGSDRAEALKRMEDAL 446
PCC  R. norvegicus 337  TECITGLDLVQEMILVAKGYPL-----RHKQ---ED------IP-ISGWAVECRVYAEDPYKSFLPSIGRLSQYQEPIHLPGVRVDSGIQPGSDISIYHDPMISKLVTYGSDRAEALKRMEDAL 446
ACC  S.cerevisiae  393  TEMVSGVNLPAAQLQIAMGIPM------HRI---SD------IR-TLYG---MNPHSASE-IDFEFKTQ-----DATKKQRRPIPKGHCTACRITSEDPNDGFKPS---GGTLHELNFRSSSNV 488
ACC  E.coli        299  TEMITGVDLIKEQLRIAAGQPL-----SIKQ---EE------VH-VRGHAVECRINAEDP-NTFLPSPGKITRFHAPGGFGVRWES-HIYAGYTVPPYYDSMIGKLICYGENRDVAIARMKNAL 405
CPS  E. coli       311  ASKATGFPIAKVAAKLAVGYTLDELMNDIT----GGRTPASFEPSIDYVVTKIPRFNFEK-FAGA-------------------------------NDRLTTQMKSVGEVMAIGRTQQESLQKALRGL 402
CPS  S.cerevisiae  329  ASKATGYPLAYTAAKIGLGYTLPELPNPIT-----KTTVANFEPSLDYIVAKIPKWDLSK-FQYV-------------------------------DRSIGSSMKSVGEVMAIGRNYEEAFQKALRQV 419
                                       αααααα      αααααααααα                                            ββββββββββ             βββββββββ                ααα      ββββββββββ        αααααααααααα
```

1999). Taken together, it suggested that the BC domain represents a conserved structural motif forming the first partial reaction subsite. Interestingly, the P-loop motif [GXXXXGK(T/S)] commonly, though not universally, found in ATP- and GTP-binding proteins (Sarate *et al.*, 1990) is not present in rat or other PC sequences. However, the sequence GGGRGMRVV between Gly[199] and Val[208] of rat PC is found without variation in all of these PC sequences and in *E. coli* ACC, and with only conservative changes in rat ACC and PCC. In the *E. coli* BC structure (Waldrop *et al.*, 1994), the corresponding residues Gly[162] to Val[171] are sufficiently mobile not to be seen in the X-ray crystal structure (Fig. 3.10). Indeed, the above region forms part of the B-domain which protrudes out of the main body and is not involved in ATP-binding as had previously been suggested (Kondo *et al.*, 1991; Cronan and Li, 1992a).

Recent analyses have revealed very extensive similarities between the folds of BC subunit of *E. coli* ACC and an other family member of ADP-forming peptide synthetases represented by D-alanine ligase, an enzyme that catalyses the ATP-dependent ligation of two D-alanine residues as part of bacterial cell wall synthesis and by glutathione synthetase, an enzyme that catalyses the ATP-dependent synthesis of glutathione from α-L-glutamyl-L-cysteine and glycine (Artymiuk *et al.*, 1996). The reactions these three enzymes catalyse are similar in some aspects, coupling the conversion of ATP to ADP to form a carbon-nitrogen bond between a carboxyl group and an amino group: biotin carboxylase ligates ATP-activated bicarbonate to the N1' ureido nitrogen of biotin to form carboxybiotin; glutathione synthetase ligates the ATP-activated carboxyl of α-Glu-Cys to the amino group of glycine to produce glutathione; and D-alanine ligase ligates the ATP-activated carboxyl group of D-alanine to the amino group of another D-alanine. The similarities between BC and glutathione synthetase and DD-ligase is wide-ranging and involves a total of 9 α-helices and 13 β-strands all with identical sequence order and topology. Remarkably, similarities were observed when the equivalant secondary structure elements and equivalent C_α atoms of these three enzymes are superimposed (Artymiuk *et al.* 1996). The similarities were also observed near the positions of the active sites of these three enzymes. The mechanism that these enzymes catalyse are also similar: glutathione synthetase and D-alanine ligase reactions proceed through an acylphosphate intermediate (Gushima *et al.*, 1983), whereas biotin

carboxylase reaction is believed to proceed through the formation of carboxyphosphate intermediate (Ogita and Knowles, 1988). Given the similar structures and the mechanism of the reaction that these three enzymes catalyse, it was proposed these enzymes might share a common evolutionary ancestor despite the amino acid sequence similarities between these two families of ATP-binding enzymes not being very high (Artymiuk *et al.*, 1996).

Recently, the X-ray crystal structure of carbamoyl phosphate synthetase, an enzyme that catalyses the formation of carbamoyl phosphate from bicarbonate, glutamine and two molecules of MgATP through the formation of a carboxyphosphate intermediate has been solved (Thoden *et al.*, 1997). This carboxyphosphate is believed to be an intermediate in the PC reaction (Ogita and Knowles, 1988; Attwood, 1995). The first three domains in each half of the large subunit where the formation of carboxyphosphate occurs, show remarkably similar structure to those observed in the BC subunit of *E.coli* ACC (Thoden *et al.*, 1997) (Fig. 3.11). This part of the molecule also displayed a similar structure to other ATP-binding proteins, D-alanine ligase (Fan *et al.*, 1994), glutathione synthetase (Yamaguchi *et al.*, 1993) and succinyl-CoA synthetase (Wolodko *et al.*, 1994). Not only is the structure of the large subunit of CPS similar to the BC subunit of *E. coli* ACC, but some residues forming part of the active site including Arg^{303}, Asn^{301} and Glu^{299} are also identical to those found in the BC subunit of *E. coli* ACC. These residues are equivalent to Arg^{292}, Asn^{290} and Glu^{288} in BC subunit of *E. coli* ACC where they have been shown to surround a phosphate molecule (Waldrop *et al.*, 1994). Those residues are also equivalent to Arg^{328}, Asn^{326} and Glu^{324} in mammalian PC. In addition to these identical residues, there are many others where the substitutions are conservative e.g. Lys or Arg, Asp or Glu, Val, Leu, Ile or Met etc (see Fig. 3.9).

3.4.1.2 Transcarboxylation domain

Except for the sequence similarities between lipoyl domain and biotinyl domains (see below), there are no other significant sequence similarities between PC and either pyruvate dehydrogenase (EC 1.2.4.1) or pyruvate decarboxylase (EC 4.1.1.1). However, there is extensive homology within the transcarboxylation domain of PC and other biotin-dependent pyruvate-binding enzymes, viz. the 5S subunit of transcarboxylase and oxaloacetate

Figure 3.11 The X-ray crystal structures of the *E. coli* CPS and biotin carboxylase subunit of *E. coli* ACC. A schematic representation of the three-dimensional structure of the large subunit of *E. coli* CPS (Thoden *et al.*, 1997)[A] and the biotin carboxylase subunit of *E. coli* ACC (Waldrop *et al.*, 1994) [B], showing the arrangement of β-strands and α-helices generated with MOLSCRIPT (Kraulis, 1991) and Raster3D (Merrit and Murphy, 1994; Bacon and Anderson, 1988). Regions of structural similarity between these two enzymes determined by superimposition using Homology/Insight (Molecular Simulations Inc. San Diego, CA] are indicated by the same colours. Red (corresponding to the residues 1-142 of CPS and residues 1-130 of *E. coli* ACC as shown in Fig. 3.9), green (residues 143-211 and residues 131-206) and yellow (residues 212-401 and residues 207-406).

A

C

N

B

C

N
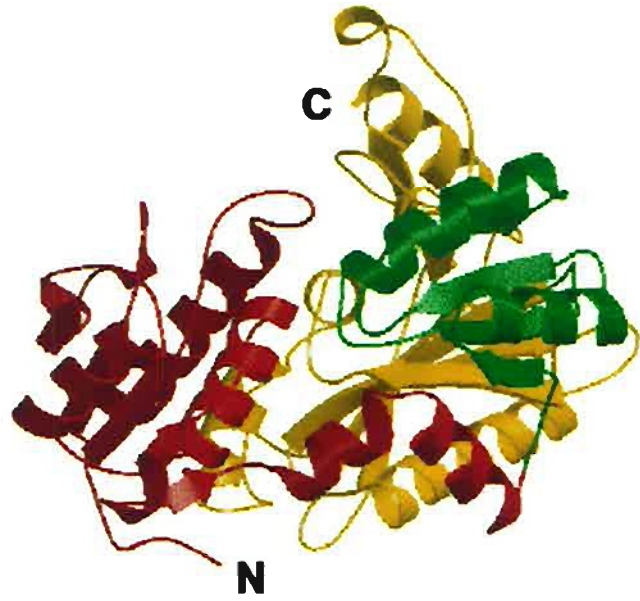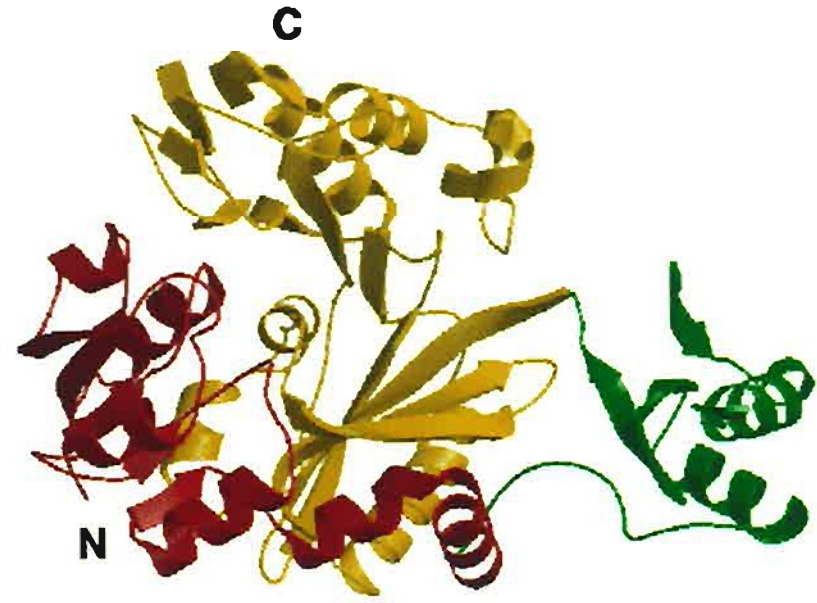
decarboxylase. Lim *et al.* (1988) previously identified the homology between residues 559 to 913 of yeast PC1 to the N-terminal region of the 5S subunit of transcarboxylase [EC 2.1.3.1] from *Propionibacterium shermanii* (Thornton *et al.*, 1993) and to the α-subunit of oxaloacetate decarboxylase from *Klebsiella pneumoniae* (Schwarz *et al.*, 1988), wherein these two enzymes are known to bind pyruvate. The residues 602 to 918 of rat PC also show significant homology to the N-terminal regions of the above enzymes. Quite remarkably, residues 605-627 (namely- ENWGGATFDVAMRFLYECPWRRL-) of rat PC conform to a consensus motif EXWGGATXDXXXRFLECPWXRL which is present in PCs from human (Wexler *et al.*, 1994; MacKay *et al.*, 1994), mosquito (Tu and Hagedorn, 1997), *S. cerevisiae* (Lim *et al.*, 1988), *Rhizobium etli* (Dunn *et al.*, 1996), *M. thermoautotrophicum* (Mukhopadhyay *et al.*, 1998) and *Bacillus subtilis* (Li *et al.*, 1998) and as well as to the oxaloacetate-metabolizing biotin-dependent enzymes of bacterial origin described above (Fig. 3.12). In the 5S subunit of *P. shermanii* transcarboxylase, it has been shown that Trp^{73} (corresponding to the underlined Trp above) is involved with the pyruvate binding site (Kumar *et al.*, 1988). In the presence of pyruvate, this tryptophan is protected from modifiation by the tryptophan-specific reagent 2,4-dinitrophenyl sulphenyl chloride, suggesting that this residue is directly involved or near the pyruvate binding site of the enzyme.

PC is known to be a metalloprotein, containing tightly bound either Mn^{2+} as in the case of vertebrate PC (Scrutton *et al.*, 1973 and Wallace and Easterbrook-Smith, 1985), or Zn^{2+} for yeast PC (Scrutton *et al.*, 1970) and this appears to play a structural role rather than a catalytic role (Carver *et al.*, 1988). Chelating of Mn^{2+} by 1,10-phenanthroline causes the loss of enzymatic activity resulting from the destabilisation of the active tetrameric structure of the enzyme (Carver *et al.*, 1988). A recent study by Dimroth and Thorner (1992) has shown that oxaloacetate decarboxylase (ODC) contains a bound Zn^{2+}. Given the high level of sequence similarity between ODC and PC, it is very likely that both enzymes would contain a putative metal binding site in the pyruvate binding domain. The motif HXH (Vallee *et al.*, 1990) or HXXEH or HEXXH (Jiang and Bond, 1992) are know to be part of the metal binding sites of carbonic anhydrase and metalloproteinase, respectively. Mozier *et al.* (1991) have shown that a 24-residue tryptic peptide of protein kinase C inhibitor-1

Figure 3.12 Multiple sequence alignment of the transcarboxylation domains of PC from different organisms, decarboxylase enzymes and 3-hydroxyl-3-methylglutaryl-CoA lyase shown by ψ-BLAST search (Altschul *et al.*, 1997) to be related, were compared using Clustal W. (Thompson *et al.*, 1994). The highly conserved amino acid residues found in different groups of enzymes are shown by shaded boxes. The putative pyruvate binding-site is asterisked. Sources: PC *H. sapiens* (human) [Wexler *et al.*, 1994; MacKay *et al.*, 1994; Walker *et al.*, 1995b], PC *Aedes aegypti* (mosquito) [Tu and Hagedorn, 1997], PC *A. terreus* [Li *et al.*, 1998], *S. cerevisiae* PC1 [Lim *et al.*, 1988], PC *R. etli* [Dunn *et al.*, 1996], PC *Bacillus subtilis* [Kunst *et al.*, 1997], PC *M. thermoautotrophicum* (Mukhopadhyay *et al.*, 1998]; TC, 5S subunit of *P. shermanii* [Thornton *et al.*, 1993}; ODC (Oxaloacetate decarboxylase α-subunit) *Klebsiella pneumoniae* [Schwartz *et al.*, 1988], ODC *Salmonella typhimurium* [Woehlke *et al.*, 1992}; 3-hydroxyl-3-methylglutaryl-CoA lyase (HMG-lyase) rat (*R. norvegicus*) [Cullingford *et al.*, 1998] and chicken (*G. gallus*) [Mitchell *et al.*, 1993].

```
PC          H. sapiens       563  LLLMDTTFRDAHQSL LATRVRTHDLKKIAP YVAHNFSKLFSMENW GGATFDVAMRFLYEC PWRRLQELRELIPNI PFQMLLRGANAVGYT  652
PC          R. norvegicus    563  LLLMDTTFRDAHQSL LATRVRTHDLKKIAP YVAHNFNNLFSMENW GGATFDVAMRFLYEC PWRRLQELRELIPNI PFQMLLRGANAVGYT  652
PC          A. aegypti       580  LLLMDTTFRDAHQSL LATRVRTHDLLKISP YVSHKFNNLYSLENW GGATFDVALRFLHEC PWERLEDMRKQIPNI PFQMLLRGANAVGYT  669
PC          A. terreus       579  CLIMDTTWRDAHQSL LATRVRTIDLLNIAH ETSHALSNAYSLECW GGATFDVAMRFLYED PWDRLRKMRKAVPNI PFQMLLRGANGVAYS  668
PC          S. cerevisiae    557  TLLMDTTWRDAHQSL LATRVRTHDLATIAP TTAHALAGRFALECW GGATFDVAMRFLHED PWERLRKLRSLVPNI PFQMLLRGANGVAYS  646
PC          R. etli          540  VLLTDTTMRDGHQSL LATRMRTYDIARIAG TYSHALPNLLSLECW GGATFDVSMRFLTED PWERLALIREGAPNL LLQMLLRGANGVGYT  629
PC          B. subtilis      534  VLLTDTTFRDAHQSL LATRIRSHDLKKIAN PTAALWPELFSMEMW GGATFDVAYRFLKED PWKRLEDLRKEVPNT LFQMLLRSSNAVGYT  623
PC          M. thermo.       4    IKVVETAFRDAHQSL LATRLRTRDMTPIAE EMDRVG--FFSLEAW GGATFDTCIRYLNED PWERLRELKEHVKRT PIQMLLRGQNLVGYK  91
ODCα        K. pneumonieae   3    VAITDVVLRDAHQSL FATRLRLDDMLPVAA QLDDVG--YRSLECW GGATFDACIRFLGED PWVRLRELKKAMPKT PLQMLLRGQNLLGYR  90
ODCα        S. typhimurium   3    IAITDVVLRDAHQSL FATRLRLDDMLPIAA ALDDVG--YGSLECW GGATFDACIRFLGED PWLRLRELKKAMPKT PLQMLLRGQNLLGYR  90
TC          P. shermanii     14   VGITELVLRDAHQSL MATRMAMEDMVGACA DIDAAG--YWSVECW GGATYDSCIRFLNED PWERLRTFRKLMPNS RLQMLLRGQNLLGYR  101
HMG-lyase   R. norvegicus    33   VKIVEVGPRDGLQ-N EKSIVPTPVKIKLID MLSEAG-----LPVI EATSFVSPKWVPQMA DHSDVLKGIQKFPGI NYPVLTPNMKGFEEA  116
HMG-lyase   G. gallus        6    VKVVEVGPRDGLQ-N EKSVVPTPVKIRLID MLSETG-----LPVI EATSFVSPRWVPQMA DHAEVMQGINKLPGV SYPVLTPNLKGFQAA  89

PC          H. sapiens       653  NYPDNVVFKFCEVAK ENGMDVFRVFDSLNY LPNMLLGMEAAGSAG GVVEAAISYTGDVAD PSRTKYSLQYYMGLA EELVRAGTHILCIKD  742
PC          R. norvegicus    653  NYPDNVVFKFCEVAK ENGMDVFRIFDSLNY LPNMLLGMEAAGSAG GVVEAAISYTGDVAD PSRTKYSLEYYMGLA EELVRAGTHILCIKD  742
PC          A. aegypti       670  NYPDNVVHKFCELSV QCGMDIFRVFDSLNY LPNLILGMEAAGNAG GVVEAAISYTGDVSD PTKKKYDLKYYTNLA DELVKAGTHILCIKD  759
PC          A. terreus       669  SLPDNAIYHFCKNAK KCGVDIFRVFDALND IDQLEVGIKAVHAAE GVVEATVCYSGDMLN P-KKKYNLEYYLALV DKIVALKPHVLGIKD  757
PC          S. cerevisiae    647  SLPDNAIDHFVKQAK DNGVDIFRVFDALND LEQLKVGVDAVKKAG GVVEATVCFSGDMLQ P-GKKYNLDYYLEIA EKIVQMGTHILGIKD  735
PC          R. etli          630  NYPDNVVKYFVRQAA KGGIDLFRVFDCLNW VENMRVSMDAIAEEN KLCEAAICYTGDILN SARPKYDLKYYTNLA VELEKAGAHIIAVKD  719
PC          B. subtilis      624  NYPDNVIKEFVKQSA QSGIDVFRIFDSLNW VKGMTLAIDAVRDTG KVAEAAICYTGDILD KNRTKYDLAYYTSMA KELEAAGAHILGIKD  713
PC          M. thermo.       92   HYPDDIVRKFIEKSY ENGVDVFRIFDALND IRNMEYAIKVAREQE AHVQGVICYT--IS- P---YHTLESYVDFA RELEALECDSVAIKD  175
ODCα        K. pneumonieae   91   HYADDVVERFVERAV KNGMDVFRVFDAMND PRNMQAALQAVRRHG AHAQGTLSYT--TS- P---AHTLQTWLDLT EQLLETGVDSVAIKD  174
ODCα        S. typhimurium   91   HYADDVVERFVERAV KNGMDVFRVFDAMND PRNMKAALQAVRSHG AHAQGTLSYT--TS- P---AHTLQTWLDLT EQLLETGVDSIAIKD  174
TC          P. shermanii     102  HYNDEVVDRFVDKSA ENGMDVFRVFDAMND PRNMAHAMAAVKKAG KHAQGTICYT--IS- P---VHTVEGYVKLA GQLLDMGADSIALKD  185
HMG-lyase   R. norvegicus    117  VAAGAKEVSIFGAAS ELFTRKNVNCSIEES FQRFDGVMQAARAAS ISVRGYVSCA--LGC PYEGKVSPAKVAEVA KKLYSMGCYEISLGD  204
HMG-lyase   G. gallus        90   VAAGAKEVSIFGAAS ELFTKKNINCSIEES LERFSEVMNAARAAS IPVRGYVSCV--LGC PYEGNISAAKVAEVS KKMYSMGCYEISLGD  177

PC          H. sapiens       743  MAGLLKPTACTMLVS SLRDRFPDLPLHIHT HDTSGAGVAAMLACA QAGADVVDVAADSMS GMTSQPSMGALVACT RGTPLDTEVPMERVF  832
PC          R. norvegicus    743  MAGLLKPAACTMLVS SLRDRFPDLPLHIHT HDTSGSGVAAMLACA QAGADVVDVAVDSMS GMTSQPSMGALVACT KGTPLDTEVPLERVF  832
PC          A. aegypti       760  MAGLLKPQAAKLLIA AIREKHPDVPIHIHT HDTSGAGVASMLACA EAGADVVDVAVDSMS GMTSQPSMGAVVASL QGTPLDTGLNLRDIS  849
PC          A. terreus       758  MAGVLKPQAARLLVG SIRERYPDLPIHVHT HDSAGTGVASMIACA QAGADAVDAATDSMS GMTSQPSIGAILASL EGTEHDPGLNSAHVR  847
PC          S. cerevisiae    736  MAGTMKPAAAKLLIG SLRAKYPDLPIHVHT HDSAGTAVASMTACA LAGADVVDVAINSMS GLTSQPSINALLASL EG-NIDTGINVEHVR  824
PC          R. etli          720  MAGLLKPAAAKVLFK ALREAT-GLPIHFHT HDTSGIAAATVLAAV EAGVDAVDAAMDALS GNTSQPCLGSIVEAL SGSERDPGLDPAWIR  808
PC          B. subtilis      714  MAGLLKPQAAYELVS ALKETI-DIPVHLHT HDTSGNGIYMYAKAV EAGVDIIDVAVSSMA GLTSQPSASGFYHAM EGNDRRPEMNVQGVE  802
PC          M. thermo.       176  MAGLISPHDAYELVR ALKEET-DLMVNLHC HCTSGMTPMSYYAAC EAGVDILDTAISPLS WGASQPPTESIVAAL RDTPYDTGLDLEILK  264
ODCα        K. pneumonieae   175  MSGILTPHAAFELVS EIKKRY-DVTLHLHC HATTGMAEMALLKAI EAGVDGVDTAISSMS ATYGHPATEALVATL AGTPYDTGLDIHKLE  263
ODCα        S. typhimurium   175  MSGILTPMAAYELVS EIKKRF-EVRLHLHC HATTGMAEMALLKAI EAGVDGVDTAISSMS ATYGHPATEALVATL AGTEHDTGLDILKLE  263
TC          P. shermanii     186  MAALLKPQPAYDIIK AIKDIRPEDADQPAL HSTTGVTEVSLMKAI EAG----VDTAISSMS LGPGHNPTESVAEML EGTGYTTNLDYDRLH  272
HMG-lyase   R. norvegicus    205  TIGVGTPGLMKDMLT AVLHEVPVAALAVHC HDTYGQALANTLVAL QMGVSVVDSSVAGLG GCPYAKGASGNLATE DLVYMLTGLGIHTGV  294
HMG-lyase   G. gallus        178  RIGIGTPGSMKEMLA AVMKEVPVGALAVHC HDTYGQALANILVAL QMGVSVVDASVAGLG GCPYAQGASGNVATE DLVYMLNGLGIHTGV  267
```

(PKCI-1) containing the sequence HVHLH, as well as a number of synthetic peptides containing various length segments of the PKCI-1 sequence spanning the HVHLH site, would bind a single Zn^{2+} ion. The consensus sequence HXHXH corresponding to the residues 769-773 of the rat PC (Fig. 3.8 and Fig. 3.12) is consistent with the motifs for these other metal-binding enzymes described above. ψ-BLAST searches (Altschul *et al.*, 1997) also reveal a significant degree of similarity with 3-hydroxyl-3-methylglutaryl-CoA lyase from rat (Cullingford *et al.*, 1998) and chicken (Mitchell *et al.*, 1993).

3.4.1.3 Biotinyl domain

The biotin attachment site of PC was first identified from sheep and avian species by peptide sequencing (Rylatt *et al.*, 1977). The conserved motif consists of the tetrapeptide, AMKM with a single biotin covalently attached to the ε-amino group of the lysine via an amide bond. With the advance of recombinant DNA technology, more sequence information from a whole range of different biotin-dependent enzymes were obtained by sequencing cDNA or genomic clones. Lim *et al.* (1988) have shown that the carboxyl-terminal portion of yeast PC shows high sequence similarity with the area surrounding the lipoyl attachment regions of *E. coli* pyruvate dehydrogenase. This sequence similarity between these two groups of enzymes suggests an analogous role for both biotin and lipoic acid in acting as a 'swinging arm' between active sites (Lim *et al.*, 1988).

For rat PC, the potential biotinyl domain was located between residues 1103 and 1178 with the biotinylated lysine located 35 residues upstream from the C-terminus, in common with many biotin-dependent enzymes (Samols *et al.*, 1988). Thampy *et al.* (1988) reported the sequence of 24 amino acid residues around the biotin attachment of rat PC by peptide sequencing. However, seven amino acid differences were noted between the previously reported peptide sequence (Thampy *et al.*, 1988) and the sequence inferred from cDNA in this thesis. The sequence inferred from cDNA in this region is also identical with other mammalian PC (Zhang *et al.*, 1993; Wexler *et al.*, 1994; and MacKay *et al.*, 1994). Therefore it is very likely that the previous reported sequence (Thampy *et al.*, 1988) was incorrect, perhaps due to impurities being present.

A number of residues within the biotinyl domain of rat PC showed significant identity with the biotinyl domain of PC from different species and other biotin-containing enzymes (Fig. 3.13), suggesting that they fold to a similar structure. The crystal structure of the C-terminal 80 residues of the biotin carboxyl carrier subunit of *E. coli* ACC (holoenzyme) by Athappily and Hendrickson (1995) revealed that this domain adopts the same basic fold as the lipoyl domains of *E. coli* pyruvate dehydrogenase (Dardel *et al.*, 1993; Green *et al.*, 1995). The structure of the holoprotein is very similar to that of the apoprotein, determined by NMR (Yao *et al.*, 1997) with small local conformational changes observed in the β turn that contains lysine residue modified in the biotin ligation reaction. Chemical modification and proteolysis studies of apo-and holo- enzymes also indicates a conformational change accompanies biotinylation (Chapman-Smith *et al.*, 1997). The recent determination by NMR of the three dimensional structure of the entire 1.3S subunit of *P. shermanii* transcarboxylase, which functions as the carboxyl group carrier of this enzyme, also showed that the C-terminal half of this subunit is folded into a compact domain consistent with the fold found in both the carboxyl carrier protein of *E. coli* ACC and in the lipoyl domains, to which this domain exhibits only 26-30% sequence similarity (Reddy *et al.*, 1998). Therefore, it is predictable that the biotinyl domain of yeast PC would fold to a similar structure to the lipoyl domains (Brocklehurst and Perham, 1993), as there are remarkable sequence similarities between these two families of proteins (Lim *et al.*, 1988).

A number of highly conserved residues flanking the biotin attachment site of different biotin-dependent enzymes, are highlighted in Fig. 3.13. Since it has long been shown that the holocarboxylase synthetase from mammals can biotinylate bacterial apocarboxylases (McAllister and Coon, 1966) and that mammalian apocarboxylases are biotinylated by bacterial biotin ligase *in vitro* (Lane *et al.*, 1964), this sequence conservation may reflect a molecular mechanism, common to all of the biotin-enzymes in interacting with biotin ligase or holocarboxylase synthetase (Chapman-Smith and Cronan, 1999). Although the AMKM motif is highly conserved across biotin-enzymes, substitution of either methionine flanking the biotinylated lysine of 1.3S biotinyl subunit of *P. shermanii* transcarboxylase and α-subunit of human PCC (Leon-Del-Rio and Gravel, 1994), had no effect on biotinylation efficiency (Shenoy *et al.*, 1988). However, substitutions of the methionines flanking the

			1	10	20	30	40	50	60	70	80	90	100	

```
                      1        10        20        30        40        50        60        70        80        90       100

PC   R. norvegicus   1099 -HFHPKALKDVKGQ-IGAPMPGKVIDV-------KVAAGAKVVKGQPLCVLSAMKMETVVTSPME-GTIRKVHVTKDMTLEGDDLILEIE            1178
PC   H. sapiens      1099 -HFHPKALKDVKGQ-IGAPMPGKVIDI-------KVVAGAKVAKGQPLCVLSAMKMETVVTSPME-GTVRKVHVTKDMTLEGDDLILEIE            1178
PC   A. aegypti      1116 -HIHPKATKGNKDQ-VGAPMPGSVIEI-------KVKVGDRVEKGQPLVVLSAMKMEMVVQSPRA-GVVKSLDISSGMKLEGEDLILTLE.           1195
PC   A. terreus      1112 -TARPKADVGDSSQ-VGAPMSGVVVEI-------RVHDGLEVKKGDPLAVLSAMKMEMVISAPHS-GKVSGLLVKEGDSVDGQDLVCKITKA.         1193
PC   S. cerevisiae   1090 -VTKSKADMHDPLH-IGAPMAGVIVEV-------KVHKGSLIKKGQPVAVLSAMKMEMIISSPSD-GQVKEVFVSDGENVDSSDLLVLLEDQVPVETKA.  1178
PC   R. etli         1074 -AVRRKAEPGNAAH-VGAPMPGVISRV-------FVSSGQAVNAGDVLVSIEAMKMETAIHAEKD-GTIAEVLVKAGDQIDAKDLLAVYGG.          1154
PC   B. subtilis     1067 -QERLKADRTNPSH-IAASMPGTVIKV-------LAEAGTKVNKGDHLMINEAMKMETTVQAPFS-GTIKQVHVKNGEPIQTGDLLLEIEKA.         1141
PC   M. thermo.       488 -IEEAEPEPVDVEGAVKSTMQGMVVKL-------KVSEGDQVNAGDVVAVVEAMKMENDIQTPHG-GVVEKIYTAEGEKVETGDIIMVIK.           568
TC   P. shermanii      44 -AGGAGAGKAGEGE-IPAPLAGTVSKI-------LVKEGDTVKAGQTVLVLEAMKMETEINAPTD-GKVEKVLVKERDAVQGGQGLIKIG.           123

     β-sheet                                →        →                 →                →          →                →

ODCα K. pneumoniae    516 -SAPAAAAPAGAGTPVTAPLAGTIWKV-------LASEGQTVAAGEVLLILEAMKMETEIRAAQA-GTVRGIAVKAGDAVAVGDTLMTLA.           596
PCCα H. sapiens       622 -KFMLEKVTEDTSSVLRSPMPGVVVAV-------SVKPGDAVAEGQEICVIEAMKMQNSMTAGKT-GTVKSVHCQAGDTVGEGDLLVELE--          702
ACC  H. sapiens       641 -KTCVFEKENDPSV-MRSPSAGKLIQY-------IVEDGGHVFAGQCYAEIEVMKMVMTLTAVES-GCIHYVKRP-GAALDPGCVLAKMQLDNP-       823
ACC  S. cerevisiae    690 -MTTLLEVENDPTQ-LRTPSPGKLVKF-------LVENGEHIIKGQPYAEIEVMKMMQMPLVSQENGIVQLLKQP-GSTIVAGDIMAIM---          767
ACC  E. coli           69 -SMEAPAAAEISGHIVRSPMVGTFYRTPSPDAKAFIEVGQKVNVGDTLCIVEAMKMMNQIEADKS-GTVKAILVESGQPVEFDEPLVVIE.           156

     β-sheet                                →        →                 →      →          →         →        →          →
```

Figure 3.13   Multiple sequence aligment of the biotinyl domains of various biotin carboxylases.   The highly conserved residues are indicated by shaded boxes. Biotinylated lysine is shown by italics. Also shown are the β-strands (arrows) observed in the crystal structure of biotin carboxyl carrier protein of *E. coli* ACC (Athappily and Hendrickson, 1995) and in the NMR structure of the 1.3 S subunit of *P. shermanii* transcarboxylase (Reddy *et al.*, 1998). Sources PC *H. sapiens* (Wexler *et al.*, 1994; MacKay *et al.*, 1994; Walker *et al.*, 1995b), PC *Aedes aegypti* (mosquito) [Tu and Hagedorn, 1997], PC *A. terreus* [Li *et al.*, 1998] *S. cerevisiae* PC1 [Lim *et al.*, 1988], PC *R. etli* [Dun *et al.*, 1996], PC *Bacillus subtilis* [Kunst *et al.*, 1997], PC *M. thermoautotrophicum* (Mukhopadhyay *et al.*, 1998]; TC, 1.3 S subunit of *P. shermanii* [Maloy *et al.*, 1979]; ODC (Oxaloacetate decarboxylase α-subunit) *Klebsiella pneumoniae* [Schwartz *et al.*, 1988]; PCCα (Propionyl-CoA carboxylase α–subunit) human [Lamhonwah *et al.*, 1986]; ACC1 human [Ha *et al.*, 1994], ACC *S. cerevisiae* [Al-Feel *et al.*, 1992]; ACC *E. coli* [Li and Cronan, 1992a].

targeted lysine of the biotin carboxyl carrier protein of *E. coli* ACC did affect the biotinylation reaction (Rech *et al.*, 1998). The roles of highly conserved residues in the yeast PC1 biotin domain peptide are currently being investigated in this laboratory (S. Polyak, personal communication). In yeast PC, it has been demonstrated that the C-terminal 104 residues of the biotinyl domain can act as an independent domain in the biotinylation reaction *in vitro* (Val *et al.*, 1995).

The central putative transcarboxylation domain in rat PC is linked to the N-terminal biotin carboxylation domain and to the C-terminal biotinyl domain by proline-rich sequences. Of the 65 proline residues in rat PC, 13 are located between residues 492 and 550, just C-terminal of the biotin carboxylation domain, whereas 11 are located between residues 941 and 1042, just C-terminal of the transcarboxylation domain. There are also 9 alanine and 7 glycine residues located between the biotin carboxylation domain and the transcarboxylation domain. A further 12 alanine and 11 glycine residues occur between the transcarboxylation domain and the biotinyl domain. However, in general they are not distributed in as close proximity to the prolines as seen in the α-subunit of oxaloacetate decarboxylase from *K. pneumoniae* (Schwarz *et al.*, 1988) and in the biotin carrier protein of *P. shermanii* transcarboxylase (Samols *et al.*, 1988). The motif PX(P/A) found approximately 30 residues upstream of the biotin binding site (except for *M. thermoautotrophicum* and *B. subtilis* PCs), has been proposed to provide flexibility for movement of the biotin prosthetic group between catalytic centres in a manner analogous to the highly mobile Pro-Ala sequences in lipoated protein (Radford *et al.*, 1989). Because PC is composed of three domains that must fold together to form a single active site, it is reasonable to expect the presence of two proline-rich regions capable of forming hinge-like structures. In the α-subunit of human PCC, it has been shown that the PMP motif (26 residues N-terminal of the target lysine) is critical for biotinylation. Deletion of this motif abolished biotinylation (Leon-Del-Rio and Gravel, 1994).

CHAPTER 4

IDENTIFICATION AND CHARACTERISATION

OF MULTIPLE TRANSCRIPTS OF RAT AND

HUMAN PC GENES

4.1 INTRODUCTION

It is clear that PC plays an important role in different arrays of metabolic pathways in mammals. It would not be surprising, therefore if there were tissue-specific isoforms of PC, as occurs with lactate dehydrogenase or other ubiquitous enzymes. Although there is evidence for the presence of tissue-specific isoenzymes of PC in sheep liver and kidney based on a kinetic studiy (Ashman *et al.*, 1972), the authors could not isolate intact isozymes from both tissues. Such evidence could also reflect impurities of the enzyme derived from these sources. Indeed, no evidence for the existence of multiple bands which reflect different mRNAs encoding PC were detected by Northern blot analysis of RNA isolated from different human tissues (Wexler *et al.*, 1994).

However, there is a number of eukaryotic genes whose primary transcripts undergo differential splicing at the 5'-end of the 5'-untranslated region (UTR) exon, thus producing heterogeneity at the 5'-end of the mRNAs (Breitbart *et al.*, 1987). A more complex situation can exist if those transcripts are derived from alternate promoters. Also, if the size of the 5'-UTRs of these multiple mRNA isoforms is relatively small compared with the coding region, conventional Northern blot analysis will not be able to discriminate between these limited size differences.

Primer extension analysis is routinely used to identify and map the 5'-end of the transcript of a gene. This technique involves hybridisation of an antisense primer near the 5'-end of RNA followed by reverse transcription reaction to cDNA. The size of the extended cDNA product(s) can then be determined by running the cDNA on high resolving gel electrophoresis. However, the nucleotide sequence at the 5'-end cannot be obtained unless the extended product(s) is cloned and sequenced. If the abundancy of the transcript of interest is also very low even though poly(A) RNA is used, a reproducible result is still difficult to obtain.

With the development of RACE-PCR as described in Chapter 3, this difficulty can be overcome. Since this technique combines primer extension with PCR to enhance the sensitivity, it is possible to determine the 5'-end of rare transcripts. With the engineered restriction sites at the 5'-end of both primers, the PCR products representing different 5'-ends of the transcripts can then be cloned and sequenced. In this chapter, RACE-PCR

was successfully employed to identify multiple transcripts encoding PC both in rat and human.

4.2 SPECIFIC METHODS

4.2.1 Northern blot analysis

Fifty micrograms of total RNA were precipitated with 0.1 volume of 3M sodium acetate pH 5.2 together with 2 volumes of cold absolute ethanol and incubated at -20°C for 1 h. The RNA pellet was then dissolved in 4.5 µl of water, 2.0 µl of 5x MOPS buffer (0.1M 3-(N-morpholino) propanesulfonic acid, 25 mM sodium acetate, 5 mM EDTA), 3.5µl of formaldehyde, 10 µl of formamide, 0.5 µl of 10 µg/ml ethidium bromide and 2 µl of RNA loading buffer, and then the RNA was denatured at 70°C for 15 min. RNA was electrophoresed on 1% agarose gel containing 2.2 M formaldehyde using 1x MOPS buffer as running buffer at 80 V for 4 h. The gel containing fractionated RNA was washed 3 x 10 min with water and transferred to the nylon membrane using 10x SSC as transfer buffer, overnight. RNA was fixed on the nylon membrane by UV-crosslinking. The membrane was incubated at 42°C overnight in 100 ml of hybridisation solution containing 50% formamide, 5x Denhardt's solution, 5x SSPE (0.9 M NaCl, 50 mM NaH_2PO_4 pH 7.7 and 5 mM EDTA), 1% (w/v) SDS and 200 µg/ml sonicated salmon sperm DNA. The membrane was then transferred to 20 ml of fresh solution containing the appropriate probe and further incubated at 42°C overnight. The membrane was then washed in 2x SSC-0.1% SDS for 2 x 15 min at room temperature, and subsequently washed in the same solution for 2 x 15 min at 42°C. The stringent wash was performed twice with 0.1 x SSC-0.1% SDS for 15 min at 42°C. The membrane was autoradiographed at -70°C for 2-7 days.

4.2.2 Multiplex reverse transcriptase polymerase chain reaction (RT-PCR)

The first strand cDNA for PCR was synthesized as follows: 10 µg of total RNA from different rat tissues were hybridized with 100 ng of cDNA synthesis primer (PC8) (5' GGGACCACTGGAACGCCTGC 3') [positions +484 to +503] in a total volume of 10 µl at 65°C for 5 min. The primed cDNA was then brought up to 30 µl with the following final concentrations: 25 units of Moloney murine leukemia virus reverse transcriptase (MMLV), 1.0 mM each of dNTPs, 64 units of RNase inhibitor, 42 mM Tris-HCl, pH 8.3, 8 mM $MgCl_2$ and 8 mM DTT. Reverse transcription was carried out at 42°C

for 1 h. The reaction was then terminated by adding 1 µl of 0.5 M EDTA pH 7.5. RNA was hydrolysed with 2 µl of 6 M NaOH at 65°C for 30 min and neutralized with 2 µl of 6 M acetic acid. cDNA was subsequently purified with BresaClean Kit and then ethanol precipitation in the presence of glycogen. The cDNA pellet was resuspended in 10µl of sterile distilled water. PCR was performed in a total volume of 50 µl containing 50 mM KCl, 10 mM Tris-HCl, pH 9, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 200 mM each of dNTPs, 1µl of 1:100 dilution of cDNA, 2 units of *Taq* DNA polymerase, 100 ng of each primer. The PCR primers were as follows: 5' GGC<u>GAATTC</u>CGATGGCAATCTCACCTCTGTTGGC 3' (MRACE II) which was directed to a common coding sequence of both classes of PC cDNAs [positions +127 to +150, with an *Eco*RI restriction site attached at the 5' end] and the specific primers 5' GCCAATGACCTCGGGTGGAGCAG 3' for the class I type of PC cDNA (oligo A) and 5' GGCTTGAGGCGACGGGGCGAAG 3' for the class II type of PC cDNA (oligo B). Samples were amplified for 35 cycles with denaturation at 94°C, 45 sec (2 min for an initial denaturation), annealing at 65°C, 45 sec and extension at 72°C, 45 sec. Aliquots (5 µl) from the RT-PCR reactions were electrophoresed on 2.5 % agarose gel. The products were confirmed by Southern blot hybridization with an internal oligonucleotide probe (5' ACAGCCCCCGTTGCCTCCCCA 3') [positions +61 to +81].

4.3 RESULTS AND DISCUSSION

4.3.1 Multiple rat tissue Northern blot analysis

Northern blot analysis was initially carried out with different rat tissues to determine whether there is any evidence for the presence of multiple PC mRNA isoforms transcribed from the PC gene. Total RNA preparations from different adult rat tissues were isolated, electrophoresed on a formaldehyde gel and probed with a combination of different PC cDNA fragments (Fig. 3.7; Chapter 3). As shown in Fig. 4.1, only a single band was detected that was approximately 4.0 kb in length, in agreement with the size of the PC cDNA. The expression of PC mRNA exhibits a tissue-specific pattern, being most abundant in liver and kidney, which are the gluconeogenic tissues. The levels of PC mRNA in these tissues are also consistent with those in human liver and kidney (Wexler *et al.*, 1994). The level of rat PC mRNA was also substantial in adipose tissue and brain, in keeping with the relatively high levels of PC activity in these tissues (Wallace, 1985). This is not surprising because adipose PC plays a role in lipogenesis in the export of mitochondrial acetyl-CoA to the cytosol as citrate for the *de novo* biosynthesis of fatty acids (Ballard and Hanson, 1970). In brain, where PC is localised in glial cells (Gamberino *et al.*, 1997) the enzyme plays an anaplerotic role in the transfer of acetyl groups from mitochondria to the cytosol for the biosynthesis of several neurotransmitter substances. Moderate expression was detected in adrenal gland, heart and lactating mammary gland. A very low level of expression was found in skeletal muscle and spleen, suggesting that the anaplerotic function of PC may be less important for normal function of these tissues.

4.3.2 Identification of multiple PC mRNA isoforms in different rat tissues by RACE-PCR

Although Northern analysis did not identify multiple bands reflecting existence of multiple PC mRNA isoforms, RACE-PCR was carried out for the reason mentioned above. Ideally, any transcripts which differ only at their 5'-UTRs can be obtained by reverse transcription using a primer derived from the coding regions. As indicated in Fig. 4.2, different cDNAs with heterogeneity at the 5'-end are subsequently ligated to a single-strand anchor adaptor. An anchor primer containing the sequence that is complementary to the anchor adaptor together with nested gene specific primers can then be used to selectively

67
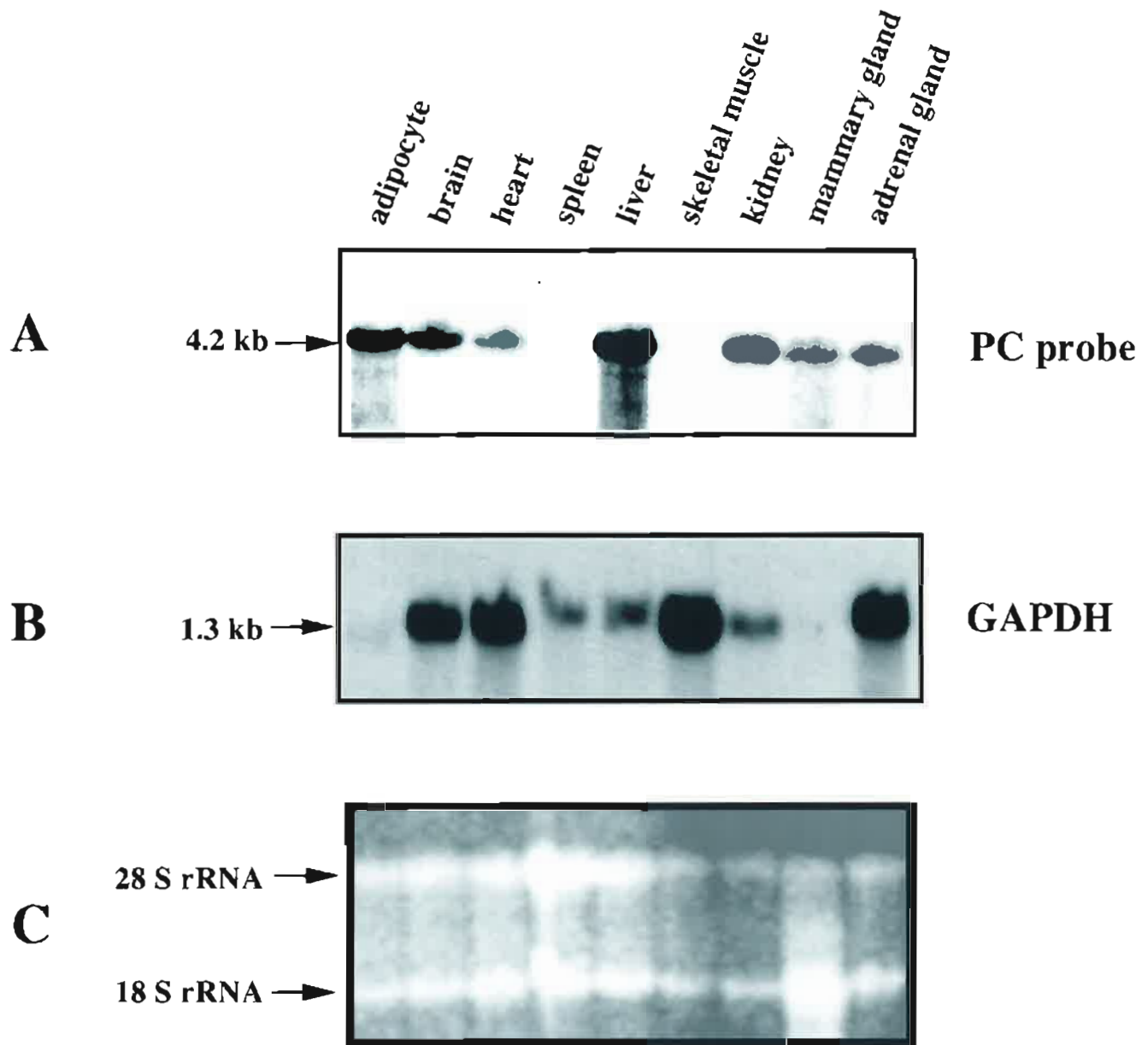
Figure 4.1   Northern blot analysis of PC mRNA in different rat tissues. Samples of total RNA (40 µg) from different rat tissues were electrophoresed, blotted onto nylon membrane and hybridised with combination of cDNA fragments encoding rat PC (A). Integrity of RNAs was assessed by stripping the membrane and reprobing with cDNA encoding rat GAPDH (B). RNA loading was assessed by comparison with the 18S and 28S rRNA bands revealed by ethidium bromide (C).

Figure 4.2 Schematic diagram of the 5'-RACE PCR procedure used to obtain 5'-ends of rat PC mRNAs. Transcripts divergent at their 5'-UTRs were reverse-transcribed with a cDNA synthesis primer. Following reverse transcription, an anchor adaptor was ligated to the 5'-end of cDNAs using T4 RNA ligase and PCR-amplified with 5'-nested primer and anchor primer. PCR products were confirmed by probing with an oligonucleotide internal to the nested PCR primer.

5' ☐ AUG start codon AAAAAA
← cDNA synthesis primer

AUG start codon
5' ▨ AAAAAA
← cDNA synthesis primer

reverse transcription

AUG
5' ☐ AAAAAA
3' ☐ 5'

AUG
5' ▨ AAAAAA
3' ▨ 5'

RNA hydrolysis

3' ▬ 5'
3' ▬ 5' 3' ▬ 5'
3' ▬ 5'

adapter ligation

3' ▬ 5'
3' ▬ 5'

PCR

anchor primer

5' ▬→
3' ▬ 5'
← 5' nested primer

5' ▬→
3' ▬ 5'
← 5' nested primer

Hybridization

*****
☐ ³²P probe

AUG
| |
5' ▬ 3'
3' ▬ 5'
| |

AUG | |
5' ▨ 3'
3' ▨ 5'

Cloning and Sequencing
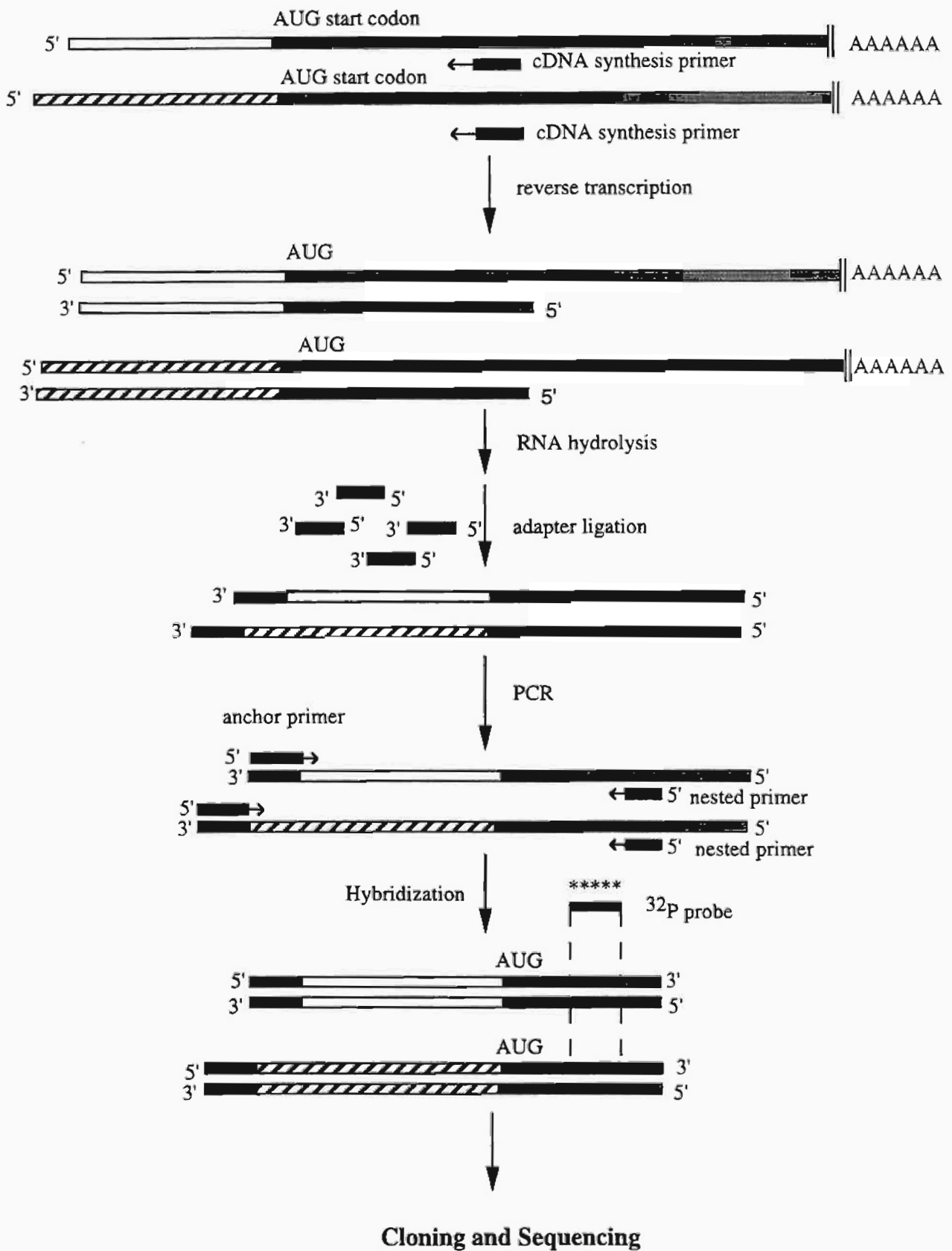
PCR-amplify the 5'-UTRs of cDNAs. The products are analysed by agarose gel electrophoresis and can be confirmed by probing with an oligonucleotide internal to the nested PCR primer (Fig. 4.2). This 5'-RACE-PCR works in the same way as that described in Chapter 3 except that the cDNA synthesis primer is usually designed to be not too far downstream from the AUG initiation codon to avoid premature termination during the reverse transcription step. 5'-RACE-PCR was performed using the same set of primers that were used to generate RACE (Anchor/MRACE II) fragment described in Chapter 3 (see Fig. 3.2) but with different RNAs extracted from four tissues e.g. liver, kidney, adipose tissue and brain. The cDNA synthesis primer (MRACE I) was designed from the coding region of rat PC cDNA (Chapter 3), and reverse transcribed following the schematic diagram in Fig. 4.2. The initial PCR profile was carried out by annealing at 60°C, however a smeared pattern was obtained after agarose gel electrophoresis (data not shown). Increasing the annealing temperature to 65°C improved the specificity of the reaction. Southern blot hybridisation with the third specific nested oligonucleotide probe corresponding to residues 24-30 of rat PC (HUM19, derived from human sequence but also conserved in rat sequence), was employed to verify the authenticity of the products. As shown in Fig 4.3, multiple bands hybridised strongly to the probe in all four tissues examined, confirming the existence of multiple transcripts. The sizes of these products were relatively small, ranging between 0.2-0.3 kb.

4.3.3 Cloning and sequencing of RACE products

Individual bands hybridising to the probe were gel purifed, cloned and sequenced. Nucleotide sequence analysis clearly showed that five distinct species of rat PC mRNAs were present. Comparison of the nucleotide sequences of these five alternative forms showed that they contained the same coding sequence i.e. 150 bp downstream from the ATG initiation codon encoding the first 50 N-terminal residues of rat PC as described in Chapter 3, but differed at their 5'-UTRs. These rat PC mRNAs isoforms were designated as rUTR A, rUTR B, rUTR C, rUTR D and rUTR E, respectively. The 5'-UTR of these PC mRNA isoforms are 222 bp, 113 bp, 95 bp, 96 bp and 59 bp in length, respectively (Fig. 4.4). These rat PC mRNA isoforms can be grouped into 2 classes according to their sequence
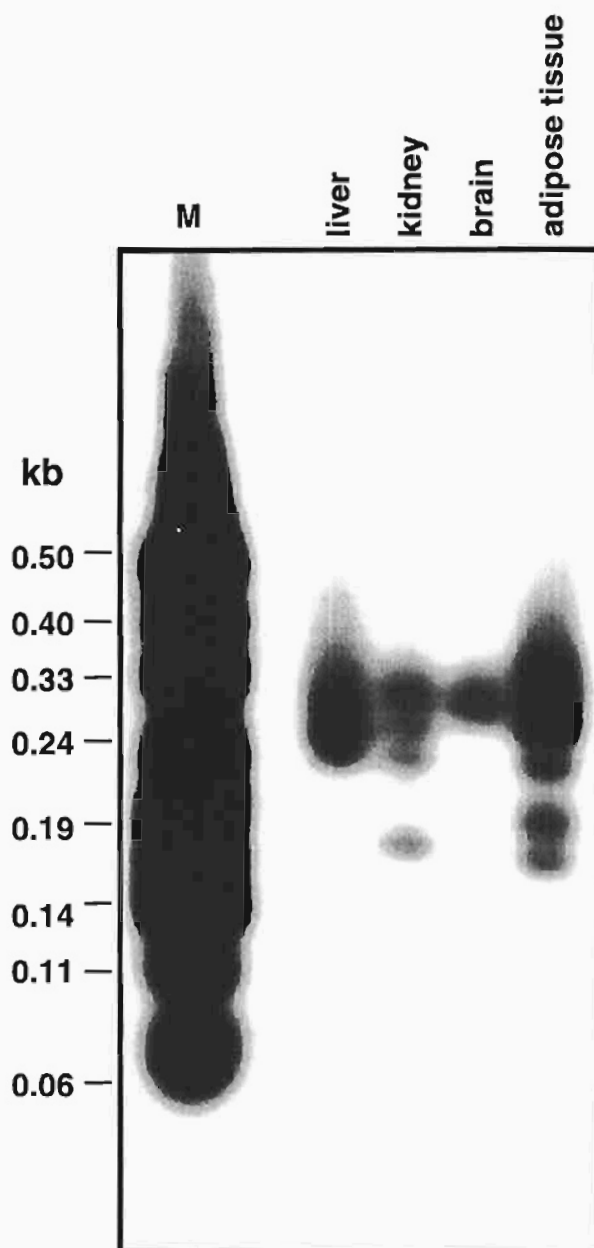
Figure 4.3 **Southern blot analysis of 5'-RACE products representing the 5'-ends of rat PC mRNAs, probed with ^{32}P-labeled HUM 19 primer (internal to MRACE II sequence).** RACE-PCR products were performed from 1:10 dilution of cDNA synthesised from total RNA extracted from rat liver, kidney, brain and adipose tissue. M, DNA marker.

Figure 4.4 The nucleotide sequences of the 5'-UTRs and coding regions of 5'-RACE-PCR clones encoding rat PC mRNAs. Sequence homology within each class of rat PC mRNA is boxed; the inferred amino acid sequence is shown as single letters. The dashed lines are introduced to maximise alignment of these sequences. Arrows represent the positions of the primer binding sites used for multiplex RT-PCR analysis.

CLASS I RAT PC mRNA

-222   OLIGO A   -200   -150

GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAG

OLIGO C   -100   -50

GTAATTGTTCCAGAACAGGTACATAGAGGAGGAGGCAAGCACACTGCACCAATCTGGCAATGAGCCGGTCCCCTGGCATGATCTCACCTGTTGTC

ACCCTTGTTTCAGGAAGAAACTTCATGAGACGATGCTGAAGTTCCAAACAGTTCGAGGGGGCCTGAGGCTCCTGGGTGTCCGCCGATCCTCC
                            M L K F Q T V R G G L R L L G V R R S S

ACAGCCCCCGTTGCCTCCCCAAATGTCCGGCGTCTGGAGTACAAGCCCATCAAGAAAGTAATGGTGGCCAACAGAGGTGAGATTGCCATC
T A P V A S P N V R R L E Y K P I K K V M V A N R G E I A I

rUTR A

-113   -100   OLIGO A   -50

GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAG

OLIGO C

GTAATTGTTCCAGAACAG---------------ATGCTGAAGTTCCAAACAGTTCGAGGGGGCCTGAGGCTCCTGGGTGTCCGCCGATCCTCC
                            M L K F Q T V R G G L R L L G V R R S S

ACAGCCCCCGTTGCCTCCCCAAATGTCCGGCGTCTGGAGTACAAGCCCATCAAGAAAGTAATGGTGGCCAACAGAGGTGAGATTGCCATC
T A P V A S P N V R R L E Y K P I K K V M V A N R G E I A I

rUTR B

-95   OLIGO A   -50

GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAG

----------------------------ATGCTGAAGTTCCAAACAGTTCGAGGGGGCCTGAGGCTCCTGGGTGTCCGCCGATCCTCC
                            M L K F Q T V R G G L R L L G V R R S S

ACAGCCCCCGTTGCCTCCCCAAATGTCCGGCGTCTGGAGTACAAGCCCATCAAGAAAGTAATGGTGGCCAACAGAGGTGAGATTGCCATC
T A P V A S P N V R R L E Y K P I K K V M V A N R G E I A I

rUTR C

CLASS II RAT PC mRNA

OLIGO B

-96

GTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG CTAGCATCTGCCTTCTGGAGA GCTGATTGCACCAATG

-50

----------------------------------ATGCTGAAGTTCCAAACAGTTCGAGGGGGGCCTGAGGCTCCTGGGTGTCCGCCGATCCTCC

M L K F Q T V R G G L R L L G V R R S S

ACAGCCCCCGTTGCCTCCCCAAATGTCCGGCGTCTGGAGTACAAGCCCATCAAGAAAGTAATGGTGGCCAACAGAGGTGAGATTGCCATC

T A P V A S P N V R R L E Y K P I K K V M V A N R G E I A I

rUTR D

OLIGO B

-59

GTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG --------------------------------------

----------------------------------ATGCTGAAGTTCCAAACAGTTCGAGGGGGGCCTGAGGCTCCTGGGTGTCCGCCGATCCTCC

M L K F Q T V R G G L R L L G V R R S S

ACAGCCCCCGTTGCCTCCCCAAATGTCCGGCGTCTGGAGTACAAGCCCATCAAGAAAGTAATGGTGGCCAACAGAGGTGAGATTGCCATC

T A P V A S P N V R R L E Y K P I K K V M V A N R G E I A I

rUTR E

homology at the 5'-UTR. rUTR A, rUTR B and rUTR C are grouped in a class I due to the presence of a common 95 bp region (boxed) while rUTR D and rUTR E are grouped in a class II due to the presence of a common 59 bp (boxed). Within class I, rUTR A is distinguished by the presence of a 109 bp insert located immediately 5' of the ATG initiation codon and 3' of the 18 bp insert which is also common to rUTR B but not to rUTR C (Fig. 4.4). Within class II, rUTR D is distinguished from rUTR E by a 37 bp insert immediately 5' of the ATG initiation codon. The sequence diversity observed among these 5 forms of rat PC mRNA was immediately upstream from the ATG initiation codon, suggesting that the splice junction between the 5'-UTR exon(s) and the first coding exon occurs at this position. None of the 5'-UTR of these mRNAs falls under the consensus sequence GCGCC(A/G)CCATG inferred by Kozak (1987).

4.3.4 Distribution of rat PC mRNA isoforms in various tissues

In order to investigate the tissue distribution of these PC mRNA isoforms isolated by RACE-PCR, a multiplex RT-PCR assay was developed such that these mRNA isoforms can be detected simultaneously. The schematic diagram of this multiplex RT-PCR is illustrated in Fig. 4.5. Ideally, reverse transcription of these five mRNA isoforms can be performed using the PC8 primer located downstream from the AUG initiation codon, as the cDNA synthesis primer because these mRNA isoforms contain the same coding sequence. Following cDNA synthesis, different cDNAs representing different isoforms of the PC mRNAs, which diverge at their 5'-UTRs, can then be PCR-amplifed using specific primers towards each class of the transcripts (oligo A for class I transcripts and oligo B for class II transcript). Since these transcripts have different insertion sequences upstream from the AUG initiation codon, the different sizes of fragments which are unique to each mRNA isoforms can be distinguished by electrophoresis on a high percent agarose gel (2.5%). The primer binding sites relative to the 5'-UTR of different PC mRNA isoforms are shown in Fig. 4.4.

As shown in Fig. 4.6 A, products of the expected size, 253 bp, 219 bp and 182 bp representing rUTR C, rUTR D and rUTR E, respectively were generated from different tissues. The rat PC mRNA isoforms exhibited tissue-specific expression. rUTR D was the

Figure 4.5 Schematic diagram of multiplex RT-PCR analysis. Different sequences at the 5'-UTRs of rat PC mRNA isoforms are shown by different hatched boxes. The positions of primers relative to the position of PC mRNAs are shown by arrows. The different sizes of PCR products obtained with different pairs of primers are also shown.
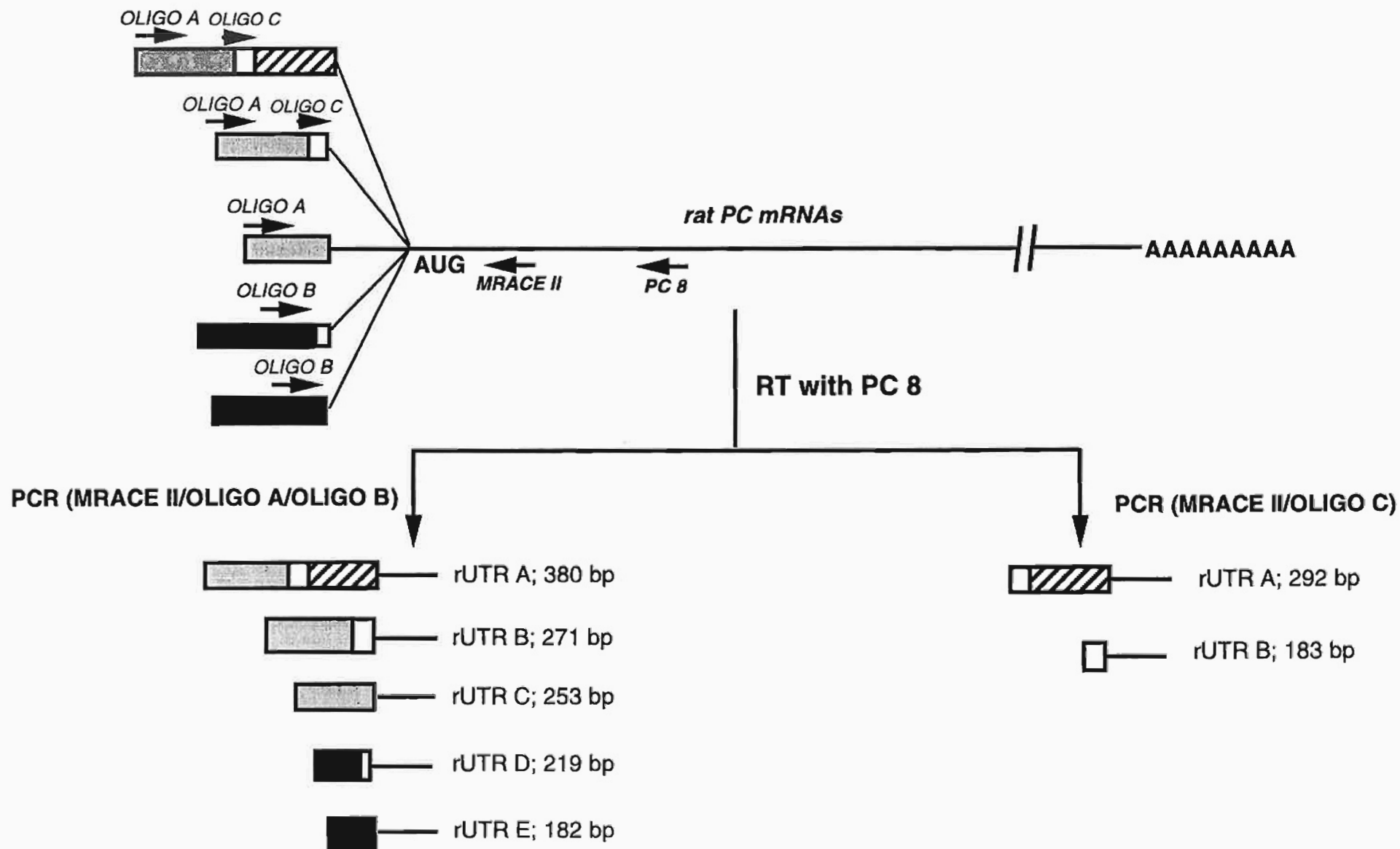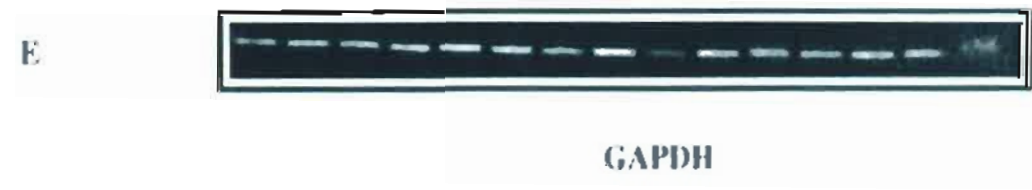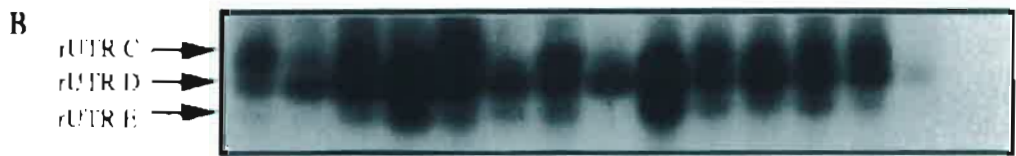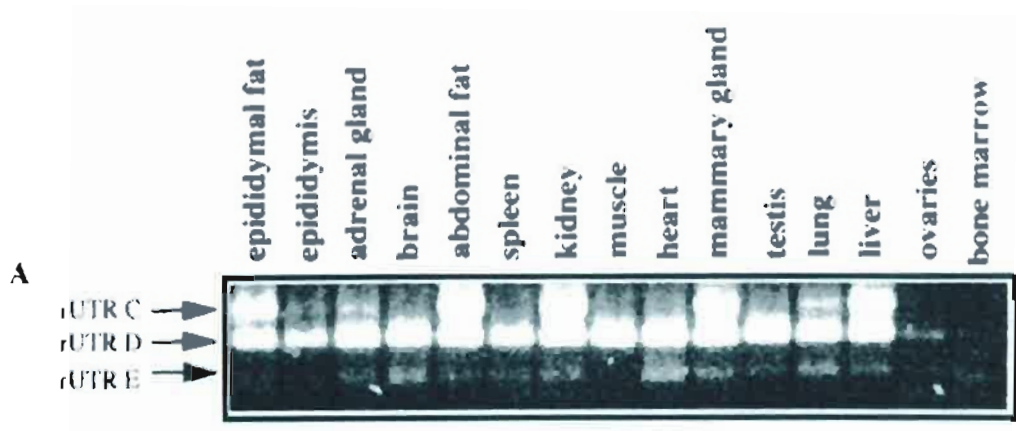
most abundant transcript in all tissues examined, except in ovaries and bone marrow where the level of expression was significantly lower. Likewise, rUTR E was also detected in most tissues but in much lower abundancy than rUTR D. Interestingly, this form was not detected in epididymis, skeletal muscle or bone marrow. rUTR C was only expressed in certain tissues, i.e. highly abundant in epididymal fat, abdominal fat, liver, kidney and lactating mammary gland but only moderately expressed in epididymis, adrenal gland and lung (Fig.4.6).

Southern blot analysis using an oligonucleotide probe (Fig. 4.6 B) against the common coding region of rat PC mRNA confirmed that these PCR products represented different PC mRNA isoforms. However, 380 bp and 271 bp products representing rUTR A and rUTR B were not detected in all tissues examined, even upon Southern blot hybridisation. This suggested that the expression of these two isoforms is much lower than the other three forms. Therefore, another primer, oligo C (Fig. 4.4) was designed such that it will bind to only rUTR A and rUTR B. As expected, 183 bp and 292 bp products representing both forms were detected by ethidium bromide staining in epididymal fat tissue but not in other tissues (Fig. 4.6 C). However, these 2 forms were detected upon Southern blot hybridisation in other tissues i.e. epididymis, abdominal fat tissue, liver, kidney, lactating and lactating mammary gland, suggesting that their concentrations are very low (Fig. 4.6 D). The tissue-specific expression of these PC mRNA isoforms in various rat tissues are summarised in Table 4.1.

From this expression pattern, it is clear that class I transcripts (rUTR A, rUTR B and rUTR C) are always co-expressed together. Likewise rUTR D and rUTR E comprising class II transcript are usually co-expressed. This suggests that expression of rat PC mRNA isoforms may be under the control of two promoters that would be differentially expressed in different tissues. It is likely that the 5'-flanking region of the sequence encoding class I transcripts is tissue-specific promoter, specifically expressed in lipogenic tissues (fat tissues, lactating mammary gland and liver) and gluconeogenic tissues (liver and kidney). On the other hand, the sequence upstream of the class II transcripts would be a housekeeping promoter which is expressed in most tissues.

Figure 4.6 Multiplex RT-PCR analysis of rat PC mRNA isoforms expressed in different tissues. Total RNAs were reverse transcribed and subjected to PCR with oligo A, oligo B and MRACE II (A), and PCR with oligo C and MRACE II (C) and analysed by electrophoresis on a 2.5% agarose gel, followed by Southern blot analysis using an internal oligonucleotide probe, HUM19 corresponding to a common coding region (B and D, respectively). In these tissues, GAPDH cDNA was also synthesised and used as an internal control for RT-PCR, (E).

A

rUTR C →
rUTR D →
rUTR E →

B

rUTR C →
rUTR D →
rUTR E →

C

rUTR A →
rUTR B →

D

rUTR A →
rUTR B →

E

GAPDH

Table 4.1 Distribution of different PC mRNA isoforms
in different rat tissues

Tissue	Class I transcripts			Class II transcripts	
	rUTR A	rUTR B	rUTR C	rUTR D	rUTR E
Abdominal fat	+	+	+	+	+
Epididymal fat pads	+	+	+	+	-
Lactating mammary gland	+	+	+	+	+
Kidney	+	+	+	+	+
Liver	+	+	+	+	+
Epididymis	+	+	-	+	-
Adrenal gland	-	-	-	+	+
Brain	-	-	-	+	+
Spleen	-	-	-	+	+
Lung	-	-	-	+	+
Heart	-	-	-	+	+
Testis	-	-	-	+	+
Skeletal muscle	-	-	-	+	-
Ovaries	-	-	-	+	-
Bone marrow	-	-	-	-	-

4.3.5 Multiple rat PC mRNA isoforms are transcribed from a single copy gene

To determine whether the identified rat PC mRNA isoforms are the product of a single gene, a cDNA probe corresponding to residues 551 to 907 of the transcarboxylation domain (Chapter 3) of the enzyme was used as a probe in a Southern blot analysis of restriction-digested rat genomic DNA. Only one restriction fragment of length 6 kb (*Bam*HI) or 12 kb (*Hind*III) was detected (Fig. 4.7). A single band was also seen when a cDNA probe corresponding to the biotinyl domain was used (Cassady, 1987). The uniqueness of fragment recognition with a cDNA probe in two different digests strongly suggests the

71
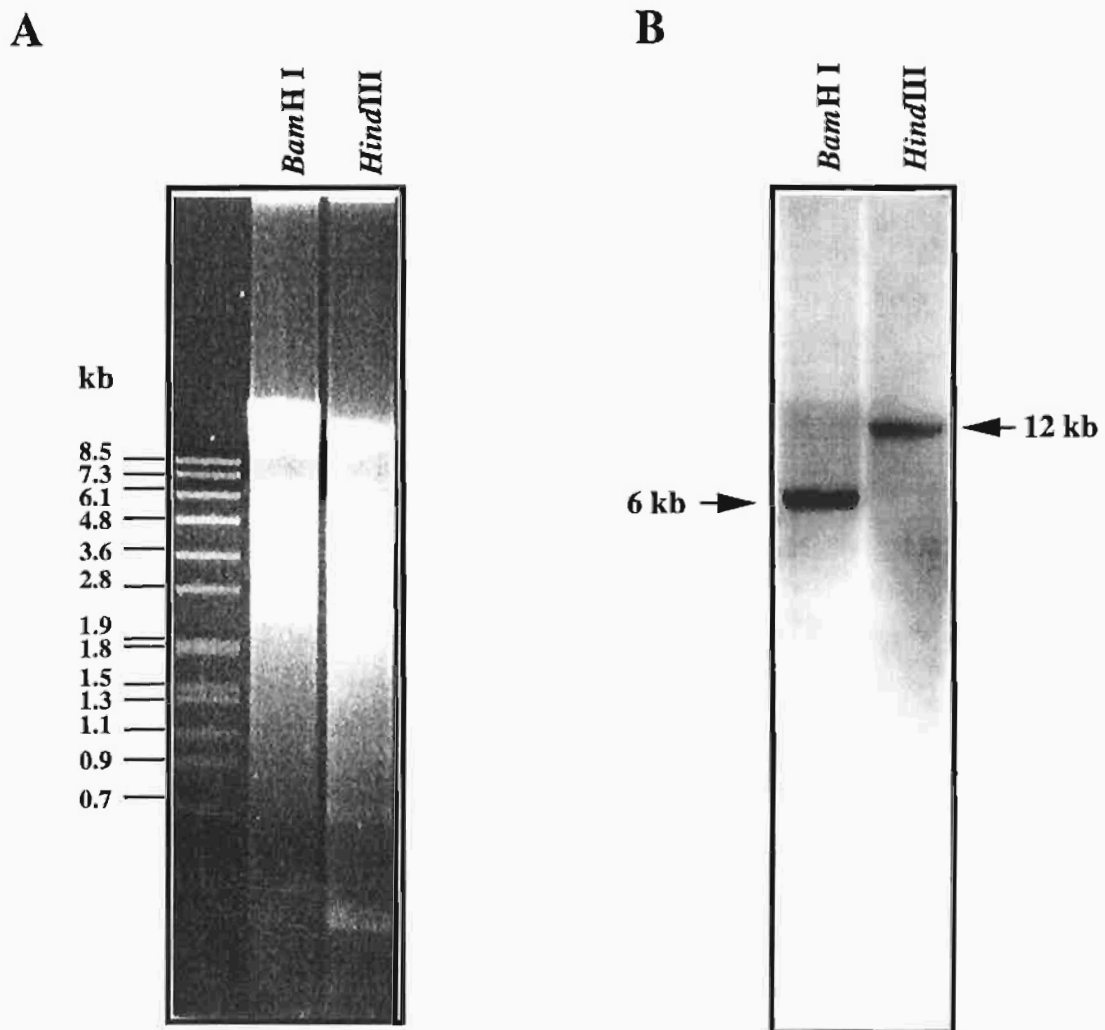
Figure 4.7   Analysis of copy number of PC gene using Southern blot of rat genomic DNA. Rat liver genomic DNA was digested with either *Bam* HI or *Hind* III and electrophoresed on 1% agarose gel (A). The DNAs were transferred to nylon membrane and probed with cDNA fragment corresponding to the transcarboxylation domain of the enzyme (Chapter 3) (B).

existence of only one copy of the PC gene per haploid genome. It is very likely that these multiple PC mRNA isoforms must arise by differential splicing of the 5'-UTR exons of the primary transcripts.

4.3.6 Identification of multiple PC mRNA isoforms in human liver by RACE-PCR

The presence of multiple transcripts of PC in different rat tissues raises the possibility of there being multiple transcripts of human PC. Additional evidence came from a previously obtained result in this laboratory by Dr M. Walker. She found several cDNA clones encoding human PC contained the same coding sequence but differed at the 5'-UTRs ranging in length from 0.5 kb to 1.5 kb. It was later found that some of these clones were chimeric and contained unrelated sequences of other genes ligated to the coding sequence of human PC. This situation is likely to have occurred during the library construction and has been reported by many investigators. To eliminate this artifact, RACE-PCR was carried out with human liver RNA.

Identification of multiple transcripts of human PC was essentially the same as that described for rat PC. In brief, total RNA of human liver was reverse transcribed with HUM15 primer designed from the published human PC cDNA sequence (Wexler *et al.*, 1994; MacKay *et al.*, 1994; Walker *et al.*, 1995b) [see Fig. 4.8] and amplified by PCR with HUM 19 and anchor primer using different dilutions i.e. 1:40, 1:80, 1:160 and 1:320 of cDNA template. Southern blot hybridisation with an internal oligonucleotide probe, HUM16 (Fig. 4.8) was carried out to verify the authenticity of the amplified products.

As showed in Fig. 4.9, different patterns of discrete bands were generated from different dilutions of cDNA. These bands were individually purified and cloned. Sequencing of the resulting clones revealed the presence of two different species of human PC mRNAs namely, hUTR A and hUTR B. Both hUTR A and hUTR B contained a common 130 bp of coding sequence and this corresponded to the N-terminal region of human cDNA (Wexler *et al.*, 1994; MacKay *et al.*, 1994; Walker *et al.*, 1995b) but differing in their 5'-UTR. hUTR A contained 201 bp while hUTR B contained 131 bp upstream from the ATG initiation codon (Fig. 4.10). Sequencing of other bands present in Fig. 4.9 revealed these to be truncated versions of either hUTR A or hUTR B. The 5'-UTR of hUTR

ATGCTGAAGTTCCGAACAGTCCATGGGGGGCCTGAGGCTCCTGGGAATCCGCCGAACCTCCACCGCCCCCGCTGCC
CACGACTTCAAGGCTTGTCAGGTACCCCCGGACTCCGAGGACCCTTAGGCGGCTTGGAGGTGGCGGGGGCGACGG
   M   L   K   F   R   T   V   H   G   G   L   R   L   L   G   I   R   R   T   S   T   A   P A A

HUM 16

TCCCCAAATGTCCGGCGCCTGGAGTATAAGCCCATCAAGAAAGTCATGGTGGCCAACAGAGGTGAGATTGCCATC
AGGGGTTTACAGGCCGCGGACCTCATATTCGGGTAGTTCTTTCAGTACCACCGGTTGTCTCCACTCTAACGGTAG
  S   P   N   V   R   R   L   E   Y   K   P   I   K   K   V   M   V   A   N   R   G   E   I   A   I

HUM 19

CGTGTGTTCCGGGCCTGCACGGAGCTGGGCATCCGCACCGTAGCCATCTACTCTGAGCAGGACACGGGCCAGATG
GCACACAAGGCCCGGACGTGCCTCGACCCGTAGGCGTGGCATCGGTAGATGAGACTCGTCCTGTGCCCGGTCTAC
  R   V   F   R   A   C   T   E   L   G   I   R   T   V   A   I   Y   S   E   Q   D   T   G   Q   M

HUM 15

Figure 4.8   Nucleotide sequence of the 5'-end of coding region of human PC cDNA sequence (Wexler *et al.*, 1994; MacKay *et al.*, 1994; Walker *et al.*, 1995b). Inferred amino acid sequences are shown by single capital letters. Primer binding sites used for cDNA synthesis and PCR are shown by arrows.
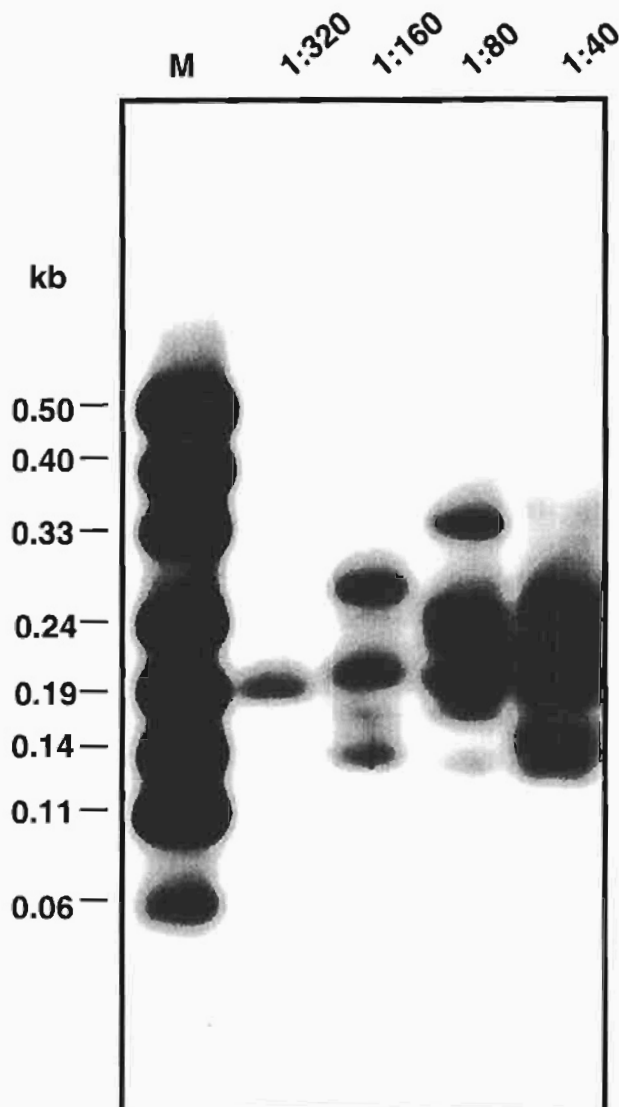
Figure 4.9   Southern blot analysis of 5'-RACE products represented human PC mRNAs were probed with the ^{32}P-labelled HUM19 primer (internal to MRACE II sequence). RACE-PCR products were generated from different dilutions of cDNA (1:40, 1:80, 1:160 and 1:320) synthesised from total liver RNA. M, DNA markers.

```
                                                         -201                    -180
                                                         GTTTGCCAAAGGAGTCCGGGCC                    hUTR A

                -160                    -140                    -120                    -100
TGACGGTGGAAGATGGGAATGGCCAAAAGAATAAACCATTTCCTCTTACTTACCGAATGAAAAGCTTGCTTCATCTTGGCGTTCAAGAC        hUTR A

        -80                    -60                    -40                    -20            -1
TCTCTAATATGATTCCACCTGTTGGGCTCTTTCTTCCATTTGCCTCCGCAGATAGTGTCTGCCTTCTGGAGAGCTGACCAAACACTAAGG      hUTR A
                                            GGGTTCTTACCCATTTAAAGTTTGACCAAACACTAAGG                kidney 1


ATG CTGAAGTTCCGAACAGTCCATGGGGGCCTGAGGCTCCTGGGAATCCGCCGAACCTCCACCGCCCCCGCTGCCTCCCCAAATGTCCGG      hUTR A
 M   L   K   F   R   T   V   H   G   G   L   R   L   L   G   I   R   R   T   S   T   A   P   A   A   S   P   N   V   R    kidney 1




                                        -131                    -100
                                        AGTTCTGGAAAGTCATTGACAGGGGATTAACCTAGAACATCA                hUTR B

        -80                    -60                    -40                    -20            -1
ATTATAAACAGGCAGCAGGCCTCAGGCCCAGACTTGGGTGTTACAGGAACAGTGTGGCCTCTCTGGAACTCTGCAGACCAACTGCCGTG       hUTR B
    GGCAGGCAGCAGGCCTCAGG-CCAGACTTGGGTGTTACAGGAACAGTGTGGCCTCTCTGGAACTCTGCAGACCAACTGCCGTG          kidney 2



ATG CTGAAGTTCCGAACAGTCCATGGGGGCCTGAGGCTCCTGGGAATCCGCCGAACCTCCACCGCCCCCGCTGCCTCCCCAAATGTCCGG      hUTR B
 M   L   K   F   R   T   V   H   G   G   L   R   L   L   G   I   R   R   T   S   T   A   P   A   A   S   P   N   V   R    kidney 2
```

Figure 4.10   The nucleotide sequences of 5'-UTRs and coding regions of 5'-RACE-PCR clones encoding human PC mRNAs. Nucleotide sequences are shown by capitals and inferred amino acid sequences are shown as single capitals. Also shown are the alignments of the 5'-UTRs of both classes of human PC mRNAs to those isolated from kidney cDNA libraries (kidney 1 and kidney 2) previously reported by Wexler et al. (1994) and MacKay et al., (1994), respectively. The dash in kidney 2 sequence is introduced to maximise alignment of these sequences.

A was identical to nucleotide -1 to -16 (relative to the initiation codon) of the 5'-UTR reported by Wexler *et al.* (1994) (Fig. 4.10, kidney 1) followed by 24 different nucleotides, and then an additional 161 nucleotides. Attempts to PCR amplify the 5'-UTR reported by Wexler *et al.* (1994) from either liver or kidney RNAs using primers corresponding to nucleotides -40 to -17 (sense) and nucleotides +355 to +357 (antisense) (Wexler *et al.*, 1994) were unsuccessful. These findings suggested that nucleotides -40 to -17 prior to the ATG initiation codon of the reported sequence (Wexler *et al.*, 1994) do not represent the 5'-UTR of human PC mRNA. The 5'-UTR of hUTR B was identical to the one described by MacKay *et al.* (1994) (Fig. 4.10, kidney 2), except for the insertion of a single nucleotide (C) at position -63, two nucleotide changes at position -82 and -83, followed by an additional 48 nucleotides. The sequence differences among these multiple forms are again observed immediately upstream from the ATG initiation codon, suggesting that the splice junction between the 5'-UTR exon(s) and the first coding exon occurs at this position. As it has previously been shown by somatic cell hybrids (Freytag and Collier, 1984) and fluorescence *in situ* hybridisation (Walker *et al.*, 1995a), that there is only one gene for PC in the human genome, these two species of human PC mRNA isoforms must arise by differential splicing from the first coding exon of the PC gene.

Unfortunately, no good quality human RNA extracted from different tissues was available at the time of this study, hence further investigation of the expression pattern of these two human PC mRNA isoforms was not carried out.

CHAPTER 5

STRUCTURAL ORGANISATION, ALTERNATE

SPLICING AND CHROMOSOMAL

LOCALISATION OF THE RAT PC GENE

5.1 INTRODUCTION

It has long been proposed that biotin carboxylases arose by fusion events within a pool of ancestral genes (Obermayer and Lynen 1976). Determination of the gene structures of mammalian biotin carboxylases provides an opportunity to explore this evolutionary relationship between various enzymes in the biotin carboxylase family. A comparison of the gene structure to different tissue-specific mRNA isoforms described in Chapter 4 would explain how the alternative splicing of the primary transcript of the gene can result in the production of these multiple transcripts with 5'-end heterogeneity. In addition, isolation of the promoter and the upstream *cis*-acting elements of the PC gene would also provide an excellent opportunity to investigate transcriptional regulation of the rat PC gene using a variety of molecular biology tools e.g. reporter gene analysis, DNase I footprinting, gel shift analysis.

Dr A.I. Cassady (1987) had previously screened a rat genomic library with a cDNA probe corresponding to the biotinyl domain of rat PC and this yielded one positive clone, λRG 1.2. Characterisation of this clone by Southern blot analysis and DNA sequencing revealed that the last 3'-end exons encoding the biotinyl domain of rat PC were located in this clone (Cassady, 1987). At the time, complete cDNA sequences corresponding to the transcarboxylase and the biotin carboxylase domains were not available. Thus determination of exons encoding these two domains could not be precisely mapped although Booker (1990) tried to map an additional 7 exons in this clone on the basis of conformity to the splice acceptor and donor consensus sequences of Cech (1983).

Long-distance PCR and Gene walking

The isolation of the gene of interest from a genomic library constructed in lambda bacteriophage is a time-consuming process and involves the screening several million plaques. Sometimes the fragments of the gene are not found in the library, no matter how many plaques are screened. Although other libraries have been developed including cosmid, yeast artificial chromosome (YAC) or even bacterial artificial chromosome (BAC) in order to reduce the number of clones to be screened, a number of investigators still face the problem of insert deletion from these libraries. Long-distance or long-range PCR has been developed

in order to eliminate this problem. LD-PCR was first described by Barnes (1994) and Cheng *et al.* (1994). This special feature of PCR requires the use of two different DNA polymerases i.e. ordinary stable polymerase e.g. *Taq* polymerase or *Tth* polymerase and a high fidelity polymerase e.g. Vent polymerase. This allows amplification of significantly longer fragments, up to 35 kb (Barnes, 1994) which is beyond the limit of ordinary *Taq* polymerase. Ideally, any fragments of the gene can be amplified from a genomic DNA template using primers that are designed from the cDNA sequence. However, more than one PCR reaction is needed if the gene of interest spans several hundred kilobases. Also, without knowledge of the intron/exon boundaries of the gene, it is difficult to design primers that do not cross intron/exon junctions. Several PCR-based methods are available for walking from a known region to an unknown region in cloned or uncloned genomic DNA. These include inverse PCR (Ochman *et al.*, 1988), randomly primed PCR (Parker *et al.*, 1991) and adaptor ligation PCR (Rosenthal and Jones, 1990; Riley *et al.*, 1990; Lagerstrom *et al.*, 1991; Jones and Winistorfer, 1993). However, these methods have not been successfully applied to walking into uncloned genomic DNA.

Recently, a PCR-based method for walking in uncloned genomic DNA has been described (Siebert *et al.*, 1995). This special feature includes several new PCR technologies i.e. adaptor ligation, vectorette PCR (Lagerstrom *et al.*, 1991), suppression PCR (Lukyanov *et al.*, 1994) and is commercially available from Clontech. Briefly, a special adaptor is ligated to the ends of DNA fragments generated by digestion of genomic DNA with five different enzymes i.e. *EcoRV*, *ScaI*, *DraI*, *PvuII* and *SspI* separately (Fig. 5.1). One end of the adaptor is blunt so that it can ligate to both ends of DNA fragments. The "vectorette" feature of the adaptor is the presence of an amine group on the 3'-end of the lower strand. This blocks polymerase catalysing the extension of the lower adaptor strand thereby preventing the generation of a primer binding site unless a defined, gene-specific primer extends a DNA strand opposite to the upper strand of the adaptor. Suppression PCR involves the use of adaptor sequences such that the upper strand is longer than the lower strand. In the event where non-specific amplification occurs and creates double-stranded DNA with a double-stranded adaptor at both ends, a "panhandle structure" will form during the annealing step due to the presence of inverted terminal repeats at both ends of the

Figure 5.1 Schematic diagram of Genome Walking kit. Genomic DNA is digested separately with five restriction enzymes that generate blunt ends followed by ligation with the adaptor sequence. One end of the adaptor sequence contains an inverted repeat sequence which upon denaturation and annealing during PCR, can form a "panhandle structure" which prevents non-specific amplification. Primary PCR using the first adaptor primer (AP1) and a gene-specific primer 1 is carried out and followed by secondary PCR using the second adaptor primer (AP2) and a second gene-specific (or nested) primer with *Tth* polymerase. PCR products can then be analysed by agarose gel electrophoresis.

Genomic DNA

ligate to adaptor | digest with different restriction enzymes

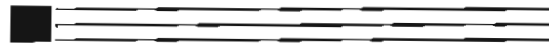EcoRV | PvuII | ScaI | SspI | DraI
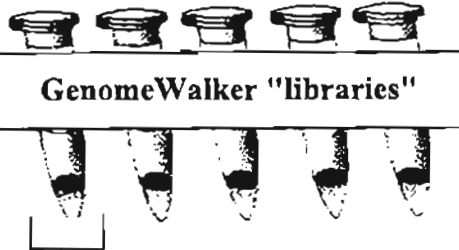
GenomeWalker "libraries"
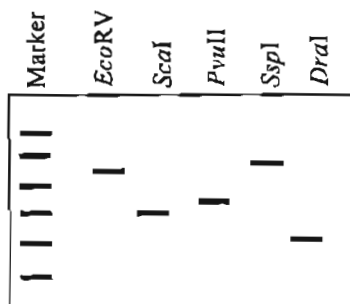
Amplify gene of interest
from all five libraries

GSP2
GSP1

5'-
N
N
-5'

AP1→
AP2→
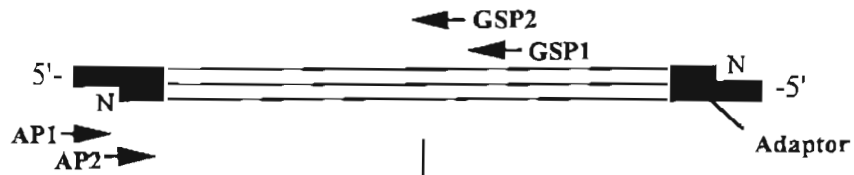
Adaptor

AP1
GSP1 — **Primary PCR**

AP2
GSP2 — **Secondary or "nested" PCR**

Agarose gel electrophoresis

Marker | EcoRV | ScaI | PvuII | SspI | DraI

adaptor. This structure is more stable than the primer-template hybrid thus suppressing exponential amplification of non-specific molecules. However, in the case where specific extension from gene-specific primer occurs, double-stranded DNA with a double-stranded adaptor at one end cannot form this structure. PCR can then proceed normally. Secondary PCR using nested primers at both ends can then be done to increase the specificity of the amplification reaction. The longest PCR fragment obtained from one of these five libraries can then be analysed. Multiple walks can be done if the PCR product does not contain the full length of the gene.

In this chapter, both this PCR-based technique and library screening were used to successfully obtain the complete structure of the rat PC gene.

5.2 SPECIFIC METHODS

5.2.1 Screening of lambda genomic libraries

A single colony of *E. coli* LE 392 was grown overnight in 20 ml of LB broth supplemented with 0.2 % maltose and 10 mM $MgSO_4$ at 37°C. Cells were pelleted by centrifugation at 5000 x g for 10 min and suspended in 5 ml of 10 mM $MgSO_4$. Two hundred microlitres of suspended cells were infected with 5 x 10^4 lambda phage particles at 37°C for 20 min. Twenty millilitres of top agar (0.7% agarose in NZCYM) were then added to the *E. coli*/phage mixture and immediately poured onto LMM agar plates. The plates were then inverted and incubated at 37°C for 6-9 h or until plaques were visible. Duplicate nylon membranes were placed over the agar plates allowing plaques to transfer for 5 min. The membranes were then soaked in 0.5 M NaOH-1.5 M NaCl for 2 x 5 min, in 0.5 M Tris pH 7.4-1.5 M NaCl for 2 x 5 min and finally washed in 10 x SSC for 5 min. The DNA on the membrane was fixed by UV-crosslinking. The hybridisation was performed at 42°C in hybridisation solution containing 40% formamide, 1% (w/v) SDS, 50 mM Tris-HCl, pH 7.4, 10% (w/v) PEG 6000 and 100 µg/ml sonicated salmon sperm DNA for 6 h. Fresh solution was then replaced with heat-denatured probes and further incubated at the same temperature overnight. The membranes were washed for 2 x 30 min in 2x SSC-0.1 % SDS, 2 x 30 min in 0.1x SSC-0.1% SDS, and 2 x 30 min at 55°C. The blots were autoradiographed at -70°C for 1-2 days. The positive clones were picked from the LMM plates and the phage were eluted overnight in phage storage buffer (10 mM Tris-HCl pH 7.4, 100 mM NaCl, 10 mM $MgCl_2$ and 0.05% gelatin). Different dilutions of phage were replated through the secondary and the tertiary screenings until a single plaque was obtained.

5.2.2 Small scale preparation of lambda bacteriophage DNA

A single colony of *E. coli* LE 392 was grown overnight in LB supplemented with 10 mM $MgSO_4$. An overnight culture was diluted with LB plus 10 mM $MgSO_4$ to A_{600} of 0.1-0.2. Ten millilitres of diluted culture were transferred to a 250 ml flask along with a single fresh plaque punched out from the LMM plate. The culture was grown at 37°C for 6 h or until the culture became clear due to the lysis of *E. coli* by phage. Two drops of chloroform were added and the lysate was centrifuged at 5000 x g for 10 min. The

77

supernatant was taken and mixed with an equal volume of phage storage buffer. The bacterial DNA was digested by adding 20 μl of 10 mg/ml DNaseI at 37°C for 1 h. The phage were precipitated by adding 1.8 ml of 4 M NaCl and 2 g of solid PEG 6000, and then incubated on ice for 1 h. The phage were recovered by centrifugation at 10,000 x g for 15 min at 4°C. The pellet was suspended in 0.3 ml phage storage buffer in a 1.5 ml microtube. PEG 6000 was removed by extraction with 0.3 ml chloroform twice followed by 0.3 ml of phenol and chloroform [1:1 (v/v)]. The phage DNA was precipitated by adding 0.1 volume of 3 M sodium acetate pH 5.2 followed by 2 volumes of cold absolute ethanol, and then kept at -20°C for 1 h. The phage DNA was pelleted by centrifugation at 10,000 x g for 15 min and washed with 70% ethanol. The pellet was dissolved in 400 μl water containing 20 mg/ml RNase A and incubated at 37°C for 30 min. The phage DNA was subsequently purified by phenol/chloroform extraction and ethanol precipitation as described above.

5.2.3 Screening of cosmid genomic library

Ten to twenty thousand bacterial colonies were plated onto nylon membranes (master filters) on 100 mm LB plates containing 50 μg/ml ampicillin and incubated for 8 h or more at 37°C until colonies were formed (1 mm in diameter). The nylon membranes were then removed and placed with the colonies side up onto a sterile Whatman paper. A new nylon membrane (replica filter) was pre-wetted on LB plates and placed over the master filter. Another sterile Whatman paper was then placed over the replica filter and pressed firmly with a glass plate. The replica filter was then removed from the master filter. A second membrane was then placed over the master filter as described above. The master and both replica filters were placed over new LB-agar plates and incubated at 37°C until colonies were visible. The replica filters with the colonies side up were sequentially placed on Whatman papers previously pre-wetted with 0.5 M NaOH for 1 min twice and with 1 M Tris-HCl pH 7.4 for 1 min. The filters were then placed on a third set of papers pre-wetted with 1 M Tris-HCl pH 7.4 and 1.5 M NaCl for 1 min before being soaked in the same solution for 5 min. The filters were air-dried and fixed by UV-cross linking.

5.2.4 Long distance PCR (LD-PCR)

LD-PCR was performed both with genomic DNA and with the Rat Genome Walker kit. For LD-PCR using genomic DNA as template, the reaction was carried out in a total volume of 50 μl containing 1X *Tth* PCR buffer (40 mM Tris-HCl, pH 9.3 at 25°C, 15 mM potassium acetate), 1.1 mM magnesium acetate, 0.2 mM of each dNTP, 0.25 μM each primer, 100 ng of genomic DNA and 1 μl of 50X Advantage™ *Tth* polymerase mix. The reaction mixture was subjected to 42 rounds of PCR amplification. The PCR profile consisted of an initial denaturation at 94°C for 1 min followed by 7 cycles of denaturation at 94°C for 30 sec, annealing and extension at 72°C for 6 min, 35 cycles of denaturation at 94°C for 30 sec, annealing and extension at 68°C for 6 min, and followed by the final extension at 68°C for 12 min.

For the reactions performed with the Rat Genome Walker kit, the PCR was carried out as follows. The primary PCR was carried out in a total volume of 50 μl containing 1X *Tth* PCR buffer, 1.1 mM magnesium acetate, 0.2 mM each dNTP, 0.25 μM AP1 primer and the first gene-specific primer (GSP1), 1 μl each of the *Dra*I, *Eco*RV, *Pvu*II, *Sca*I or *Ssp*I libraries, and 1 μl of 50X Advantage™ *Tth* polymerase mix. The PCR profile consisted of 7 cycles of denaturation at 94°C for 25 sec, annealing and extension at 72°C for 6 min, 36 cycles of denaturation at 94°C for 25 sec, annealing and extension at 67°C for 6 min, and followed by the final extension at 67°C for 12 min. One microlitre of the primary PCR product was then diluted to 1:50 with water, and 1 μl was used as template for the secondary PCR using the conditions as described above except the primers were AP2 and the second gene-specific primer (GSP2). The PCR program consisted of 5 cycles of denaturation at 94°C for 25 sec, annealing and extension at 72°C for 6 min, 22 cycles of denaturation at 94°C for 25 sec, annealing and extension at 67°C for 5 min, and the final extension at 67°C for 12 min.

5.3 RESULTS

5.3.1 Organisation of the coding exons

The nucleotide sequence of the λRG 1.2 previously determined by Booker (1990) was directly compared with the complete cDNA sequence reported in Chapter 3. This identified 7 additional exons located upstream from the last 3'-exon encompassing the biotinyl domain of the rat PC (Cassady, 1987). These 10 exons corresponded to 2/3 of the rat PC cDNA sequence (Fig. 5.2). However the exons encoding the biotin carboxylation domain of the rat PC are not present in this lambda clone.

The first attempt to isolate a segment of the PC gene from rat genomic DNA by PCR was carried out using primers randomly chosen from the cDNA sequence since no information on the PC gene structure of any mammalian species was available at the time. The first pair of primers including PC9 (corresponding to positions +61 to +81 relative to an initiation codon; sense primer) and PC8 (corresponding to positions +484 to +503; antisense primer) and a second pair of primers including PC12 [corresponding to positions +484 to +503, sense primer (complementary to PC8)] and PC7A (corresponding to positions +826 to +847) were used to PCR amplify genomic DNA using ordinary *Taq* polymerase. Upon amplification, single bands 0.75 kb and 0.60 kb product were generated from both pairs of primers respectively. Using PC9 and PC7A as primers in another PCR reaction, a 1.35 kb fragment which corresponded to the sum of the above fragments was obtained, suggesting that they are specific products (Fig. 5.3A). This 1.35 kb fragment was purified from the gel and directly sequenced in both directions. Comparison of its sequence to the cDNA sequence revealed that this fragment encompassed 6 exons and 5 introns (Fig. 5.3B) corresponding to half of the biotin carboxylation domain. The positions of the above primer binding sites which were revealed after sequencing are also shown in Fig. 5.3B.

The second attempt to PCR-amplify genomic DNA using the same sense primer (PC 9) and an antisense primer directed against the 5'-end of λRG 1.2 clone with long-range enzyme (Advantage *Tth* polymerase mix) failed to give any product, even though the extension time in each cycle was increased up to 20 min, as recommended by the manufacturer for amplification up to 20-40 kb fragments. This suggested the presence of very large intron(s) between the last exon located in the 1.35 kb fragment and the first exon
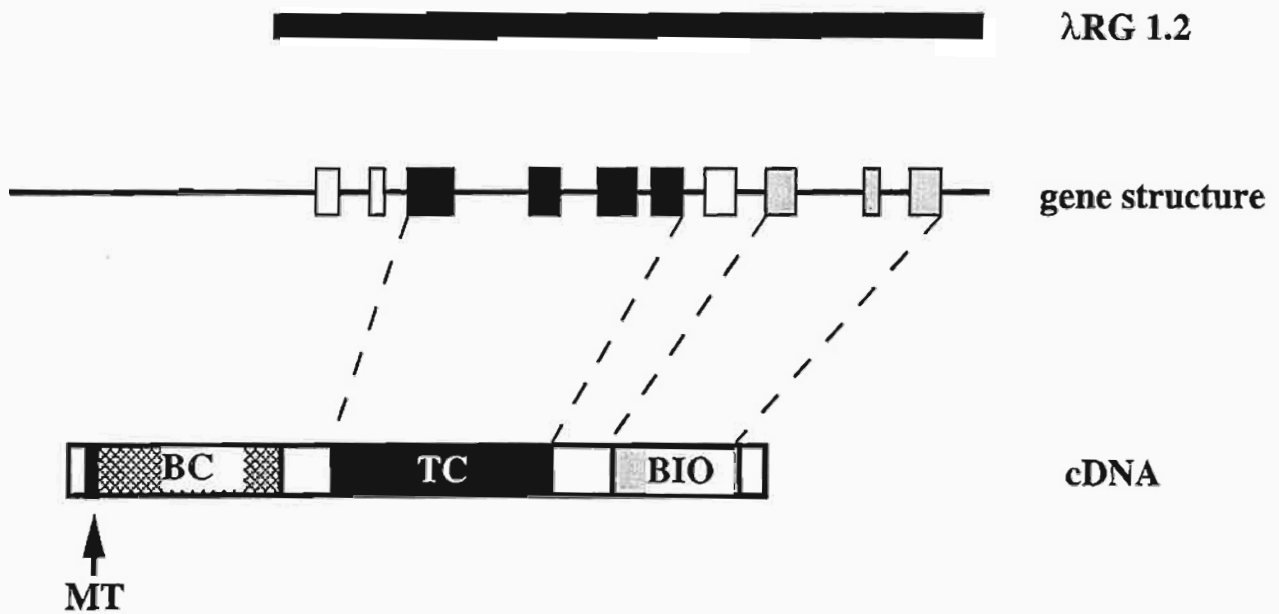
Figure 5.2   Schematic diagram of the lambda genomic clone, λRG 1.2 previously isolated and sequenced by Cassady (1987) and Booker (1990) relative to the gene structure and cDNA of rat PC. Boxes represent exons relative to cDNA structure. BC, biotin carboxylation domain; TC, transcarboxylation domain; BIO, biotinyl domain; MT, mitochondrial targetting sequence.
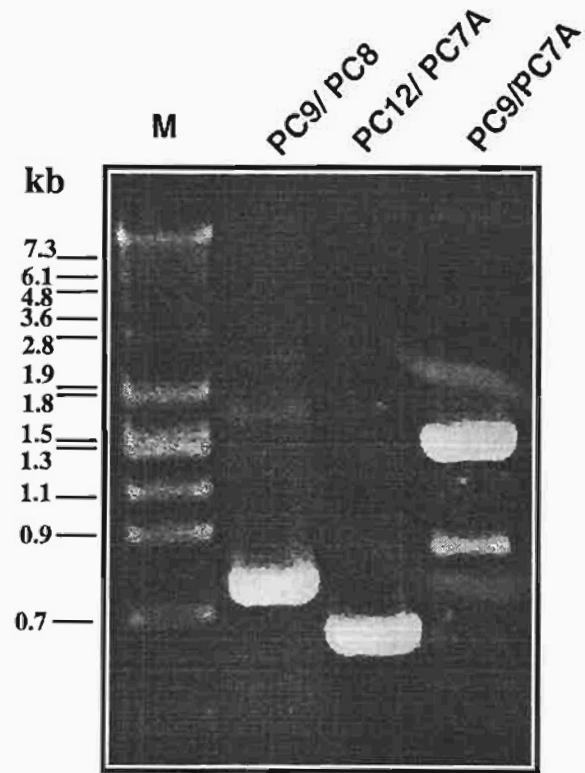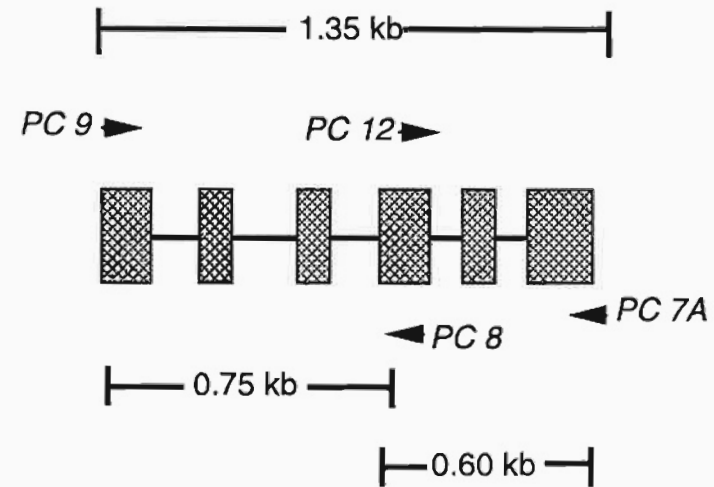
Figure 5.3   Isolation of 0.85 kb, 0.65 kb and 1.5 kb genomic fragments containing 5 exons using different pairs of PCR primers derived from cDNA sequence. A, PCR products using different combinations of PCR primers analysed on 1.5% agarose gel., B, position of exons on genomic DNA and primer binding site revealed upon sequencing PCR products in A.
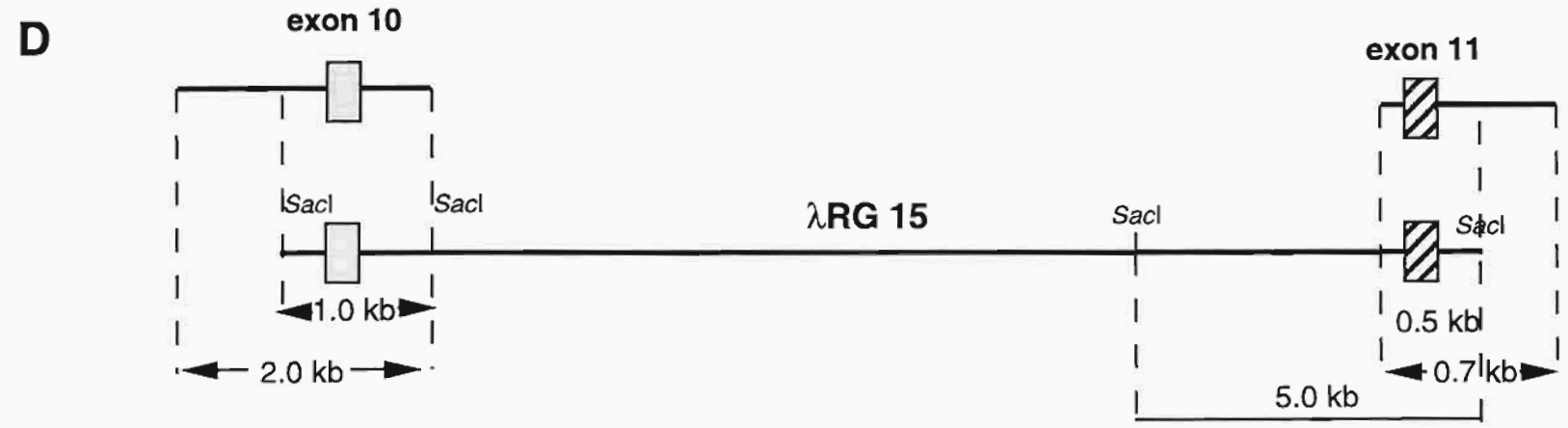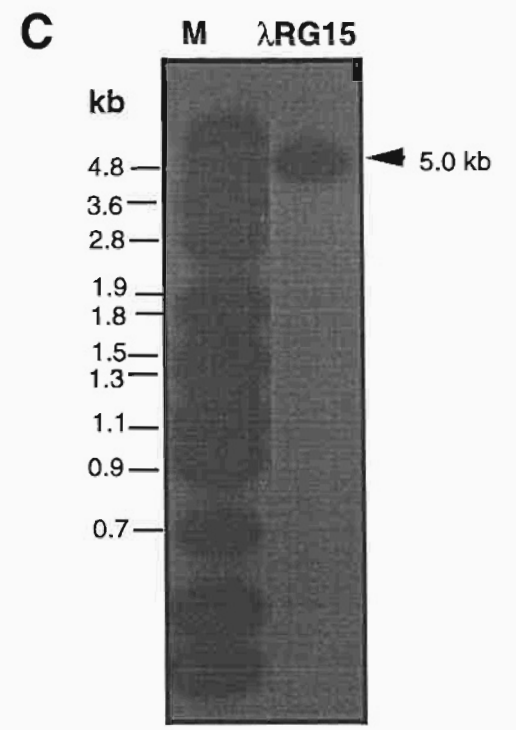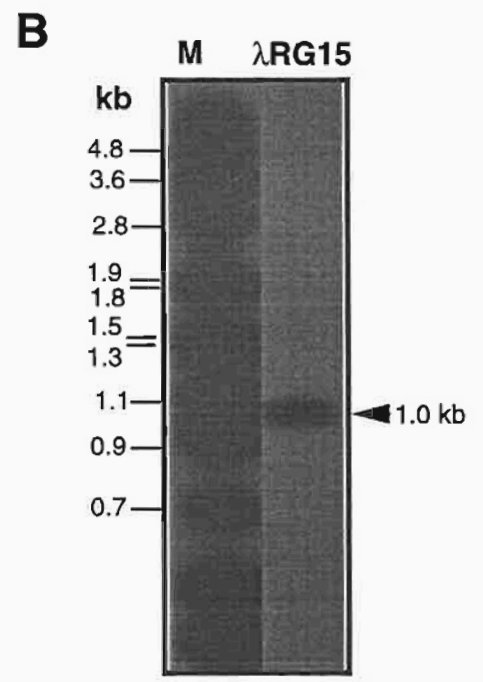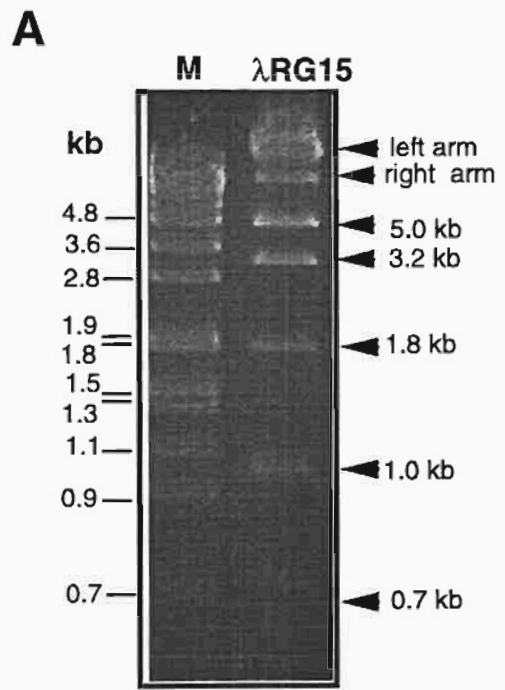
located in the λRG 1.2 clone. Since the first exon of λRG 1.2 clone was completely determined and corresponded to a fragment of the enzyme between T457 and L504, the 3'-end of the upstream exon must therefore end at K456. A new antisense primer (PC18, 5'-CTTTACACCTCGGACACGGAACTCC-3'), corresponding to the segment between residues E449 and Lys 456 (cDNA positions +1344 to +1368), was designed and used together with the PC9 sense primer in long-distance PCR. This would avoid amplifcation of very large intron that might be located in front of the first exon of the λRG 1.2 clone if that was the case. Upon amplification, an 8.0 kb fragment was generated with minor bands of 2.0 kb to 0.5 kb (Fig. 5.4A). Secondary PCR with nested primers [PC17, complementary to PC7A (Fig. 5.4A)] and PC 19 [corresponding to the segment between T441 and A448 (positions +1320 to +1343), 5'-GCCAGGGCTCTGCTCATCTTGGTG-3'] was carried out using the primary PCR product as template. A single strong band of 6.65 kb was generated, suggesting specific amplification. Both the 8.0 kb and 6.65 kb PCR fragments were purified from the gel and partially sequenced using primers from both ends. Comparison of their nucleotide sequences with the cDNA sequence of PC confirmed that they are specific products. The 6.65 kb PCR fragment contained 3 additional exons downstream from the first 6 exons identified within the 1.35 kb fragment. The positions of the above primer binding sites which were revealed after sequencing are also shown in Fig. 5.4B.

However, the gap representing the intron between the 3'-end of the 8.0 kb fragment and 5'-end of the λRG 1.2 was still missing. Therefore, library screcning was undertaken to search for this fragment. Approximately 750,000 plaques were screened with a mixture of the 3'-end fragment of the 8.0 kb PCR product and the 5'-end of λRG 1.2 as probes. The aim was to isolate any clones that contained sequences corresponding to the 3'-end of the 8.0 kb PCR product and the 5'-end of λRG1.2 clone. One positive clone, λRG 15 was found to hybridise to both probes after tertiary screening. Phage DNA from this plaque was digested with *Sac*I which released the insert sizes of 5.0, 3.2, 1.8, 1.0 and 0.7 kb from the left and right arms of lambda phage (Fig. 5.5A) and followed by Southern blot hybridisation using both the 3'-end of the 8.0 kb fragment and the 5'-end of λRG 1.2 as probes separately. As shown in Fig. 5.5B and 5.5C, the 5.0 kb fragment hybridiscd to the 5'-end of λRG 1.2 clone while the 1.0 kb fragment hybridised to the 3'-end of the 8.0 kb PCR product. Partial
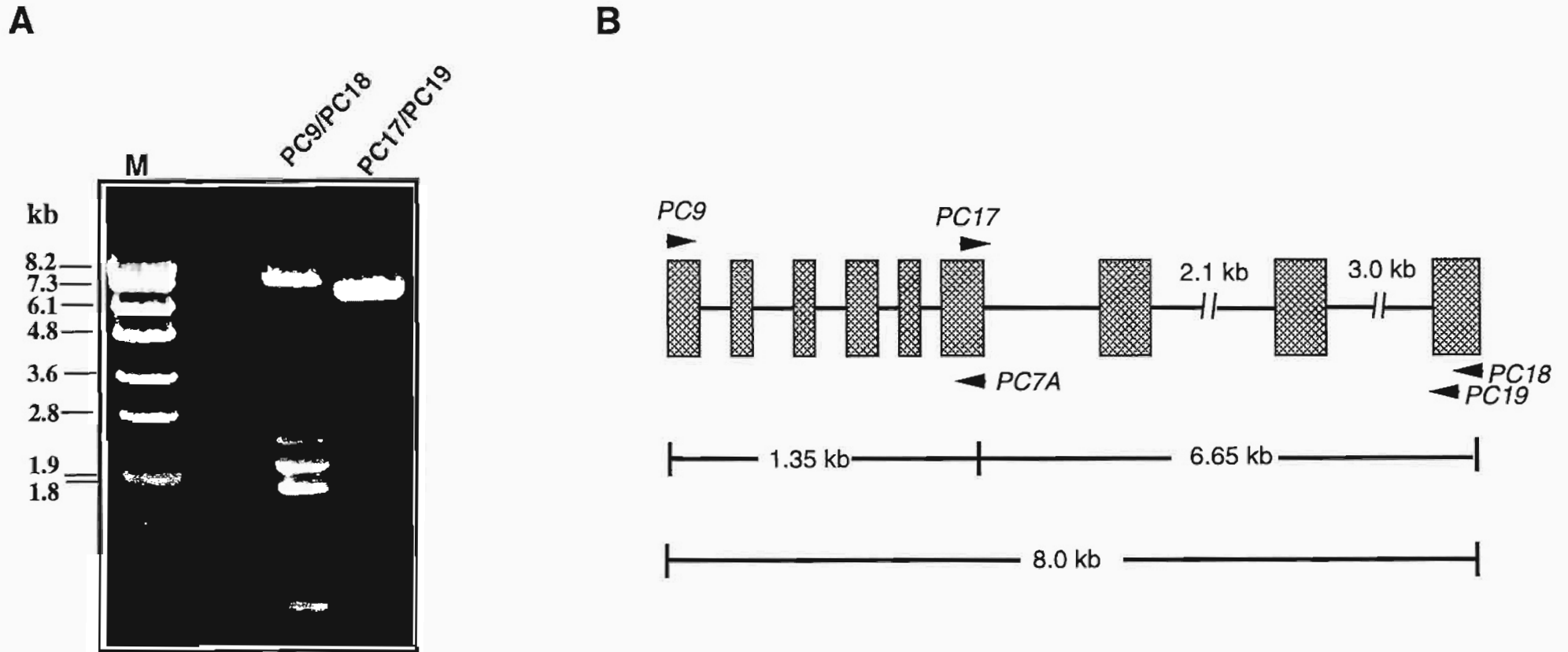
Figure 5.4   Isolation of the 8.0 kb genomic fragment containing 9 exons using long distance PCR. A, PCR products amplified using different pairs of PCR primers analysed on 1.5% agarose gel electrophoresis. B, positions of exons and PCR primers on the genomic DNA revealed upon sequencing PCR products in A.

Figure 5.5 Southern analysis of the λRG 15 clone. Phage DNA isolated from this clone was digested with *Sac*I and analysed by 0.7% agarose gel electrophoresis (A) and by Southern blot hybridisation with either a 2.0 kb probe amplified from the 3'-end of the 8.0 kb PCR product (B) or a 0.7 kb probe amplified from the 5'-end of the λRG 1.2 clone (C). Overlapping regions of this clone and both probes are also indicated (D).

A

M λRG15

kb

◄ left arm
◄ right arm

4.8 —
3.6 —
2.8 —

◄ 5.0 kb
◄ 3.2 kb

1.9 —
1.8 —

◄ 1.8 kb

1.5 —
1.3 —

1.1 —

◄ 1.0 kb

0.9 —

0.7 —

◄ 0.7 kb

B

M λRG15

kb

4.8 —
3.6 —

2.8 —

1.9 —
1.8 —

1.5 —
1.3 —

1.1 —
0.9 —

◄ 1.0 kb

0.7 —

C

M λRG15

kb

4.8 —
3.6 —

2.8 —

1.9 —
1.8 —

1.5 —
1.3 —

1.1 —

0.9 —

0.7 —

◄ 5.0 kb

D

exon 10

exon 11

SacI   SacI   λRG 15   SacI   SacI

1.0 kb

0.5 kb

2.0 kb

5.0 kb

0.7 kb

sequencing from both ends of this clone using primers that are complementary with the left and right arms of lambda DNA (sequences shown in Chapter 2), showed that the insert of this clone overlapped with the 3'-end of the 8.0 kb PCR product and the 5'-end of λRG 1.2 clone by approximately 1.0 kb and 0.5 kb, respectively (Fig. 5.5D). The 5'-end of this clone contained only one exon which overlapped with the last exon identified in the 8.0 kb PCR product and adjoins a very large intron approximately 10 kb followed by a single exon located at the 3' end which also overlapped with the first exon of λRG 1.2 clone.

When all of these fragments isolated by PCR, LD-PCR and library screening were assembled and compared with the full length cDNA sequence, the complete structure of the coding exons was obtained. As shown in Fig. 5.6, the coding region of the rat PC gene is comprised of nineteen exons spanning approximately 23 kb. Exon 2 is the first coding exon beginning immediately with the ATG initiation codon. This exon spans 138 bp downstream, encoding the mitochondrial targetting sequence and part of the biotin carboxylation domain. The biotin carboxylation domain and the transcarboxylation domain are encoded by exons 2-10 and exons 13-16, respectively (Fig. 5.6). The last three exons, exons 18-20, encode the biotinyl domain of the enzyme as well as the 3'-untranslated region including the polyadenylation signal. The polypeptide segment linking the biotin carboxylation domain and the transcarboxylation domain is encoded by exons 11 and 12, while that linking the transcarboxylation domain and the biotinyl domain was encoded by exon 17. All coding exons sequences are in agreement with the cDNA sequence reported in Chapter 3. Different DNA fragments generated by PCR and LD-PCR with different pairs of primers are summarised in Table 5.1.

The sizes of the introns are relatively small, being less than 1.0 kb except for introns I, J and K which were 3.0 kb, 2.1 kb and 10.0 kb, respectively (Table 5.2). The sequences surrounding the intron-exon boundaries are shown in Table 5.2 together with the amino acid ranges encoded by these exons. The intron-exon splice junction sequences matched the consensus sequences: each intron begins with a GT dinucleotide and ends with an AG dinucleotide, sequences thought to be necessary for correct RNA splicing (Breathnach and Chambon, 1981).

Figure 5.6 Structure of the rat PC gene. A, overlapping lambda phage clones (λRG 1.2, λRG 15 and λRG 2) and PCR products (A, B, C, D and E) and non-overlapping fragments (F and G) spanning over 40 kb of the gene are indicated. B, organization of exons and introns of the gene. The positions of exons are shown in boxes and numbered. The initiation codon (ATG) and stop codon (TGA) are shown. C, the structure of PC cDNA (Chapter 3). The regions encoding the three functional domains are indicated. MT, mitochondrial targetting sequence; BC, biotin carboxylation domain; TC, transcarboxylation domain; BIO, biotinyl domain.
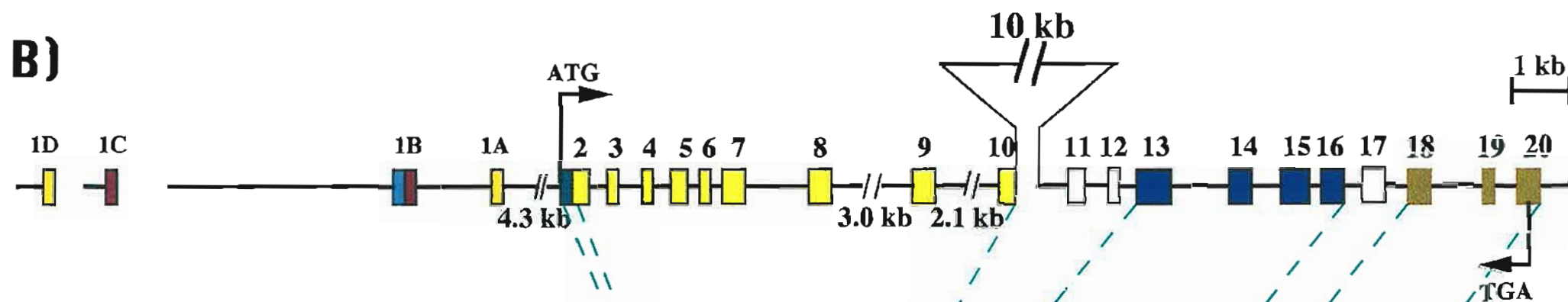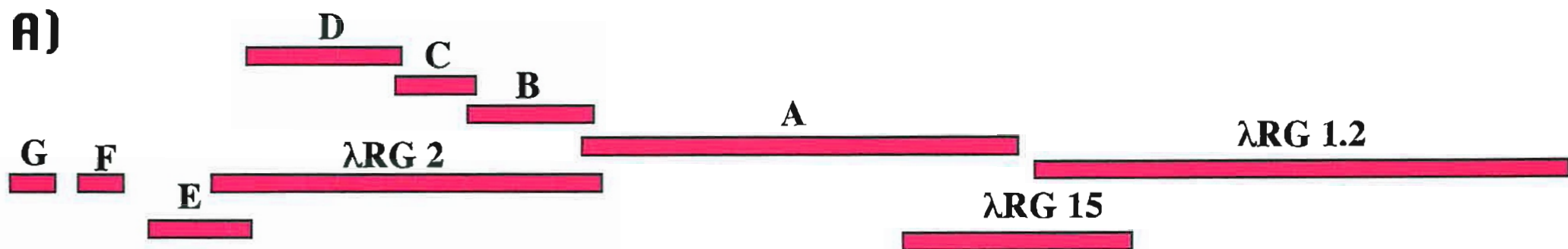
A)

D C B A λRG 1.2

G F λRG 2

E λRG 15

B)

10 kb

1D 1C 1B 1A ATG 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

4.3 kb 3.0 kb 2.1 kb TGA

BC TC BIO

MT

C)

40 kb

Table 5.1

Oligonucleotides used for PCR, LD-PCR and Southern blot analysis.

Underlines indicate the restriction sites (GAATTC, *Eco*RI; GGATCC, *Bam*HI, GGTACC, *Kpn* I) introduced at the 5' end to facilitate cloning.

oligonucleotide name	fragments amplified by PCR or LD-PCR (kb)	oligonucleotide sequence
PC 9		5'-ACAGCCCCCGTTGCCTCCCCA-3'
PC18	A (8.0)	5'-CTTTACACCTCGGACACGGAACTCC-3'
RAGE 1[1]		5'-GATACCCAGCTCTGTGCAGGCACGAAACAC-3'
RAGE 2[2]	B (5.5)	5'-GGC<u>GAATTC</u>CACCATTACTTTCTTGATGGGCTTGTA-3'
RAGE 5		5'-CTCCTAGCTGGCATCCAGGTTCTC-3'
Oligo A	C (1.5)	5'-GCCAATGACCTCGGGTGGAGCAG-3'
PC 24[1]		5'-TACCTTGTACCAGCAGCCAAACACC-3'
PC 25[2]	D (4.0)	5'-ATCGTGCTGCACACAGAGGACGTGTT-3'
PC 34[1]		5'-GGATAGGCGGCAAGATTCCTGAAGCAGC-3'
PC 35[2]	E (3.6)	5'-AAA<u>GGATCC</u>AACCCAGCCAGGAGTGAGGTGGAG-3'
PC 38[1]		5'-CATTGGTGCAATCAGCTCTCCAGAA-3'
PC 39[2]	F (1.2)	5'-AGCTCTCCAGAAGGCAGATGCTAG-3'
PC 32[1]		5'-CTTCGCCCCGTCGCCTCAAGCCGT-3'
PC 33[2]	G (1.1)	5'-CCGTGGCCGCCGCCTCTGACGG-3'
Oligo B	-	5'-GGCTTGAGGCGACGGGGCGAAG-3'
Oligo C	-	5'-CGTCTCATGAAGTTTCTTCCTG-3'

[1,2] indicate GSP1 and GSP 2 used for LD-PCR.

5.3.2 5'-Untranslated region exons and alternative splicing

As described in Chapter 4, multiple rat PC mRNA isoforms with 5'-end heterogeneity were identified and exhibited tissue-specific expression. To understand the origin of these heterogeneous mRNAs and to determine the mechanisms of PC gene expression, the 5'-end of the rat PC was also isolated using PCR-based gene walking as described earlier.

The first walk towards the 5'-end, upstream from the first coding exon, was carried out using 1st and 2nd gene-specific primer (GSP1 and GSP2), namely RAGE1 and RAGE2. These two primers were designed from nucleotide sequences within exon 2. Upon secondary PCR, single strong bands were generated from *Dra*I, *Pvu*II, *Sca*I and *Ssp*I libraries with size of 0.9 kb, 4.8 kb, 5.0 kb and 0.6 kb, respectively, but not the *Eco*RV library (Fig. 5.7A). These four fragments were hybridised to an internal oligo probe (PC9) confirming specificity of the nested PCR (Fig. 5.7B). To verify if any of these fragments contained 5'-UTR exons corresponding to the 5'-UTRs of rat PC mRNA identified by

Table 5.2
Exon-intron boundaries of the rat PC gene.

Lowercase letters indicate the intron sequence, and uppercase letters indicate the exon sequence. Positions of splice sites are relative to the cDNA sequence (Chapter 3).
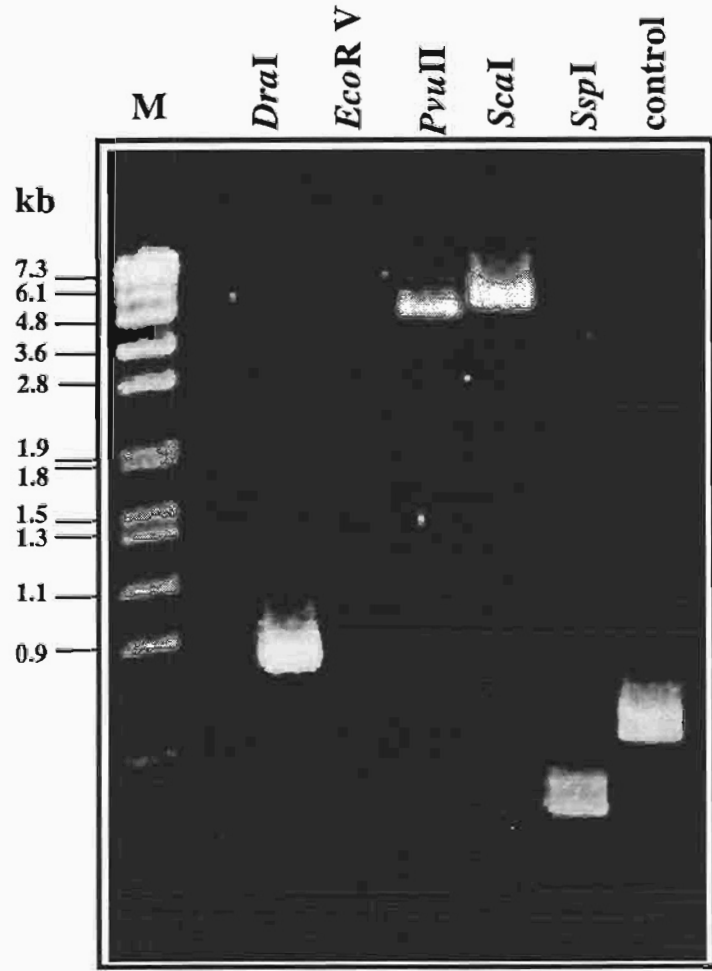
Exon		Exon/intron boundaries		
Number	Residues	5'-Donor site	Intron name/size	3'Acceptor site
1D	5'UTR	GCGAAG-	ND	cttcagCTAGCC
1C	5'UTR	CCAATG-	ND	cctcagATGCTG
1B	5'UTR	GAACAGgtatgg	A/~4.3 kb	tattagGTACAT
1A	5'UTR	GAGACGgtacga	B/~1.3 kb	cctcagATGCTG
2	M_1-G_{46}	ACAGAGgtaaat	C/116 bp	tcctagGTGAGA
3	G_{46}-K_{107}	GCCAAGgtagag	D/173 bp	ttgcagGAGAAT
4	E_{108}-G_{173}	CTGCAGgtgaca	E/99 bp	ccaacagGCGTTC
5	G_{173}-E_{211}	TACGAGgtaacg	F/131 bp	ccctagGAGCTG
6	E_{212}-L_{250}	TCCTAGgtgagt	G/177 bp	tcccagGGGAC
7	G_{251}-Q_{301}	AAGCAGgtgagg	H/~1.0 kb	gtgcagGTTGGC
8	V_{302}-D_{341}	TACAGAgtgagc	I/~3.0 kb	ttgcagTGTGGA
9	V_{342}-E_{395}	ATTGAGgtaggt	J/~2.1 kb	ccagagGTTTTC
10	V_{296}-K_{456}	GTAAAGgtgaga	K/~10 kb	ttgcagACCAAC
11	T_{457}-L_{504}	ACCTTGgtttgc	L/250 bp	ccccagGACAC
12	G_{505}-I_{534}	CCATAGgtagg	M/180 bp	accctagGCCCA
13	G_{535}-K_{661}	CTTCAAgtaagc	N/450 bp	tgccagGTTCTG
14	F_{662}-K_{741}	ATTAAGgtgcca	O/128 bp	tcccagGACATG
15	D_{742}-T_{824}	ACACAGgtagga	P/180 bp	ccacagAGGTAC
16	E_{825}-K_{906}	ATCAAGgtgagc	Q/90 bp	cttcagGTGACA
17	V_{907}-K_{966}	TCTAAGgtaggg	R/76 bp	cctcagGTGCTA
18	V_{967}-E_{1049}	TTTGAGgtcagt	S/202 bp	ctccagGTTGAG
19	V_{1050}-K_{1096}	ATGAAGgtacag	T/83 bp	ctgcagGAGATG
20	E_{1097}-E_{1178}, 3' UTR	GCCCTCC	-	-

ND, not determined

RACE-PCR, Southern blot analysis was carried out. Different oligonucleotide probes (A, B and C) directed against various regions of the different 5'-UTRs of rat PC mRNAs (see Fig. 5.8) were used to localise potential exons. Oligo C hybridised strongly to the 4.8 kb and the 5.0 kb fragments amplifed from *PvuII* and *ScaI* libraries (Fig. 5.7C). Since oligo C did not

Figure 5.7 Isolation of exon 1A of the rat PC gene by PCR based-Gene Walking. Different sizes of PCR fragments were generated from different restriction libraries (*Dra*I, *Eco*RV, *Pvu*II, *Sca*I and *Ssp*I) upon secondary PCR amplification and analysed by electrophoresis on a 1.5% agarose gel (A) followed by Southern blot analysis using different oligonucleotide probes directed against the coding region (B) and the 5'-UTRs of different rat PC mRNAs (C and D).
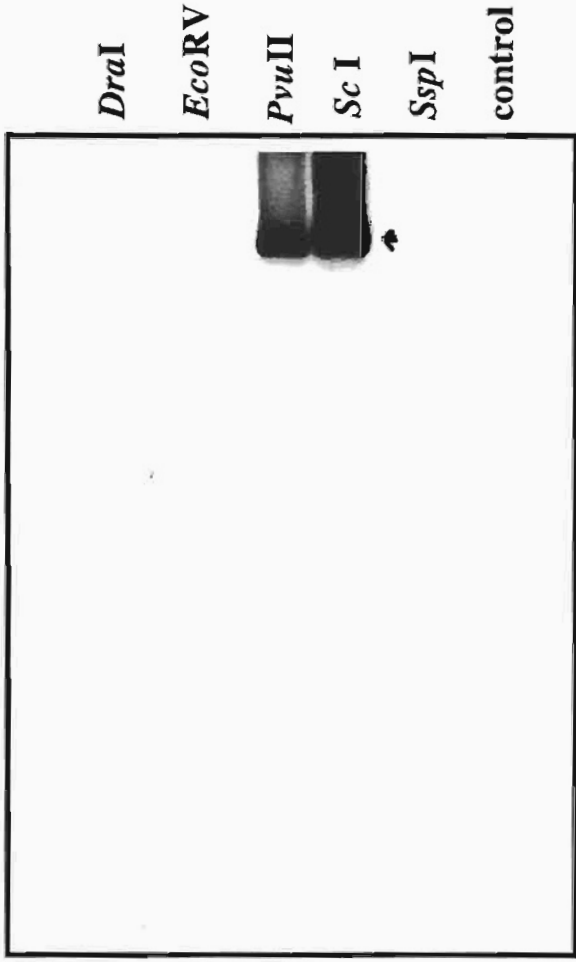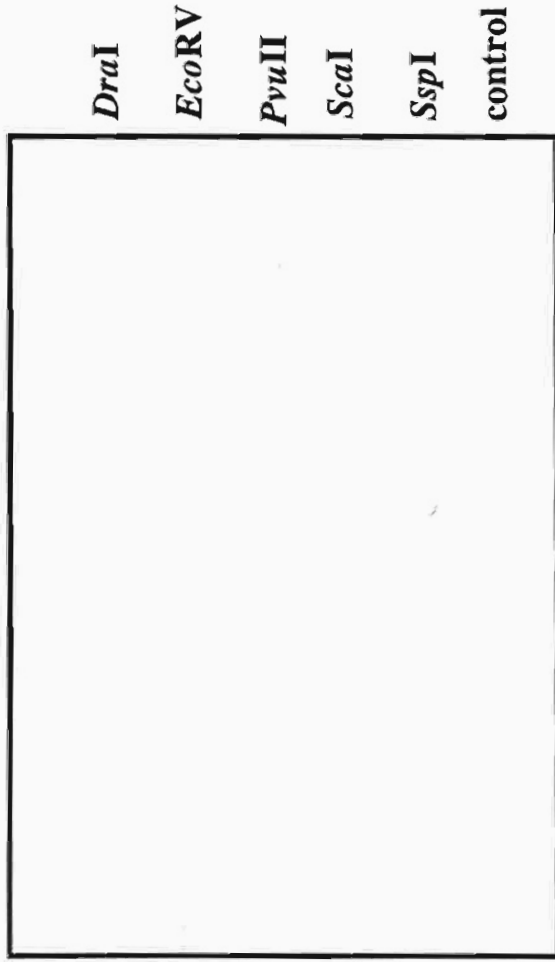
C

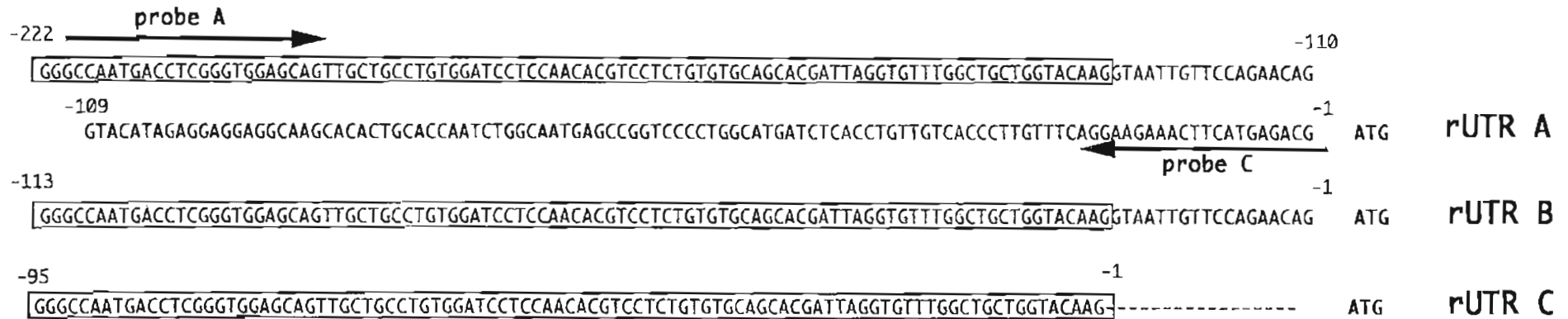DraI EcoRV PvuII Sc I SspI control

D

DraI EcoRV PvuII ScaI SspI control

CLASS I rat PC cDNA

```
        probe A
-222  ━━━━━━━━━━━━━━━━━━━▶                                                                        -110
GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAGGTAATTGTTCCAGAACAG

   -109                                                                                    -1
     GTACATAGAGGAGGAGGCAAGCACACTGCACCAATCTGGCAATGAGCCGGTCCCCTGGCATGATCTCACCTGTTGTCACCCTTGTTTCAGGAAGAAACTTCATGAGACG   ATG   rUTR A
                                                                         ◀━━━━━━━━━━━━━━━
                                                                              probe C
-113                                                                                        -1
GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAGGTAATTGTTCCAGAACAG   ATG   rUTR B

-95                                                                              -1
GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAG----------------   ATG   rUTR C
```

CLASS II rat PC cDNA

```
                                    probe B
                        -96  ━━━━━━━━━━━━━━━━▶  -37                                    -1
                      GTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAGCTAGCATCTGCCTTCTGGAGAGCTGATTGCACCAATG   ATG   rUTR D

                        -59                                          -1
                      GTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG---------------------------------------   ATG   rUTR E
```

Figure 5.8   Positions of oligonucleotide probes used to localise 5'-UTR exons. Only 5'-UTRs of different PC mRNA isoforms (rUTR A, rUTR B, rUTR C, rUTR D and rUTR E) are shown. Different oligonucleotide probes directed against different portions of the 5'-UTRs are shown by arrows.

hybridise to the 0.9 kb and the 0.6 kb fragments derived from *Dra*I and *Ssp*I libraries, this indicated that the first non-coding exon is located at least 0.9 kb upstream from the first coding exon. In contrast, oligos A and B did not hybridise to any of these fragments (Fig. 5.7D) suggesting that these 4 fragments do not contain further exons corresponding to the 5'-UTRs. Nucleotide sequence analysis of the 5.0 kb fragment showed later that it contained only one exon namely exon 1A, located 4.3 kb upstream from the first coding exon. This exon corresponded to positions -1 to -109 (relative to the AUG initiation codon) of the 5'-UTR of rUTR A. It is likely that the next upstream exon would correspond to the region -110 to -222 of the 5'-UTR of rUTR A (or -1 to -113 of rUTR B and -1 to -95 of rUTR C) (see Fig. 5.8). Therefore PCR was carried out using primers derived from the 3'-end of the 5.0 kb fragment (RAGE 5) and from the most 5'-end of rUTR A (oligo A). The oligo A binding site is also common with rUTR B and rUTR C. As shown in Fig. 5.9, a single strong band of 1.5 kb fragment (C in Fig. 5.6) was generated with this pair of primers. Nucleotide sequence analysis of this fragment revealed the presence of an additional upstream exon, namely exon 1B which corresponded to the regions -110 to -222 of rUTR A and -1 to -113 of rUTR B and -1 to -95 of rUTR E, as expected.

Further 5'-end fragments were obtained by PCR-based gene walking using PC24 and PC25 designed from the 5'-end of the above 1.5 kb fragment C (Fig. 5.6) containing exon 1B, as GSP1 and GSP2 primers yielded single discrete bands of 1.1 kb, 2.0 kb, 1.0 kb and 4.0 kb from the *Dra*I, *Pvu*II, *Sca*I and *Ssp*I libraries but not from the *Eco*RV library (Fig. 5.10A). Southern analysis using oligo probe B (Fig. 5.8) showed that it did not hybridise to any of these four fragments suggesting that these PCR products did not contain additional 5'-UTR exons (Fig. 5.10B). Walking towards the 5'-end of the rat PC gene using either the PCR-based Genome Walker kit with primers designed from the 5'-end of the 4.0 kb PCR product derived from *Ssp*I library (flanking sequence of exon 1B) [fragment D, Fig. 5.6] or screening the lambda genomic library using exon 1B flanking sequence as probe, identified further PCR fragments or lambda clones [fragment E and λRG 2, Fig. 5.6] extending towards the 5'-end of the gene already over 10 kb. However, none of these hybridised with probe B. Two possible explanations can be drawn from this experiment, either the presence of a very large intron lies between exons 1B and further upstream 5'-
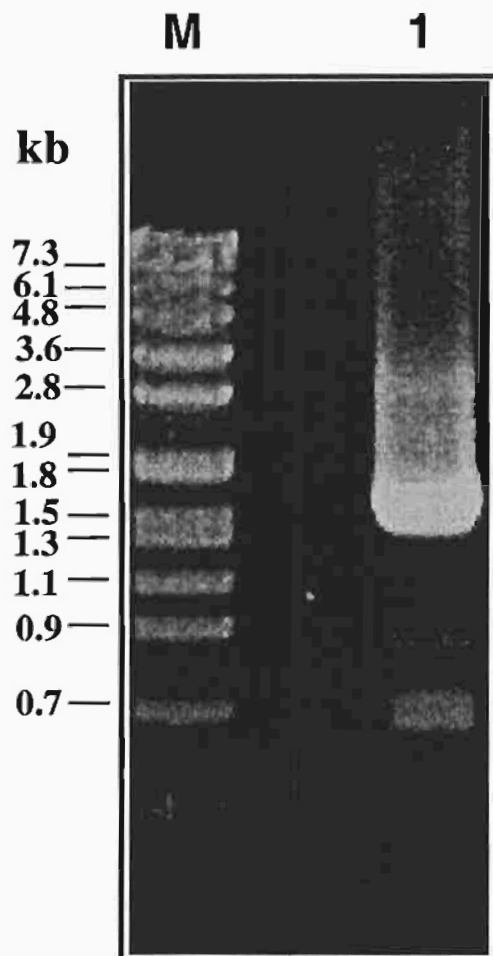
Figure 5.9   Isolation of exon 1B of the rat PC gene by PCR using a primer designed from the most 5'-end of class I PC mRNA (oligo A) or most 5'-end of class II PC mRNA (oligo B) and 5'-end of 5.0 kb fragment amplified from *Sca*I library in Fig. 5.7.
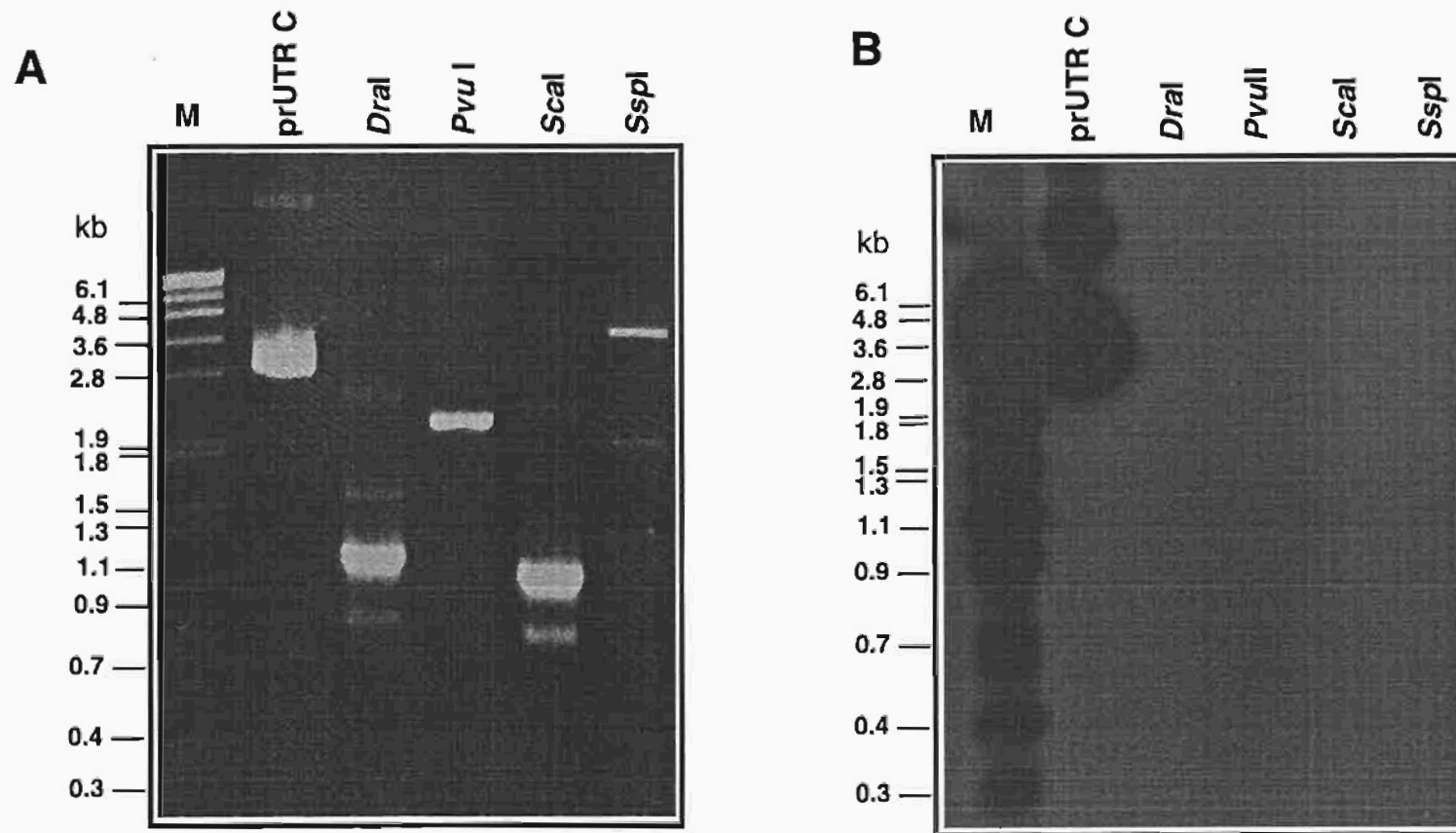
Figure 5.10 Isolation of the 5'-flanking region of exon 1B (putative proximal promoter) of the rat PC gene by PCR based-Gene Walking. Different sized PCR fragments were generated from different restriction libraries (*Dra*I, *Pvu*II, *Sca*I and *Ssp*I) upon secondary amplification using AP2 and a GSP2 primer designed from the 5'-end of the 1.3 kb PCR product in Fig. 5.9. prUTR C, plasmid containing oligo B sequence (+ve control)

UTR exon(s), or the probe B crosses intron/exon boundaries. From sequence aligment of the 5'-UTRs of rUTR D and rUTR E, it is likely that there are two more upstream exons: one encoding regions -1 to -37 of rUTR D but spliced out in rUTR E, and another encoding regions -38 to -96 of rUTR D and -1 to -59 of rUTR E (see Fig. 5.8). Since probe B corresponded to both regions, hybridisation of the probe to an intron/exon boundary of the gene can be ruled out.

A rat genomic library constructed in cosmid vector pWE 15 and containing very large inserts up to 50-100 kb, was then screened. Approximately 2,000,000 cosmid clones were screened with either the 4.0 kb PCR fragment containing exon 1B (fragment D, Fig 5.6) or oligoprobe B. This attempt again failed to detect any positive clone, suggesting this is a poor quality of cosmid library.
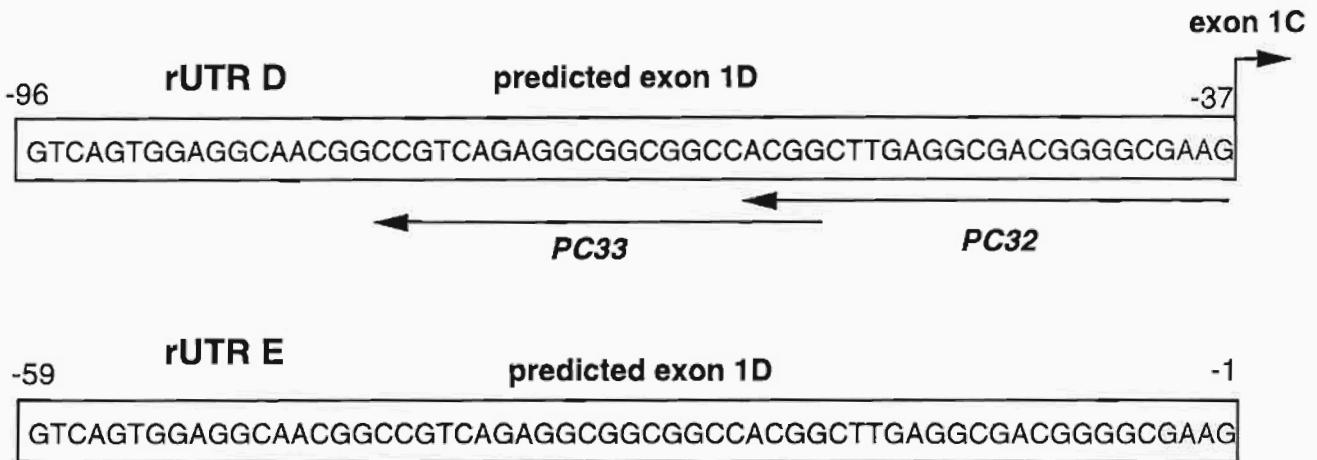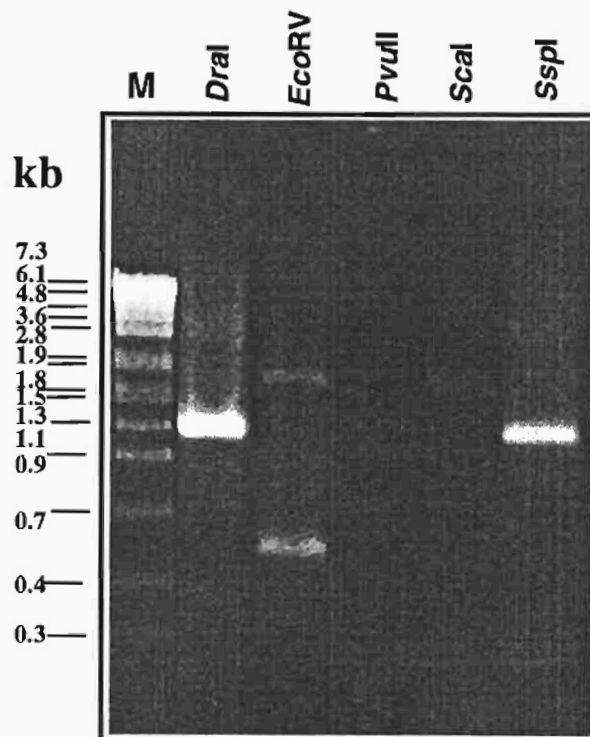
Given the difficulty getting overlapping fragments or clones which contained the rest of the gene, an alternative short-cut PCR method has been used such that only genomic fragments containing the sequences corresponding to the 5'-UTR of rUTR D and rUTR E will be isolated from Rat Genone Walker kit. The first gene-specific primer (GSP1) (PC38) was designed from positions -1 to -24 of 5'-UTR of rUTR D and used with AP1 for the primary PCR. This was followed by secondary PCR using the AP2 primer and second gene-specific primer (GSP2) (PC39) designed from positions -14 to -37. The 5'-end of this primer has 12 bp overlapping with the 3'-end of PC38 (see Fig. 5.11A). As indicated in Fig. 5.11B, discrete bands of sizes 0.35 kb, 1.7 kb, 0.2 kb and 0.32 kb were generated from *Dra*I, *Eco*RV, *Pvu*II and *Ssp*I but not *Sca*I libraries upon secondary PCR. Nucleotide sequence analysis from the 5'-end of the longest fragment derived from the *Eco*RV library showed that this 1.7 kb fragment did not contain sequence corresponding to positions -38 to -96 but corresponded to positions -14 to -37 of 5'-UTR of rUTR D. Indeed, 2 bp upstream from the 3'-end of the PC39 binding site corresponded to an AG dinucleotide, the conserved sequence of the intron/exon splice junction (Breathnach and Chambon, 1981). A DNA fragment of the identical size and sequence was also obtained from PCR-amplified genomic DNA using a primer derived from the 5'-end of the 1.7 kb PCR fragment amplified from the *Eco*RV library and PC38 primer (data not shown). However, exact size of this intron was not determined.

A

predicted exon 1D

predicted exon 1C   -1

CGGGGCGAAG|CTAGCATCTGCCTTCTGGAGAGCTGATTGCACCAATGA|TG

-36   M
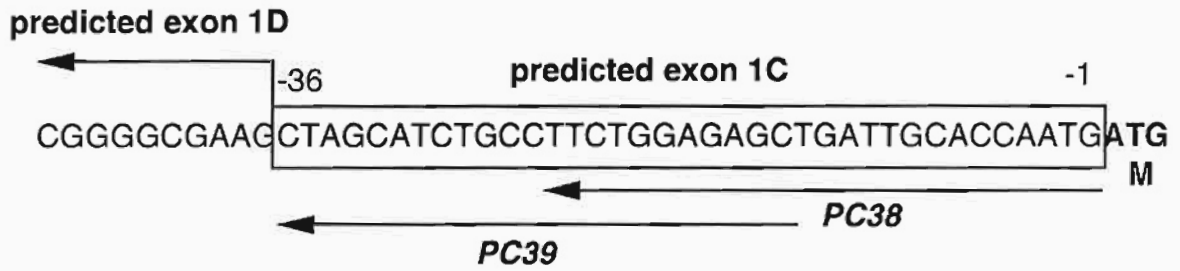
PC38

PC39

B
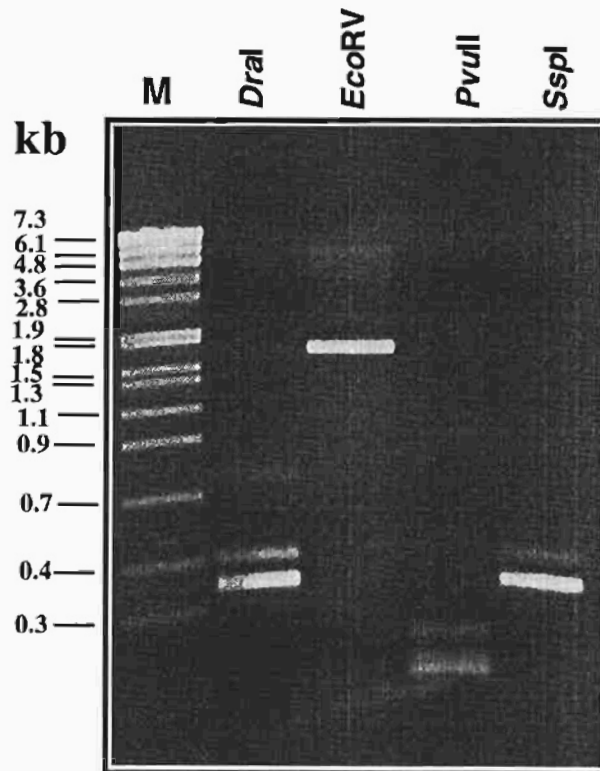
Figure 5.11   Isolation of exon 1C of the rat PC gene by PCR based-Gene Walking. A, Positions of PC38 and PC39 binding sites relative to 5'-end of rUTR D (predicted exon 1C) used as GSP1 and GSP2 in PCR. B, Different sized PCR fragments were generated from different restriction libraries (DraI, EcoRV, PvuII and SspI) upon secondary amplification using primers, AP2 and PC39.

Finally, the genomic fragment encompassing the last exon corresponding to the most 5'-end of UTRs of rUTR D and E was isolated by the same strategy as that described for exon 1C. The first gene-specific primer (GSP 1) (PC32) was designed from positions -37 to -60 of the 5'-end of rUTR D (corresponding to positions -1 to -24 of 5'-end of rUTR E) (see Fig. 5.12A) and was used with AP1 primer for the primary PCR. This was followed by a secondary PCR using the AP2 primer and a second gene-specific primer (GSP 2) (PC33) designed from positions -57 to -78 of 5'-UTR of rUTR D (corresponding to positions -21 to -42 of the 5'-end of rUTR E). As shown in Fig. 5.12B, similar sizes of fragment (1.1 kb) were generated from *Dra*I and *Ssp*I but not from the other libraries. Nucleotide sequence analysis of these two fragments showed that they contain sequences corresponding to the positions -37 to -96 and -1 to -59 of the 5'-end of rUTR D and rUTR E, respectively. A DNA fragment of the same size and sequence was also obtained from PCR-amplified genomic DNA using a primer derived from the 5'-end of the 1.1 kb PCR fragment amplified from the *Dra*I library, and PC33 primer (data not shown). However, the intron/exon boundaries at the 3'-end of exons 1C and 1D were missing due to an inability to get an overlapping fragment either by PCR or by screening genomic libraries.

As indicated in Fig. 5.13, a direct comparison of the nucleotide sequences of four 5'-UTR exons to the 5'-UTRs of PC mRNA isoforms, identified by RACE-PCR in Chapter 4, allowed an explanation of how the alternative splicing of these upstream non-coding exons can result in the production of multiple transcripts with 5'-end heterogeneity. As shown in Fig. 5.14, class I transcripts (rUTR A, rUTR B or rUTR C) were generated by joining exon 1B to one or more of the downstream exons. rUTR A was generated by joining exons 1B, 1A, to the first coding exon (exon 2), while rUTR B was generated by joining exon 1B directly to exon 2, and skipping exon 1A. Interestingly, rUTR C was generated by the same mechanism as rUTR B except that an internal donor site within exon 1B (see Fig. 5.13) was used in joining to exon 2 directly. These two functional donor sites within exon 1B were each followed by a dinucleotide GT and were thus consistent with the consensus splice junction (Breathnach and Chambon, 1981). The use of an internal donor within the 5'-UTR exon has also been reported for the 3-hydroxy-3-methylglutaryl-CoA reductase gene in which four internal donor sites in the same exon can be spliced to the same

A

exon 1C

rUTR D   predicted exon 1D

-96   -37

GTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG

PC33   PC32

rUTR E   predicted exon 1D

-59   -1

GTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG

B

Figure 5.12   Isolation of exon 1D of the rat PC gene by PCR based-Gene Walking. A, Positions of PC32 and PC33 binding site relative to 5'-UTRs of rUTR D and rUTR E (predicted exon 1D) used as GSP1 and GSP2 in PCR. B, Different sized PCR fragments were generated from different restriction libraries (*Dra*I, *Eco*RV, *Pvu*II, *Sca*I and *Ssp*I) upon secondary amplification using the primers, AP2 and PC33 .

Figure 5.13 Alignment of exons 1D, 1C, 1B, 1A and 2 of PC gene with 5'-UTRs of rUTR A, rUTR B, rUTR C, rUTR D and rUTR E transcripts. The upper line indicates the genomic sequence and the lower line indicates the cDNA sequence. The alternative 5' donor sites within exon 1B are boxed. Lowercase letters indicate the intron sequence. The amino acid residues encoded from the first coding exon are shown by capital letters.

Exon 1D

GTTCCCTGTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG

· ·

GTTCCCTGTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTTGAGGCGACGGGGCGAAG

Exon 1C

cttcagCTAGCATCTGCCTTCTGGAGAGCTGATTGCACCAATG

· ·

CTAGCATCTGCCTTCTGGAGAGCTGATTGCACCAATG

Exon 1B

attctgcGGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACA`AGGT`AATTGTTCCAGAAC`AGgt`atg

· ·

GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACGTCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAGGTAATTGTTCCAGAACAG

Exon 1A

tattagGTACATAGAGGAGGAGGCAAGCACACTGCACCAATCTGGCAATGAGCCGGTCCCCTGGCATGATCTCACCTGTTGTCACCCTTGTTTCAGGAAGAAACYTCATGAGACGgtacga

· ·

GTACATAGAGGAGGAGGCAAGCACACTGCACCAATCTGGCAATGAGCCGGTCCCCTGGCATGATCTCACCTGTTGTCACCCTTGTTTCAGGAAGAAACYTCATGAGACG

Exon 2

cctcagATGCTAAAGTTCCAAACAGTTCGAGGGGGCCTGAGGCTCCTCGGTGTCCGCCGATCCTCCACAGCCCCCGTTGCCTCCCCAAATGTCCGGCGTCTGGAGTACAAGCCCATC

· ·

ATGCTAAAGTTCCAAACAGTTCGAGGGGGCCTGAGGCTCCTCGGTGTCCGCCGATCCTCCACAGCCCCCGTTGCCTCCCCAAATGTCCGGCGTCTGGAGTACAAGCCCATC

 M L K F Q T V R G G L R L L G V R R S S T A P V A S P N V R R L E Y K P I

AAGAAAGTAATGGTGGCCAACAGAGgtaaat

· ·

AAGAAAGTAATGGTGGCCAACAGAG

 K K V M V A N R

Figure 5.14 Alternative splicing of 5′ untranslated region exons of the rat PC gene. Class I (rUTR A, rUTR B and rUTRC) and class II (rUTR D and rUTR E) transcripts are generated by alternative usage of exons 1D, 1C, 1B and 1A. Exons 1D, 1C, 1B, 1A, 2 and 3 are shown as boxes. Arrow indicates the ATG initiation codon. The proximal promoter (P1) and distal promoter (P2) that mediate the production of multiple mRNA isoforms are shown.

5' end of Rat PC gene
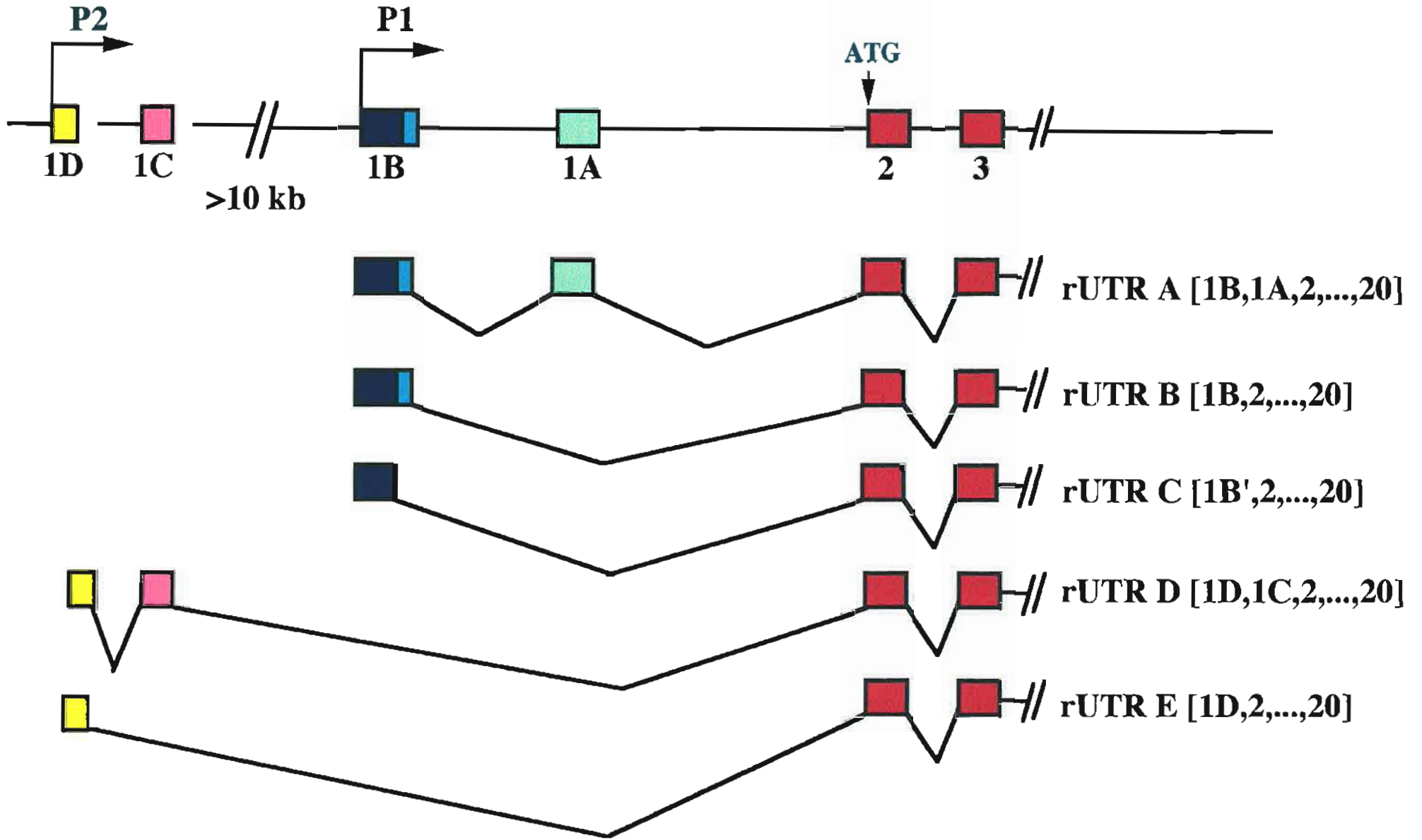
3'-acceptor site (Reynolds *et al.*, 1985). On the other hand, the class II transcripts (rUTR D and rUTR E) were generated as follows: rUTR D was generated by joining exons 1D, 1C, and 2 while rUTR E was generated by joining exon 1D directly to exon 2, and skipping exon 1C.

Taken together with the tissue-specific expression of different PC mRNA isoforms, it is very clear that there are two different primary transcripts, alternatively transcribed from two promoters. The first transcript is initiated from the distal promoter (P2) flanking exon 1D, and undergoes differential splicing to produce two mature transcripts, rUTR D and rUTR E. These transcripts are ubiquitously expressed. On the other hand, the second transcript appears to initiate from the proximal promoter (P1) flanking exon 1B, and undergoes differential splicing to produce three mature transcripts, rUTR A, rUTR B and rUTR C. These transcripts were restricted to gluconeogenic tissues (liver and kidney) and lipogenic tissues (adipose tissue, lactating mammary gland and liver) (see Chapter 4)

5.3.4 Chromosomal localisation of the rat PC gene

In collaboration with Dr. Graham Webb, Department of Animal Science, The University of Adelaide, fluorescence *in situ* hybridisation (FISH) was carried out in order to map the PC gene on the rat chromosomes (Webb *et al.*, 1997). In brief, chromosomes were derived from culture of the skin fibroblasts of a newborn male Wistar rat. A 4.0 kb *EcoRI-BamHI* fragment encompassing exons 11 to 18 of rat PC gene was used as probe and hybridised *in situ* to the metaphase chromosomes. As indicated in Fig. 5.15, the probe hybridised over a broad, bright R-band, distally located on rat chromosome 1 which is the largest chromosome. By comparison with standard ideograms of the chromosomes of *Rattus norvegicus* (Rønne *et al.*, 1987), the target band was identified on the long arm at 1q4. The sub-bands 1q41-43 could not be distinguished, but the location of 16 of the 21 grains photographed, indicated that the PC gene is likely to be in sub-band 1q43 (Fig. 5.15).
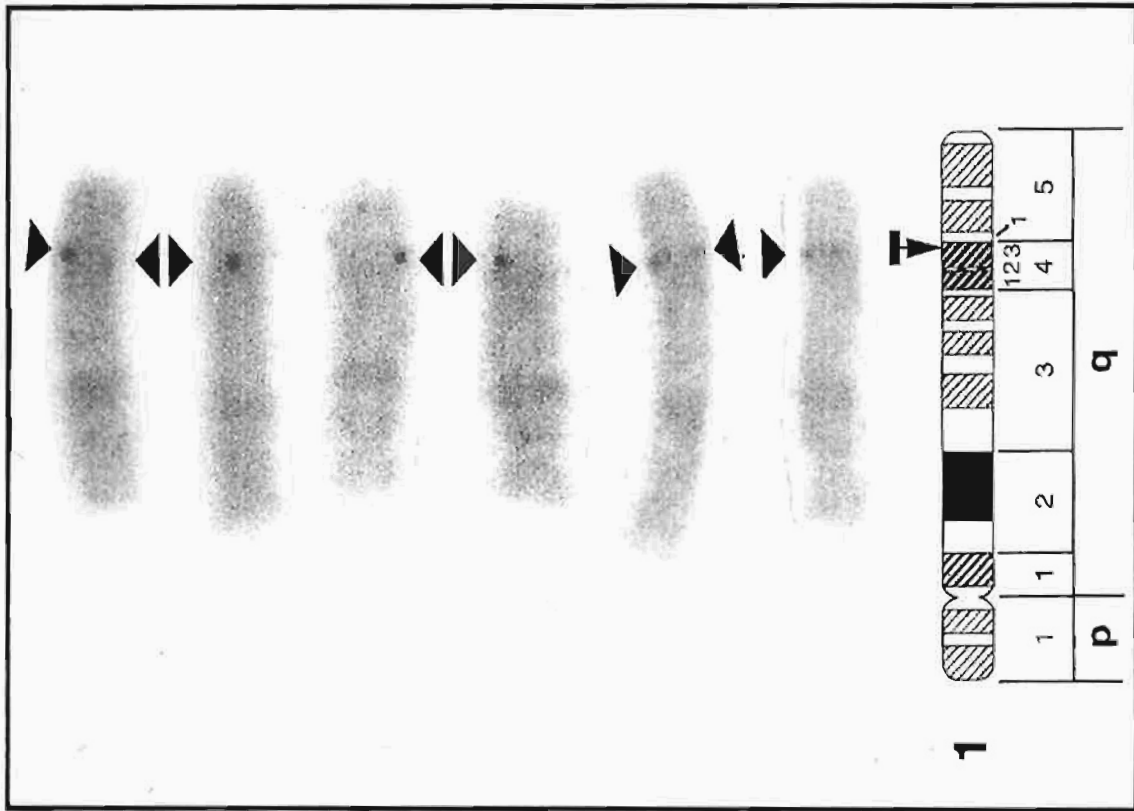
Figure 5.15 *FISH* **of PC probe to rat chromosome 1 from different slides.** The two left chromosome1 are from the same cell, as is the next pair. The ideograms shows the bands conventionally labeled but drawn as observed with PPD11 banding. The arrow to the ideogram indicates the probable localisation of the PC gene, and the bar across it shows the range of all signal grains photographed. The subband 1q42 is shown with a dotted line on the ideogram; it was not seen in this study.
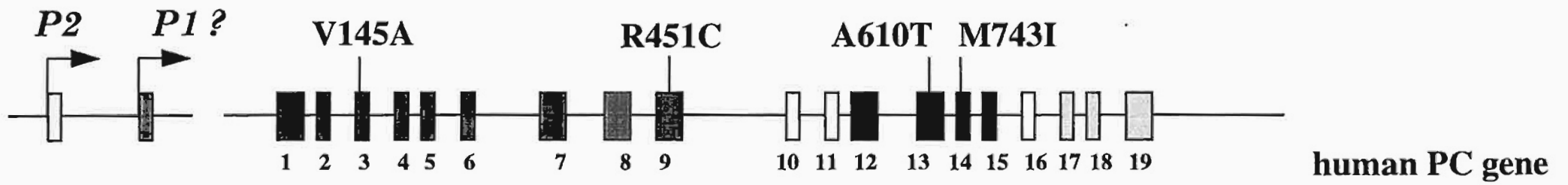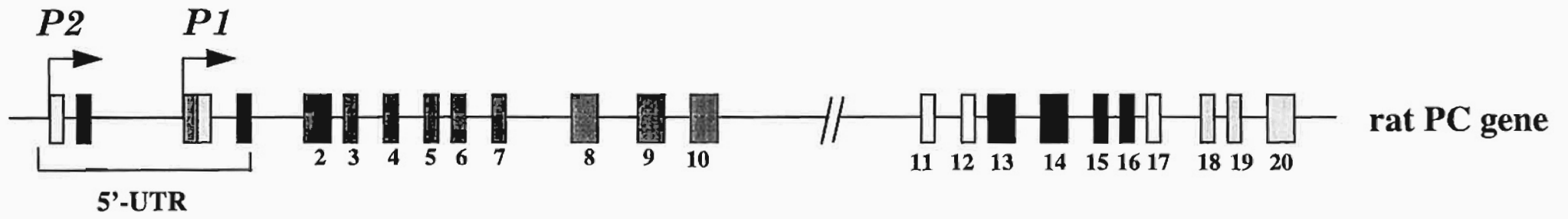
5.4 DISCUSSION

In this chapter, the complete structure of the rat PC gene has been isolated by screening genomic libraries and by performing PCR. The entire gene spans over 40 kb and consists of nineteen coding exons and four 5'-UTR exons. Recently, the human PC gene structure has been reported (Carbone *et al.*, 1998). The organisation of intron/exon boundaries is highly conserved between rat and human genes i.e. both genes contain nineteen coding exons and eighteen introns from the start codon (Fig. 5.16). Both rat and human PC genes contain a large intron separating exons 10 and 11 of rat (intron K) and exons 9 and 10 of human (intron 9), which occurs at the boundary of the biotin carboxylation domain and the transcarboxylation domain of PC. However, in rat this intron is about 10 kb while in human this intron is about 3 kb. This difference in the size of this intron accounts for the larger size of the rat PC gene relative to the human gene. The placement of a relatively large intron between the exons encoding the biotin carboxylase and the transcarboxylase domains of both rat and human PC genes, suggests a close relationship between exon boundaries and protein domains. In rat, this region is also consistent with the highly susceptible chymotrypsin cleavage site that separates these two domains (Chapter 3).

Each of the coding exons was comparably sized in the rat and the human genes, with the exception of exon 13 (relative to equivalent exon 12 in the human gene) which is larger in rat. In the rat gene, the putative pyruvate-binding motif,

[-ENWG<u>G</u>ATFDVAMRFLYECPWRRL-] (residues 605-627) is encoded by this exon while in the human gene, this motif was interrupted by intron 12 (corresponding to the underlined G) thus splitting another half of pyruvate binding-site to exon 13. In the rat PC gene, the corresponding intron is located downstream by approximately 156 nucleotides, such that this motif remains uninterrrupted.

Clearly there are many proteins in which some structural part of the molecule resembles a part of other proteins in the same family or even in an unrelated family, and where this appears to have resulted from gene duplication and rearrangement (Doolittle, 1995). This enables the rearrangment of protein domains with different functions. In the case of the biotin carboxylase family, it has been proposed that this group of enzymes has evolved into complex multifunctional proteins from smaller monofunctional precursors

Figure 5.16 Comparison of genomic structure of rat and human PC genes.
The human PC gene consists of 19 coding exons (Carbone *et al.*, 1998) and is organised in
the same manner as in the rat gene which includes two alternate promoters (P1 and P2).
Two alternate promoters (P1 and P2), located upstream from the first coding exon of human
PC gene, are likely to control alternate transcription of a single PC gene (Chapter 4).
Different point mutations on the human PC gene, recently reported to be responsible for
some forms of the PC deficiency (Wexler *et al.*, 1998; Carbone *et al.*, 1998), are also
shown. Boxes represent exon sequence. The cDNA structure is also shown; BC, biotin
carboxylation domain; TC, transcarboxylation domain; BIO, biotin carboxyl carrier domain.
5'-UTR, 5'-untranslated region exon.

rat PC gene

P2 P1

5'-UTR

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

human PC gene

P2 P1?

V145A R451C A610T M743I

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

cDNA

5' — BC TC BIO — 3'

through successive gene fusions (Obermayer and Lynen, 1976), perhaps via recombination and rearrangement of primordial genes encoding different functional domains of biotin carboxylases as shown in Fig. 5.17. In Eubacteria like *B. stearothermophilus* (Kondo *et al.*, 1997)and *R. etli* (Dunn *et al.*, 1996) or a lower eukaryote like yeast, there would seem to have been a fusion of the ancestral genes encoding the biotin carboxylase (BC), transcarboxylase [TC (Pyr)] and biotin carboxyl carrier (BCCP) components (in order as for PC) encoding a single polypeptide. In Archaebacteria like *M. thermoautotrophicum* (Mukhopadhyay *et al.*, 1998) which is distantly related to Eubacteria, there appears to have been a fusion of genes encoding TC (Pyr) and BCCP components encoding a 75 kDa biotinylated subunit (PYCB or β-subunit). In contrast, the gene encoding a 52 kDa non-biotinylated subunit (PYCA or α-subunit) was located approximately 727 kb or about half a genome apart from the gene encoding PYCB. The amino acid sequence of the PYCA subunit corresponded to that of the biotin carboxylase domain, whereas the amino acid sequence of the PYCB subunit corresponded to those of the transcarboxylation and the biotin carboxyl carrier domains of PC from a number of species which possess a single polypeptide in each subunit. However, in higher eukaryotes, there may have been an interruption of these primordial or ancestral genes encoding different components of biotin carboxylase by introns (Palmer and Logson, 1991) during evolution. It is widely accepted that the introns that are present in in eukaryotic genes can enhance the rate of evolution through recombination events between intron sequences of different genes followed by divergence of the duplicated gene thereby creating new combinations of independent, folded protein domains (Watson *et al.*, 1987).

The same scenario might be predicted with the ACC and PCC encoding genes. For *E. coli* ACC, the gene encoding the BCCP subunit has been found to be fused to the biotin carboxylase subunit, and cotranscribed as a bicistronic mRNA (Li and Cronan, 1992a) whereas the carboxytransferase subunit (TC:ACC) is encoded by another two genes (Li and Cronan, 1992b). However in eukaryotes, homologues of these three subunits of ACC are located within the same polypeptide (Lopez-Casillas *et al.*, 1988) and hence form three discrete domains rather than subunits. This suggests fusion of the genes encoding these three subunits during evolution of eukaryotes. It is noted that there is a rearrangement of the

gene encoding the biotin carboxylase and the biotin carboxyl carrier components in *E. coli* ACC and mammalian ACC. In *E. coli*, the gene encoding the biotin carboxyl carrier subunit is located upstream from the biotin carboxylase subunit whereas in mammals, the gene encoding the biotin carboxylase domain is located upstream from the biotin carboxyl carrier domain followed by the carboxyltransferase domain on a single mRNA. Since PC and ACC use a different acceptor molecule in the transcarboxylation reaction (pyruvate and acetyl-CoA respectively), this would explain the marked difference in the amino acid sequence in this region of the enzyme. This further suggests that the exons encoding these two parts of both enzymes were derived from different ancestral genes [TC (Pyr) and TC (ACC) in Fig. 5.16]. In contrast, the amino acid sequences within the BC and BCCP domains of both enzymes have a high sequence similarity as would be expected from the same reaction mechanism they catalyse, suggesting that the genes encoding these two domains were derived from the same ancestral genes.

The same prediction could also be expected with mammalian PCC but with a more complicated pattern. PCC consists of two subunits, the α-subunit containing the biotin carboxylase and the biotin carboxyl carrier domains, which show strong sequence similarities to the same regions of PC and ACC, while the β-subunit contains the carboxyltransferase domain. In humans these two subunits are located on different chromosomes (Lamhonwah *et al.*, 1986). This could suggest that the ancestral genes encoding the biotin carboxylase and the biotin carboxyl carrier domains are recombined on the same locus of the chromosome thus forming an α-subunit encoding the PCC gene whereas another ancestral gene encoding carboxyltransferase exons [TC (PCC)], which differs from those of PC and ACC, is shuffled on another chromosome. It is noted that the organisation of the biotin carboxylase and the biotin carboxyl carrier domains on the mRNAs encoding ACC and the α-subunit of PCC are in the same order i.e. the biotin carboxylation domain is located upstream from the biotin carboxyl carrier domain. It seems possible that different recombination events of these ancestral genes have created a variety of biotin carboxylase enzymes with diverse functions.

Although there are segments of the transcarboxylation domain of rat PC that share a high degree of amino acid sequence similarity to that of the 5S subunit of *P. shermanii*

Figure 5.17 Schematic diagram represents possible evolutionary pathway of various biotin carboxylase enzyme through gene fusion. Exons encoding different functional domains found in the ancestral gene are shown by boxes. BC, biotin carboxylase, TC, transcarboxylase; BCCP, biotin carboxyl carrier protein. mRNA encoding different biotin carboxylases are also shown.
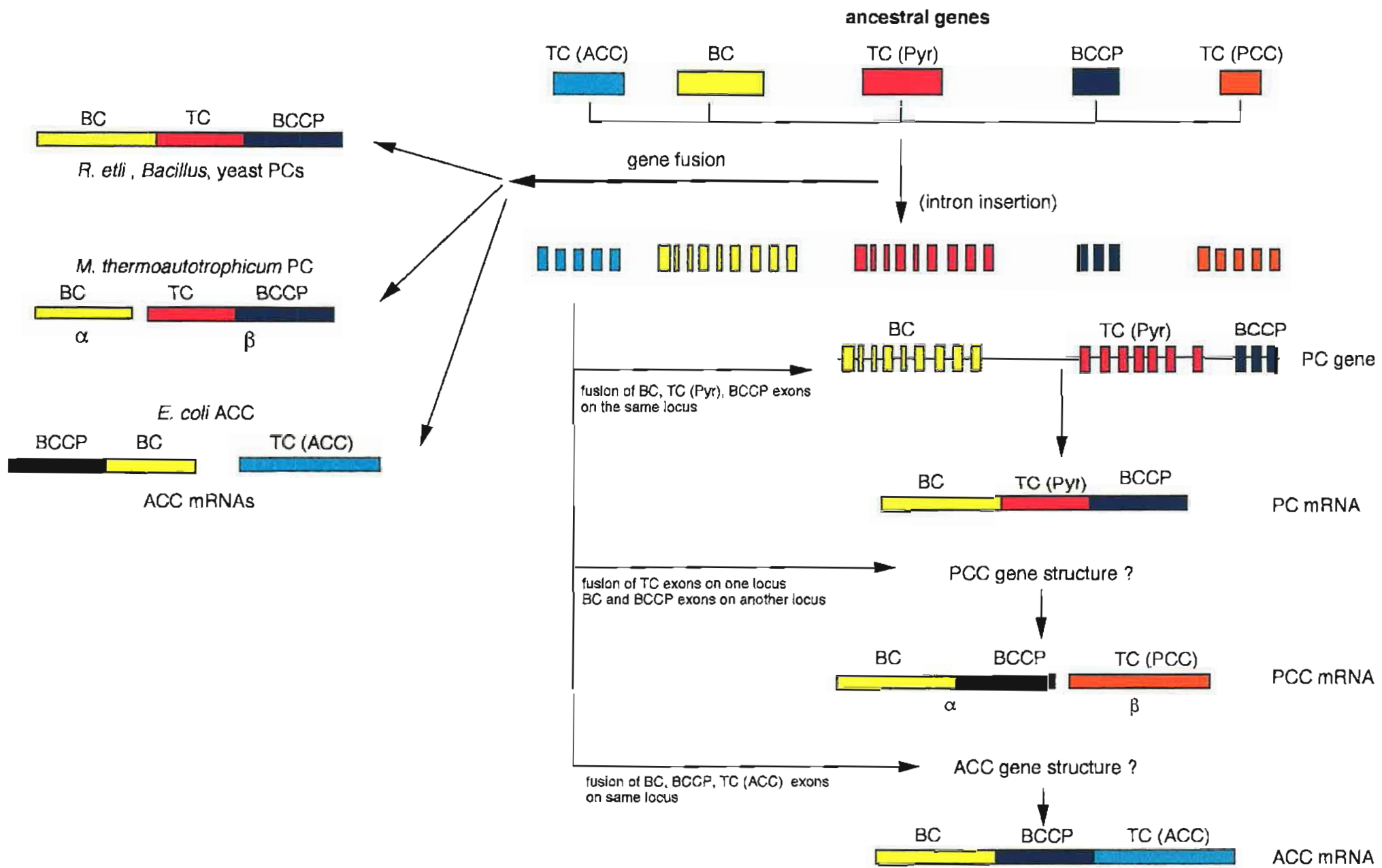
transcarboxylase (Thornton *et al.*, 1993), the overall sequence similarity is not high. It would appear that the remainder of the PC gene is not modular, but more data regarding the functional and structural units within PC are required. Whether the biotin carboxylase family is derived from duplication and rearrangement of common ancestral genes, determination of exon/intron structures of other biotin carboxylases will greatly address this question. No complete structure of another mammalian biotin carboxylase gene has yet been published. Only three studies have reported the isolation of the 5'-UTR exons of rat (Luo *et al.*, 1989), sheep (Barber *et al.*, 1995) and chicken (El Khadir-Mounier *et al.*, 1996) ACC, thus limiting the comparison of exon/intron organisation of PC to those of other biotin carboxylases.

The genes encoding PC, glutathione S-transferase (placental, pi type, GST3), and glycogen-related muscle phosphorylase (PYGM) appear to be syntenic in the rat, mouse and humans. PC and GST3 were mapped at band 1q43 in rat (Levan *et al.*, 1992) and very close to the centromere of chromosome 19 in mouse (Poirer and Guenet, 1996) and all three genes were mapped at band 11q13 in human (Board *et al.*, 1989; Szepetowski *et al.*, 1992; Walker *et al.*, 1995a). Since PYGM gene has been mapped to 143 centimorgan on rat chromosome 1 (Jacob *et al.*, 1995), and 0 centimorgan on mouse chromosome 19 (Poirer and Guenet, 1996), these data suggest that the PC gene is likely to map between 135 and 155 centrimorgan in the rat. The only other gene which has been mapped to 1q43 in the rat, and to 11q13 in humans, encodes protein phosphatase 1, catalytic subunit, alpha, PPP1CA, which maps to chromosome 7 in the mouse (Show *et al.*, 1996). This indicates that PC, GST3 and PYGM genes are near the interface of segments in the rat derived from mouse chromosomes 19 and 7.

The use of alternate promoters in controlling gene expression has been reported in many genes. In most cases, alternate promoters are located in front of coding exons and thus alternate transcription from different promoters would affect the primary structure of the proteins to be translated. These different protein isoforms can have different properties or different subcellular localisation. The biological role of such changes can be dramatically amplified when the protein isoforms thus produced are themselves important regulatory molecules such as transcription factors, hormone receptors, and ion channels (Smith *et al.*, 1989). However, in some other genes, including rat PC, alternative splicing occurs at the

5'-UTR thus maintaining the primary structures of enzymes. The presence of two tissue-specific promoters of the rat PC gene suggests that its expression can be modulated at the transcriptional level. Tissue-specific transcription factors (*trans*-acting factors) that are present in particular cell types are known to play a major role in activating or repressing particular genes through binding to specific sequences located near the promoter.

Regulation of rat PC expression is perhaps not only limited in the transcriptional step since transcripts from the same promoter do undergo differential splicing pathways i.e. some mature transcripts exclude particular exon sequences from the others. This further suggests that regulation of PC expression could also be mediated through a post-transcriptional step. It is known that different 5'-UTRs of mRNA which result from alternative splicing can affect either stability (Brawerman, 1987; Raghow, 1987) or translational efficiency of the transcripts (Kozak, 1988). Investigation of the physiological roles of alternate promoters of rat PC and the role of 5'-UTR of different rat PC mRNA isoforms will be described in later Chapters of this thesis.

Although the structural organisation of the human PC gene has been reported by Carbone *et al.* (1998), the authors were not able to isolate the 5'-UTR exons of the gene which correspond to the 5'-UTR of human PC mRNA reported in Chapter 4. However, it is very likely that a set of 5'-UTR exons upstream of the coding exons parallel to the one described in rat gene will give rise to alternatively spliced 5'-UTR versions of human PC mRNAs and may be located several kilobases upstream from the first coding exon as in the rat gene.

It is well defined that splicing of primary transcripts (or nuclear pre-mRNA transcripts) involve the assembly of essential factors known as small nuclear RNAs together with small nuclear riboproteins to form a spiceosome (Maniatis and Reed, 1987; Guthrie, 1988). There is evidence to suggest that the factors found in different cell types and stages of development acting in concert with essential splicing factors described above determine alternate splicing pathways of the primary transcripts (Breitbart *et al.*, 1987; Leff *et al.*, 1987; Tsurushita *et al.*, 1988) although the nature of this difference in the splicing environment found in cell-type or developmental-specific stages is still unclear.

CHAPTER 6

PARTIAL CHARACTERISATION OF

ALTERNATE PROMOTERS OF THE RAT PC

GENE

6.1 INTRODUCTION

Precise spatial and temporal control of gene expression is essential for the development and survival of all living organisms. Immense efforts are being invested into unravelling the complex molecular mechanisms which regulate the 'on-off' switches inherent in all genomes. Regulation of gene expression relies upon the recruitment at the promoter of the gene of transcription factors, which in turn modulate transcription under different environmental, metabolic or developmental states of the cell. Isolation and characterisation of the sequence at the 5'-end of the gene will therefore provide information regarding the regulatory elements that mediate interaction with transcription factors. Availability of a number of molecular biology techniques e.g. reporter gene analysis, site-directed mutagenesis, gel mobility shift and DNase protection assays, allows investigators to dissect particular sequences that play a crucial role in transcriptional regulation.

Studies in Chapter 4 described the identification of multiple rat PC transcripts exhibiting 5'-end heterogencity which were expressed in a tissue-specific manner. This suggests that rat PC mRNAs are transcribed from alternate promoters. Isolation of the 5'-end of the rat PC gene in Chapter 5, clearly demonstrated that these multiple mRNAs are generated by alternate splicing. Since class I transcripts (rUTR A, rUTR B and rUTR C) contain an identical sequence at their most 5'-end, and likewise with class II transcripts (rUTR D and rUTR E), it is very likely that the flanking sequences at the 5'-end of exons 1B and 1C are active as promoters, and thus are responsible for initiating the transcription of two primary transcripts which then undergo differential splicing. To test whether these two flanking sequences resemble typical eukaryotic promoters, DNA sequence analyses were performed. Reporter gene analysis was also carried out by excising these putative promoter fragments and fusing them to the luciferase reporter gene. This was followed by 5'-end deletion analysis to identify the minimal promoter region required for basal transcription.

6.2 SPECIFIC METHODS

6.2.1 Transient transfection of reporter gene constructs

Reporter gene constructs were transfected into the cells by the electroporation method. Briefly, 5×10^6 cells for COS-1 were suspended in 0.5 ml of cold electroporation

buffer containing 20 mM Hepes, pH 7.5, 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM dextrose. Cells were mixed with 10 pmole of each construct, 2 µg of pRSV-βGal and 250 µg of carrier DNA (salmon sperm DNA) and transferred to a 1.0 ml cuvette (Bio-Rad) and subsequently electroporated in the case of COS-1 cells, with a 250-volt pulse at 960 microfarads using the Gene Pulser (Bio-Rad). Transfected cells were then maintained in the same medium at 37°C for 24 h. Transfection of HepG2 cell was essentially the same as described for COS-1 cells except that 10 µg of pRSV-βGal and 500 µg of carrier DNA were used. Transfection of CHO-K1 cells was also performed in a similar manner to that for HepG2 cells except that the electroporation buffer was substituted with PBS and the cells electroporated with a 1800-volt pulse at 25 microfarads. Transfected HepG2 and CHO-K1 cells were maintained at 37°C for 48 h in DMEM or F12/Ham's media supplemented with 10% fetal calf serum respectively.

6.2.2 Luciferase and β-galactosidase Assays

Transfected cells were harvested from Petri dishes by scraping with a pipette and transferred to a 10 ml tube. Cells were harvested by centrifugation at 1300 x g for 5 min, washed with PBS before being suspended in 100 µl of 1x cell culture lysis buffer (Promega) and frozen. Lysates were thawed and centrifuged at 13,000 x g for 5 min. An aliquot of supernatant equivalent to 100 µg of total protein was added to a 5 ml test tube and placed in the luminometer (Berthold model LB 9502 luminometer). Three hundred microlitres of luciferase reaction buffer (39 mM glycyl glycine, 23.3 mM MgSO4, 7.8 mM ATP pH 7.0) and 300 µl of 111 µM beetle luciferin were then injected, and the luciferase activity measured.

β-galactosidase activity was assayed using *o*-nitrophenyl-β-D-galactopyranoside (ONPG) as substrate (Herbomel *et al.*, 1984). Briefly, 100 µg of protein were added to 0.5ml of reaction buffer containing 100 mM phosphate buffer, pH 7.3, 10 mM KCl, 1 mM $MgCl_2$ and 50 mM β-mercaptoethanol) followed by adding 100 µl of 2 mg/ml ONPG. The reactions were incubated at 37°C for 1 h and terminated by adding 1 ml of 1M Na_2CO_3. β-galactosidase activity was determined as A_{420}/µg protein/h x 100. To normalise the transfection efficiency of each experiment, the luciferase activity was divided by the

β-galactosidase activity expressed as units per microgram of protein. The maps of luciferase and β-galactosidase expression plasmids are shown in Fig. 6.1 and 6.2, respectively.
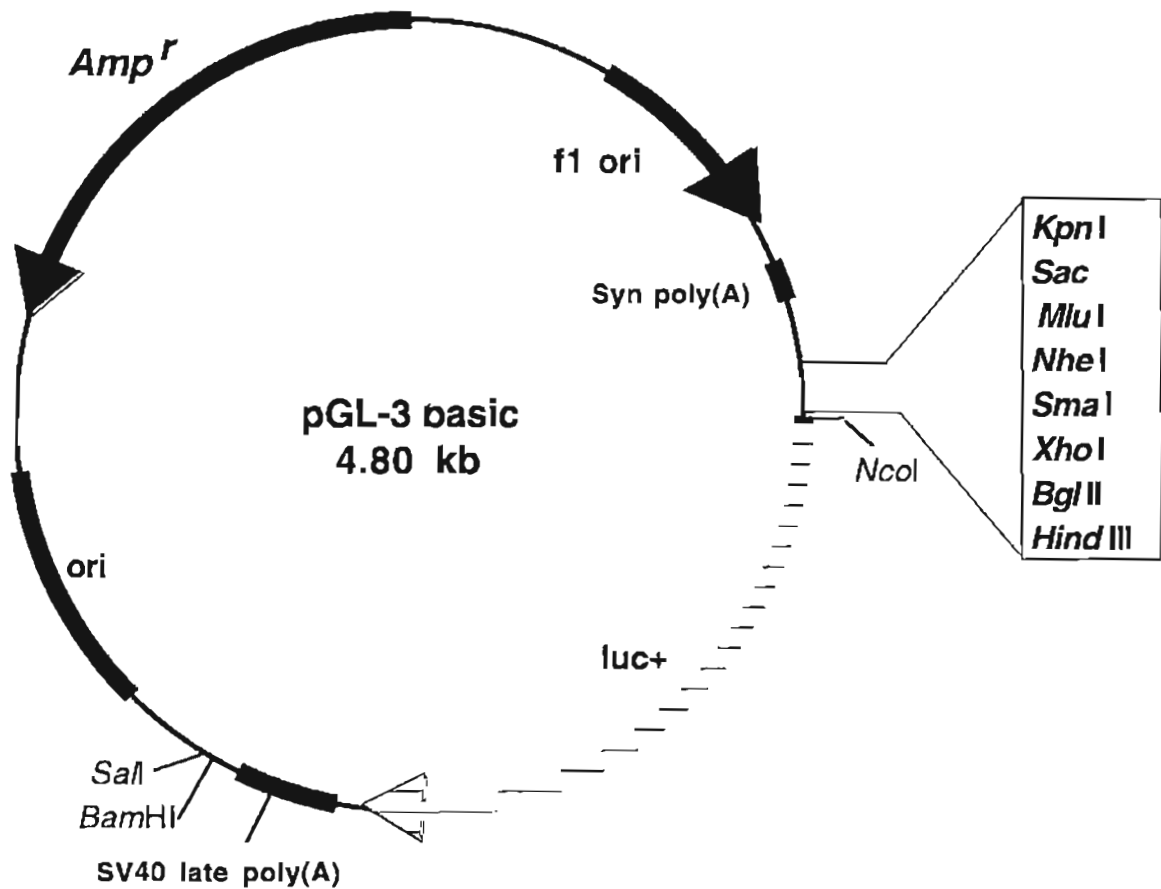
Figure 6.1   Map of pGL-3 basic plasmid. The multiple cloning site is located 5'-upstream from the cDNA encoding the firefly luciferase cDNA (luc+) followed by SV40 late poly adenylation signal. The origin of replication (*ori*) and ampicillin resistance gene (*Amp'*) allow plasmids to be propagated and selected in *E. coli. f1 ori* allows production of single stranded DNA for mutagenesis. The plasmid also contains a synthetic poly(A) upstream from the multiple cloning site to terminate spurious transcription which may initiate within the vector backbone.
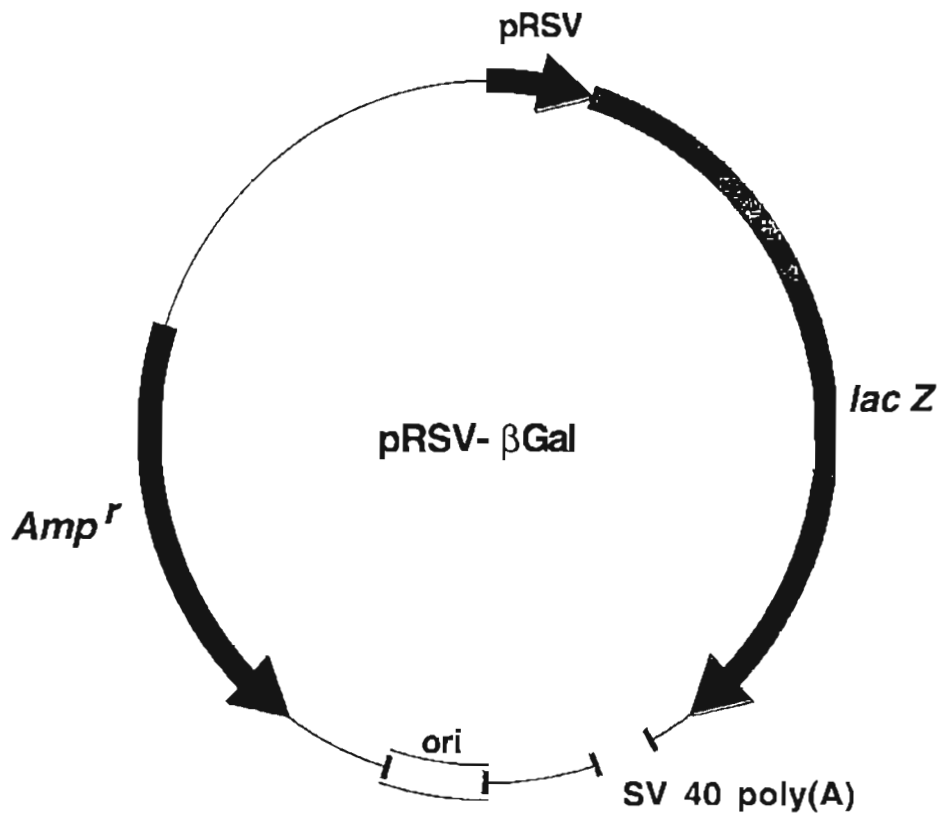
Figure 6.2 Map of pRSV-βgal plasmid. The lacZ gene encoding β-galactosidase driven by the Rous Sarcoma virus promoter (pRSV) is shown. Also shown are the origin of replication (ori), Ampicillin resistant gene (*Amp^r*) and SV40 polyadenylation signal [SV 40 poly(A)].

6.3 RESULTS

6.3.1 Nucleotide sequence analysis of the rat PC proximal promoter

To identify the putative promoter and *cis*-acting elements that flank exon 1B, the 1.1-kb fragment isolated from the *DraI* library (Fig. 5.10A, Chapter 5) was cloned into pBluescript (p*DraI* #6) and sequenced in both directions. As shown in Fig. 6.3, no canonical TATA box or CAAT box was observed in the first 100 bp upstream from the transcription initiation site, although an inverted CAAT box was observed at -220. Near the transcription initiation site (+1), the motif 5'-ATTCTGC⁺IGGGCCA-3' very closely resembled the initiator element HIP-1 box (housekeeping initiator protein 1) with consensus sequence 5'-ATTCN$_{1-30}$GCCA-3' (Means and Farnham, 1990). Computer-assisted analysis revealed a number of potential transcription factor binding sites for AP2, Sp1, cAMP-responsive element binding protein (CREB), nuclear factor 1 (NF 1), hepatic nuclear factor 4 (HNF-4), c-Myb, c-Myc, and PEA 3 (Faisst and Meyer, 1992). Interestingly, there was an insulin-responsive element (IRE) that overlaps the Sp1 binding site located at position -138. This motif was first identified as IRE-A element in the promoter of glyceraldehyde-3-phosphate dehydrogenase gene (Nasrin *et al.*, 1991). At position -198, the sequence closely matched that of the fat-specific element 1 (FSE1) [5'-TCAGGGCCAGGAACTG-3'] of the fatty acid synthase gene (Hsu *et al.*, 1996). There were also eleven copies of the unusual pyrimidine motif, TCCCC or TCCCCC arranged as direct or inverted repeats throughout the proximal promoter of the rat PC gene (Fig. 6.3).

6.3.2 Construction of the promoter-reporter gene fusion plasmids

The reporter gene used in this experiment was the firefly luciferase, cloned in the pGL-3 basic vector (Fig. 6.1). This plasmid contains no promoter or *cis*-acting element, thus any DNA fragment that exhibits promoter activity would drive the expression of the luciferase upon transfection into appropriate mammalian cell lines. A series of 5'-flanking proximal promoter deletion constructs ranging in length from -1153 to -116 were generated. The p*DraI*#6 plasmid containing the 1.1 kb upstream flanking sequence of exon 1B was used as parental plasmid for constructing the promoter-reporter gene fusion plasmid. Due to the presence of an internal *Kpn*I site within the insert, subcloning of a whole insert with

Figure 6.3 Nucleotide sequence of the 5'-flanking sequence of exon 1B (proximal promoter) of rat PC. The transcription initiation site is designated as +1. Putative transcription factor binding sites are shown. Exon sequences are shown in italics. HIP-1 box (Means and Farnham, 1990) is shown in a black box. Different putative transcription factor binding sites for NF-1, c-Myb, c-Myc, HNF-4, CREB, Sp1, AP2, PEA3 (Faisst and Meyer, 1992) and FSE1 (Hsu *et al.*, 1996) are underlined. Putative IRE (Nasrin *et al.*, 1991) is shown by lines above the sequences. The unusual motif TCCCC or TCCCCC or their inverted repeats are shown by open boxes.

-1153 TTGGAGTCTCAACCGTTGTCAACTATGAGGACTTTAGGATAACCATCTGAC
          c-Myb

-1102 CAAGTACTCCCCTAATAGGGTACTGTGCTGCTAATGGGAGCCTGCCTCCC

-1052 ATTCCCCGTACCCCTCTCTGCATGCTATCATTTGGGTGACAGGAGACATC

-1002 TGGAACAGTGCCAGCCTGAAGGGCTGTATCTGGAAACACCTTCTAAGAGA
          NF-1   c-Myc

-952 CATGCTGTGAGGATGGTCACATTAGGATGCCAGGACTTGAGCTTGGACTG

-902 ATCGAGTTTAGTGCCCAAAGCAGGGATTCTCTCTGGCCTGGAGGGAATCA

-852 CACTCTTCAGTTGCCCAGTCTTCTTACCGTTTAGGTCCTCTGCTTCTCCC
          c-Myb

-802 CCTCCTGATCCCCAAGCTTTATGACCCCCTGAGTTGTACAGAGGGTTTCG

-752 CCCCCTCAGCCTTTAGAGTGTGCTCAGGTGTAGCTTGGCCAACCTTGTCT

-702 GCTTCCTTCTCAGACCCTCGAGCTTAAGAGAGCTGTCACAGCATCTGTGA
        PEA3

-652 GGGCCTCTAATGGGACAGTGATCAGTGGGTCCACAACTGTCCTGCAGCAG
                                            c-Myb

-602 GTTTCCCAGGCTGATACGTCCTCAGTACAGGCCTGTGAGGGCAGTGGGAT

-552 TCAGTGAGGGAAAGAAAGGCTAGATTAGCAGTGGAGATGCTTCAGGAACA

-502 GAGTGGTGTGTTCTCTTGTCTGTGACAGCAAGGCCTTGAATTGTGTCTGG

-452 CAGTGAATTAGGGGAAAGGCTGATTGCAGCAGAGGCCTCCTGTGATTGTA

-402 GTAGTGAACCAACCTTCAGGGTAATGTGACCCTTGCCCTCCATCATGTCT
                                    HNF-4

-352 GGCACAGTCCCTCCCAAGCAACCTGTCTAGGTACCATTCTACGTTCTGAG

-302 TGACGTTCTTCTGCCTTAACTTCAAGTGGAAGCATGTCCAAAATCCACCA
        CREB

-252 CAAAAGTCTTCAGCTGCTTTCTAGCAGATATTGGCTTGTAATCCTGCATA

-202 TCTCTGAGGGCCCAGGGAACAGGCTCTAGGCCACTGCTCATCTACTTTCT
          FSE1   IRE

-152 GCATGACCCACTGCCCCGCCTCTCTGGTCCGGCCCTCCTAGGCAGCCAGT
                      Sp1

-102 GGCCCTTGATGGCCTCAGGAGGCGGCGCCCCTTTCCCCTGATTTCAGCCTT

-52 GGCTACAGCCCAGGCCTCCAGCAGTCTAGTGCTGGAGAACTTTGTATTCT
              AP-2
    +1
-2 GCGGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAAC
      HIP-1

*Kpn*I and *Bam*HI as a single fragment was not possible. As indicated in Fig. 6.4, p*Dra*I#6 was therefore excised as two fragments as *Kpn*I-*Sph*I and *Sph*I-*Bam*HI fragments, respectively, by digesting with *Kpn*I, *Sph*I and *Bam*HI, and cloning into the *Kpn*I-*Bgl*II sites of pGL-3 basic. This resulted in the pGL-P1 construct which contained the full length promoter. The first deletion construct was made by digesting p*Dra* I#6 with *Hind*III then religated and this eliminated the 5'-end *Hind*III fragment. The insert was then excised as two fragments, *Kpn*I-*Pst*I and *Pst*I-*Bam*HI fragments with *Kpn*I, *Pst*I and *Bam*HI digestions and cloned into *Kpn*I-*Bgl*II sites of pGL-3 basic to generate pGL-P1Δ*Hind*III construct. Further deletion was carried out by digesting the pGL-P1Δ*Hind*III construct with *Xho*I and this resulted in the pGL-P1Δ*Xho*I construct. A further deletion construct was generated by digesting p*Dra*I#6 with *Kpn*I and *Bam*HI and cloned into *Kpn*I-*Bgl*II sites of pGL-3 basic and resulted in the pGL-P1Δ*Kpn*I construct. Since there were no further restriction sites downstream from the *Kpn*I site, further deletion constructs were generated by PCR using genomic DNA as template. DelB (5'-TCCC<u>GGTACC</u>ATGGCTGCCAGTGGCCCTTGAT-3') or DelA (5'-TCCC<u>GGTACC</u>ATGGCTGCCAGTGGCCCTTGAT-3') primers with a *Kpn*I restriction site attached at the 5'-end (underlined) were designed, and together with PC28 primer were used in PCR to generate 2 further deletion constructs. Both PCR products were then digested with *Kpn*I and *Bam*HI and cloned into the *Kpn*I-*Bgl*II sites of pGL-3 basic, resulting in pGL-P1ΔA and pGL-P1ΔB constructs, containing only 153 bp and 116 bp of 5'-flanking sequence upstream from the transcription initiation site, respectively.

6.3.3 Expression of the deleted proximal promoter-reporter gene fusion plasmids

The African green monkey kidney cell line 1 (COS-1), was chosen as a host for transfecting the chimeric gene constructs for several reasons. First, since COS-1 cells were derived from kidney, the same cell type in which PC is expressed (gluconeogenic tissue). Second, this cell line was transformed with the large SV-40 T antigen thus expressing large amounts of this antigen and since the pGL-3 basic vector contains the SV-40 antigen receptor this allows its replication in COS-1 cells. This cell line is also fast-growing and highly transfectable using a number of transfection methods.

Figure 6.4 Generation of chimeric gene constructs containing deleted rat PC proximal promoter fused to luciferase reporter gene. pDraI#6 was used as the parental plamid to generate 5'-nested deletions which were cloned at the multiple cloning sites in front of the luciferase reporter gene in the pGL-3 basic plasmid. This yielded plasmids, pGL-P1, pGL-P1ΔHindIII, pGL-P1ΔXhoI and pGL-P1ΔKpnI. Shorter deletion constructs were generated by PCR from pDra I with del A or del B primer together with PC28 primer and cloned in pGL-3 basic plasmid, resulting in pGL-P1ΔA and pGL-P1ΔB. Different restriction sites present in the proximal promoter are shown.
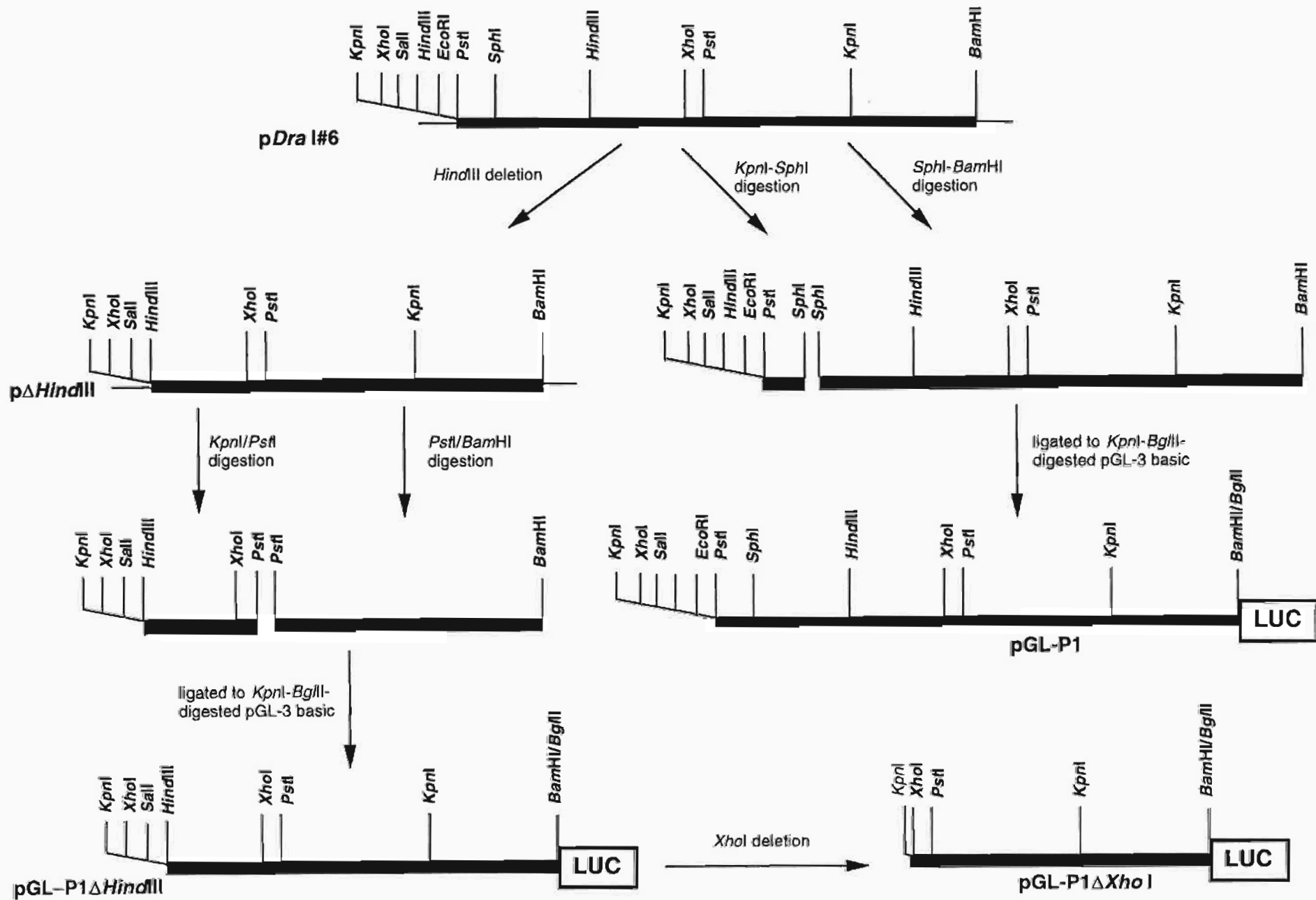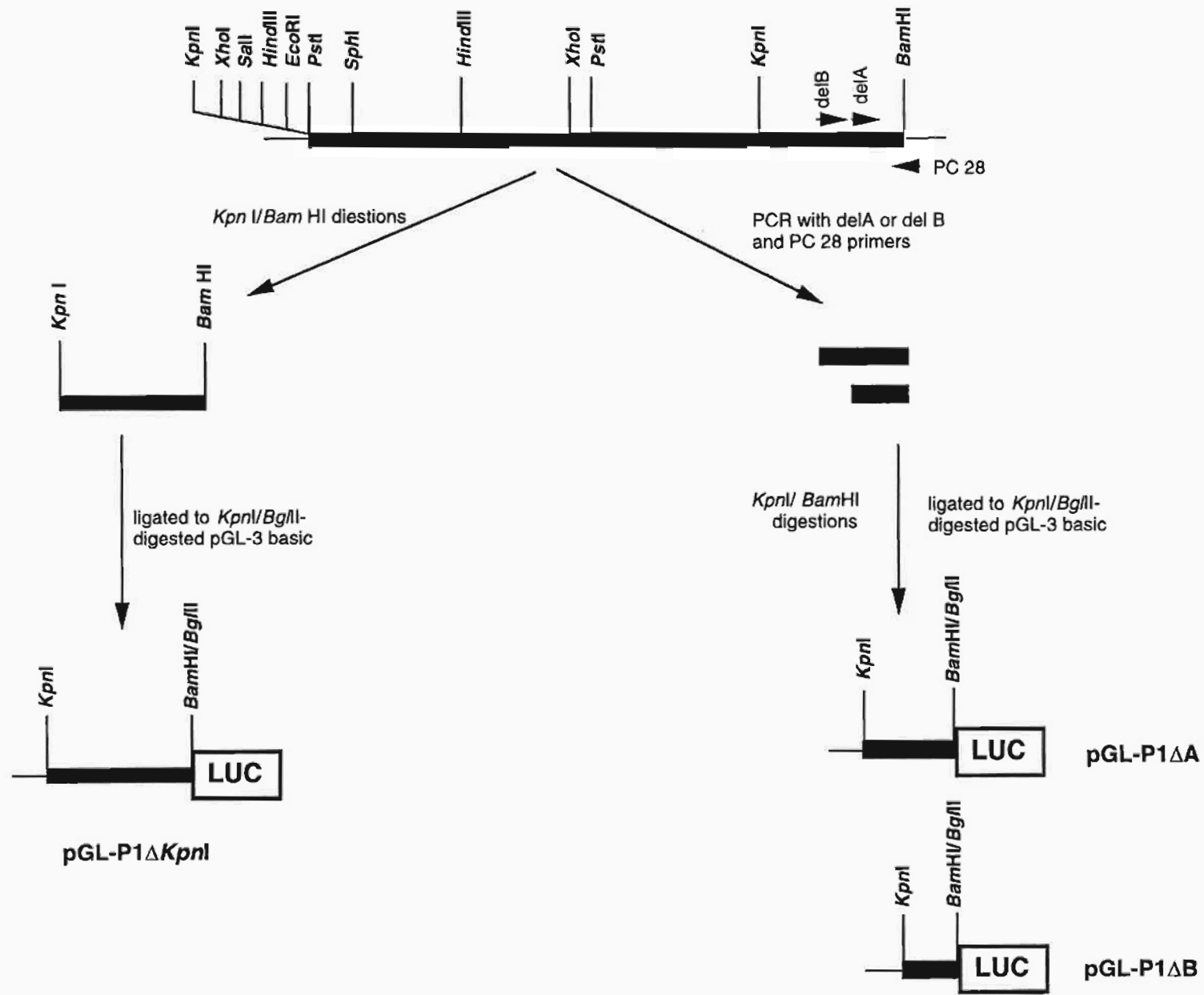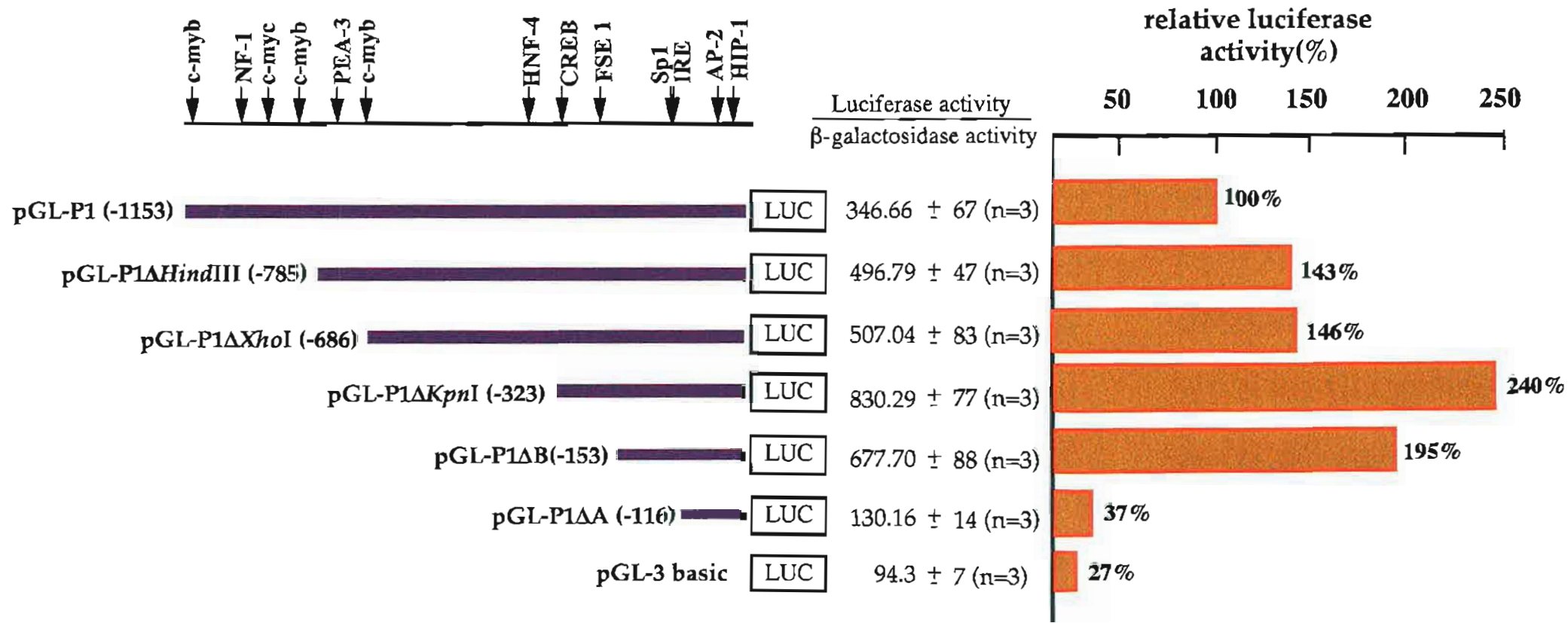
Six constructs were transiently transfected into COS-1 cells by electroporation. As indicated in Fig. 6.5, expression of the longest construct, pGL-P1 was substantially higher than the promoterless construct (pGL-3 basic), confirming that this promoter fragment was active. However, deletion of the region between -1153 to -785 (pGL-3-P1Δ*Hind*III construct) resulted in an increase in luciferase activity up to 143% relative to pGL-P1. Further deletion of the region between -785 to -686 (pGL-P1Δ*Xho*I) caused only a small increase in relative promoter activity, whereas deletion of the region to -323 (pGL-P1Δ*Kpn*I construct) lead to a dramatic increase in activity up to 240% relative to pGL-P1. These results suggested that negative regulatory elements reside in sequences between -1153 and -323. Additional deletions between -323 and -153 (pGL-P1ΔA), lead to a relative decrease in promoter activity to 195%. However, further deletion to -116 (pGL-P1ΔB) resulted in a marked decrease in relative activity to only 37%, which was very close to that of the promoterless construct suggesting the core promoter is located within the first 153 bp upstream from the transcription initiation site (Fig. 6.5). Examination of the DNA sequence in this region (-153 to -116) revealed the presence of only one potential Sp1 binding site although the HIP-1 box was still intact. This Sp1 site may be an important element for transcription of the proximal promoter.

6.3.4 Nucleotide sequence analysis of the distal promoter

To identify the putative promoter and *cis*-acting elements that flank exon 1D, the 1.1kb fragment isolated from *Ssp*I library (Fig. 5.12B, Chapter 5) was cloned into pGEM3Z (pGEM-1.2) and sequenced in both directions. As indicated in Fig. 6.6, nucleotide sequence analysis of the 1.1 kb flanking region of exon 1D identified three copies of CAAT boxes located at positions -64, -94 and -224 relative to the second transcription initiation site identified by RACE-PCR in Chapter 4. Again, no classical TATA box was found in the first 100 bp of this promoter. Several putative transcription factor binding sites including c-Myc, Sp1, AP1, AP2, PPAR and PuF (Faisst and Meyer, 1992) were indentified in this region. A potential IRE (Nasrin *et al.*, 1991) located at positions -228 to -236, which again overlapped the Sp1 binding site, was also found in this promoter. These potential

Figure 6.5 Analysis of promoter activities of deletion constructs containing different lengths of proximal promoter fused to the luciferase reporter gene. These constructs were transiently expressed in COS-1 cells. The luciferase activity of each construct was normalised in comparison with co-expressed β-galactosidase activity. The relative luciferase activities shown are the means ± the standard deviations for tripicate determinations. Relative luciferase activities are shown as a percentage of the activity in relative to pGL-P1 construct, which was arbitrarily set to 100%. The different putative transcription factor binding sites relative to the promoter are also shown.

Transcription factor binding sites (left to right): c-myb, NF-1, c-myc, c-myb, PEA-3, c-myb, HNF-4, CREB, FSE 1, Sp1, IRE, AP-2, HIP-1

Construct	Luciferase activity / β-galactosidase activity	relative luciferase activity (%)
pGL-P1 (-1153)	346.66 ± 67 (n=3)	100%
pGL-P1ΔHindIII (-785)	496.79 ± 47 (n=3)	143%
pGL-P1ΔXhoI (-686)	507.04 ± 83 (n=3)	146%
pGL-P1ΔKpnI (-323)	830.29 ± 77 (n=3)	240%
pGL-P1ΔB(-153)	677.70 ± 88 (n=3)	195%
pGL-P1ΔA (-116)	130.16 ± 14 (n=3)	37%
pGL-3 basic	94.3 ± 7 (n=3)	27%

Figure 6.6 Nucleotide sequence of the 5'-flanking sequence of exon 1D (distal promoter). The transcription initiation site is designated as +1. Putative transcription factor binding sites are shown. Exon sequences are shown by italics. Different putative transcription factor binding sites including c-Myc, Sp1, AP1, AP2, PPAR, PuF (Faisst and Meyer, 1992) are underlined. Three potential CCAAT boxes are shown by bolded underlines. A putative IRE (Nasrin *et al.*, 1991) is shown by a line above the sequences.

```
-1151  AAATCCGTATTCTCCTAATATTCTCCTATATTGGATTCATGCAAGGTAGG

-1101  GGTCCTGATAAACGGATAAAGGGAGGACCAGAAGGATGGAATTGCCTTTT

-1051  TTCTTTTTCTTTTTTTAAGAGAAGCCATATTGCAGAGGCCCACCTTGAAC
                        c-Myc

-1001  TCGAAATCCTCCAGCTTCAGTCTTCCAAGTCCGGTGTTACATTATCTGTG

-951   TCACTTAGCCCAGTGGCAAGTTCTTCCCAGCAGAGTTGGAGGTTAGTAAG
          Sp1

-901   CAATTATTGTTTAACGCCATGTTCTCATAATTCCTGATTGTCCTGTTAGG

-851   AACTGGTCTATTTTTTTCTTCTTGGCCTGATCAATGATGATGGAGGACTT

-801   ATTTGTTTACTGTCTGGTCTCCTATCACTGGAAGGCTGGACTTCAGAGTT

-751   GCTTTGGTGACTCCCTTTCGTTAGCTGAACCGAACCCTCTAGAAAGAAAA
              AP1

-701   TCTTTAAGGTTCTGAGCAGCCCGGTTTTGCCGATCAGGCCACTGAGCTCA

-651   GAGACGTGAAGCTGATGCTCCAAGGTCACTCAGAGGCGGGCTCCTGACCA
                                PPAR        Sp1

-601   CCATCCCCATTCACACAGCTTCAAAAGGACGCAGACACACAACCCCAAGT

-551   GGTACCGGCGCTCTCTAATTCCTCGTGCCCTTAGCTAACGTTTCTCCTGC

-501   GACCTCTTCTGTATCTGCTAAAGAGTACGTGCATCTGGCAGCACCGCACC

-451   GCTCCAACCCCCGTGCATCTGTTATGGTTGCGGTTTGAAGAGACACGTGC
                                                 c-Myc

-401   TACTCGAGTGAATGAATTGCAGGTCCCCACCCCCGCAGACAGCCAATCAG
                             PuF, AP1

-351   GATGAGCTTCCAGACGGACCACTTCCGCCTATTGCGGCCTCTGGCCGCGC
                          Sp1

-301   CAGACGGCCCACGTGAGGCTTACAAGGAAAGGGAGTAGGCGGTGGCGCGT
                c-Myc        IRE

-251   AGAGGGTGGGCTCTTCCCGCCTCTTCACCAATCAGAAGGTCTTCTGGGGA
           PuF        Sp1

-201   CGCGAGGGGCGGAGTCGACGTGGAGGCGCCAACGCTTAAACGCGAGGGGT
           Sp1

-151   AGGTGGGTGCAGGGAGGAGACCTGAAAGGGGGAGCGGTGAATGATAGGGT

-101   CATTCATCCAATCTTTGGAGGGGATGGGCTGTCTCAACCAATGGAAAGTC
               AP2

-51    TTACGGCGGAGCCAGTGCTGCAGCAAGTTTGGTTCGGCGGAGACCGCAGC
       +1

-1     TGTTCCCTGTCAGTGGAGGCAACGGCCGTCAGAGGCGGCGGCCACGGCTT
```

transcription factor binding sites seemed to cluster together within the first 400 bp in contrast with that seen in the proximal promoter.

6.3.5 Construction of the distal promoter-reporter gene fusion plasmids

The pGEM-1.2 plasmid, containing the 1.1 kb upstream flanking sequence of exon 1D was used as a parental plasmid for constructing the promoter-reporter gene fusion plasmids. Due to the presence of *Kpn*I and *Sac*I within the 1.1 kb fragment, a whole insert could not be excised as a single fragment and subcloned directly into pGL-3 basic vector. As indicated in Fig. 6.7, the 5'-end *Sal*I-*Xba*I fragment was initially subcloned into *Sal*I-*Xba*I-digested pBluescript to provide a polylinker that is compatible with that of the pGL-3 basic vector. This fragment was then excised with *Kpn*I and *Xba*I together with the *Xba*I-*Bam*HI which was subcloned from pGEM-1.2. These two fragments were then ligated into *Kpn*I-*Bgl*II-digested pGL-3 basic and this resulted in the pGL-P2 construct which contained the full length promoter. Other 5'-end deletion constructs were generated by digesting pGEM-1.2 with *Sac*I and *Bam*HI or *Kpn*I and *Bam*HI, and ligating the excised 3' fragments to either *Sac*I-*Bgl*II or *Kpn*I-*Bgl*II-digested pGL-3 basic, respectively. Further deletion constructs were created at the *Xho*I and *Sal*I sites respectively. These two constructs were generated by digesting pGEM-1.2 with *Xho*I and *Bam*HI or *Sal*I and *Bam*HI respectively, and cloning into pBluescript to provide the polylinker that is compatible to that of pGL-3 basic. The inserts of these two constructs were then excised from pBluescript with *Kpn*I and *Bam*HI and ligated into *Kpn*I/*Bgl*II sites of pGL-3 basic, thereby forming pGL-P2Δ*Xho*I and pGL-P2Δ*Sal*I constructs respectively. The last construct, containing the shortest fragment, was generated by digesting pGEM-1.2 with *Pst*I followed by religation of the vector to generate the pGEM-1.2Δ*Pst*I construct. Finally, the insert of this construct was excised with *Sac*I and *Bam*HI and cloned into *Sac*I and *Bgl*II sites of pGL-3 basic, and resulted in the pGL-P2Δ*Pst*I construct.

6.3.6 Expression of the deleted distal promoter-reporter gene fusion plasmids

Similar studies to those described in section 6.3.3 were also carried out to identify the minimum sequence required for basal transcription from the distal promoter in COS-1

Figure 6.7 Generation of chimeric gene constructs containing deleted rat PC distal promoter fused to luciferase reporter gene. pGEM-1.2 was used as the parental plasmid to generate 5'-nested deletions cloned into the multiple cloning site upstream of the luciferase reporter gene in the pGL-3 basic plasmid. This yielded plasmids pGL-P2, pGL-P2ΔSacI, pGL-P2ΔKpnI, pGL-P2ΔXhoI, pGL-P2ΔSalI and pGL-P2ΔPstI containing different lengths of promoter fused to luciferase reporter gene. Different restriction sites present in the proximal promoter are shown.

pGEM-1.2

SacI SalI PstI SalI · XbaI SacI · KpnI · XhoI · SalI · PstI BamHI SacII NcoI

SalI/XbaI digestion · XhoI/BamHI digestion · SalI/BamHI digestion

SalI · XbaI

XhoI · SalI · PstI BamHI

SalI · PstI BamHI

ligated to SalI-XbaI-digested pBluescript

XbaI-BamHI digestion

ligated to XhoI-BamHI-digested pBluescript

ligated to SalI-BamHI-digested pBluescript

KpnI XhoI SalI · XbaI

KpnI XhoI · SalI · PstI BamHI

KpnI XhoI SalI · PstI BamHI

KpnI/XbaI digestion

ligated to KpnI-BglII-digested pGL-3 basic

ligated to KpnI-BglII-digested pGL-3 basic

XhoI SalI · XbaI XbaI SacI · KpnI · XhoI · SalI · PstI BamHI

ligated to KpnI-BglII-digested pGL-3 basic

KpnI XhoI · SalI · PstI BamHI/BglII LUC

KpnI XhoI SalI · PstI BamHI/BglII LUC

XhoI SalI · XbaI SacI · KpnI · XhoI · SalI · PstI BamHI/BglII LUC

pGL-P2

pGL-P2ΔXhoI

pGL-P2ΔSalI

pGEM-1.2

*Sac*I/*Bam* HI digestions

Kpn I/*Bam* HI digestion

*Pst*I deletion
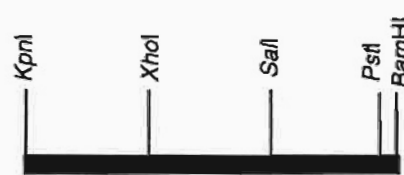
pGEM-1.2Δ*Pst*I

ligated to *Sac*I-*Bgl*II-digested pGL-3 basic

ligated to *Kpn*I-*Bgl*II-digested pGL-3 basic

ligated to *Sac*I-*Bgl*II-digested pGL-3 basic
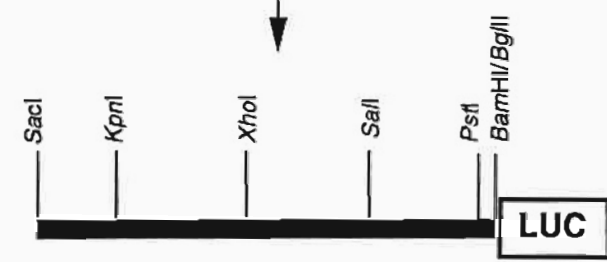
pGL-P2Δ*Sac*I
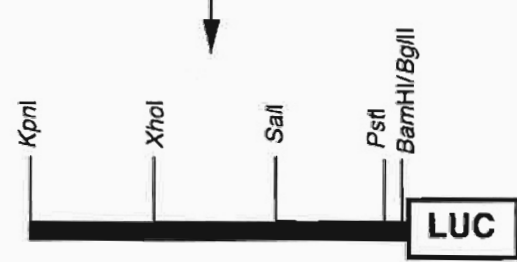
pGL-P2Δ*Kpn*I
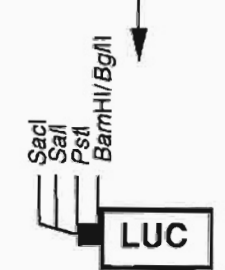
pGL-P2Δ*Pst*I

cells. Six constructs containing different lengths of the distal promoter (Fig. 6.8), were transiently transfected into COS-1 cells. Expression of the longest construct was much higher than the promoterless construct. Deletion of the region between -1151 to -653 (pGL-P2ΔSacI) resulted in a decrease in relative promoter activity to 68%. However, further deletions of the region -653 to -550 (pGL-P2ΔKpnI) and -550 to -400 (pGL-P2ΔXhoI) recovered the relative activities up to 117% and 161% respectively, suggesting that repressor(s) may be located within these regions. Progressive deletion to -187 (pGL-P2ΔSalI) led to a decrease in relative activity to 85% (Fig. 6.8). Further deletion to -35 (pGL-P2ΔPstI) caused a further loss of promoter activity to only 3.5%, suggesting that the core promoter is located within the first 187 bp. This included the two CCAAT boxes and AP2 binding site.

6.3.7 Expression of the proximal promoter and the distal promoter in different cell lines

Comparison of luciferase activities detected from both promoters in COS-1 cells showed that the distal promoter drives the expression of the reporter gene at a higher level than the proximal promoter. However, this could reflect a difference in the transfection efficiency of cells of different passage number of cell culture since transfection experiments of both promoters were not performed at the same time. Therefore, the highest expressing constructs of the proximal (pGL-P1ΔKpnI) and the distal promoters (pGL-P2ΔXhoI) were each transiently transfected into different cell lines including COS-1, HepG2 and CHO-K cells in the same experiments and compared. As shown in Fig. 6.9, expression from the distal promoter was 8-9-fold higher than the proximal promoter in COS-1 cells, whereas the expression from the distal promoter was about 4-5-fold higher than the proximal promoter in both HepG2 and CHO-K1 cell lines. This difference in fold-response of the distal promoter over the proximal promoter could reflect different availability of basal transcription factors present in the different cell types, nevertheless the distal promoter was stronger than the proximal promoter in these three cell lines.

Figure 6.8 Analysis of promoter activities of deletion constructs containing different lengths of distal promoter fused to the luciferase reporter gene. These constructs were transiently expressed in COS-1 cells. The luciferase activity of each construct was normalised in comparison with co-expressed β-galactosidase activity. The relative luciferase activities shown are the means ± the standard deviations for tripicate determinations. Relative luciferase activities are shown as a percentage of the activity in relative to pGL-P2 construct, which was arbitrarily set to 100%. The different putative transcription factor binding sites relative to the promoter are also shown.
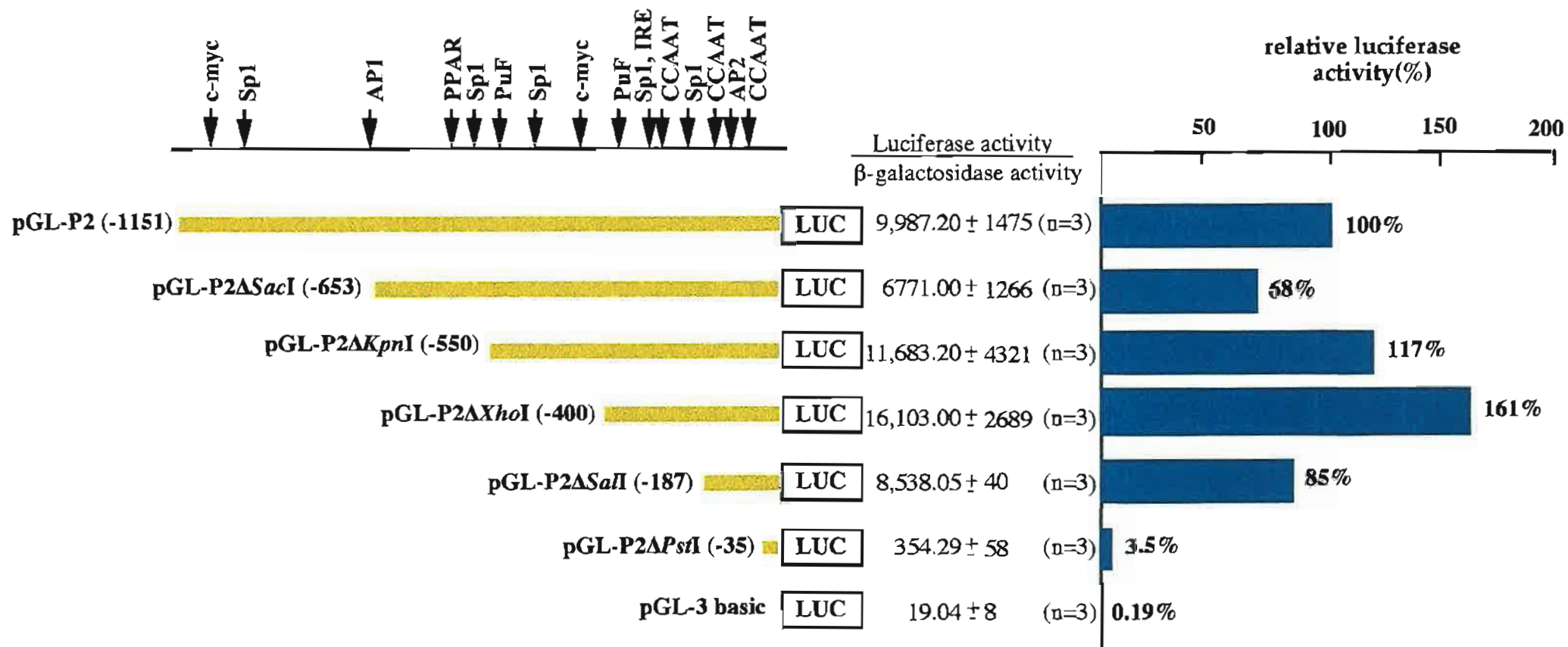
Figure 6.9 Expression of luciferase from pGL-P1ΔKpnI construct (proximal promoter) and pGL-P2ΔXhoI construct (distal promoter) transiently transfected into COS-1, CHO-K1 and HepG2 cells. The luciferase activity of each construct was normalised in comparison with co-expressed β-galactosidase activity. The relative luciferase activities shown are the means ± the standard error of means of tripicate determinations. Relative luciferase activity of the pGL-P1ΔKpnI construct is shown as a percentage of the activity in relative to pGL-P2ΔXho I construct, which was arbitrarily set to 100%.
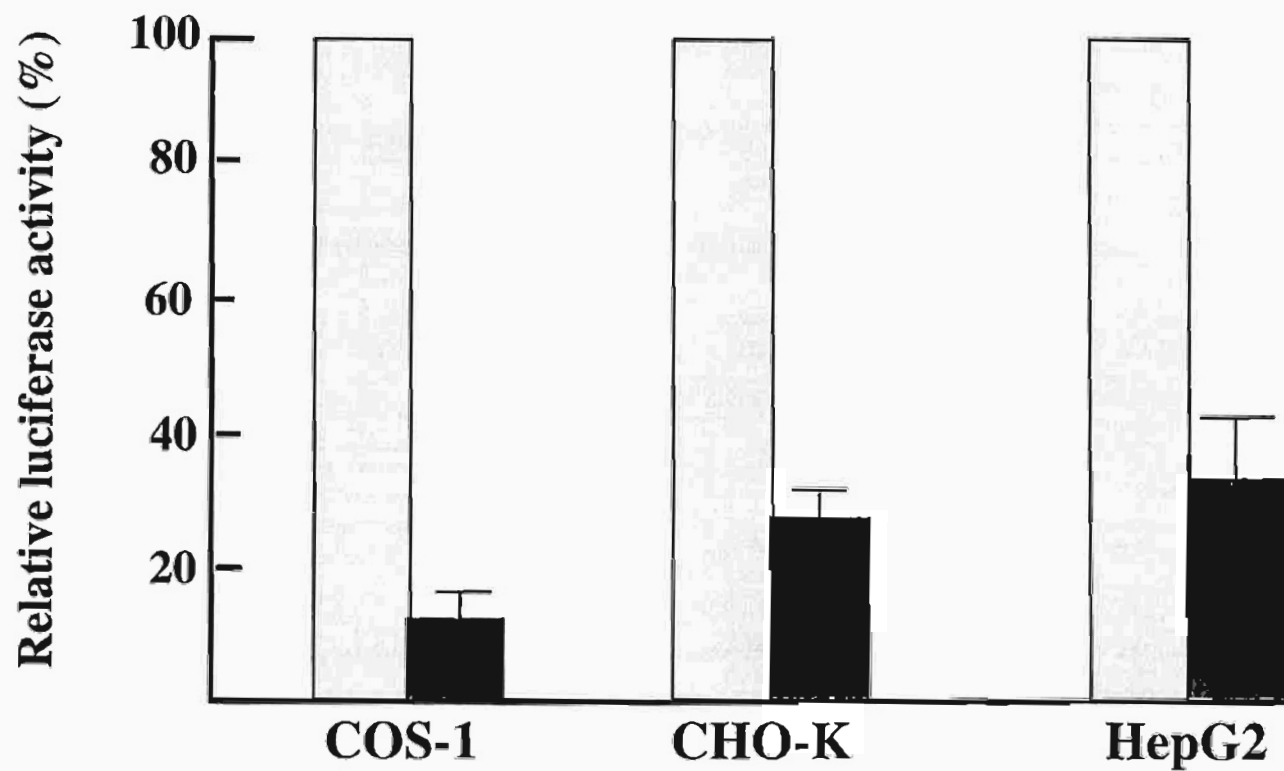
6.3.8 Regulation of PC promoters by insulin

The presence of a potential insulin-responsive element (IRE) located in both promoters of the rat PC gene raised the possibility that PC transcripts may be regulated by insulin through the binding of an insulin-responsive protein that could bind to this potential IRE and modulate PC transcription. To test this, both full length promoters (pGL-P1 and pGL-P2) were transiently transfected into COS-1 cells. The transfected cells were then recovered in serum-containing media (DMEM plus 10% fetal calf serum) for 24 h, since the serum contains a number of undefined substances including growth factors, hormones and certainly insulin. To avoid misinterpretaion of the result due to the endogenous insulin present in the serum, the transfected cells were maintained in serum-free media containing different concentrations of insulin i.e. 0.1 nM, 1.0 nM, 10 nM and 100 nM over the next 24 h. The pRSV-βgal expressing β-galactosidase and pGL-3 promoter containing SV40 promoter were also included in the experiment to serve as negative control. As shown in Table 6.1, insulin inhibited expression of the luciferase reporter gene from the pGL-P1 construct (proximal promoter) transfected into COS-1 cells in a dose-dependent manner. A small response to insulin was observed at the lowest concentration (0.1 nM), and progressively more inhibition was observed when the concentration of insulin was increased to 1.0 nM and 10.0 nM, respectively. The maximum inhibition (50%) of the luciferase activity was noted when the transfected cells were maintained at 100 nM concentration of insulin with a half maximum effect seen at 1 nM concentration. In contrast, insulin did not affect expression of the luciferase from the distal promoter construct (pGL-P2). COS-1 cells transfected with either pRSV-βgal or pGL-3 promoter were not affected by insulin (data not shown). This suggested that insulin specifically inhibited luciferase expression from the proximal promoter.

TABLE 6.1

Effect of insulin on luciferase expression from the pGL-P1 and pGL-P2 chimeric gene constructs.

COS-1 cells were transfected with pRSV-βGal and pGL-P1 or pGL-P2 and plasmids, and cultured in serum-free media containing 0, 0.1, 1.0, 10.0 and 100.0 nM insulin for 24 h. Relative promoter activities of both constructs are represented as fold response to the control value (no insulin) and results are mean $\pm$ the standard deviations for triplicate determinations.

Construct	Insulin (nM)				
	0	0.1	1.0	10.0	100
pGL-P1	1.0±0.22	0.88±0.25	0.75±0.21	0.63±0.18	0.48±0.15
pGL-P2	1.0±0.28	1.0±0.21	0.97±0.19	1.1±0.30	0.95±0.18

6.4 DISCUSSION

Regulation of rat PC expression at the gene level, revealed a complex transcriptional mechanism involving alternate usage of two promoters. The proximal promoter is responsible for the production of rUTR A, rUTR B and rUTR C transcripts and is restricted to gluconeogenic tissues and lipogenic tissues, and thus may be considered to be a tissue-specific promoter. In contrast, the distal promoter is responsible for the production of rUTR D and rUTR E transcripts, the transcripts that are expressed in most tissue types, and thus exhibits the characteristics of a housekeeping promoter. Nucleotide sequence analysis of the proximal promoter revealed no canonical TATA box, a sequence that is typical of eukaryotic promoters. The TATA box is usually located near positions -30 to -25 relative to the transcription initiation site (Benoist and Chambon, 1981), and is known to be recognised by a TATA binding protein (TBP) followed by the assembly of the transcription initiation complex which precisely positions RNA polymerase II to initiate transcription (Roeder, 1996). Mutation of the TATA box to other sequences has been shown to affect correct transcription initiation and yield a low level of transcription for a number of genes (Breathnach and Chambon, 1981), indicating the importance of this sequence.

Despite the critical role of the TATA box, a number of genes regulated by promoters lacking a TATA box, designated 'TATA-less', is now known to be very large (Azizkhan *et al.*, 1993). One of these groups is the 'housekeeping genes', so called because they encode proteins required for cellular metabolism. Although these genes do not contain a TATA box, TBP has been shown to be required for their *in vitro* transcription (Pugh and Tjian, 1990; Smale *et al.*, 1990; Pugh and Tjian, 1991), indicating that transcription by RNA polymerase II can be positioned through other motifs, in addition to the TATA box. The initiator element motif, which was first described in the mouse terminal deoxynucleotidyl transferase TATA-less gene, has been shown to be sufficient to promote a low level of specific transcription *in vitro* (Smale and Baltimore, 1989; Roeder, 1991). Such a motif has been shown to be mediated through binding of another factor known as the TATA-binding protein-associated factor to direct basal transcription (Martinez *et al.*, 1994). However, no such element was found in both promoters of the rat PC gene. Indeed, near the transcription initiation site, the sequence resembling a HIP-1 box found in the TATA-less promoter of mouse dihydrofolate

reductase (DHFR) gene (Means and Farnham, 1993) was identified in the rat PC proximal promoter. This HIP-1 box in mouse DHFR promoter has been shown to position the protein called 'housekeeping initiator binding protein 1 (HIP-1)". Binding of this protein to the HIP-1 box positions RNA polymerase II to start transcription (Means and Farnham, 1993). The HIP-1 motif has also been found in the DHFR genes of hamster (Mitchelle *et al.*, 1986) and human (Chen *et al.*, 1984). Other non-TATA box containing genes have also been found to have this sequence motif, including genes encoding SV40 late protein (Ghosp *et al.*, 1978), hypoxanthine phosphoribosyl transferase (Melton *et al.*, 1986), Ki-RAS (Hoffman *et al.*, 1987), 3-phosphoglycerate kinase (Singer-Sam *et al.*, 1984), osteonectin (McVey *et al.*, 1988), interferon regulatory factor 1 (Miyamoto *et al.*, 1988), SURF-1 (William and Fried, 1986) and transcobalamin II (Li *et al.*, 1995) (Fig. 6.10).

The presence of multiple GC boxes located upstream of the transcription initiation site is also known to be another feature of TATA-less promoter genes (McKnight and Tjian, 1986) and the multiple GC box plays a role in transcription (Blake *et al.*, 1990). In the mouse DHFR gene, transcription cannot occur if the upstream Sp1 binding site is deleted (Means and Farnham, 1993). Although multiple GC boxes were found upstream of the HIP-1 box of the proximal promoter of rat PC gene, only one copy matches the consensus sequence. Deletion of this potential Sp1 binding site (pGL-P1ΔP1) markedly abolished transcription of the reporter gene even though the HIP-1 box was still intact, suggesting a critical role of Sp1 in the transcription of the rat PC gene.

Other putative transcription factor binding sites for AP2, which is known to regulate the efficiency of transcription, presumably by altering the rate or conformation of polymerase attachment (Imagawa *et al.*, 1987; Mitchell *et al.*, 1987) were also found to be located between the Sp1 and HIP-1 binding sites in the proximal promoter of the rat PC gene. The presence of a putative liver-enriched transcription factor binding site for HNF-4 and a fat-specific element 1 (FSE1), is consistent with the gluconeogenic and lipogenic roles of this promoter, which is only transcriptionally active in these tissue types. It is not suprising to see the presence of a putative cAMP-responsive element (CREB) in the proximal promoter of the rat PC gene as it is known that other gluconeogenic enzyme genes are up-regulated by cAMP (Pilkis and Granner, 1992). The cAMP-responsive element in conjunction with

ATTC $(N)_{1-30}$GCCA   consensus sequence

ATTTC $(N)_2$GCCA   DHFR

ATTTC $(N)_{19}$GCCA   insulin regulatory factor

TTTC $(N)_2$GCCA   HPRT

ATTC $(N)_1$GCCA   3-phosphoglycerate kinase

ATTC $(N)_2$GCAG   osteonectin

ATTC $(N)_{27}$GCCA   transcobalamin II

ATTC $(N)_5$GCCA   pyruvate carboxylase

Figure 6.10   Housekeeping initiator binding protein-1 (HIP-1) box found in different genes.

CCAAT/enhancer binding proteins have been identified in the promoter of the PEPCK promoter and have been shown to mediate transcription of the PEPCK gene (Liu *et al.*, 1991; Roesler *et al.*, 1993; Roesler *et al.*, 1994; O'Brien *et al.*, 1994). It has been suggested that a cAMP-mediated pathway may be an important mechanism in response to cadmium-induced PC gene expression in the rat (Chapatawa *et al.*, 1980; Chapatawa *et al.*, 1982) and more recently in *C. elegans* (Liao and Freedman, 1998).

The presence of TCCCC or TCCCCC or their inverted repeats (11 copies) throughout this promoter raises the possibility that they might act as the repressor. As deletion progressed, these motifs were removed. However, computer analysis failed to identify any negative elements in the transcription factor data bases that match up to this motif.

Very much like the proximal promoter, a TATA box was not identified in the distal promoter. The DNA sequence near the transcription initiation site did not resemble any other alternative initiator elements in the databases. Nevertheless, when this promoter was cloned upstream from the reporter gene, it can drive luciferase reporter gene expression at a very high level suggesting that transcription initiation can function through other motifs located within this promoter region. As with other TATA-less promoters, multiple Sp1 binding sites were located within the 1 kb upstream from the transcription initiation site. Most interestingly, one of these Sp1 binding sites overlapped a putative IRE, as found in the proximal promoter. In contrast with the proximal promoter, three copies of CCAAT boxes were found in the distal promoter. The CCAAT box is known to be another common element which is usually located between 80 and 100 bp upstream of the transcription initiation site of eukaryotic encoding proteins (Bucher, 1990) and may be present in one or a few copies (van Huijsduijnen *et al.*, 1990). Many studies have demonstrated the importance of CCAAT boxes to promoter function, both for constitutively expressed genes and developmentally regulated genes (Maniatis *et al.*, 1987; Knight *et al.*, 1987). To date, several transcription factors have been isolated that bind to the CCAAT motif. The most well characterised is the CCAAT-binding factor (CBF or CP1 or NF-Y) (Dorn *et al.*, 1987; Chodosh *et al.*, 1988; Maity *et al.*, 1988) which recognises CCAAT boxes in a large number of other genes (Santoro *et al.*, 1988; Landschulz *et al.*, 1988). Mutations in the CCAAT box

cause a several-fold decrease in transcriptional activity both *in vivo* and *in vitro* (Maity and Crombrugghe, 1998). Binding of CBF to its recognition motif may regulate transcription from various promoters by cooperative interactions with other transcription factors that bind to a specific promoter (Maity and Crombrugghe, 1998) but the mechanisms are not yet understood. Deletion analysis demonstrated that the CCAAT box at -224 is not required for basal transcription, while two proximal CCAAT boxes located at -94 and -64 and may be an AP2 binding site were required for basal transcription.

The distal promoter is a stronger promoter than the proximal promoter when assayed in different cell types, consistent with the *in vivo* expression data presented in Chapter 4, i.e. transcripts generated from the distal promoter are expressed in wide variety of tissues. Therefore, it is very likely that the distal promoter of the rat PC gene acts as a housekeeping promoter which only requires general transcription factor(s) present in a wide variety of cell types.

In mammals, insulin is the principal hormone controlling blood glucose and acts by stimulating glucose influx and metabolism in muscle and adipocytes and inhibiting gluconeogenesis by the liver (Granner and Pilkis, 1992). Insulin has been shown to affect PC expression *in vivo*. In diabetic rats, in which the activity of hepatic PC was 2 times higher than in control animals, administration of insulin decreased PC activity to the control level (Weinberg and Utter, 1980). The action of insulin is mediated through the insulin receptor, a transmembrane glycoprotein with intrinsic protein tyrosine kinase activity. It is known that binding of insulin to its receptor triggers signal transduction pathways. This involves phosphorylation of the insulin receptor substrate 1 (IRS-1), which in turns activates several downstream effectors including transcription factors (White and Kahn, 1994). The IRE was originally identified in the promoter of the human GAPDH gene and was comprised of two distinct binding regions i.e. IRE-A and IRE-B. Insulin-responsive binding protein (IRE-ABP), which bound to the IRE-A, has been identified as a member of the HMG box transcription factor family (Nasrin *et al.*, 1991). IRE-ABP also exhibits high sequence similarity to the testis-determining factor, SRY gene, within the HMG box region. Furthermore, SRY protein was also capable of binding to IRE-A, suggesting that this motif is a candidate SRY-response element (Nasrin *et al.*, 1991).

Although both promoters contain identical putative IRE sequences, 5'-CCCGCCTCT-3' only the proximal promoter responded to insulin, suggesting that insulin action may be directed through another unidentified element located only in the proximal promoter. Alternatively, promoter context may be important for functional activity of this sequence. Nevertheless, selective inhibition by insulin of the proximal promoter, which is active in gluconeogenic tissues, indicates that PC expression can be negatively regulated at the transcriptional level.

Insulin also down-regulates expression of another gluconeogenic enzyme gene, PEPCK both *in vivo* and *in vitro* (Andreone *et al.*, 1982; Granner *et al.*, 1983; Sasaki *et al.*, 1984). Analysis of the rat PEPCK promoter has demonstrated that multihormonal regulatory regions are required to manifest the full response to insulin, cAMP and glucocorticoids (Liu *et al.*, 1991; Lucas *et al.*, 1991; Giralt *et al.*, 1991; Imai *et al.*, 1990; Forest *et al.*, 1990; O'Brien *et al.*, 1990; O'Brien *et al.*, 1991). The IRE found in the PEPCK promoter (Forest *et al.*, 1990; O'Brien *et al.*, 1990) is different from that found in the GAPDH promoter, suggesting that both elements identified in these two genes bind different IRE binding proteins. In the PEPCK gene, the distal IRE has been found to coincide with a glucocorticoid accessory factor element (AF2) (O'Brien *et al.*, 1990) which could provide a mechanism for the dominant negative effect that insulin has on glucocorticoid-stimulated PEPCK gene transcription (Magnuson *et al.*, 1987). Recently, IRE-binding proteins bound to an IRE in the PEPCK promoter have been identified to be a family of C/EBP proteins and an unidentified 20-kDa factor termed p20-C/IBP. These proteins also bound to a cAMP-responsive element on the PEPCK promoter with the same pattern to that of the IRE but with different conformations (O'Brien *et al.*, 1994). There are several different IREs identified in other genes whose products are regulated by insulin, including insulin-like growth factor binding protein-1 (Suwanickul *et al.*, 1993), α-amylase (Johnson *et al.*, 1993), glucagon (Philippe, 1991) and *c-fos* (Boxer *et al.*, 1989), but none of these IRE motifs are found in the proximal promoter of rat PC. A summary of IRE sequences of these genes are shown in Fig. 6.11.

In conclusion, a number of putative transcription factor binding sites for both ubiquitous and tissue-specific factors and IRE have been identified in the proximal and distal

Gene	IREs	IRE-binding proteins	reference
PEPCK	5'-TGGTGTTTTGACAAC-3'	c/EBP, p20-c/IBP	O'Brien *et al.* (1994)
α-amylase	5'-GTTTATTTTGCGTGA-3'	unknown	Johnson *et al.* (1993)
c-fos	5'-GGATGTCCATATTAGGACATCT-3'	unknown	Boxer *et al.* (1989)
hIGF-BP1	5'-CAAAACAAACTTATTTTG-3'	unknown	Suwanickul *et al.* (1993)
Glucagon	5'-AGTAGTTTTTCACGC-3'	unknown	Phillippe (1991)
TSH receptor	5'-CTTGTTTGGATGGAGAGTTGCCTAGGCAA-3'	unknown	Shimura *et al.* (1994)
GAPDH	5'-AACTTT<u>CCCGCCTCT</u>CAGCCTTTGAAAG-3'	IRE-ABP	Nasrin *et al.* (1991)
phosphoribosyl transferase	5'-<u>CCCGCCTCT</u>CAGCC-3'	unknown	Yoshikawa *et. al.* (1988)
PC	5'-TGC<u>CCCGCCTCT</u>CTGGTCCC-3' 5'-TCTT<u>CCCGCCTCT</u>TCACCAAT-3'	unknown	

Figure 6.11   Insulin responsive elements (IRE) found in different genes.

promoters of the rat PC gene. Although preliminary deletion analysis on both of these promoters have suggested possible functional roles for some of these binding sites, functional analysis is still required. Therefore, future experiments would involve site-directed mutagenesis, footprinting and gel retardation assay.

CHAPTER 7

REGULATION OF RAT PC EXPRESSION *IN*

VIVO

7.1 INTRODUCTION

The anaplerotic role of PC is to meet the cell's demand for oxaloacetate. The regulation of glucose metabolism is very important during different pathophysiological states i.e. starvation, the weaning period and in diabetes. An animal cells must have control mechanisms in order to maintain glucose homeostasis during such periods. At the onset of gluconeogenesis after the birth of a rat, the marked increases in gluconeogeneic enzymes i.e. PEPCK, PC, fructose-1, 6-bisphosphatase and glucose-6-phosphatase are observed primarily in the liver. This coordinate increase in the levels of these enzymes can ensure gluconeogenesis will occur effectively. Although the mechanisms involved in PEPCK induction have been studied intensively and shown to be regulated at the transcriptional level (Hanson and Patel, 1994), the mechanisms that control the production of PC during postnatal development are unknown. The studies on the isolation of cDNA (Chapter 3 and 4) of rat PC indicated that although there is only one copy of the PC gene revealed by fluorescense *in situ* hybridisation (Chapter 5) and Southern analysis (Chapter 4), it generates several isoforms of PC mRNA, which diverge at their 5'-UTRs but share the same open reading frame encoding a 1178-residue polypeptide. Two tissue-specific promoters are responsible for the production of five mature transcripts. Transcripts generated from the proximal promoter are restricted to gluconeogenec and lipogenic tissues, whereas transcripts generated from the distal promoter are expressed in a wide variety of tissues (Chapter 4). Key questions that remain unanswered are which promoter of the rat PC gene is activated or repressed under different physiological conditions and whether the 5'-UTRs of different PC mRNAs can modulate their translation.

To gain a further insight into the molecular mechanisms that regulate PC expression *in vivo* and *in vitro*, the role of trancriptional regulation during postnatal development, in genetically obese rats and in different cultured cells have been investigated. To assess the physiological role of the two promoters of the PC gene, the PC mRNAs have been analysed by means of an RT-PCR assay that identifies each one of the multiple forms comprising the PC mRNA population. The roles of the different 5'-UTRs of the several rat PC mRNAs on their translation were also investigated.

7.2 SPECIFIC METHODS

7.2.1 Animals

Sprague-Dawley rats were bred and housed at a constant temperature of 25°C in the animal house, University of Adelaide, and were subjected to treatments approved by the Animal Ethics Committee of the University of Adelaide. Studies were performed on rat fetuses at 20 days of gestation, on 1 and 7 day old suckling rats, and on 28 day old weaned rats fed with complete standard chow. Litters of 8 or 12 weeks old Zucker rats were obtained from Monash University, Clayton, Victoria. Tissues were quickly removed from the sacrificed animals, snap frozen in liquid nitrogen and stored at -80°C.

7.2.2 Cell culture

Different rat cell lines were routinely cultured as follows. BRL 3A (ATCC: CRL 1442), Reuber H35 (ATCC: CRL 1548), McA-RH7777 (ATCC: CRL 1601), L6 myoblast (ATCC: CRL 1458), rat mammary gland carcinoma cell line (obtained from Institute of Medical and Veterinary Science, Adelaide, Australia), FTO3 (obtained from the Institute of Cell and Tumour Biology, Heidelberg, Germany), fetal rat liver hepatoma cell line (FRL4.1; gift from Dr G. Yeoh, Department of Biochemistry, University of Western Australia, Nedlands) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum, 100 µg/ml streptomycin, and 100 units/ml penicillin at 37°C in a humidified atmosphere of 5% CO_2.

Pancreatic islets were isolated from well-fed Sprague-Dawley rats weighing 300 g as described previously (MacDonald *et al.*, 1995a). Islets were immediately homogenised in RNAzol B (Tel-Test Inc., Friendswood, TX) or maintained for 24 h in RPMI 1640 culture medium containing 10% fetal calf serum and homogenised in RNAzol B. The pancreatic islet RNA samples were prepared in the laboratory of Professor M.J. MacDonald, Children's Diabetes Center, University of Wisconsin Medical School, Madison, Wisconsin and transported to Adelaide in dried ice. Rat insulinoma (INS-1) cells were also obtained from Professor M.J. MacDonald and maintained in Adelaide in RPMI 1640 medium containing 50 µM β-mercaptoethanol, 1 mM pyruvate, 10 mM Hepes buffer, pH 7.35 and 10% fetal calf serum.

7.2.3 Ribonuclease Protection Aassay

The pRPC3 plasmid was constructed by subcloning a 0.3 kb *SacI-PstI* fragment of RACE (Anchor/PC 2) PCR product amplified from rat PC cDNA (Chapter 3) to *SacI-PstI* digested pBluescript. The pβ-actin plasmid was constructed by RT-PCR as follows : rat brain RNA was annealed to a cDNA synthesis primer designed from the coding region of rat β-actin mRNA [5'-TCTTCATGGTG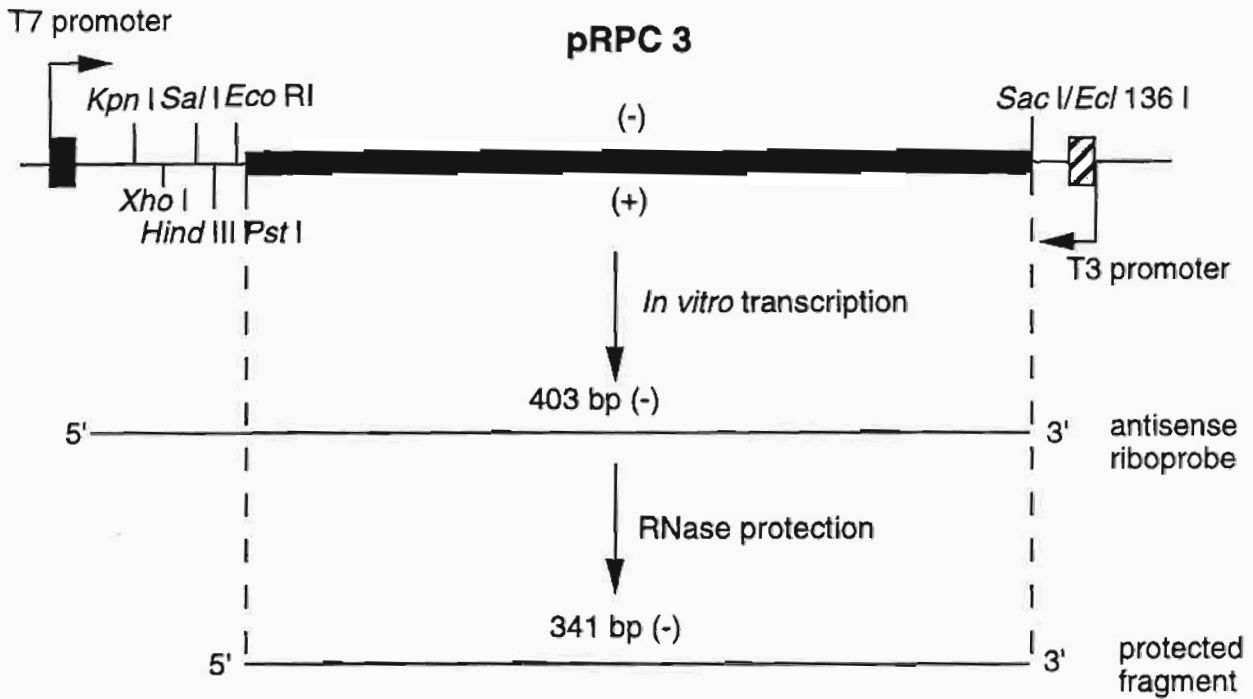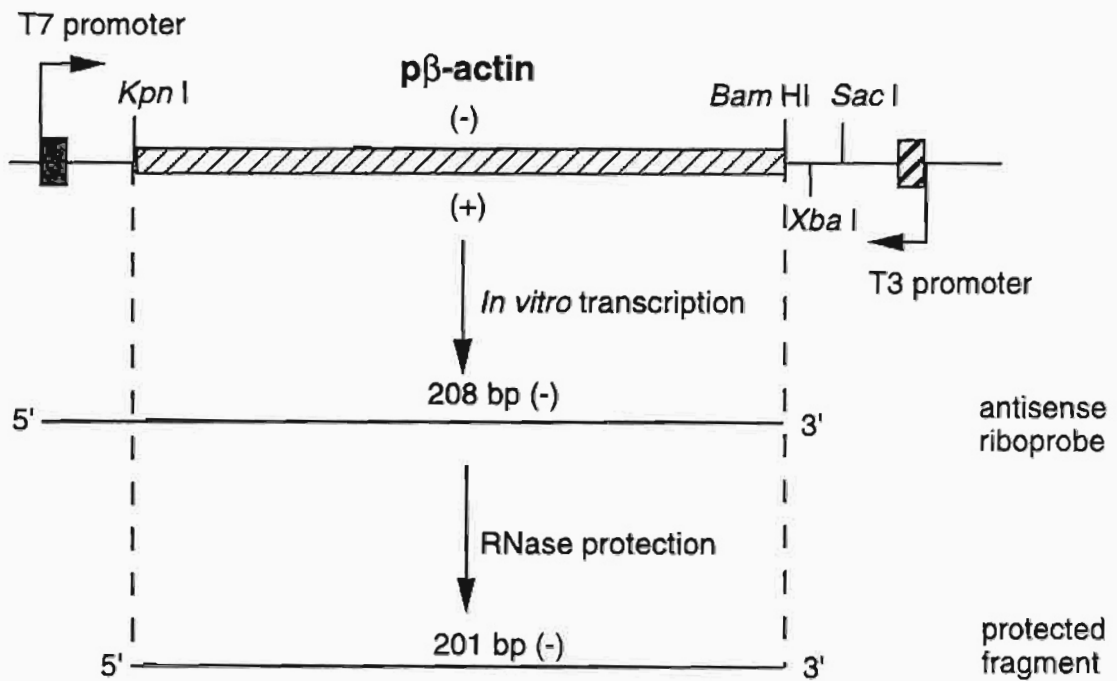CTAGGAGCCAG-3'] (Nudel *et al.*, 1983) and reverse transcribed. The cDNA was subjected to PCR with a sense primer [β-actin (+)] (5'-AACTT<u>GGATCC</u>CCTGGAGAAGAGCTATGAGCTG-3'; with a *Bam*HI restriction site attached at the 5'-end, underlined) and an antisense primer [β-actin (-)] (5'-TCCCG<u>GGTACC</u>AGACAGCACTGTGTTTGGCA-3', with a *Kpn*I restriction site attached at the 5'-end, underlined) (Nudel *et al.*, 1983). The PCR products were double-digested with *Kpn*I and *Bam*HI, cloned into *Kpn*I-*Bam*HI digested pBluescript and sequenced.

The above plasmids were linearised by digestion with *Ecl*136 and *Bam*HI respectively and purified. The antisense riboprobes were synthesised using the MaxiScript kit (Ambion, TX) with T7 RNA polymerase in the presence of 50 μCi of [α-^{32}P UTP] (8000 mCi/mmole). Full length probe was gel purified and eluted. The ribonuclease protection analysis was performed using a ribonuclease protection assay kit (Ambion, TX). Briefly, 10 μg of total RNA were hybridised with approximately 1×10^5 cpm of each probe at 45°C for 18 h. Probe and unhybridized RNAs were digested with RNase A and T1 for 30 min at 37°C. The 201 bp protected fragment representing β-actin mRNA and the 341 bp protected fragment representing the common coding region of different PC mRNAs (see Fig. 7.1) were denatured and separated by electrophoresis on a 6% acrylamide-8 M urea gel at 350 V for 2 h. The gel was dried and placed in the PhosphorImager screen overnight. The intensities of hybridisation bands were quantitated by PhosphorImager analysis using the ImageQuant Software (Molecular Dynamics). Uneven loading of RNA samples analysed was corrected by comparison with β-actin bands quantitated by PhosphorImager analysis.

Figure 7.1 Schematic diagram of the generation of riboprobes used for the ribonuclease protection analysis. (A), pRPC 3 plasmid containing 0.3 kb *PstI-SacI* fragment of rat PC cDNA cloned in pBluescript. The plasmid was linearised with *EcI*136I and transcribed *in vitro* in the presence of [α-^{32}P]UTP with T7 RNA polymerase. This yielded a 403 bp antisense riboprobe (including the polylinker region of pBluescript) which upon hybridisation to PC mRNAs and digestion with RNase generated a 341 bp protected fragment. (B), pβ-actin plasmid containing 0.2 kb *KpnI-Bam*HI fragment of rat β-actin cDNA cloned in pBluescript. The plasmid was linearised with *Bam*HI and transcribed *in vitro* in the presence of [α-^{32}P]UTP with T7 RNA polymerase. This yielded a 208 bp antisense riboprobe (including the polylinker region of pBluescript) which upon hybridisation to β-actin mRNA and digestion with RNase generated a 201 bp protected fragment.

A

T7 promoter

pRPC 3

Kpn I *Sal* I *Eco* RI

Sac I/*Ecl* 136 I

Xho I

Hind III *Pst* I

(-)

(+)

T3 promoter

In vitro transcription

403 bp (-)

5'   3'   antisense riboprobe

RNase protection

341 bp (-)

5'   3'   protected fragment

B

T7 promoter

pβ-actin

Kpn I

Bam HI *Sac* I

(-)

(+)

Xba I

T3 promoter

In vitro transcription

208 bp (-)

5'   3'   antisense riboprobe

RNase protection

201 bp (-)

5'   3'   protected fragment

7.2.4 Reverse transcriptase polymerase chain reaction (RT-PCR)

Detection of expression of different PC mRNA isoforms by RT-PCR was performed essentially the same as that described in Chapter 4, except that the β-actin mRNA replaced GAPDH mRNA as an internal control for cDNA synthesis and PCR reactions. The cDNA synthesis primer and PCR primers used to amplify β-actin mRNA were the same set as those used to synthesise pβ-actin (see section 7.2.3). The PCR products were analyzed on 2.5% agarose gel electrophoresis.

7.2.5 Preparation of tissue homogenate and PC assay

One gram of frozen liver was ground to a powder in liquid nitrogen and homogenised in 4 volumes of extraction buffer (0.25 M sucrose, 0.1 mM EDTA, 20 mM Tris-HCl, pH 7.2, 0.5 mM DTT and 0.5 mM PMSF) at 4°C. The homogenate was freeze-dried and reconstituted in 3 ml of 50 mM Tris-acetate pH 7.0, 5 mM ATP, 5 mM $MgCl_2$ and 0.5 mM EDTA. The homogenate was centrifuged at 20,000 x g for 20 min at 4°C and the supernatant containing PC was kept at -80°C. For adipose tissue, 1-2 g of epididymal fat pads were homogenised as described above and fat was removed from the homogenate by centrifugation at 10,000 x g for 5 min at 4°C. The aqueous extract was removed from the cell debris, freeze-dried and subsequently reconstituted as described above. Twenty micrograms of total protein were assayed for PC activity. The PC activity was assayed by a radioactive CO_2-fixation assay previously described (Ballard and Hanson, 1967; Ballard *et al.*, 1970) except that the reaction was coupled with the conversion of oxaloacetate to malate by 2 units of malic dehydrogenase in the presence of 0.1 mM NADH. The reaction was assayed at 30°C for 2 min. One unit of PC is defined as the conversion of 1 μmole of $NaH^{14}CO_3$ to malate per minute and the PC activity is defined as milliunits per milligram of total protein.

7.2.6 SDS-PAGE and Western analysis

Total protein from liver homogenate (20 μg) was subjected to reducing discontinuous SDS-PAGE on a 4% stacking gel and 7.5% polyacrylamide separating gel (Laemmli, 1970). The separated proteins were transferred to nitrocellulose filters using a semi-dry electroblotter (Multiphor II Novablot, Pharmacia, Uppsala, Sweden). Membranes were blocked with 1% bovine serum albumin in 10 mM Tris-HCl pH 7.5, 150 mM NaCl

and 0.05% Tween 20. The membrane was reacted for 2 h with 1:20,000 dilution of anti-chicken PC rabbit IgG polyclonal antibodies which cross-react with rat PC (Rohde *et al.*, 1991). The secondary anti-rabbit antibodies conjugated with alkaline phosphatase were then reacted with the primary antibodies for 2 h. The immunoprecipitate bands were visualised by adding nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indoyl phosphate (BCIP) and quantitated by laser densitometer (Molecular Dynamics). To correct variations between blots due to different incubation times with NBT and BCIP, a known amount of purified rat PC was also run as an internal control such that the intensities of immunoprecipitated bands between different experiments could be compared.

7.2.7 In vitro translation

In vitro translation was performed with rabbit reticulocyte lysates. The cDNA encoding the biotin carboxylation (BC) domain of the rat PC was generated by RT-PCR using a forward primer directed against residues 1 to 9 of rat PC with a *Bam*HI restriction site attached at the 5'-end (underlined)

(5'-CCAAA<u>GGATCC</u>ATGCTGAAGTTCCAAACAGTTCGAGG-3') and a reverse primer directed against residues 482 to 489 together with stop codons (bold) and *Kpn*I restriction site (underlined)

(5'-TCCC<u>GGTACC</u>**TTATCA**CTGGAACAGCTCGGGGTTCTCATCG-3'). RT-PCR conditions were the same as described above. The PCR product was then digested with *Bam*HI and *Kpn*I and cloned into *Bam*HI and *Kpn*I sites of pGEM3Zf (+/-) and sequenced. The 5'-UTRs of rUTR C, D and E which had previously been cloned in pBluescript (Chapter 4) were then excised with *Eco*RI and *Bsu*36I and inserted in frame in front of the coding region of the above construct. Finally, the 3'-UTR and poly(A) tail of rat PC cDNA were excised from λRL 2.35 clones (Chapter 3) and fused to the 3'-end of the above constructs at *Kpn*I site. This resulted in the final constructs containing different 5'-UTRs of rUTR C, D and E of the rat PC mRNAs (Chapter 4) respectively, fused to the BC domain followed by the 3'-UTR and poly(A) tail. A schematic diagram of the construction of these plasmids is shown in Fig. 7.2. The RNAs were transcribed from 1.5 µg of either circular or linearised plasmids, with T7 RNA polymerase in the presence of m7Gppp(5'). Free m7Gppp(5')G was removed by LiCl precipitation. Capped mRNAs (0.2 µg) were then

subjected to translation *in vitro* with rabbit reticulocyte lysate (Promega). Briefly, the reaction contained 25 µl of rabbit reticulocyte lysate, 1 µl of 1mM amino acid mixture minus methionine, 20 µCi of [^{35}S] methionine, 40 U of RNase inhibitor, capped mRNA and adjusted to a final volume of 50 µl with water. The reaction was incubated at 30°C for 60 min. The synthesised proteins were analysed by reducing discontinuous SDS/PAGE (12.5% separating gel, 4% stacking gel). The dried gel was exposed to a PhosphorImager screen overnight before being quantitated as above.

7.2.8 Isolation of polysome bound RNAs

Rat livers were homogenised in 3 ml of buffer A containing 10 mM Hepes, pH 7.4, 100 mM KCl, 5 mM MgCl$_2$, 2% Triton X-100, 100 µg/ml cycloheximide and 20,000 units of RNase inhibitor. Nuclei and unbroken cells were removed by centrifugation at 10,000x g at 4°C for 5 min. One ml of cytosolic supernatant was then overlaid on 11 ml of 15% to 45% sucrose gradient in buffer A and centrifuged at 40,000 rpm at 4°C for 2 h in a Ti60 rotor (Beckman). One ml fractions were collected manually from the bottom to the top of the tube. As a control, an equal portion of cytosolic supernatant was adjusted to 20 mM EDTA but lacking MgCl$_2$. RNA in each fraction was extracted with phenol/chloroform and following ethanol precipitation, was subjected to RT-PCR analysis.

Figure 7.2 Generation of the BC domain of the rat PC fused to its 5'-UTRs, 3'-UTR and poly(A) tail. RT-PCR was used to generate the BC domain which was subsequently fused to different 5'-UTR, and to a common 3'-UTR and poly(A) tail of rat PC cDNA (Chapter 3). PCR primers are shown by arrows, stop codons introduced at the 3'-end of antisense primer are indicated by asterisks. The coding region of mRNA encoding three domains, i.e. the biotin carboxylation domain (BC), the transcarboxylation domain (TC) and the biotinyl domain (BIO) are also shown.

BamHI Bsu36I

5' [BC] [ ] [TC] [ ] [BtO] [ ] 3' mRNA

** KpnI

PCR/cloned

Bsu36I

EcoRI BamHI KpnI SaII

[BC]

EcoRI/Bsu36I

EcoRI Bsu36I EcoRI

5'-UTR [ ]

EcoRI/Bsu36I

λRL2.35

EcoRI Bsu36I KpnI SaII

5'-UTR [BC]

KpnI KpnI

3'-UTR, poly(A)

KpnI KpnI

EcoRI Bsu36I KpnI KpnI SaII

5'-UTR [BC] 3'-UTR, poly(A)

pRAA-BC, pRAB-BC, pRAC-BC
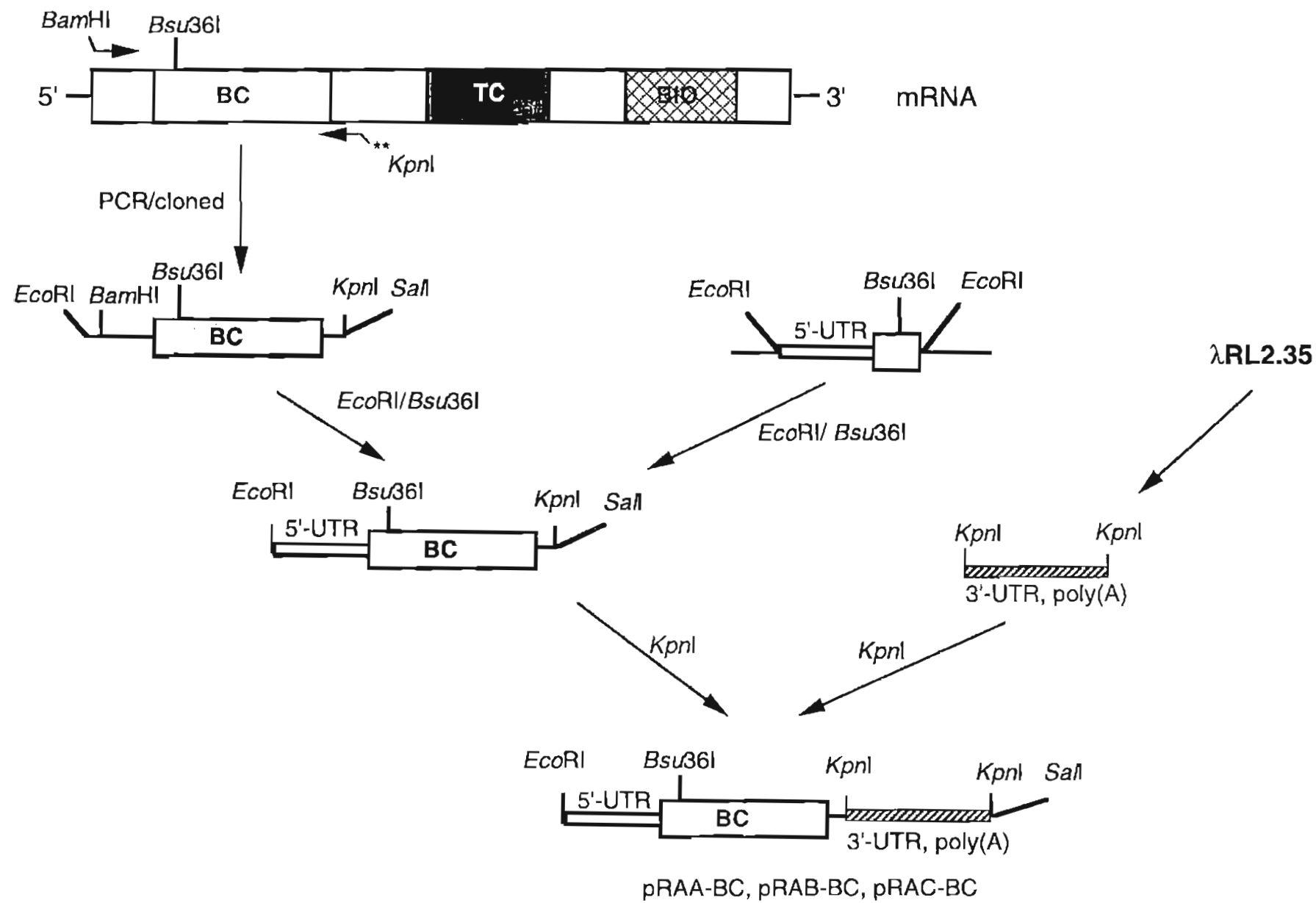
7.3 RESULTS

7.3.1 Establishment of linearity of the RNase protection assay and Western analysis for detecting PC mRNAs and PC protein

Initially, the amounts of total RNA (5, 10, 15 and 20 μg) used in the RNase protection assay were titrated with excess amounts of either β-actin or PC riboprobes (50,000 cpm each) to ensure the maximum sensitivity and linearity of the assay system. As shown in Fig. 7.3 A, B and C, linear relationships between the intensities of protected probe bands and different amounts of total RNAs were observed, both with β-actin and PC probes indicating the validity of the system. However, lower amounts of RNAs would be required if both β-actin and PC transcripts could be detected simultaneously by adding both probes in the same reaction tube. This would also minimise variation between samples. As indicated in Fig 7.3 A and D, detection of both transcripts by adding β-actin and PC probes in the same assay tube still gave reliable results, i.e. linearity between intensities of the protected fragments and amounts of total RNA was still observed. Ten micrograms of total RNA were used throughout the following experiments, as this amount of RNA gave enough sensitivity to be detected on both the PhosphoImager screen and X-ray film.

The Western analysis for PC in liver homogenates is based on an immunological reaction of polyclonal antibodies (raised against chicken PC, but also cross reactive with rat PC) to PC immobilised on nitrocellulose after SDS-PAGE, followed by an addition of goat anti-rabbit IgG antibodies conjugated with alkaline phosphatase. As shown in Fig. 7.4, there is linearity between the intensities of the immunoreactive bands (~130 kDa) over different amounts of total protein analysed. Twenty micrograms of total protein were used throughout the following experiments.

7.3.2 Developmental regulation of PC expression in postnatal rat liver

Previous studies have shown that PC activity in fetal rat liver is barely detectable but increased markedly during the suckling period (Ballard and Hanson, 1967; Yeung *et al.*, 1967). To determine whether this regulation occurs at the transcriptional, translational or post-translational steps, the levels of PC protein were measured by Western immunoblot and by PC activity assay, while the levels of PC transcripts were measured using both a

Figure 7.3 Establishment of linearity of the RNase protection analysis for PC and β-actin mRNAs. Fifty thousands cpm (count per minute) of each probes were hybridised with different amounts of total RNA (5, 10, 15 and 20 μg) separately or in the same tube. The protected fragments representing PC and β-actin transcripts were analysed on 6% acrylamide-8M urea gel and exposed to the PhosphorImager screen overnight (**A**). The intensities of the protected fragments on the PhosphorImager screen were quantitated using ImageQuant software and plotted in **B** (β-actin probe), **C** (PC probe) and **D** (β-actin and PC probes). Linear equations obtained from these assays are also shown.

A

β-actin   PC   β-actin+PC   PC   β-actin

M   5   10   15   20   5   10   15   20   5   10   15   20

bp

404
331 ◄ PC
242
190 ◄ β-actin

B

$y = 2.2849e+4 + 3.4924e+4x \quad R^2 = 0.990$

Integrate units

β-actin

total RNA (µg)

C

$y = -1150.8 + 1.1075e+4x \quad R^2 = 0.992$

Integrate units

PC

total RNA (µg)

D

$y = 2.5573e+4 + 3.2704e+4x \quad R^2 = 0.992$

Integrate units

□ PC
◆ β-actin

$y = 3811.2 + 1.0867e+4x \quad R^2 = 0.999$

total RNA (µg)
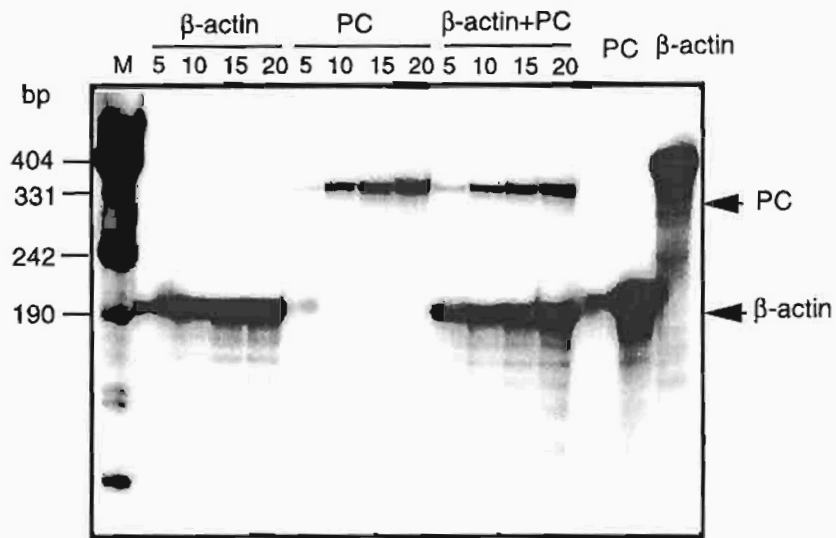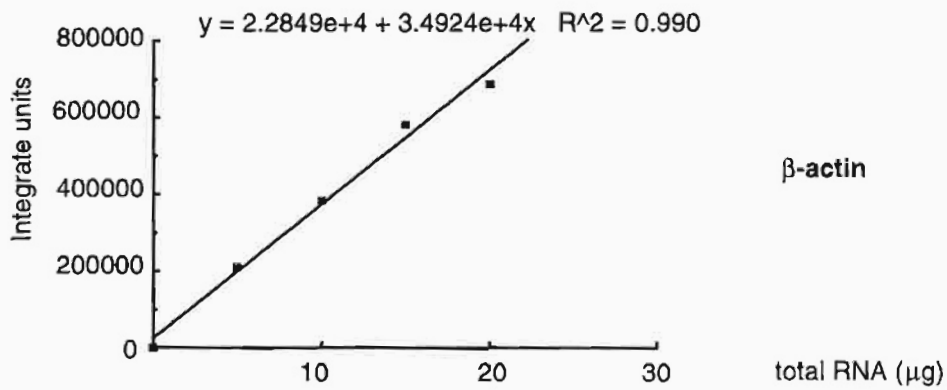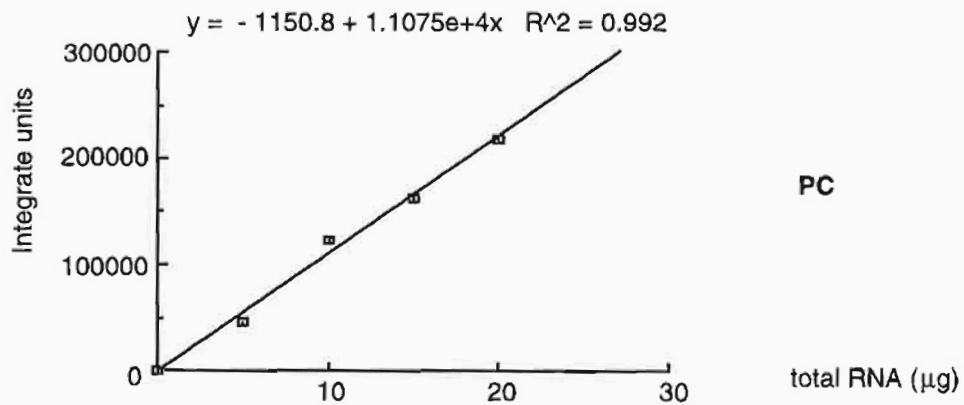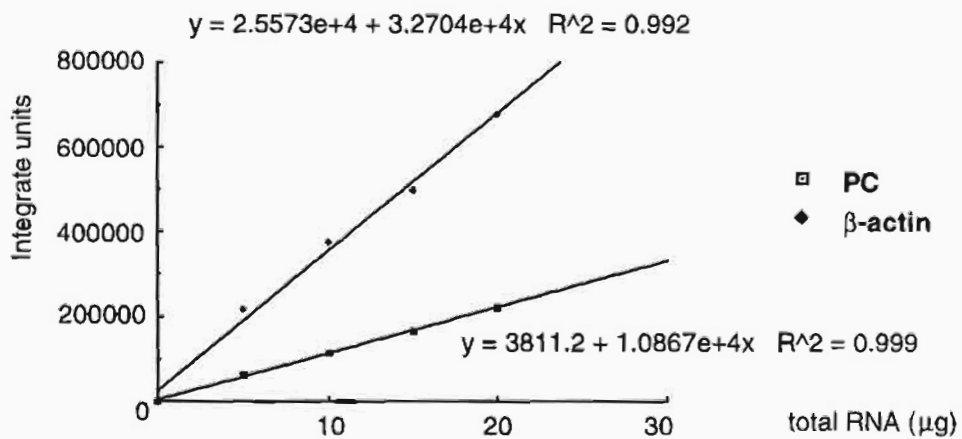
Figure 7.4 Establishment of linearity of the Western blot for PC proteins.
(A) Different amounts of total protein in a liver homogenate were separated on reducing SDS-PAGE. The separated proteins were transferred to nitrocellulose and reacted with 1:20,000 dilution of anti-chicken PC polyclonal antibodies followed by addition of goat anti-rabbit antibodies conjugated with alkaline phosphatase. The intensities of the immunoreactive PC bands were visualised by adding NBT and BCIP and quantitated by densitometer. The results are plotted in (B), where the linear equation obtained from the quantitation is also shown.

A

liver homogenate (total protein)

M   10 µg   20 µg   50 µg   100 µg   160 µg   rPC

116 kDa — ————————————————————————— ← PC

B

$y = 35.340 + 2.3322x \quad R^2 = 0.982$
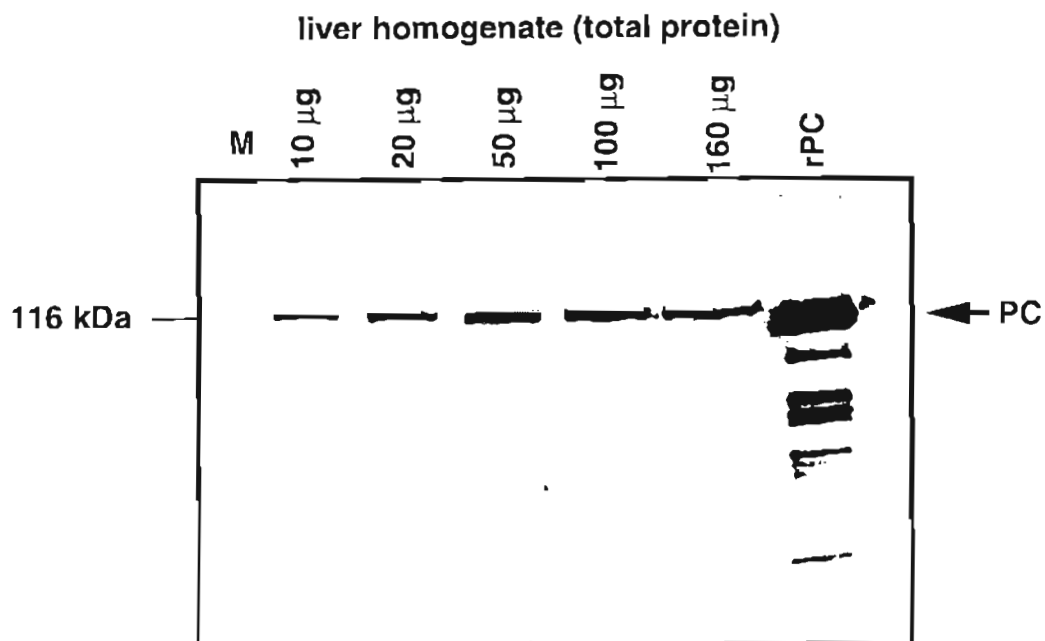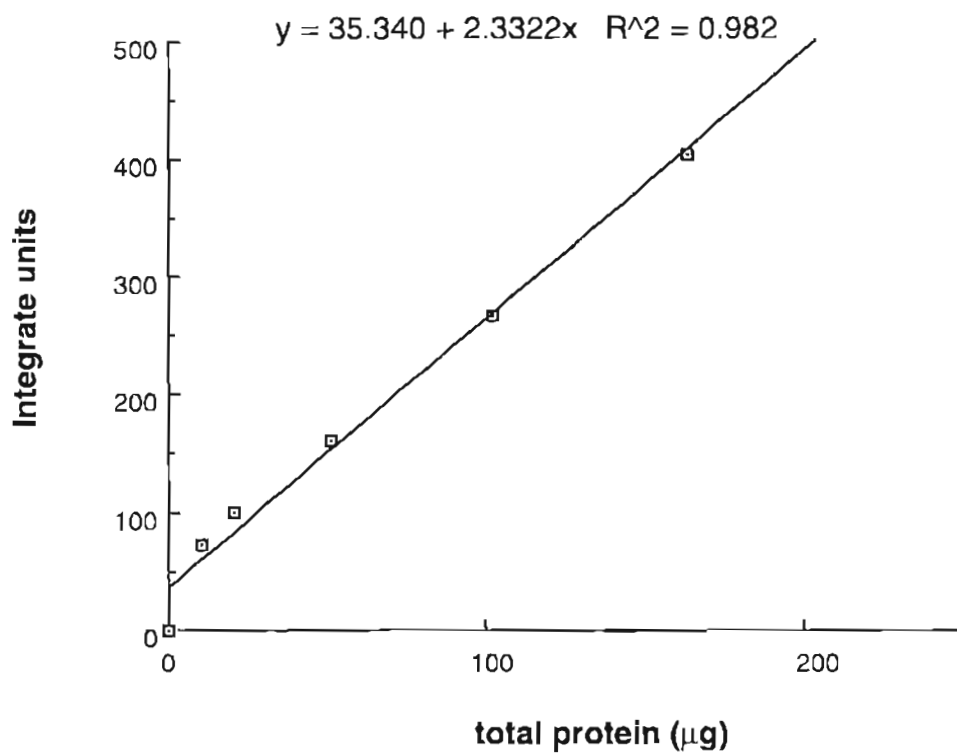
ribonuclease protection assay and RT-PCR in different aged rats i.e. late gestation (2 days before birth), suckling period (1 day and 7 day pups), weaned rats (28 days) and adults.

As seen in Fig. 7.5 A and 7.5 C, the levels of PC protein were very low in 20-day fetal rat liver but rapidly increased 4-fold within 1 day after birth and by 8-fold at day 7. However, PC levels had decreased by the time of weaning to approximately adult levels. The increase in immunoreactive PC protein during the suckling period is concomitant with a similar increase in PC activity (4-fold and 8-fold increase) as indicated in Fig. 7.5 C, indicating that post-translational modification is not responsible for the increase in PC activity. The amount of immunoreactive PC protein was also decreased to the same level in weaned rats and in adults as that seen in 1 day pups (see Fig. 7.5 A and 7.5 C). Ribonuclease protection assay using a riboprobe synthesised from the common coding sequence of different forms of PC mRNA clearly showed that although both PC protein and PC activity were low in fetal liver, its transcripts were readily detectable (Fig. 7.5B). The abundance of PC transcripts was increased by 2-fold and 4-fold in 1 day and 7 day pups respectively (Fig. 7.5 C). However, the increases in the amounts of PC transcripts are not as dramatic as those observed with the increases in PC protein (4-fold and 8-fold), suggesting that translational control might play a role in accelerating the rate of enzyme synthesis. The abundance of PC transcripts is also decreased during weaning and in adults as seen with the level of PC protein and PC activity (Fig. 7.5 B and 7.5 C).

Because the pool of rat PC mRNAs is comprised of multiple transcripts generated from alternate promoters (Chapter 4), the RNase protection assay results only reflect an overall transcriptional activity of the gene. Since the sizes of the 5'-UTRs of different rat PC mRNA isoforms are relatively small compared to the common coding region of the messages (4.0 kb), conventional Northern analysis using a probe synthesised from the coding region cannot differentiate between these subtle size differences. Therefore, semi-quantitative RT-PCR was also performed to identify which promoter of the PC gene was being used to control the expression. Two sense primers directed against the most 5'-UTR of class I and class II PC transcripts (Fig. 4.5, Chapter 4) and an antisense primer directed to a common coding region of the different transcripts were used to selectively amplify by PCR the 5'-ends of the different PC cDNA species. Different amounts of RNAs were titrated in a RT-

Figure 7.5 Western immunoblot analysis of PC, its activity and ribonuclease protection assay of PC mRNAs in rat livers during postnatal development. Livers from rats of different ages (20 day gestation (n=4), 1 day pups (n=6), 7 day pups (n=6) and 28 day pups (n=5) and adult (n=3), were removed and liver homogenates and RNAs were prepared. (A) Total proteins from liver homogenates (20 µg) were analyzed by Western immunoblot probing with anti-chicken PC antibodies (23). (B) Total RNAs were subjected to a ribonuclease protection assay using an antisense riboprobe synthesized from the common coding region of PC and β-actin riboprobe (internal control). Only two samples from each group are shown in this assay (C) Amounts of PC protein, and PC transcripts were estimated by scanning densitometry from their respective Western immunoblots (A) and ribonuclease protection assay (B). PC activities in different ages of rats were also assayed and their fold increases are also shown on the graph. Results are expressed as relative fold increase over the value for the 20 day gestation stage. They are means ± standard errors of means. T, birth.

A

20 day gestation

1 day pups

7 day pups

28 days

adults

B

| 20 day gestation | 1 day pups | 7 day pups | 28 days | adult |

PC

β-actin

C

- PC protein
- PC activity
- PC transcripts

Relative fold increase (arbitary units)

Age (days)

-2 T +1 +7 +28 adult
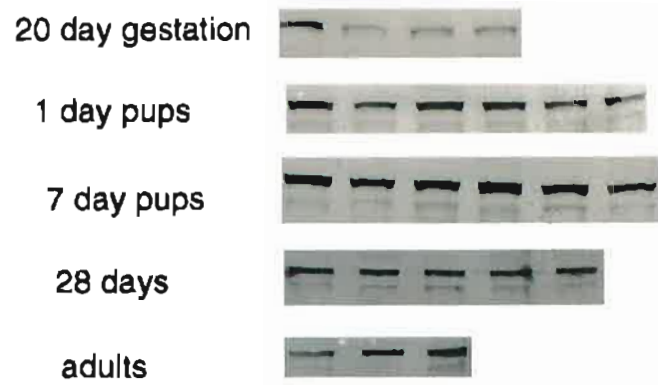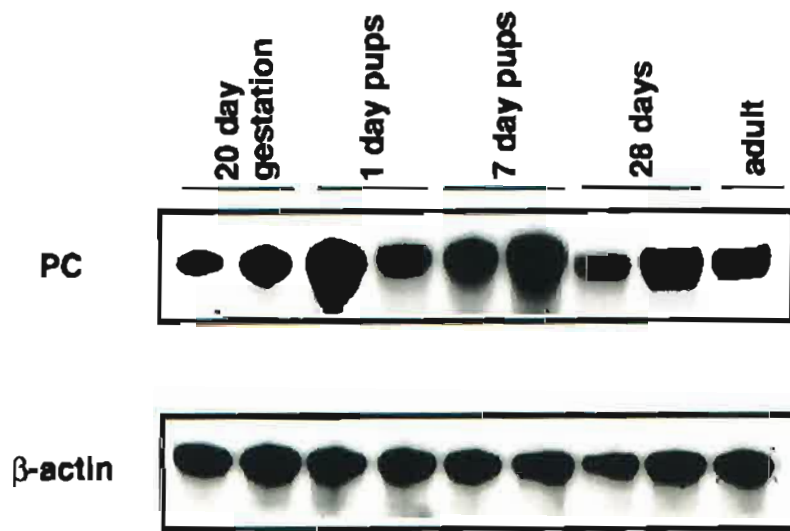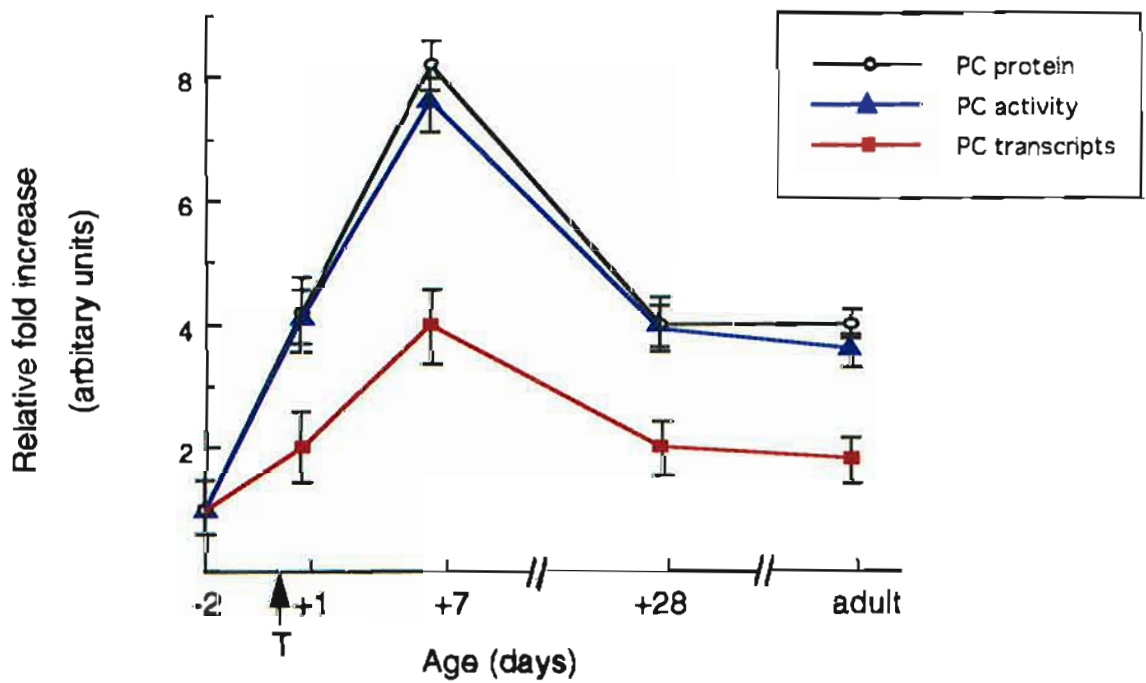
PCR assay and it was found that using a 1:100 dilution of cDNA synthesized from 10 μg of starting RNA gave reliable results (data not shown). The different sizes of PCR products expected are shown in Fig. 4.5. As shown in Fig. 7.6, in 20-day fetal liver, rUTR D, the major form transcribed from the distal promoter was detectable in a greater abundance than rUTR C, transcribed from the proximal promoter. The level of rUTR C was increased in 1-day and 7-day pups respectively. It is interesting to note that although rUTR D was accumulated in adult concomitant with a decrease of rUTR C (Fig. 7.6), the level of PC and its activity were decreased (Fig 7.5A and 7.5C). This suggested that the translation of rUTR D produced during such a period is less efficient than rUTR C. However, rUTR A and B, minor forms transcribed from the proximal promoter were not detectable and this is consistent with the results described in Chapter 4.

7.3.3 Overexpression of PC in genetically obese rats

In genetically obese Zucker rats, it has been reported that PC levels are 2-5 fold increased in the adipose tissue of obese animals (*ob/ob*) (Lynch *et al.*, 1992). To investigate whether this increase in the amount of PC is due to an enhanced transcriptional activity of the gene, ribonuclease protection assays and RT-PCR were performed as previously described above. In adipose tissue of obese rats, PC transcripts were increased 5- to 6-fold higher than those of their lean litter mates, whereas β-actin message was not affected (Fig. 7.7 A). RT-PCR clearly showed that the increase in PC mRNA levels was mainly due to an increase in rUTR C transcript, with a decrease in rUTR D and rUTR E transcripts between the two groups of rats (Fig. 7.7 B). In livers isolated from the obese rats, PC transcripts were also increased. A four-fold difference was observed between the two groups of animals (see Fig 7.7 A). RT-PCR analysis revealed the increase in hepatic PC mRNA again resulted from an increase in rUTR C (see Fig 7.7 B) with little change in rUTR D and rUTR E.

An increase in PC transcripts in both adipose tissue and liver was concomitant with an increase in the amount of PC detected by Western immunoblot. As shown in Fig. 7.7 C, the levels of adipose PC in obese rats were 4-5-fold higher than those of their lean litter mates but immunoreactive hepatic PC in obese rats was only 2-3-fold higher than in their lean litter mates.
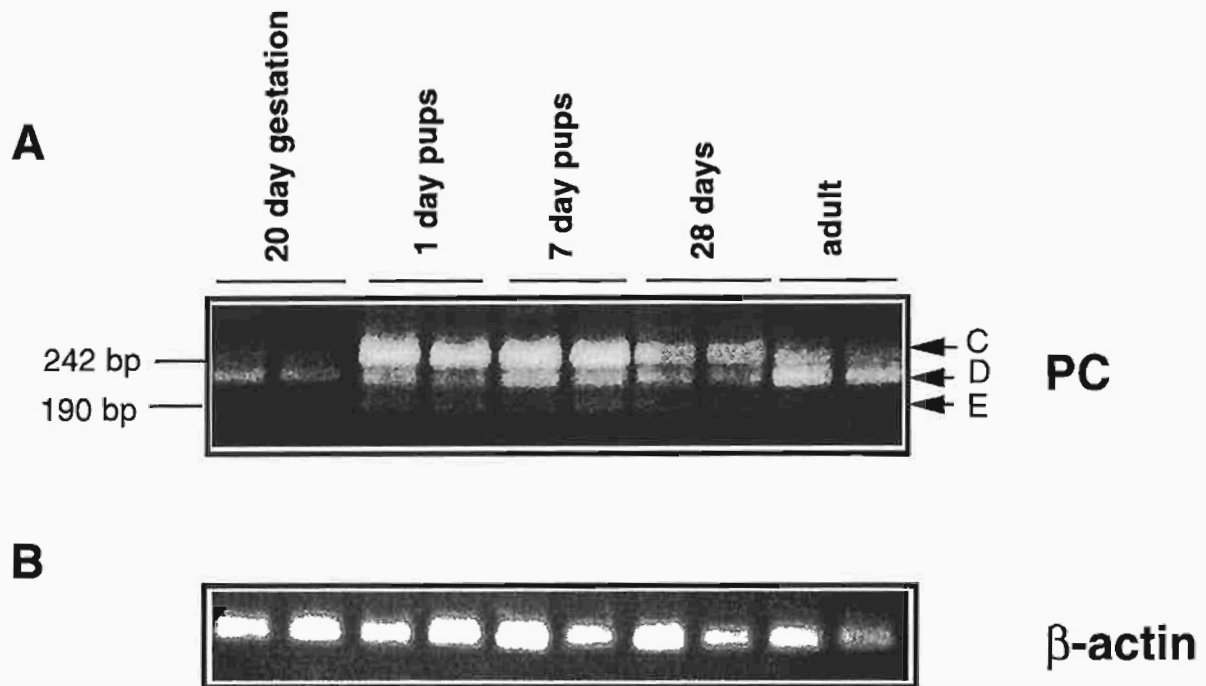
Figure 7.6   RT-PCR analysis of different species of PC mRNA isoforms in rat liver during postnatal development. Total RNAs were extracted from rat livers of different ages as described in Fig. 7.5 and subjected to RT-PCR using (**A**) PC primers; *C, D* and *E*, for rUTR C, rUTR D and rUTR E transcripts, respectively and (**B**) β -actin primers (internal control).

RNase protection

A

liver   adipose tissue

lean rats   obese rats   lean rats   obese rats

PC

β-actin

RT-PCR

B

liver   adipose tissue

lean rats   obese rats   lean rats   obese rats

242 bp —

190 bp —

C
D
E

PC

β-actin

Western immunoblot

C

liver   adipose tissue

lean rats   obese rats   lean rats   obese rats
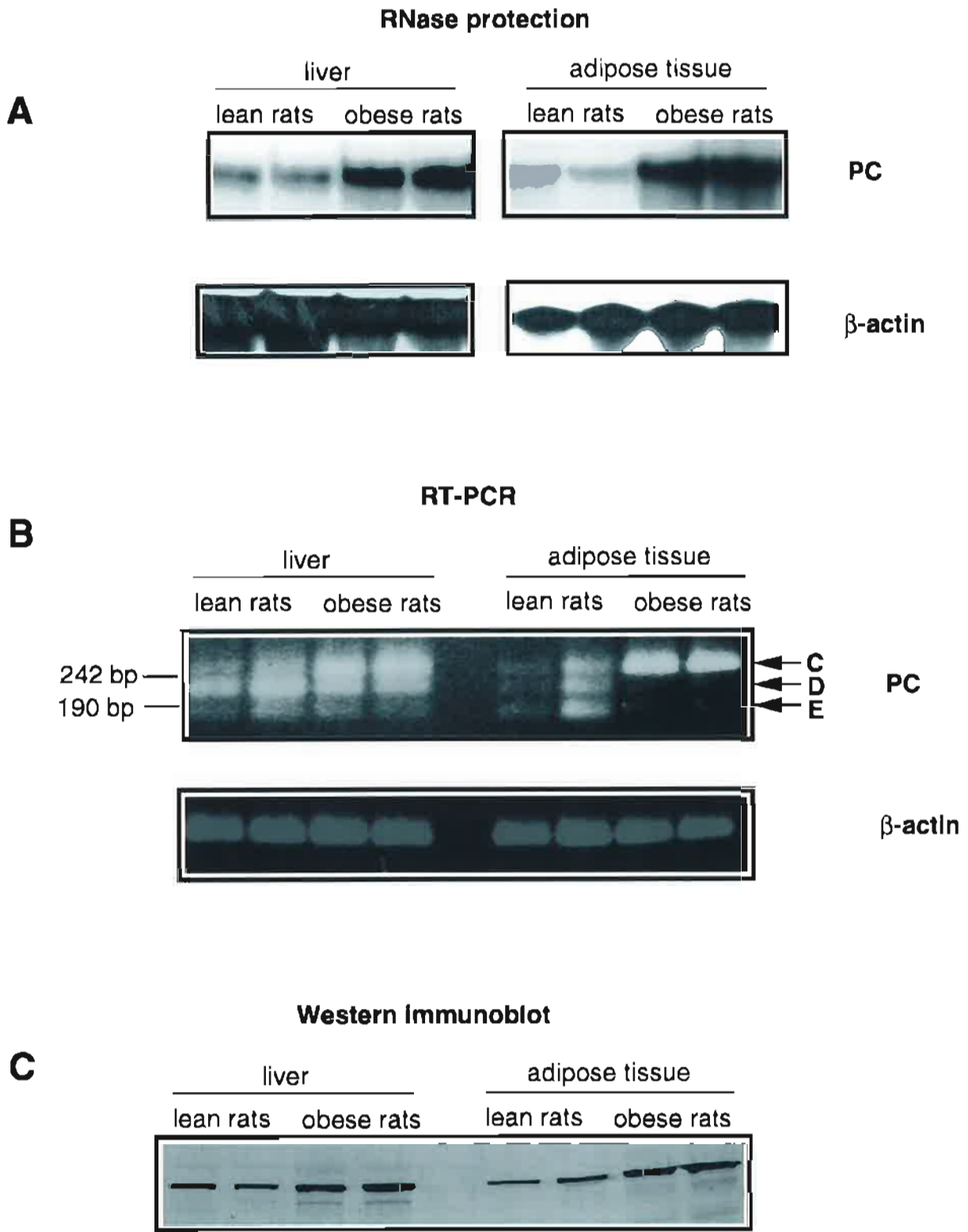
Figure 7.7   **Expression of PC in lean and obese Zucker rats.** Livers and adipose tissues from 12 week lean (n=4) and obese (n=4) rats were analyzed by (**A**) ribonuclease protection assay, (**B**) RT-PCR, and (**C**) Western immunoblot. The results for only two animals of each group are shown in the figure. C, D and E for rUTR C, rUTR D and rUTR E transcripts respectively.

7.3.4 Expression pattern of PC transcripts in cultured cells and in pancreatic islets

Although PC was expressed in every cell line as detected by Western immunoblot (data not shown), the expression patterns of PC transcripts in liver cell lines were different from that of an intact liver. rUTR C generated from the proximal promoter was not detectable even though 100 times more cDNAs were used in RT-PCR, but rUTR D and rUTR E transcribed from the distal promoter were detectable by RT-PCR (Fig. 7.8 A). It is unlikely that the primer binding site was absent or deleted since all of these cell lines were independently derived from different sources. It is likely that the proximal promoter driving expression of rUTR C in liver is not functional under these cell culture conditions.

PC expression has been shown to be activated upon glucose-induced insulin secretion in pancreatic islets (MacDonald, 1995a) and insulinoma (INS-1) cells (Brun *et al.*, 1996). RT-PCR was performed on RNA extracted from these two cell types. Only rUTR D and rUTR E were detected in both cell types (see Fig. 7.8 B) suggesting that the distal promoter plays a role in anaplerosis. To test whether glucose can increase transcription of the distal promoter, INS-1 cells were initially maintained in RPMI glucose-free medium containing 1 mM glucose for 48 days. The cells were then maintained in the same glucose-free media supplemented with 1, 10, and 20 mM glucose. As shown in Fig. 7.8 B, glucose dramatically increased the level of rUTR D and rUTR E when the INS-1 cells were switched from a 1 mM to a 10.0 mM concentration (Fig. 7.8 B).

7.3.5 5'-UTRs of different PC transcripts modulate their translation

A possible role of the different 5'-UTRs could be to determine the translational efficiency of the mRNA species. I carried out experiments using both *in vitro* translation with a reticulocyte lysate system and an *in vivo* polysome profile analysis.

Due to the lack of convenient restriction sites to join them together, there was considerable difficulty in constructing a full length rat PC cDNA clone from the different lambda clones and PCR fragments. Therefore, three constructs were generated that contained different 5'-UTRs (95-bp, rUTR C; 96-bp, rUTR D and 59-bp, rUTR E), followed by an open reading frame encoding the 490 residues of the biotin carboxylation (BC) domain plus the 3'-UTR and the poly(A) tail of rat PC cDNA (see Fig. 7.2 and 7.9 A).
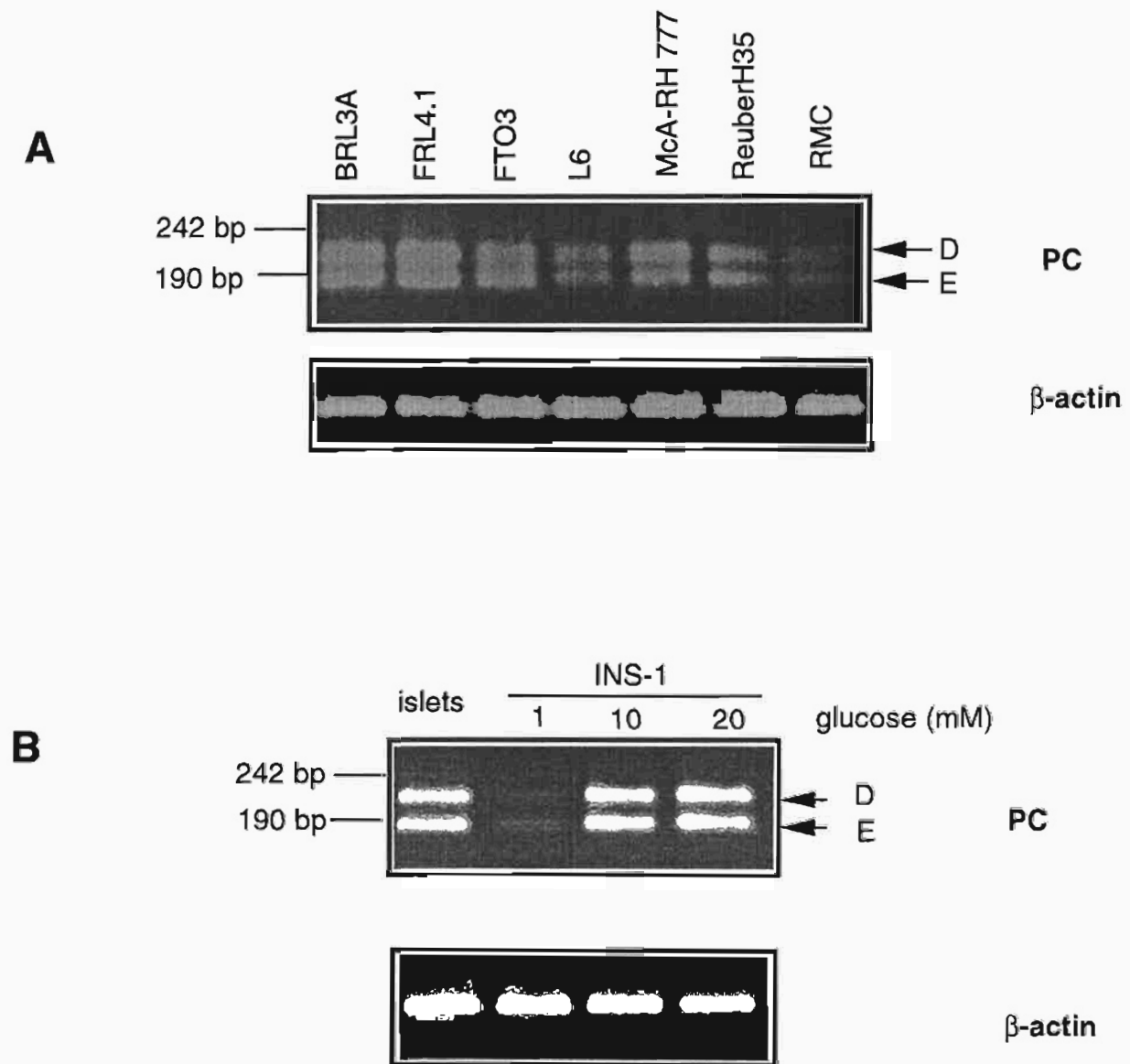
Figure 7.8   RT-PCR analysis of PC mRNA isoforms in different rat cell lines. Total RNAs from (A) liver cells (BRL3A, FRL 4.1, FTO3, McA-RH 777 and Reuber H35), myoblast (L6) and mammary gland carcinoma cells (RMC), (B) primary culture of pancreatic islets grown for 48 h in RPMI supplemented with 10 mM glucose and INS-1 cells maintained in RPMI supplemented with 1 mM, 10 mM and 20 mM of glucose respectively, were analyzed by RT-PCR with PC primers and β-actin primers (internal control). D and E, rUTR D and rUTR E, respectively.

When translated with reticulocyte lysate, each mRNA transcribed from the above constructs yielded a single protein band, of molecular weight approximately 51 kDa, as predicted (Fig. 7.9 B). The amounts of the 51 kDa band being translated from rUTR C and rUTR E were comparable. However, rUTR D yielded the same protein band but with a 5-6-fold less efficiency than rUTR C and rUTR E.

A polysome profile analysis of different PC mRNA isoforms of 7-day pup livers is shown in Fig. 7.10. rUTR C was distributed at the bottom to the middle of the gradient, associated with polysome fractions (lanes 1 to 7), suggesting that this transcript is translationally active. Likewise, rUTR E was mostly found in the bottom to the middle of the gradient (lanes 3 to 7) but with less abundancy than rUTR C. In contrast, rUTR D is found primarily in the upper portions of gradient, indicating a lower degree of translation for this transcript. It is therefore likely that most PC protein will be synthesised from rUTR C which is the predominant form being produced during the immediate postnatal gluconeogenic period. The release of ribosomes from the mRNAs by EDTA treatment caused the all transcripts to be found near the top of gradient (Fig. 7.10, lanes 11 and 12).

Figure 7.9 5'-UTR of rat PC transcripts affect translation efficiency *in vitro*.

A, the synthetic RNAs corresponding to rUTR C, D and E transcripts, respectively used for the *in vitro* translation assay with reticulocyte lysates, are represented schematically as follows: Blue represents open reading frame (ORF) encoding for the biotin carboxylation (BC) domain common to all transcripts, the beige region represents the common 3'-UTR and the poly(A) tail downstream of the BC domain. The stop codon is indicated by an asterisk. Different 5'-UTRs are indicated by different colours. **B**, the above transcripts were subjected to *in vitro* translation with rabbit reticulocyte lysates. The labeled peptides being synthesized were analyzed by SDS-PAGE and autoradiographed. M, ^{14}C labeled markers; C, D and E, rUTR C, rUTR D and rUTR E, respectively.

A
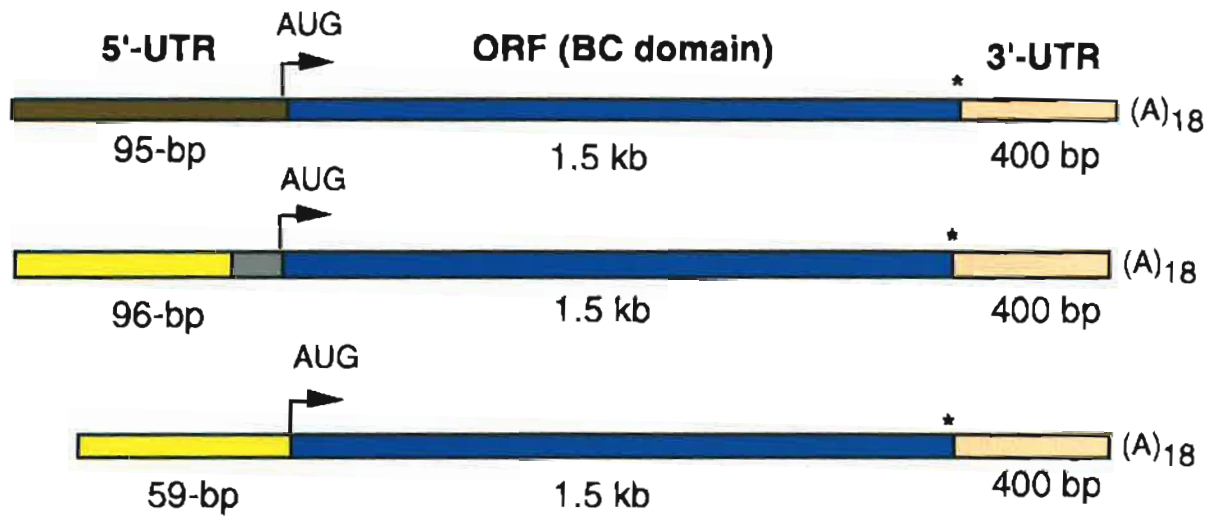
B
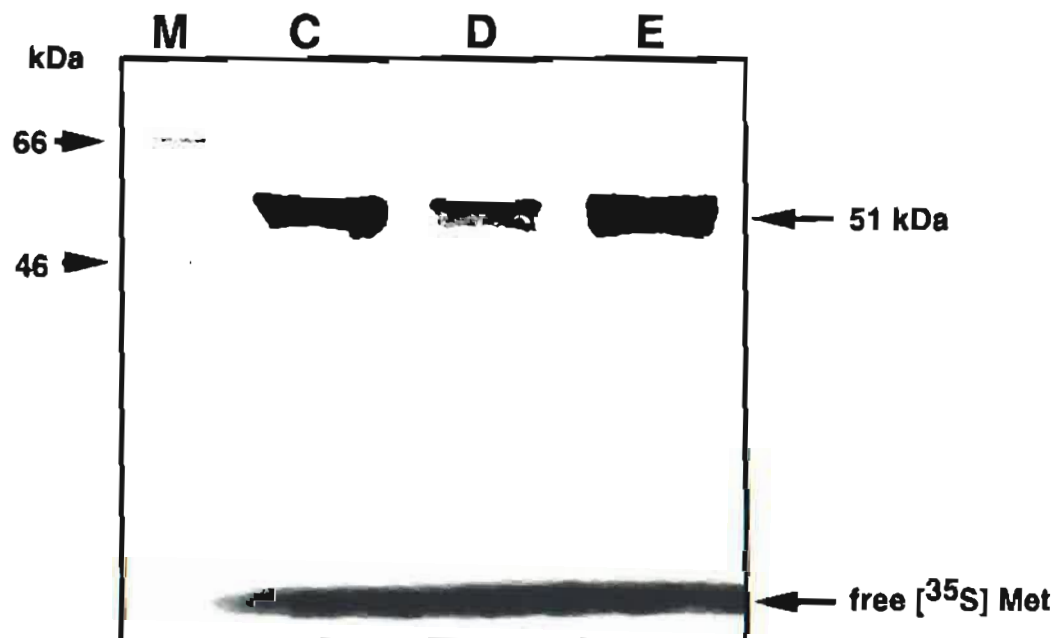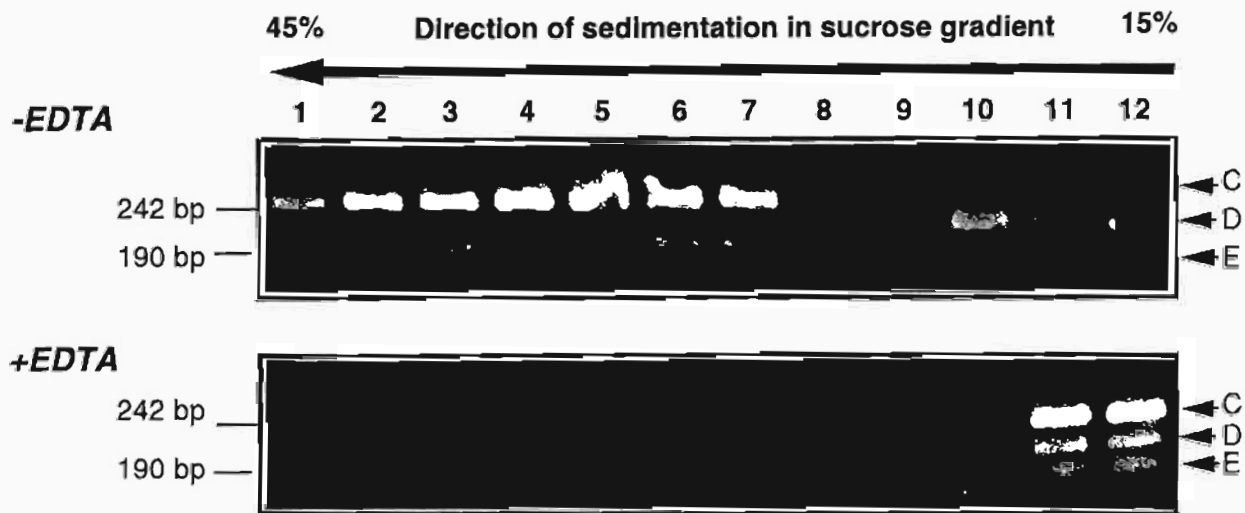
Figure 7.10 Polysome profile analysis of different PC mRNAs in livers. The polysomes were isolated from 7-day pup livers by sucrose density gradient centrifugation in the absence or presence of EDTA. 1 ml fractions were collected manually from the bottom (45% sucrose) to the top (15% sucrose) of the tube (lanes 1 to 12). RNA from each gradient fraction was subjected to RT-PCR using PC primers. Lane 1 represents the bottom of gradient; Lane 12 represents the top of the gradient (dissociated polysomes). C, D and E, for rUTR C, rUTR D and rUTR E respectively.

7.4 DISCUSSION

It is well known that gluconeogenesis is markedly increased in the liver during fasting and diabetes (Pilkis *et al.*, 1988). The suckling period constitutes another physiological situation associated with an active gluconeogenesis in liver and kidney (Girard *et al.*, 1992). Gluconeogenesis does not begin in the livers of mammals until immediately after birth (Ballard and Oliver, 1965); the maternal circulation provides glucose for the developing fetus. Of the four gluconeogenic enzymes present in liver (PC, PEPCK, fructose-1,6-bisphosphatase and glucose-6-phosphatase), PC and PEPCK are present in the liver at negligible levels before birth but appear rapidly after birth (Ballard and Hanson, 1967; Yeung *et al.*, 1967) consistent with the onset of gluconeogenesis. The increase in PC activity has also been observed in livers of the developing guinea pig, rabbit, pig, dog, cow, sheep and human (Barritt, 1985). Regulation of PC gene expression during postnatal development is a complicated process, involving switching of two promoters. rUTR C generated from the proximal promoter, is the major form which is increased during the suckling period. This is concomitant with an increased amount of PC and its activity. During this period it has been shown that gluconeogenesis occurred at the highest rate (Ballard and Oliver, 1965). This suggests that the proximal promoter of PC gene is activated to supply the demand of cells producing PC to fully participate in gluconeogenesis during such a period. However, during the weaning period and in adults when the gluconeogenic rate has decreased, the proximal promoter activity was decreased and transcript C declined. Therefore, it can be concluded that changes in the level of PC protein, which appear to be principally due to alterations in transcription of the gene constitute a long-term mechanism for the regulation of gluconeogenesis during postnatal development.

It is interesting to note that the mechanism which results in an increase in PEPCK activity during development is very similar to that of PC. PEPCK mRNA begins to accumulate in fetal rat livers on the day before birth and rapidly rises in concentration during the first day after birth (Garcia-Ruiz *et al.*, 1978; Bevenisty and Reshef, 1987; McGrane *et al.*, 1990). This increase in PEPCK mRNA level is due mainly to the enhanced rate of gene transcription (Loose *et al.*, 1986; McGrane *et al.*, 1990). The rate of PEPCK synthesis in the livers of rats during the perinatal period is slower than the synthesis of PEPCK mRNA,

suggesting that translational regulation plays an important role in the expression of PEPCK in the liver at birth (van Roon *et al.*, 1988). A change in the chromatin structure of the PEPCK gene during development has been suggested to initiate expression of PEPCK. The PEPCK gene appears to be in a compact conformation which is resistant to nuclease digestion (a feature of an inactive genes), in the fetal liver of 19 day old rats but turns to a relaxed structure which is susceptible to nuclease just prior to the onset of PEPCK transcription at birth (Bevenisty and Reshef, 1987).

Recent studies have also shown that glucose-6-phosphatase is developmentally regulated. The mRNA and the activity of glucose-6-phosphatase levels were already present in 21-day-old rat fetal livers but markedly increased during the suckling period and declined during the weaning period (Chatelain *et al.*, 1998). It is noteworthy that the level of glucose-6-phosphatase mRNA had begun to increase at the same time i.e. both mRNA and activity reached the maximum level at day 1. However, the level of mRNA rapidly declined onwards before increasing again during the weaning period, while the enzymatic activity continued to increase until day 5 and then slowly decreased, reaching a plateau during the weaning period (Chatelain *et al.*, 1998). The expression pattern of glucose-6-phosphatase during development seems to be different from that of PC in which PC mRNA, protein and activity showed the same trend throughout postnatal development, suggesting the long half life of PC mRNAs. Nevertheless, a cooperative increase in PC, PEPCK and glucose-6-phosphatase activities in a similar fashion i.e. a marked increase just after birth, a further increase during the suckling period and a decline during the weaning period, could perhaps ensure the effective operation of gluconeogenesis during postnatal development. The hormonal mechanism which mediates the increase in PC, PEPCK and glucose-6-phosphatase in liver at birth is not clearly defined. At least one of the following events occurs within the first hours after birth that is likely to regulate expression of these 3 gluconeogenic enzymes: 1) a rapid and marked fall of blood glucose which was supplied to the fetus by maternal circulation and a low portal glucose concentration caused by a low carbohydrate intake. This results in an increase of cAMP level via elevated levels of glucagon and epinephrine and a low level of insulin, 2) an increased supply of plasma fatty

acids to the liver due to the hydrolysis of milk triglycerides (Girard *et al.*, 1973; Girard *et al.*, 1992).

In genetic obesity, PC expression is increased accompanied by increased expression of other lipogenic enzymes (Lynch *et al.*, 1992). Although adipose tissue is the major lipogenic tissue, liver is also another organ where lipogenesis occurs (Ballard and Hanson, 1970). I have demonstrated that both PC transcripts and protein were also increased in livers of 12 weeks old obese rats but not in 8 weeks old obese rats. However, the effects were not as marked as in adipose tissue. Thus, it is likely that lipogenesis in the liver of the obese rats was not as active in week 8 as in week 12. Lynch *et al.* (1992) also could not detect any differences in PC between the obese and lean rats at 8 weeks of age. rUTR C is the major transcript which is increased in both liver and adipose tissue of the obese rats, suggesting that the proximal promoter is transcriptionally activated during the lipogenic period. The presence of a putative fat-specific element 1 (FSE-1) on the proximal promoter (Chapter 6) suggests that overexpression of PC might also be mediated through this motif, as occurs with the fatty acid synthetase gene (Hsu *et al.*, 1996). Insulin selectively inhibits the expression of the proximal promoter of rat PC. However, the activity of this promoter is increased in the adipose tissue of the obese Zucker rat (Fig. 7.8A) even though blood insulin levels are known to be dramatically increased in this model of diabetes (Clark *et al.*, 1983; Polonsky, 1995). Activation of the proximal promoter of PC, rather than its inhibition is another manifestation of the insulin resistance of the obese Zucker rat. The role of PC in lipogenesis has also been reported in mouse adipocytes. Transcriptional regulation of PC was first reported during differentiation of mouse 3T3-L1 fibroblasts to mature adipocytes (Angus and Lane, 1981; Freytag and Collier, 1984; Zhang *et al.*, 1995). This increase in PC mRNA during such a period is concomitant with parallel changes in the level of PC protein (Weinberg and Utter, 1979; Weinberg and Utter, 1980) and activity (Mackall and Lane, 1977). Whether the mouse PC gene contains an adipose-specific promoter, as in rat, is still unknown. However, the addition of cAMP to cultures of 3T3-L1 adipocytes caused a marked decrease in PC mRNA (Zhang *et al.*, 1995) and PC activity (Freytag and Utter, 1980; Zhang *et al.*, 1995) but not PC protein suggesting that cAMP regulates PC by

decreasing transcription of the PC gene and or enhancing degradation rate of PC mRNA (Zhang *et al.*, 1995).

Glucose homeostasis is regulated in the endocrine pancreas through opposite effects of glucose on insulin and glucagon secretion. Pancreatic β-cells are fuel sensors that adjust the rate of insulin secretion to the rate at which they metabolise glucose (Newgard and McGarry, 1995; Matchinsky, 1996). The major transduction pathway involves utilisation of ATP-sensitive K^+ channels of the plasma membrane to transduce biochemical into biophysical signals. Thus, glucose metabolism causes closure of these K^+-ATP channels, which results in membrane polarization, opening of voltage-dependent Ca^{2+} channels, and acceleration of Ca^{2+} influx. The resulting rise in cytoplasmic Ca^{2+} concentration triggers exocytosis of insulin granules (Ashcroft *et al.*, 1984). An anaplerotic role of PC in rat pancreatic islets (MacDonald, 1995a; Schuit *et al.*, 1997) and INS-1 cells (Brun *et al.*, 1996) has recently been proposed to be important for glucose-induced insulin secretion due to the abundant expression of this enzyme in these cells. Higher physiological concentrations of glucose rapidly stimulate insulin secretion by the islets and also cause an increase in PC expression in the islet (MacDonald, 1995a; MacDonald *et al.*, 1991). The high level of PC permits the rapid formation of oxaloacetate which serves as a substrate for the pyruvate/malate shuttle across the inner mitochondrial membrane thus providing a source for the generation of cytosolic NADPH by means of the malic enzyme (MacDonald, 1995b). Rapid entering of pyruvate to the TCA cycle in pancreatic β-cells also provides a large amount of ATP. It has been suggested that the rise in ATP/ADP ratio both in human and rat pancreatic β-cells may serve as a secondary messenger to trigger exocytosis of insulin granules (Detimary *et al.*, 1998). In this study, only transcripts D and E were detected in pancreatic islets and INS-1 cells, suggesting that the distal promoter of the rat PC gene is responsible for anaplerosis in these cells. Glucose also increased PC transcription from the distal promoter when INS-1 cells were maintained in 10 mM glucose. The same concentration of glucose was also reported to enhance the PC transcription rate in pancreatic islets when assayed using the nuclear run-on technique (MacDonald, 1995a). In the promoters of other glucose-induced genes, such as L-type pyruvate kinase and S14 genes, *cis*-acting element namely c-*myc*-like binding sequence, has been shown to mediate glucose

induction (Carthew *et al.*, 1985; Sladek *et al.*, 1990). It will be very interesting to determine whether the distal promoter contains glucose-responsive element(s). In the description of the nucleotide sequence of the distal promoter in Chapter 6, two copies of the motif, CACGTG, resembling a c-*myc*-like binding sequence were identified. Whether these motifs confer glucose responsiveness functional analysis of this promoter remains to be elucidated.
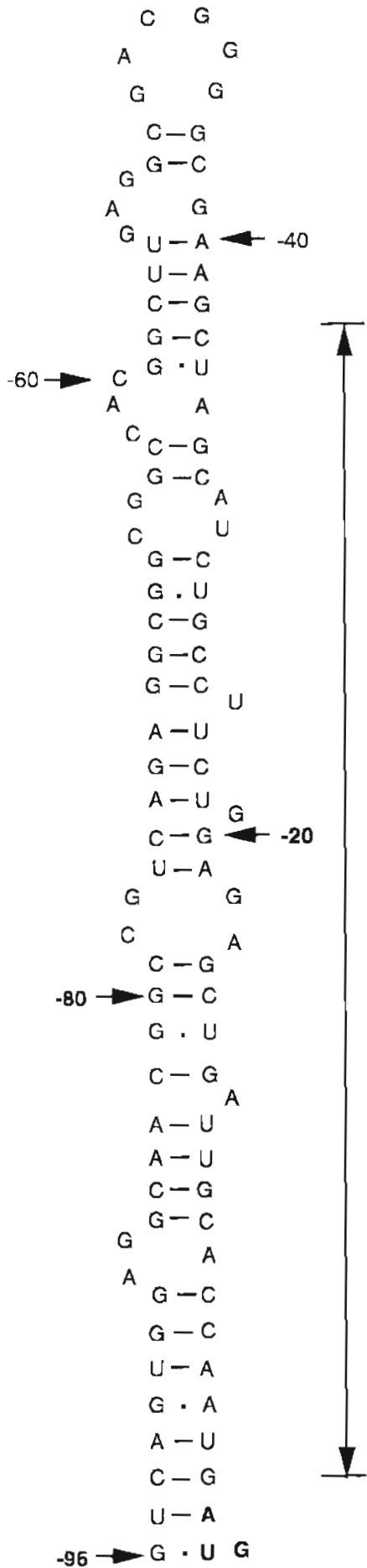
Hormonal regulation of endogenous PEPCK (Sasaki *et al.*, 1984; Sasaki and Granner, 1988) and glucose-6-phosphatase (Lange *et al.*, 1994; Argaud *et al.*, 1996; Chatelain *et al.*, 1998) gene expression have been extensively studied in rat by using hepatocytes as a model, since there are less variable factors compared with animals . However, this has proven to be impossible with PC because of the difference in gene expression pattern observed between intact liver and hepatocytes. The absence of the rUTR C transcript, produced from the glucogenic promoter in every liver cell line examined indicates that under these cultured conditions, the regulation of PC expression is different from *in vivo* conditions. This finding suggests that the environment in the culture media may not be equivalent to that found in intact liver. Since the proximal promoter is turned on during gluconeogenesis *in vivo*, this finding is consistent with the observation that under cell culture conditions, gluconeogenesis occurs only in primary cultures of hepatocytes and perfused rat livers (Yamada *et al*, 1980; Exton and Park, 1967). This perhaps explains the results obtained from the transient transfection experiment of the proximal and the distal promoter in liver and kidney cells (HepG2 and COS-1) in Chapter 6, i.e. the distal promoter drives the expression of the reporter gene at a much higher level than the proximal promoter under basal conditions. Those cell lines are likely to undergo dedifferentiation under cell culture conditions and perhaps have lost tissue-specific transcription factors that may be required to activate transcription of the proximal promoter.

It is apparent that the proximal promoter of the rat PC gene is inducible during gluconeogenesis and lipogenesis, whereas the distal promoter is important for anaplerosis. During the suckling period when gluconeogenesis is accelerated, the proximal promoter is activated, resulting in an increase in the amount of rUTR C. *In vitro* translation and *in vivo* polysome profile analysis clearly showed that the translation efficiency of rUTR C transcript is higher than that of rUTR D transcript, and hence would result in more PC protein being
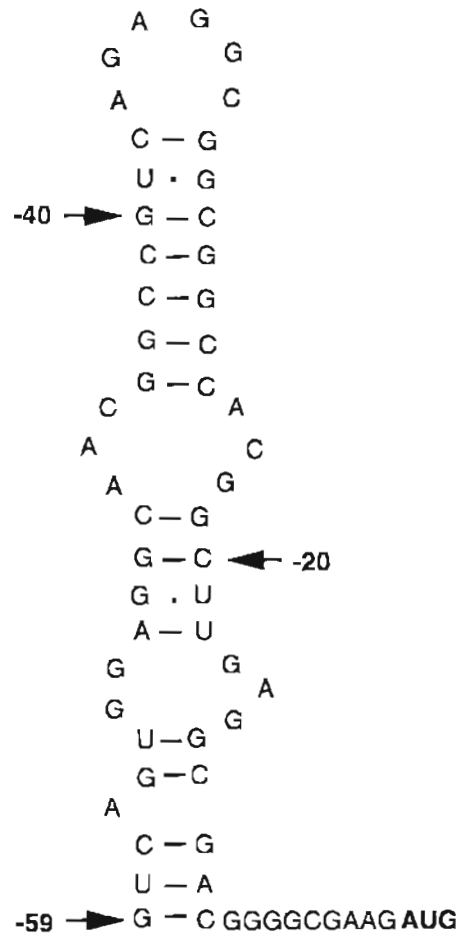
produced during this period. Activation of transcription from the proximal promoter instead of the distal promoter of PC gene might be a key mechanism to allow an independent regulation. Studies described in Chapter 6 demonstrated that both promoters contain different putative transcription factor binding sites and responded to insulin differently. Since the multiple PC transcripts only differ in their 5'-UTRs, the different translational efficiencies of these mRNA isoforms must be mediated through their 5'-UTRs.

All known eukaryotic cellular mRNAs undergoes a series of interactions with various proteins soon after nascent transcripts have been synthesised. Early events of great significance for the stability and function of the transcript, are the capping of the 5'-terminal nucleotide and the polyadenylation of the 3'-end generated by cleavage at predetermined sites. Translation initiation occurs when cap-binding proteins bind to the 5' cap structure of the mRNA and facilitate the assembly of the 40S initiation complex which in turn migrates along the 5'-UTR until reaching the first initiation codon. The elongation process can then proceed (Kozak, 1989). According to this model, a very stable secondary structure at the 5'-UTR is potentially inhibitory to translation, presumably because they could obstruct cap-binding by the 40S initiation complex and the movement of the scanning process along the 5'-UTR (Kozak, 1991; Kozak, 1992). Structures of low stability are apparently readily accessed and do not have any effect on the translation of the mRNA. In contrast, a stable structure in the coding region does not inhibit the progress of elongating 80S ribosomes (Kozak, 1989). Computer-assisted analysis of secondary structure prediction, revealed that the 5'-UTR of rUTR D can potentially form a stable hairpin structure due mainly to pairing between sequences derived from exon 1C and 1D. However, removal the sequence derived from exon 1C sequence, which is naturally spliced out in rUTR E, abolishes this structure (see Fig. 7. 11). Whether the secondary structure of rUTR D contributes to low translation efficiency must await *in vitro* mapping data and phylogenetic studies. Although the results obtained from *in vitro* translation and polysome analysis suggest that sequences at the 5'-UTRs of different PC mRNAs modulate translation, *trans*-acting factors found in the cytoplasm cannot be ruled out. It has been proposed that a *trans*-acting factor(s) involved in the translational regulation of ferritin (Theil *et al.*, 1990), aconitase (Schalinske *et al.*, 1998) and 5-aminolevulinate synthase-2 (ALAS-2) (Cox *et al.*, 1991) mRNAs, known as the iron-

Figure 7.11 Predicted secondary structures of the 5'-ends of rUTR D and rUTR E of rat PC. The Zucker and Stiegler (1981) procedure was used to construct the predicted secondary structure of 5'-UTR of rUTR D and rUTR E of rat PC mRNAs. The 37 bp sequence derived from exon 1C is indicated and the AUG start codon is in bold.
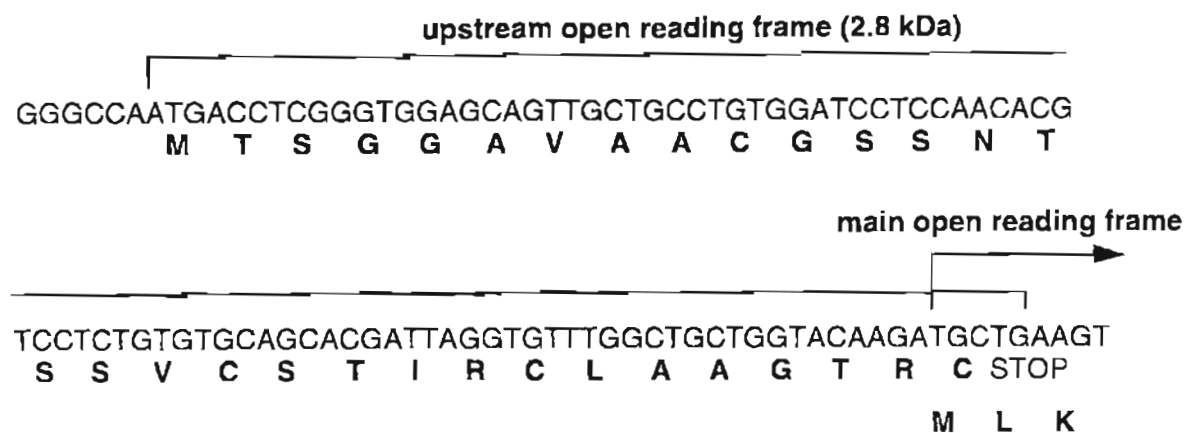
```
                  C   G
               A       G
               G       G
                  C — G
               G  G — C
               A       G
               G  U — A        ◄—— -40
                  U — A
                  C — G
                  G — C
-60 ——►    C      G · U
           A           A
            C  C — G
               G — C
            G       A
            C       U
               G — C
               G · U
               C — G
               G — C
               G — C  U                                    37 bp exon 1C
               A — U
               G — C
               A — U
               C — G  G        ◄—— -20
               U — A
            G       G
            C       A
               C — G
-80 ——►        G — C
               G · U
            C — G
               A
               C — G
               A — U
               A — U
               C — G
               G — C
            G       A
            A  G — C
               G — C
               U — A
               G · A
               A — U
               C — G
               U — A
-96 ——►        G · U  G
```

rUTR D

ΔG = -33.5 kJ/mol

```
                A   G
             G       G
             A       C
                C — G
                U · G
-40 ——►         G — C
                C — G
                C — G
                G — C
             C  G — C  A
            A        C
            A    G
                C — G
                G — C        ◄—— -20
                G · U
                A — U
            G       G   A
            G       G
               U — G
               G — C
            A
               C — G
               U — A
-59 ——►        G — C GGGGCGAAG AUG
```

rUTR E

ΔG = -16.0 kJ/mol

responsive element binding protein, carries out its negative effect not only by binding to specific sequences located at the 5'-end of the transcript but also by stabilising the 5' secondary structure with consequent inhibition of ribosome entry. It is also possible that putative *trans*-acting factor(s) present both in rabbit reticulocyte lysate and in liver cells, specifically interact with specific sequences located in the 5'-UTRs of different PC mRNAs, resulting in different translation efficiencies.

The efficiency of mRNA translation also depends on the local sequence context surrounding the initiation codon. $^{-3}$(A/G)CCAUG^{+4}G emerges as the consensus sequence for translation initiation (Kozak, 1987). It is interesting to point out that such a sequence is apparently absent in any of rUTR C, D or E ($^{-3}$AAGAUG^{+4}C, $^{-3}$AUGAUG^{+4}C and $^{-3}$AAGAUG^{+4}C respectively), but translation initiation occurs precisely at the authentic AUG codon. This suggests that translation initiation can occur through other motifs which may be less optimal for translation initiation. It is known that the presence of an AUG located upstream of the authentic AUG initiation codon in the 5'-UTR is another factor that can also affect the translation efficiency of the main open reading frame of an mRNA (Kozak, 1987). Of particular interest, rUTR C and rUTR E contain upstream AUG codons which could be used to initiate translation. Translation from an upstream AUG (position -89) of rUTR C would however produce a very short peptide of mass 2.8 kDa (see Fig. 7.12). However, no such evidence was observed with the *in vitro* translation experiment, suggesting that the sequence surrounding the upstream AUG may not be effective. Alternatively, that short peptide may not be stable upon translation. Many mammalian leaders bearing upstream open reading frames have been reported, but little is known about their functional significance (Kozak, 1991). In instances where the two open reading frames overlap, it has been shown that ribosomes terminate at the end of the upstream open reading frame and then migrate backward to the initiation codon of the downstream open reading frame where they reinitiate translation at the authentic codon (Gunnery *et al.*, 1996). In contrast, the upstream AUG found in rUTR D (**AUGAUGC**) occurs inframe and is located immediately prior to the authentic codon. Translation initiation occuring from this upstream AUG would add an extra methionine at the N-terminal of the enzyme.

upstream open reading frame (2.8 kDa)

GGGCCAATGACCTCGGGTGGAGCAGTTGCTGCCTGTGGATCCTCCAACACG
　　　　M　T　S　G　G　A　V　A　A　C　G　S　S　N　T

main open reading frame

TCCTCTGTGTGCAGCACGATTAGGTGTTTGGCTGCTGGTACAAGATGCTGAAGT
S　S　V　C　S　T　I　R　C　L　A　A　G　T　R　C STOP
　　　　　　　　　　　　　　　　　　　　　　　　　　M　L　K

Figure 7.12 Schematic diagram of translation from two open reading frames (upstream and main reading frames) of rUTR C transcript.

Post-transcriptional regulation has been reported to play an important role in controlling gene expression of many genes. This mode of regulation occurs in the cytoplasm of eukaryotic cells at the level of mRNA stability, subcellular localisation of mRNA and translation (Kozak, 1991).

Translational regulation which is mediated through the 5'-UTR resulting from alternate splicing has also been reported in one of the glycolytic enzymes, enolase, which catalyses the interconversion of 2-phosphoglycerate to phosphoenolpyruvate. Two mRNA isoforms of enolase each bearing different 5'-leader sequences, have been reported in human (Giallongo *et al.*, 1993), mouse and rat (Oliva *et al.*, 1995) muscle. Two human muscle enolase mRNAs, differing at their 5'-UTRs, have been shown to be translated with different efficiencies (Oliva *et al.*, 1995). Another example is a related biotin-containing enzyme, ACC, which plays an important role in fatty acid synthesis and shares many common regulatory mechanisms with PC. For example, multiple ACC mRNA isoforms having the same coding sequence but differing in their 5'-UTRs, have been identified in rat (Lopez-Casillas *et al.*, 1989, Lopez-Casillas and Kim, 1989). These arise by differential splicing at the 5'-end of transcripts from the single rat ACC gene which is alternatively transcribed from 2 distinct promoters (Luo *et al.*, 1989) that are highly regulated under different physiological conditions (Lopez-Casillas *et al.*, 1991, Zhang and Kim, 1995). Recent studies have also proposed that ACC is an important enzyme providing malonyl-CoA as a coupling factor necessary for glucose-induced insulin secretion in pancreatic islets (Brun *et al.*, 1996; Schuit

et al., 1997). As with PC, only the housekeeping promoter of the ACC gene is functional and activated in pancreatic islets upon glucose-induced insulin secretion (Zhang and Kim, 1995).

CHAPTER 8

EXPRESSION AND PURIFICATION OF

RECOMBINANT HUMAN PC

8.1 INTRODUCTION

In our laboratory, cDNA encoding PC from several species including yeast, rat and human are readily available to express as a recombinant form. Yeast PC has already been expressed for structure/function studies (Nezic, M.G., personal communication). However, human enzyme rather than the rat enzyme, would also prove to be valuable if expressed in a recombinant form for the following reasons.

PC deficiency in humans is an autosomal recessively inherited disease, causing a severe to mild lactic acidemia and resulting in early death or severe psychomotor disturbance (Robinson *et al.*, 1996). Despite this significant importance, human PC has not received appreciable attention. This is due mainly to the limitation of source material derived from biopsy or autopsy samples. Although several human liver and kidney cell lines are available, the expression of PC in these cell lines is extremely low (Chandler and Ballard, 1985). Dr M.E. Walker had previously isolated partial cDNA clones encoding human PC (Walker *et al.*, 1995b) and thus has made it possible to express this enzyme by recombinant means to replace native material. Earlier attempts in this laboratory to produce yeast PC in *E. coli* failed to produce functional enzyme, suggesting the lack of adequate folding machinery in bacteria. Given the availability of various mammalian expression vectors, it should be possible to produce the recombinant PC in mammalian cell lines. In this chapter, a full length human PC cDNA clone has been assembled and cloned into different mammalian expression vectors followed by transfecting into mammalian cells. Stably transfected clones expressing very high levels of recombinant human PC were successfully obtained. This recombinant material was subsequently purified for further characterisation.

Constitutive and inducible mammalian expression systems

Two major expression systems ie. the constitutive and inducible expression plasmids were initially examined to see which system can achieve the higher expression level. In some cases where overexpression of some proteins can be toxic to cells, the use of an inducible expression system can provide an alternative means to control the levels of toxic proteins.

The inducible expression system that was used in this work was the Tet-On Gene Expression System (Clontech). This system was first described by Bujard, Gossen and colleagues (Gossen and Bujard, 1992; Gossen *et al.*, 1996). The Tet-On gene expression system allows high-level, regulated gene expression in response to varying concentrations of tetracycline or its derivative, doxycycline (Dox). The Tet Expression Systems are based on two regulatory elements derived from the tetracycline resistance operon of the *E. coli* Tn10 transposon. The "regulatory plasmid", pTRE expresses a fusion protein of the mutated tetracycline repressor protein (rTetR) and VP16 activation domain of herpes simplex virus under the control of the cytomegaloviral promoter (P_{CMV}). Integration of this plasmid into the host genome results in the constitutive expression of rTetR-VP16 fusion protein. In the absence of tetracycline in the media, transfection of this cell line with the "response plasmid" which contains the gene of interest cloned downstream of a compound promoter consisting of the tetracycline-responsive element (TRE) and the minimal immediate early promoter of cytomegalovirus cause no expression of gene of interest due to the mutation in the TetR protein. However, in the presence of tetracycline in the culture media, tetracycline binds to the mutated TetR protein which becomes capable of binding to the TRE element. This binding brings the VP16 activation domain near the P_{CMV} and thereby activates transcription of the gene of interest (Fig. 8.1). The levels of expression can also be controlled by adding different concentrations of tetracycline in the media.

The constitutive expression systems that were used are comprised of two different bicistronic expression vectors. One was the pIRES1*neo* from Clontech and another was the pEFIRES-puro developed by Dr. S.M. Hobbs (Hobbs *et al.*, 1998) from the Cooperative Research Centre for Cancer Therapeutics, Institute of Cancer Research, London, U.K Both plasmids work on the same principle i.e. the gene of interest can be cloned at the multiple cloning site located upstream from the antibiotic resistance gene. Transcription from a single promoter located upstream of the gene of interest produces a bicistronic mRNA containing two open reading frames of two proteins (Fig. 8.2). The presence of an internal ribosome entry site allows the translation of two open reading frames from a single bicistronic mRNA (Jang *et al.*, 1988; Jackson *et al.*, 1990). This system minimises the number of colonies to be screened to find functional clones.
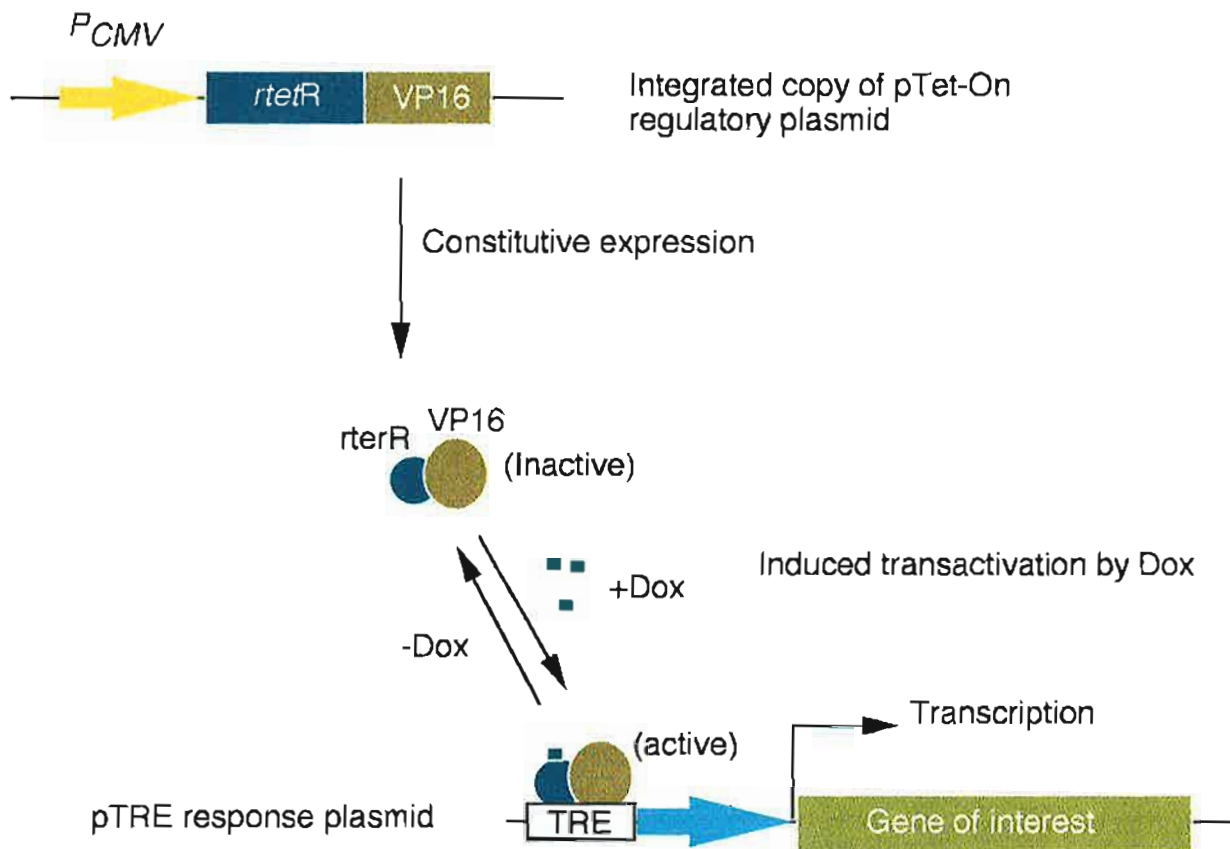
Figure 8.1 The Tet-On expression system (Clontech). The tet-controlled transcriptional activator (tTA) is a fusion of mutated tetracycline repressor (rTetR) to the VP16 activation domain (AD) of herpes simplex virus. This expression cassette is driven by a CMV promoter and is constitutively expressed. In the absence of tetracycline, this complex cannot bind to the tetracycline-responsive element (TRE) in the pTRE response plasmid. However, in the presence of tetracycline derivative (Dox, doxycycline), the tTA becomes active and thereby transactivates transcription of any genes that are cloned upstream of the TRE. Transcription of the gene of interest can be shut down by removing Dox from the culture medium.
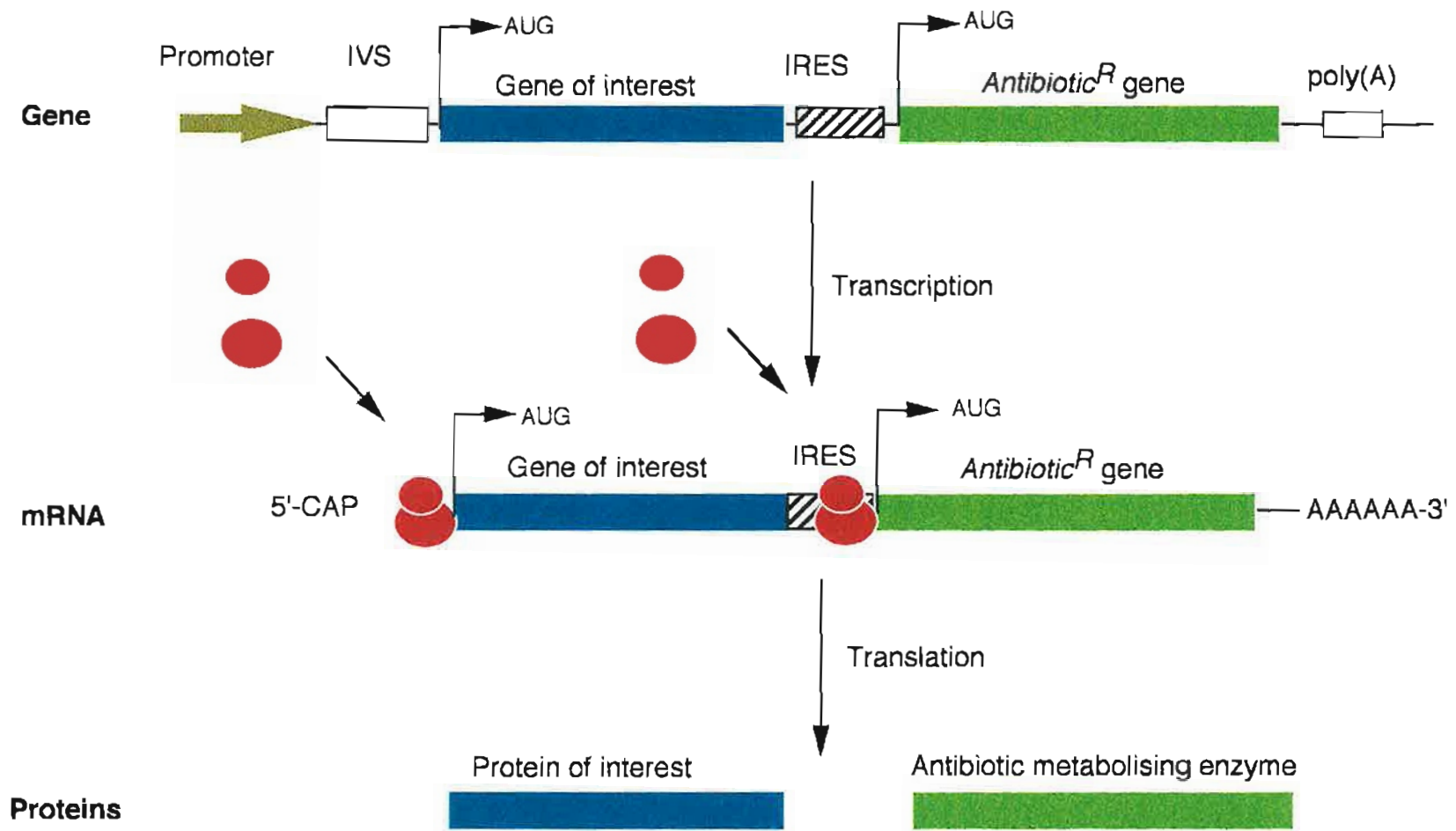
Figure 8.2 Schematic diagram of the translation of a bicistronic mRNA. The gene of interest is cloned upstream of the antibiotic resistance gene under the control of a single promoter. The internal ribosome entry site (IRES) permits a protein of interest and antibiotic selection marker to be translated from the same mRNA. IVS, synthetic intron.

8.2 SPECIFIC METHODS

8.2.1 Transient and stable transfection of 293T cells

Human embryonic kidney cells, 293T or 293T Tet-On were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and 10 μg/ml gentamycin. Cells were grown to 80-90% confluence in 175 cm² flasks at 37°C in a humidified atmosphere of 5% CO_2. Cells were trypsinized and transfected with 50 μg of plasmids (pBI-PC, pIRES-PC or pEF-PC) and electroporated as previously described in Chapter 6. Transfected cells were recovered in DMEM supplemented with 10% fetal calf serum at 37°C for 48 h before the lysates were prepared and analysed by SDS-PAGE and Western blot. For the Tet-On inducible system (pBI), after transfection, cells were recovered in DMEM media plus 10% fetal calf serum for 24 h. Different concentrations of doxycycline were then added in the same medium and further incubated at 37°C for 24 h before the expression level was monitored by SDS-PAGE and Western analyses.

For the generation of stable clones, transfected cells were plated on 10 cm² petri dishes and initially selected by adding 1 μg/ml puromycin to the culture medium for 1-2 weeks. Higher expressing clones were selected by gradually increased the concentrations of puromycin to 2, 5, 10, 20, 30 and 50 μg/ml and maintaining for 2-3 weeks. Pooled clones that resisted high concentrations of puromycin were expanded and analysed. Cell lysates were prepared and freeze-dried as previously described in Chapter 7 and analysed by SDS-PAGE and Western blot analysis to monitor the expression level.

8.2.2 Preparation of PC from mitochondrial extracts

Stably transfected clones expressing human PC were routinely grown in the same culture media as described above but supplemented with 0.5 mg/l d-biotin and 10 μg/ml puromycin. Cells were washed twice with phosphate buffer and centrifuged at 1,300 rpm for 5 min. The cell pellet was then suspended in 4 volumes of 0.25 M sucrose, 0.1 mM EDTA, 0.5 mM PMSF and 1 mM DTT, and homogenised at 4°C for 2-3 min. The unbroken cells and nuclei were removed by centrifugation at 600 x g at 4°C for 20 min. The supernatant containing mitochondria was further centrifuged at 15,000 x g at 4°C for 30. The pellet was then washed once with 4 ml of 0.1 mM EDTA pH 7.0 containing 0.5 mM PMSF and centrifuged at 15,000 x g for 30 min at 4°C. The mitochondrial pellet was

suspended in 2 ml 0.1 mM EDTA pH 7.0 and freezed-dried overnight. Dried mitochondria can be kept at -20°C or used immediately by dissolving in 5 ml of 50 mM Tris-acetate pH 7.0, 5 mM ATP pH 7.0, 5 mM $MgCl_2$ and 0.5 mM EDTA and stirred at 4°C for 30 min. Undissolved materials were removed by centrifugation at 20,000 x g for 10 min. PC was precipitated from the supernatant by adding 100% saturated ammonium sulfate solution pH 7.0 to a final concentration of 40% saturation. The precipitate was then dissolved in 0.5 ml buffer A (25 mM potassium phosphate buffer 7.2, 20 mM ammonium sulfate, 1 mM EDTA and 0.1 mM DTT) and excess residual ammonium sulfate removed by passing the solution through a 2 ml Sephadex G-25 column (Pharmacia, Upjohn) equilibrated in buffer A. The desalted sample in buffer A was further purified by mononeric avidin-Sepharose chromatography.

8.2.3 Monomeric avidin-Sepharose chromatography

Desalted samples (1-5 mg of total protein) were applied at a flow rate of 0.2 ml/min onto a 8 cm x 0.5 cm monomeric avidin column (kind gift from Dr. P. Attwood, Department of Biochemistry, University of Western Australia) which had been equilibrated with running buffer (25 mM potassium phosphate buffer 7.2, 0.2 M potassium chloride, 1 mM EDTA and 0.1 mM DTT). The column was washed with several column volumes of running buffer until no more proteins were eluted off. PC was then eluted with 1 mg/ml d-biotin in buffer A. The biotin elution solution was left to interact with avidin-bound PC on the column for 1 h before elution was continued at a flow rate of 0.2 ml/min. Eluted PC was subsequently precipitated with an equal volume of saturated ammonium sulphate, and stored in 1.6 M sucrose-0.1 M N-ethylmorpholine-HCl, pH 7.2. Eluted PC was analysed by SDS-PAGE.

8.2.4 PC assay

PC activity was assayed by a spectrophotometric method previously described by Scrutton and White (1974) with some modifications. Briefly, the assay was carried out at 37°C for 10 min in 1 ml volume containing 100 mM Hepes (K^+) pH 7.8, 2 mM Na_2ATP, 3 mM $MgCl_2$, 5 mM sodium pyruvate, 100 mM $KHCO_3$, 0.1 mM acetyl-CoA and 0.15 mM NADH and 2 units of malate dehydrogenase (10 μg/unit). PC activity was determined by monitoring the decrease in the A_{340} of NADH. Since 1 mole of NADH is

oxidized per mole of pyruvate carboxylated, the rate of reaction for a 1 cm light path and 1 ml reaction volume is calculated as:

decrease in A340/min/6.22 µmoles/min

where 6.22×10^3 is the millimolar extinction coefficient for NADH at 340 nm (Dawson *et al.*, 1969).
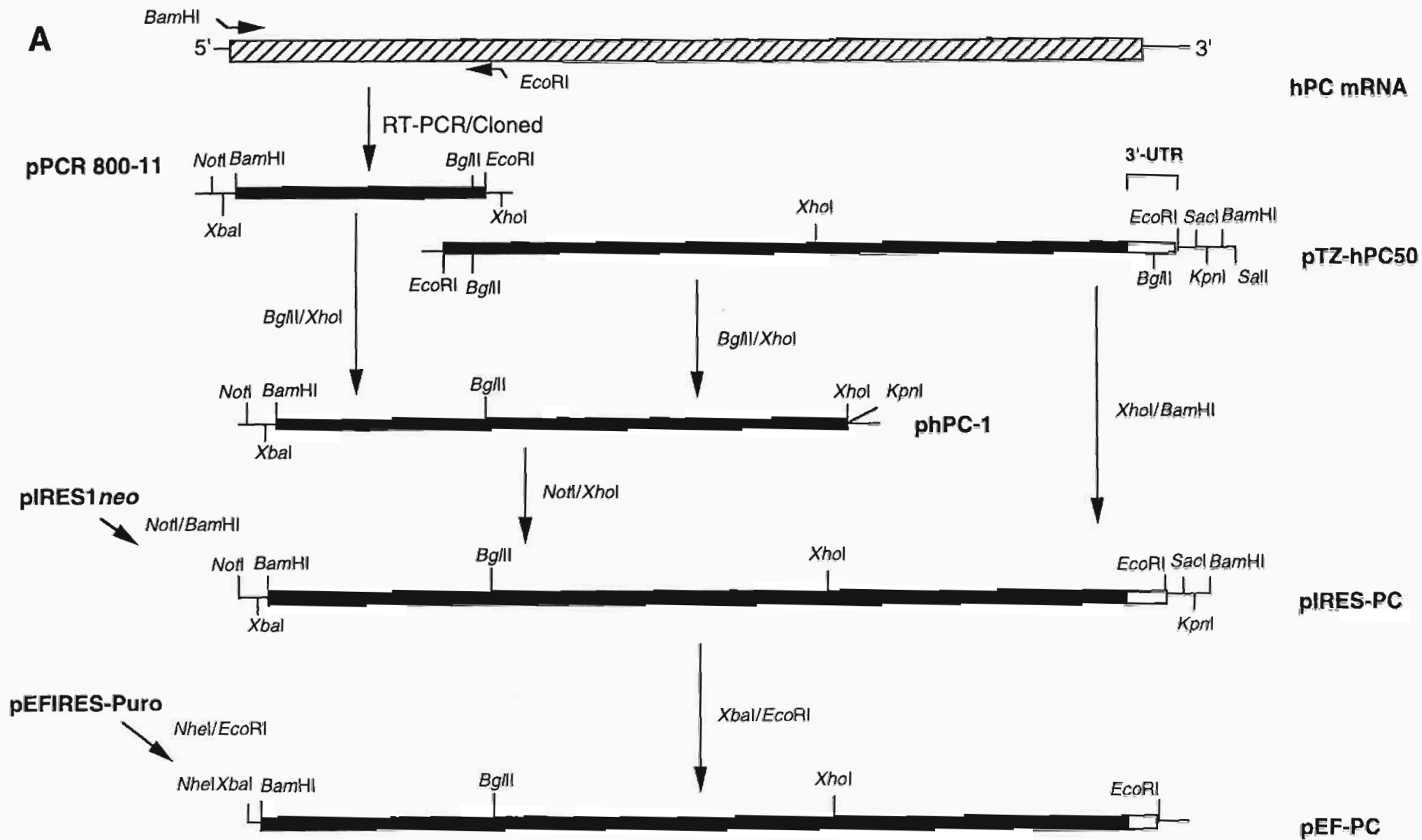
8.3 RESULTS

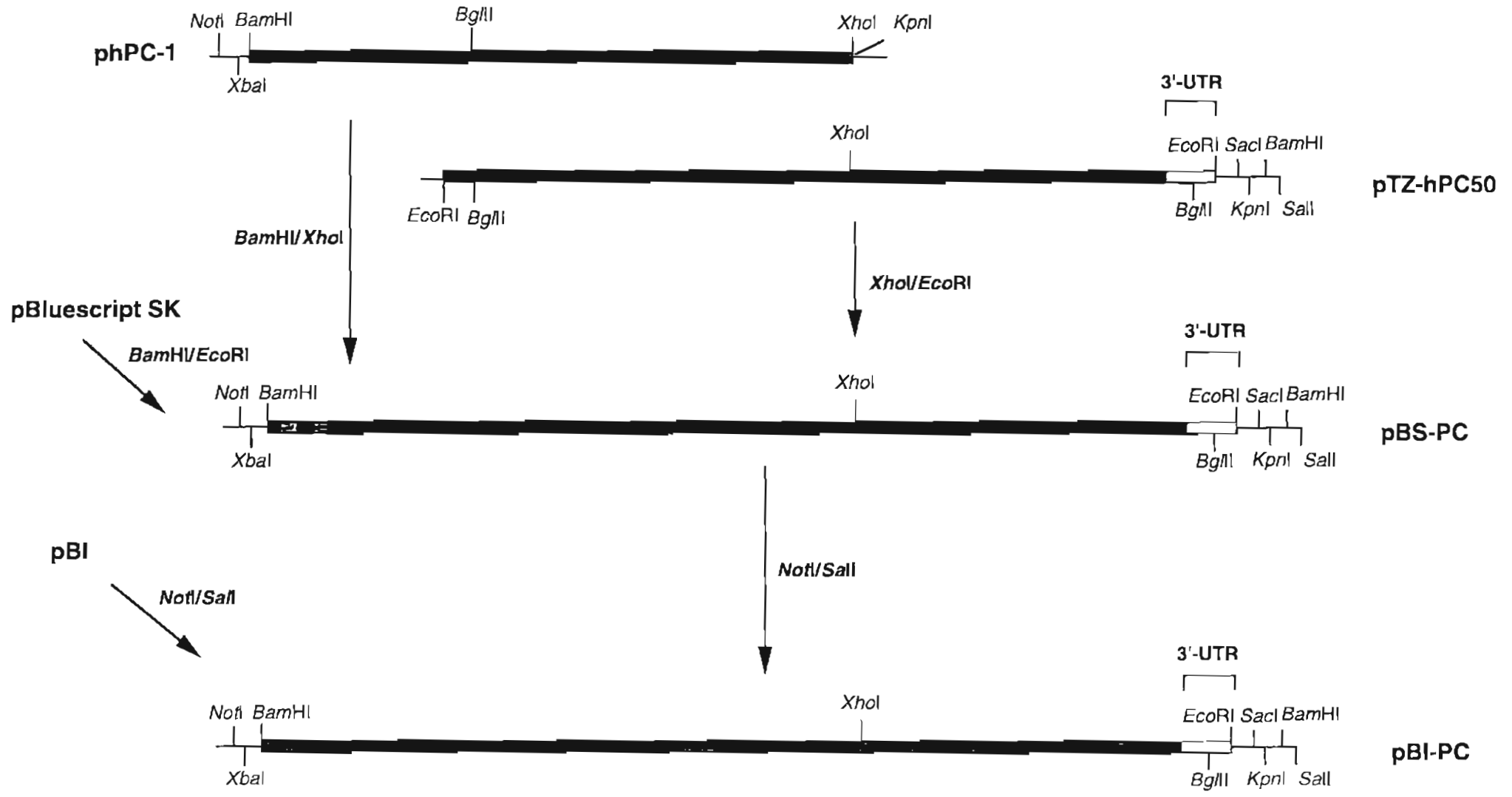### *8.3.1 Generation of a full length human PC clone*

The cDNA encoding the first 270 N-terminal residues of the human PC was generated by RT-PCR using HUM05 (5'-GGCG<u>GAATTC</u>GGTGCAGGA-TGTTCCCATACTGG-3') primer with a *Eco*RI restriction site attached at the 5'-end, underlined) as a cDNA synthesis primer followed by PCR using the same cDNA synthesis primer (HUM05) and HUM09 primer (5'-GGT<u>GGATCC</u>AGT**TTGACCAAAC-ACTAAGGATGC**-3', bold, start codon) with a *Bam*HI restriction site attached at the 5'-end, underlined) (Walker *et al.*, 1995b). These primer binding sites relative to the human PC mRNA is shown in Fig.8.3 A. The 0.8 kb PCR product was purified, digested with *Eco*RI and *Bam*HI and cloned into pBluescript SK yielding the plasmid phPCR 800-11. Sequencing was then performed to ensure no error was introduced by *Taq* polymerase. The presence of a *Bgl*II site at the 3'-end of pTZ-hPC 50 made it not possible to excise an insert from pTZ-hPC50 as a single fragment by digesting with *Bgl*II (at 5'-end) and *Eco*RI (3'-end). Therefore, the 5'-end of pTZ-hPC 50 was first inserted at the 3'-end of pPCR 800-11 by excision with *Bgl*II and *Xho*I, and this yielded phPC-5 plasmid which contains half of the coding region. The second step was carried out by digesting phPC-5 with *Not*I and *Xho*I and pTZ-hPC 50 with *Xho*I and *Bam*HI. These two fragments were then cloned into the *Not*I-*Bam*HI cloning site of pIRES1*neo* plasmid, and this yielded pIRES-PC which contained the full length of human PC cDNA (Fig. 8.4). A whole PC cDNA insert was also subcloned from pIRES1*neo* with *Xba*I and *Eco*RI, and cloned into *Nhe*I (produces compatible end with *Xba*I) and *Eco*RI of pEFIRES-puro plasmid (Fig. 8.5)
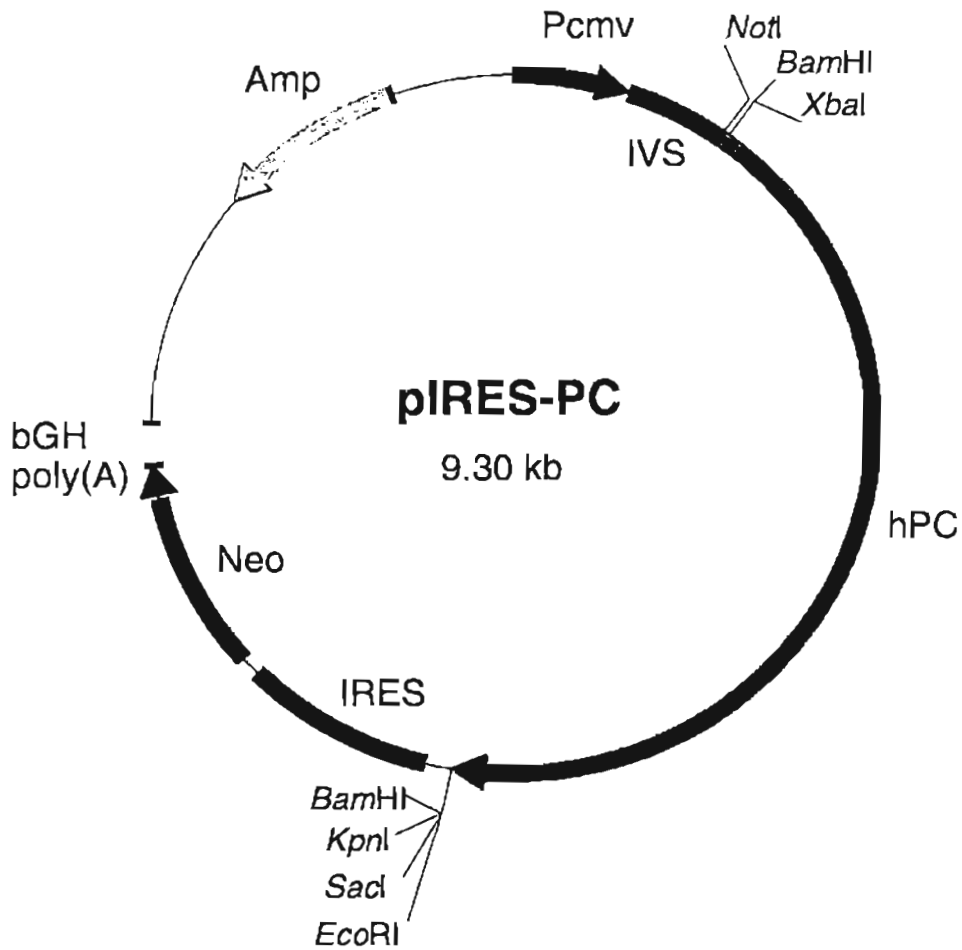
To incorporate the full length human PC cDNA into a tetracycline-inducible plasmid (pBI), the full length human PC cDNA was initially cloned into pBluescript to provide the polylinker regions which are compatible with the cloning sites in the pBI plasmid. This was carried out by ligating a *Bam*HI-*Xho*I fragment from phPC-5 and a *Xho*I-*Eco*RI fragment from pTZ-hPC50 and ligating them into the *Bam*HI-*Eco*RI-digested pBluescript, yielding plasmid pBS-PC (Fig. 8.3B). Full length human PC cDNA was then excised from pBS-PC by digestion with *Not*I and *Sal*I and cloned to *Not*I-*Sal*I-digested pBI, yielding the plasmid pBI-PC (the map of this final construct is shown in Fig. 8.6).

Figure 8.3 Schematic diagram of the construction of full length human PC cDNA in different mammalian expression vectors. A, RT-PCR was used to generate the 5'-end of cDNA encoding human PC cDNA. This PCR product was then sequenced and ligated to pTZ-hPC50 which contains most of the coding region of human PC by two step cloning. pIRES-PC and pEF-PC were the final constructs which contain full length human PC in two different bicistronic vectors. B, generation of full length human PC cDNA cloned into the Tet-on expression plasmid (pBI) to yield the pBI-PC construct.

A

BamHI
5' ▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨ 3'   **hPC mRNA**
EcoRI

↓ RT-PCR/Cloned

pPCR 800-11

NotI BamHI   BglII EcoRI
|————————————————|
XbaI   XhoI

3'-UTR
                              EcoRI SacI BamHI
pTZ-hPC50
                    Xhol
EcoRI BglII   BglII KpnI SalI

BglII/XhoI   BglII/XhoI   Xhol/BamHI

NotI BamHI   BglII   Xhol KpnI
|—————————————————————————| **phPC-1**
XbaI

↓ NotI/XhoI

pIRES1neo
  ↘ NotI/BamHI

NotI BamHI   BglII   Xhol   EcoRI SacI BamHI
|————————————————————————————————| **pIRES-PC**
XbaI   KpnI

↓ XbaI/EcoRI

pEFIRES-Puro
  ↘ NheI/EcoRI

NheI XbaI BamHI   BglII   Xhol   EcoRI
|————————————————————————————————| **pEF-PC**

B

phPC-1
*Not*I *Bam*HI   *Bgl*II   *Xho*I *Kpn*I
*Xba*I

3'-UTR
*Xho*I
*Eco*RI *Sac*I *Bam*HI
pTZ-hPC50
*Eco*RI *Bgl*II   *Bgl*II *Kpn*I *Sal*I

*Bam*HI/*Xho*I
*Xho*I/*Eco*RI

pBluescript SK
*Bam*HI/*Eco*RI

*Not*I *Bam*HI   *Xho*I   3'-UTR
*Eco*RI *Sac*I *Bam*HI
pBS-PC
*Xba*I   *Bgl*II *Kpn*I *Sal*I

pBI
*Not*I/*Sal*I   *Not*I/*Sal*I

*Not*I *Bam*HI   *Xho*I   3'-UTR
*Eco*RI *Sac*I *Bam*HI
pBI-PC
*Xba*I   *Bgl*II *Kpn*I *Sal*I

Figure 8.4   Map of pIRES-PC expression plasmid. Full length human PC cDNA (hPC) was inserted at *Not*I and *Bam*HI sites of pIRES1*neo* plasmid. The insert can be isolated from the vector by digesting with *Bam*HI. IRES, internal ribosome entry site; *Neo*, neomycin phosphotransferase gene; bGH poly(A), bovine growth hormone polyadenylation signal; Amp, ampicillin resistance gene; Pcmv, cytomegalovirus promoter; IVS, synthetic intron that enhances stability of the mRNA (Huang and Gorman, 1990).
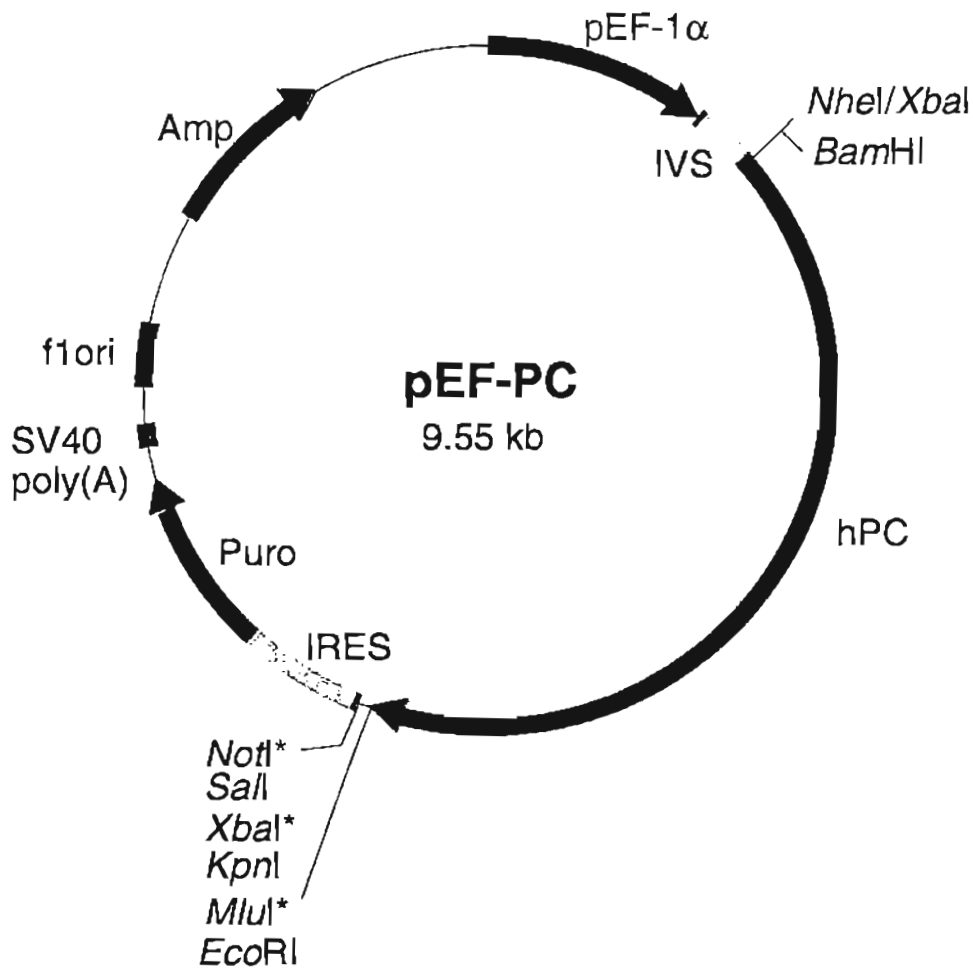
Figure 8.5   Map of pEF-PC expression plasmid. Full length human PC cDNA (hPC) was inserted at *Nhe*I and *Eco*RI sites of pEFIRES-Puro (Hobbs *et al.*, 1998). The insert can be isolated from the vector by digesting with *Bam*HI and *Eco*RI.   IRES, internal ribosome entry site; *Puro*, puromycin acetyl transferase gene; SV40 poly(A), SV40 polyadenylation signal; Amp, ampicillin resistance gene; Pcmv, cytomegalovirus promoter; IVS, synthetic intron that enhances stability of the mRNA (Huang and Gorman, 1990). Asterisks indicate unique multiple cloning sites present in the pEFIRES-puro. f1ori, allows production of single stranded DNA for mutagenesis.
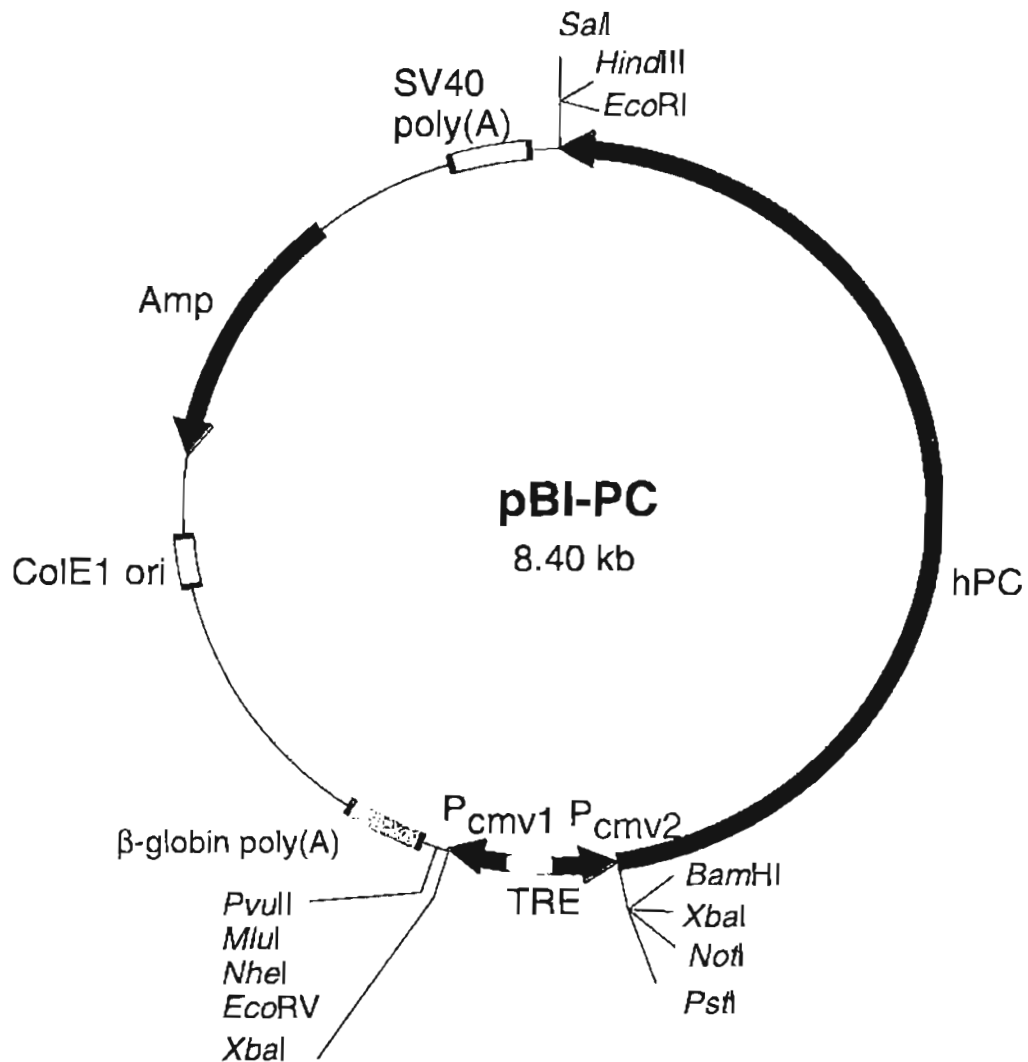
Figure 8.6   Map of pBI-PC expression plasmid. Full length human PC cDNA (hPC) was inserted at *Not*I and *Sal*I sites of pBI under the control of cytomegaloviral promoter (*Pcmv2*). The insert can be isolated from the vector by digesting with *Bam*HI. TRE, tetracycline-responsive element; SV40 poly(A), SV40 polyadenylation signal. The presence of CMV promoter on the opposite strand (*Pcmv1*) allows other genes to be cloned downstream of this second promoter. β-globin poly(A), β -globin polyadenylation signal.

8.3.2 Transient expression of human PC from different expression vectors

8.3.2.1 Tetracycline-inducible expression vector

pBI-PC, "response plasmids" were transfected into the 293T cell line stably transfected with a "regulatory plasmid" by electroporation. Transfectants were induced by adding different concentrations of doxycycline (0.5, 1.0, 2.0 and 5.0 μg/ml) in the culture media for 48 h before the proteins were prepared and analysed by SDS-PAGE and Western analysis. As shown in Fig. 8.7, the expression level of PC derived from transfection of 293T with pBI-PC was higher than that of non-transfected 293T cells. A slight increase in the level of PC expression was seen when the transfectants were grown in higher concentration of doxycycline. However, these transfectants only expressed PC at a level about 3-4-fold higher than non-transfected 293T cells. Two additional bands which reacted to avidin were identified on the basis of their molecular weight as other two endogenous biotin enzymes, propionyl-CoA carboxylase (PCC) [79 kDa] and methylcrotonyl-CoA carboxylase (MCC) [75 kDa] (Wood and Barden, 1977; Lau *et al.*, 1980). The presence of these PCC and MCC bands in other human cell lines (Chandler and Ballard, 1985) and in rat pancreatic islets (MacDonald, 1995a) have also been reported.

8.3.2.2 Bicistronic vectors

pEF-PC and pIRES-PC were separately transfected into 293T cells. Transient expression of PC was monitored by SDS-PAGE and Western analysis upon incubating the transfectants in the culture media for 48 h. As indicated in Fig. 8.8, the expression level of PC derived from transfection of 293T cells with pEF-PC was about 2-3-fold higher than those transfected with pIRES-PC, as judged by Western analysis from two independent transfection experiments. The level of PC produced by 293T cells transfected with pEF-PC was about 10-fold higher than the non-transfected 293T cells.

8.3.3 Expression of hPC from stably transfected 293T cells

Generation of stable cell lines expressing recombinant proteins is superior to transient expression i.e. proteins are continuously produced over time as the expression plasmid is stably integrated into the host chromosome(s). The results obtained from transient
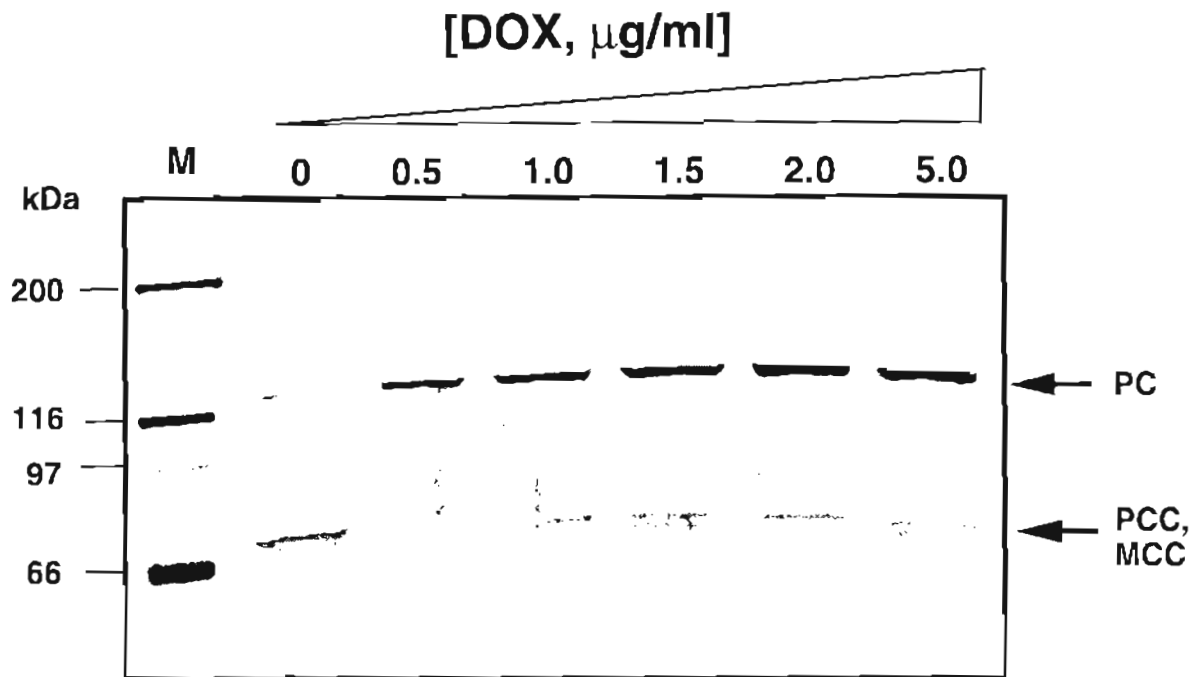
Figure 8.7 Transient expression of 293T Tet-On cells transfected with pBI-PC plasmids. Soluble proteins were extracted from cells and analysed by SDS-PAGE followed by Western analysis using avidin-conjugated with alkaline phosphatase as probe. PC expression can be induced by adding different concentrations of doxycycline (DOX) in the culture media for 48 h after transfection. PC, pyruvate carboxylase; PCC, propionyl-CoA carboxylase; MCC, methyl crotonyl-CoA carboxylase, M, biotinylated markers.
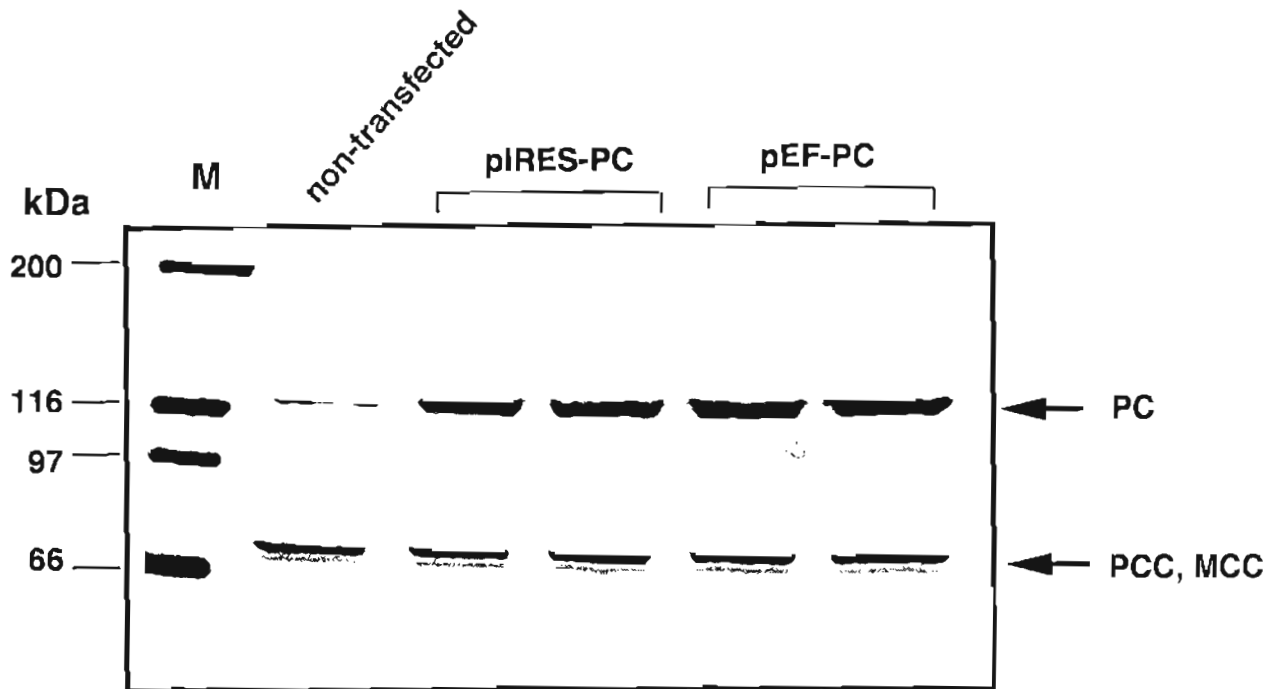
Figure 8.8   Transient expression of 293T cells transfected with either pIRES-PC or pEF-PC plasmids for 48 h. Soluble proteins were extracted from cells and analysed by SDS-PAGE followed by Western analysis using avidin-conjugated with alkaline phosphatase as probe. PC, pyruvate carboxylase; PCC, propionyl-CoA carboxylase; MCC, methyl crotonyl-CoA carboxylase, M, biotinylated markers. Results from wo independent transfection experiments are shown.

expression described in section 6.3.2, have demonstrated that the pEF-PC construct resulted in expression of recombinant human PC at higher levels than the pIRES-PC and pBI-PC constructs. For this reason, pEF-PC was used to transfect 293T cells to create a stable cell line that constitutively expresses recombinant human PC.

Four identical transfection experiments were carried out as described in methods. After transfection they were transferred into four petri dishes (plates 1, 2, 3 and 4) and transfectants were initially selected by adding 1 μg/ml of puromycin in the culture media. After being cultured in this media for 3 weeks, more than 50% of cells had died off, and resistant colonies had started to form. At this stage, two different strategies were used to select highly expressing clones by (i) isolating pooled clones and (ii) isolating pure clones. Since hPC and PAC proteins are cotranscribed on the same mRNA, any clones that are able to grow in high concentrations of puromycin added to the media should also express high levels of hPC.

i) **For isolating pooled clones** (plates 1 and 2), puromycin was successively increased to 2, 5, 10, 20, 30, 50 and 100 μg/ml in culture media. Resistant colonies were routinely maintained in such concentrations of puromycin for 5 days before the next concentration was added. This allows any colonies that express a higher level of PAC to be expanded while any colonies that poorly expressed the puromycin resistance gene will die off. In the meantime, the levels of PC expression from these pooled clones were also monitored by trypsinizing 1/4 of them and analysing by SDS-PAGE/Western analysis. As shown in Fig. 8.9A, pooled clones derived from plate 1 expressed more PC when they were maintained in higher concentrations of puromycin (2, 5 and 10 μg/ml respectively). In contrast, pooled clones derived from plate 2 did not show this trend i.e. they expressed PC at the same level in these three concentrations of puromycin. It is interesting to note that at this stage, other endogenous biotin carboxylases, i.e. PCC and MCC, were not detected in these pooled clones. Instead, another two minor bands which were slightly smaller than PCC and MCC were observed. These two bands were proportional to overexpressed PC bands and were not present in non-transfected cells suggesting that they were partial breakdown products of PC and also contain biotin moiety. Densitometer scanning analysis indicated that this breakdown product was negligible (less that 5% of the 120 kDa PC band).
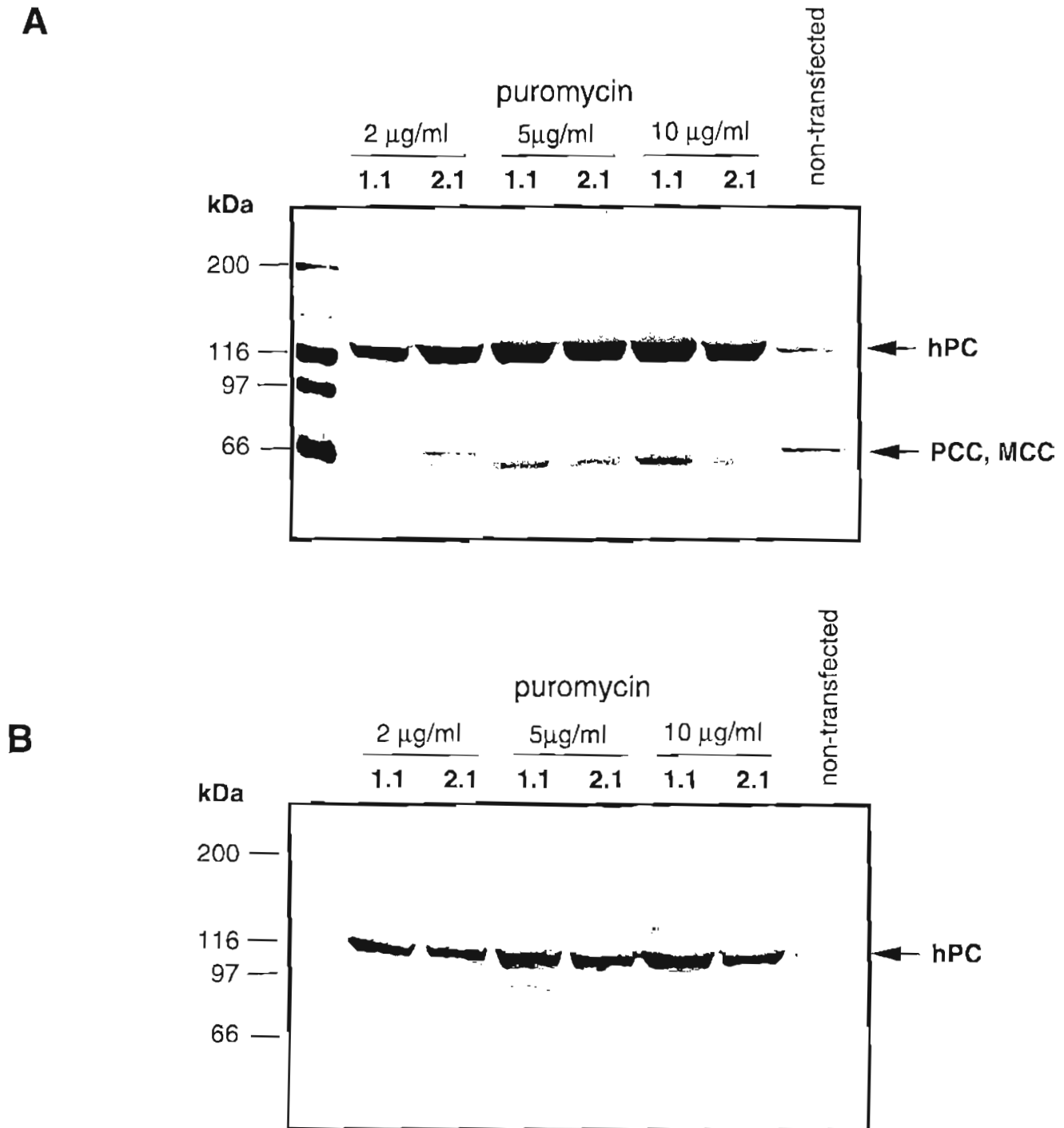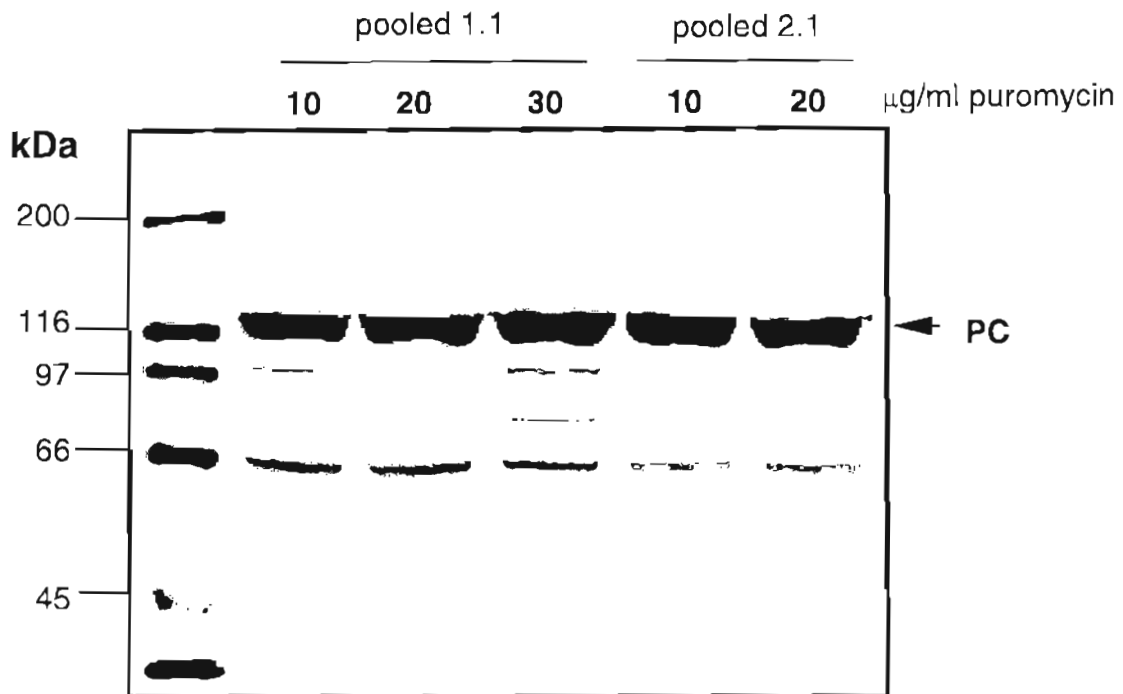
Figure 8.9   Western analysis of pooled clones stably transfected with pEF-PC plasmid. Transfectants were grown in the medium containing 1 µg/ml puromycin. Selection of highly expressing PC clones were carried out by gradually increase the concentrations of puromycin in the media (2, 5 and 10 µg/ml). Soluble proteins from these pooled clones were extracted and analysed by SDS-PAGE followed by Western analysis with (A) avidin-alkaline phosphatase and with (B) chicken PC polyclonal antibodies . PC, pyruvate carboxylase; PCC, propionyl-CoA carboxylase; MCC, methyl crotonyl-CoA carboxylase, M, biotinylated markers.

Western analysis with anti-chicken PC antibodies (1: 20,000 dilution) revealed strong cross reactivity between human PC and the antibodies suggesting chicken PC antibodies recognise similar epitopes on the human enzyme (see Fig. 8.9B). The anti-chicken PC antibodies also reacted with two other faint bands, one was about 97 kDa which is also present in non-transfected cells while another one was about 130-140 kDa. However, the latter band was not present in non-transfected cells and it perhaps represents the unprocessed precursor (uncleaved mitochondrial targetting sequence at the N-terminal) of recombinant human PC due to its overexpression in the cells. At higher concentrations of puromycin (20 and 30 μg/ml), only slightly increased levels of PC expression were observed (Fig. 8.10). In contrast, expression levels of PC detected from pooled clones derived from plate 2 (10 and 20 μg/ml puromycin) remained unchanged (Fig. 8.10). However both cell lines were not able to grow at higher concentrations of puromycin (50 and 100 μg/ml) suggesting a toxic effect of the drug.

Since pooled clones derived from plate 1 expressed the highest level of hPC, these clones were expanded and passaged. To test whether PC expression levels could be regulated by the concentration of puromycin, these cell lines were grown in different concentrations of puromycin from 0 to 30 μg/ml for 7 days. As shown in Fig. 8.11, different concentrations of puromycin did not induce PC expression, as similar levels of PC were detected when the cells were grown even in the absence of puromycin. Constitutive expression of PC in these cell lines is rather due to the presence of a high copy number of transgenes integrated into the chromosome of 293T cells. It is noteworthy that at this stage in contrast to the situation in non-transfected cells, the PC band has become a prominent species in transfected cells as can be seen in the Coomassie stained gel. It is also apparent that PCC and MCC were again not detectable. This appears to be a feature of the overexpression of PC which competes for biotinylation of endogenous biotin carboxylases. This pattern was still observed even though extra biotin was routinely added to the culture media (2.0 mg/L).

ii) For the isolation of pure colonies, after maintaining transfectants in the medium containing 1 μg/ml puromycin for 3 weeks. Twenty four out of fifty colonies were picked from the plate and transferred to 6-well plates and grown in medium containing 2

Figure 8.10   Western analysis of pooled clones stably transfected with pEF-PC plasmid. Transfectants were grown in the medium containing 10 μg/ml puromycin. Selection of highly expressing PC clones were carried out by gradually increasing the concentrations of puromycin in the media (20 and 30 μg/ml). Soluble proteins from these pooled clones were extracted and analysed by SDS-PAGE followed by Western analysis with avidin-alkaline phosphatase. PC, pyruvate carboxylase; PCC, propionyl-CoA carboxylase; MCC, methyl crotonyl-CoA carboxylase, M, biotinylated markers.
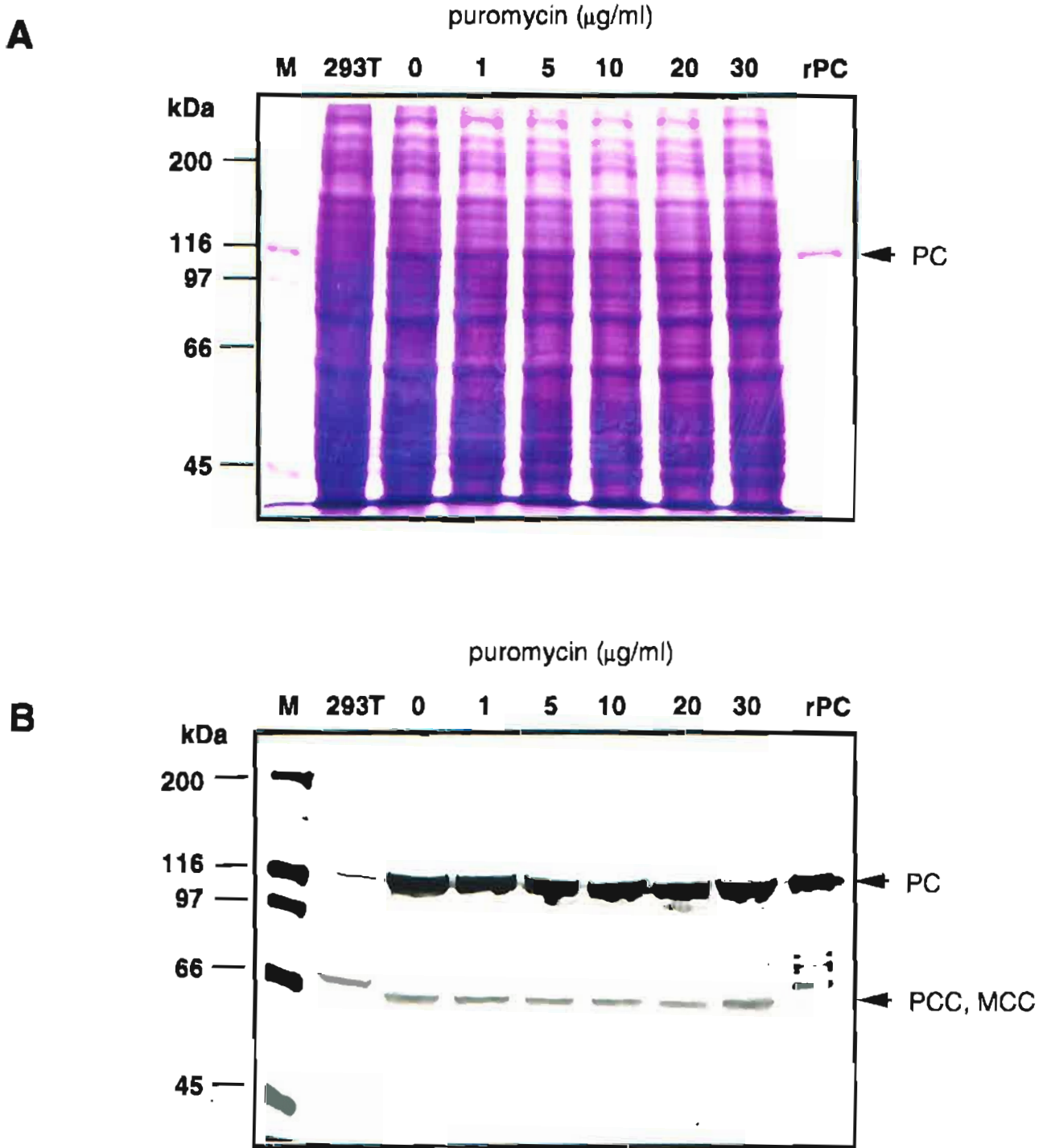
Figure 8.11 Constitutive expression of human PC from pooled clones 1.1. Pooled clones 1.1 were expanded and grown in different concentrations of puromycin (1, 5, 10, 20 and 30 µg/ml). Soluble proteins were extracted and analysed by SDS-PAGE and stained with Coomassie blue (A) and Western analysis probing with avidin-alkaline phosphatase (B). rPC, purified rat PC (Booker, 1990); PCC, propionyl-CoA carboxylase; MCC, methyl crotonyl-CoA carboxylase, M, biotinylated markers.

µg/ml puromycin for 5 days. At this stage, sixteen colonies survived and were then maintained in medium containing 10 µg/ml puromycin. After maintaining these 16 colonies in this medium, only two colonies, clones 12 and 22 survived and had grown rapidly. However, both clones failed to grow in higher concentrations of puromycin (20 and 30 µg/ml). SDS-PAGE and Western analysis indicated that these two clones expressed PC at a similar level to the pooled clones derived from plate 1 (1.1) (Fig. 8.12). The level of PC protein expressed from these clones was about 25-30-fold higher than non-transfected cells as judged by Western blot probing with avidin-alkaline phosphatase.

PC activity assays were performed with crude lysates prepared from these four cell lines (pooled clones 1.1, pooled clones 1.2, pure clones 12 and 22) using the radiochemical method as described in Chapter 7. The results showed that these clones incorporated $^{14}HCO_3$ into malate with about 30- fold higher activity than non-transfected cells in parallel with the amount of PC band detected by Western blot with avidin alkaline phosphatase, suggesting that these four cell lines produced active PC.

Table 8.1 Relative PC activity* of different 293T clones overexpressing PC.

Clone	relative PC activity*
non-transfected 293T	1
1.1	33
2.1	30
12	35
22	27.5

*determined by radiochemical assay

8.3.4 Purification and characterisation of recombinant hPC

Mitochondria were isolated from 20 ml packed volume of 293T cells (80 x 175cm^2 culture flasks with 90-100% confluence) using differential centrifugation. To facilitate subsequent extraction of their matrix proteins, the mitochondria were freeze-dried as described previously. SDS-PAGE analysis indicated that PC (120 kDa) constituted a major
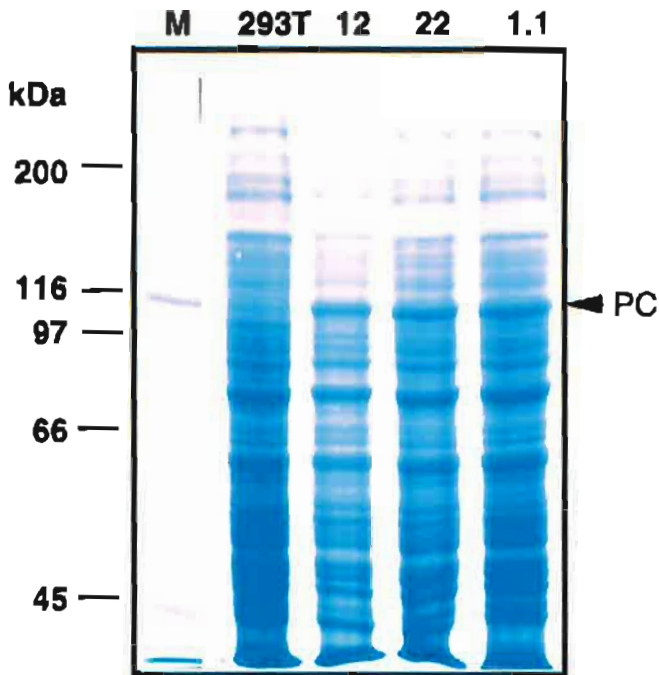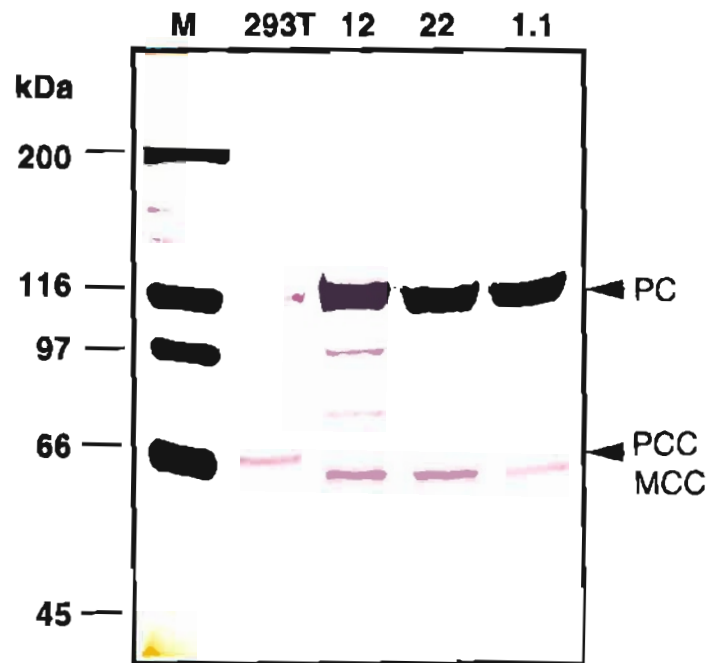
Figure 8.12   Western analysis of single colonies (clones 12 and 22). These two clones were grown in the media containing 10 μg/ml of puromycin. Soluble proteins were extracted and analysed by SDS-PAGE and stained with Coomassie blue (A) and by Western analysis with avidin-alkaline phosphatase (B). PC, pyruvate carboxylase; PCC, propionyl-CoA carboxylase; MCC, methyl crotonyl-CoA carboxylase, M, biotinylated markers.

protein (70% of total protein) found in 40% ammonium sulphate precipitate fraction as shown in Fig. 8.13. The specific activity of PC in this step was about 8-10 U/mg protein. The specific activity obtained from this step is comparable with that reported from partially purifed PC from liver (30% purity) [Scrutton and White, 1974] suggesting that this recombinant enzyme is as active as the native enzyme. Further purification of recombinant hPC was carried out by avidin-Sepharose chromatography. This technique is based on interaction between biotin which is covalently attached near the C-terminal of the enzyme to avidin immobilised on the Sepharose matrix. Since the binding of native avidin to biotin is a very strong interaction, monomeric avidin was used instead to reduce this tight interaction (Henrikson *et al.*, 1979). Five to ten milligrams of the desalted materials obtained after ammonium sulphate precipitation were applied to the column. The A_{280} profile detected after eluting PC with 2 column volumes of d-biotin yielded a single peak as indicated in Fig. 8.14. PC eluted from the column showed a single band on SDS-PAGE with the specific activity of 20 U/mg protein. However, the yield of PC is relatively poor, with only 20% recovery (see table 1).

Table 8.2 Purification of recombinant human PC from pooled clones 1.1 stably transfected with pEF-PC expression plasmid.

Purification step	protein (mg)	specific activity* (U/mg protein)	Yield (%)
mitochondrial extraction 40% $(NH_4)_2SO_4$ precipitation	10	10	100
avidin-sepharose	2	20	20

*determined by spectrophotometric assay

A
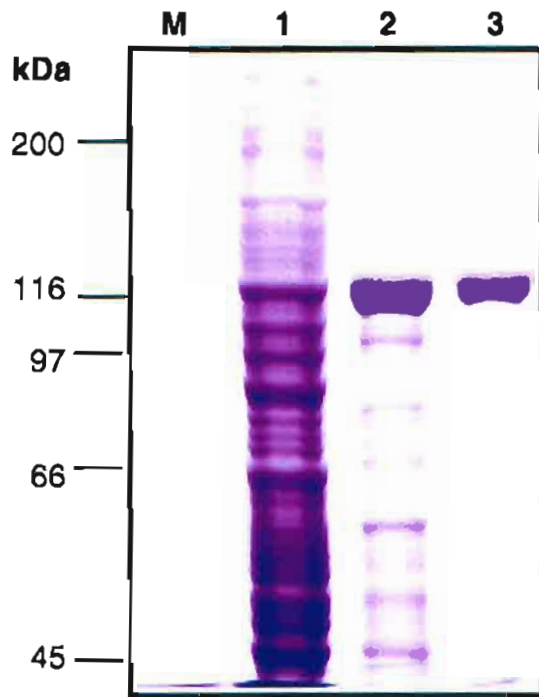
B
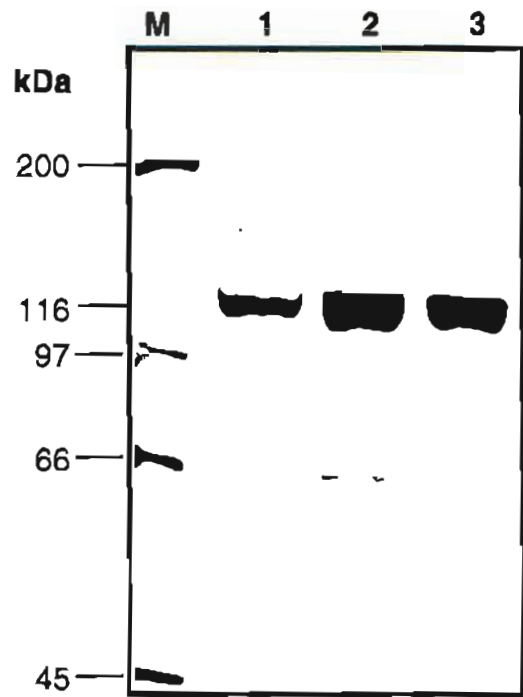
Figure 8.13   SDS-PAGE (A) and Western analysis (B) of purification of recombinant human PC from clones 1.1. M, biotinylated marker; lane 1, soluble material from whole cell lysate; lane 2, 40% ammonium sulfate precipitate of mitochondrial proteins before monomeric avidin affinity chromatography; lane 3, material after monomeric avidin affinity chromatography.
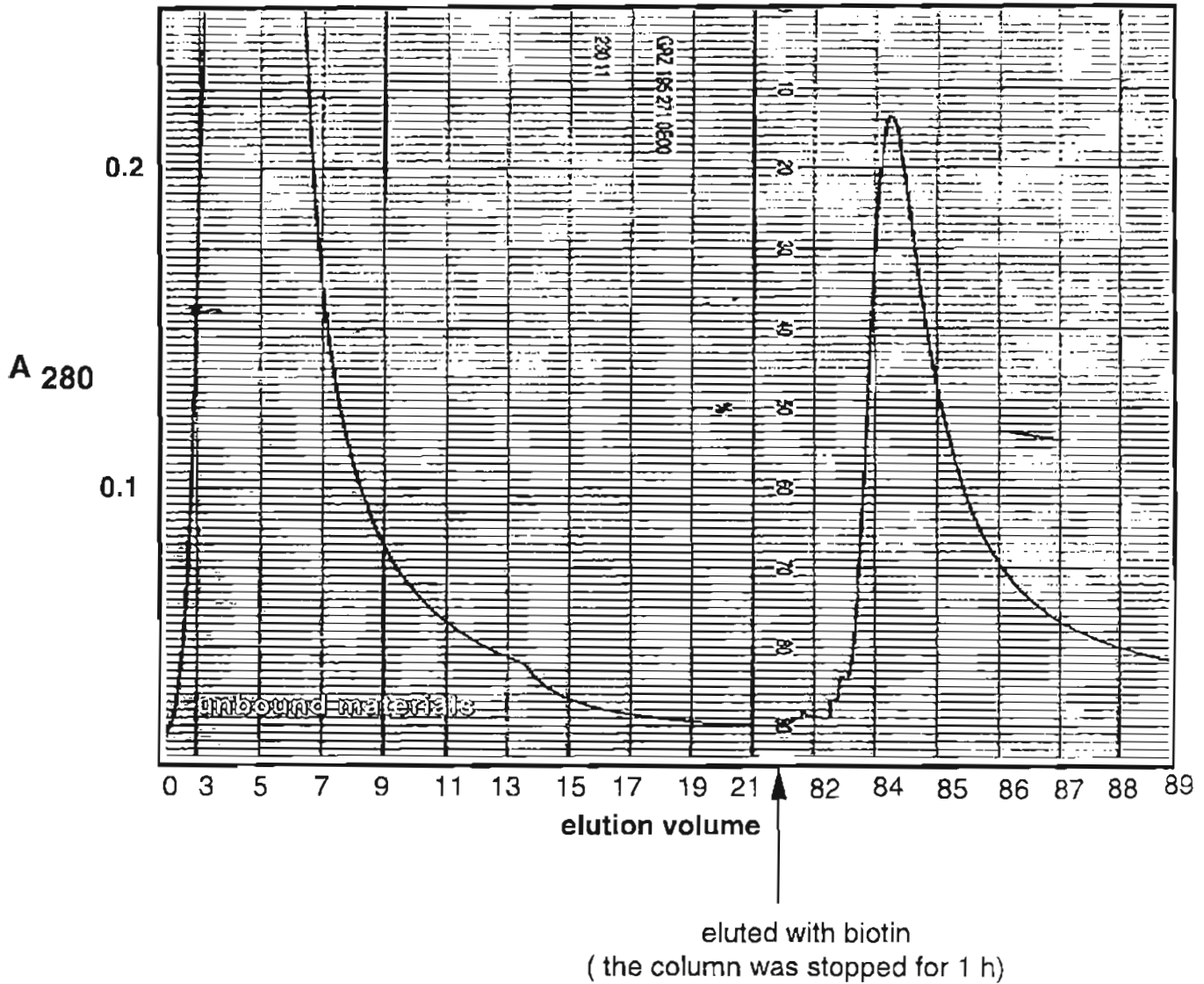
Figure 8.14 Monomeric avidin-Sepharose chromatography of recombinant human PC after ammonium sulphate precipitation (40% saturation) of mitochondrial extract. Ten micrograms of this material were applied to a monomeric avidin-Sepharose column at a flow rate of 0.2 ml/min. After a period of equilibration in the biotin-containing elution buffer (1mg/ml in running buffer), PC was eluted from the column with 1 mg/ml biotin in running buffer (arrow) and concentrated by precipitation with 50% ammonium sulphate.

8.4 DISCUSSION

In this Chapter, recombinant human PC was successfully produced at a very high level of expression which was achieved by using a highly efficient expression system. Conventional plasmids designed for stable expression of foreign proteins in eukaryotic cells commonly contain two separate expression cassettes, one for the protein of interest and another for a selectable marker such as the transposable element (Tn5 *neo* gene). Disruption of the circular plasmid occurs during genomic integration. If this takes place within the expression cassette for the protein of interest leaving the marker cassette intact, the transfected cells will survive but will not express the protein of interest. Loss of the protein of interest as a result of deletion of cDNA encoding it, with time can also occur as the selective pressure is only exerted on the drug resistant marker (Rees *et al.*, 1996). These problems can be overcome if the genes are fused within a single transcription cassette designed to produce a bicistronic message, linking expression of the required protein and the selectable marker at the RNA level (Rees *et al.*, 1996). The use of bicistronic vectors with the selectable marker downstream of an internal ribosome entry site sequence eliminates false positives in which transfected cells express drug resistance but not the protein of interest, as can occur with conventional dual cassette vectors (Rees *et al.*, 1996; Hobbs *et al.*, 1998). In addition, expression of the protein of interest can be maintained over a period of time by maintaining transfected cells in the media containing antibiotic as both proteins are simultaneously produced from the same mRNA. The use of the human elongation factor 1α promoter to drive the expression of human PC has been shown from the transient expression experiments to be more effective than the cytomegalovirus promoter. The elongation factor 1α promoter is highly active leading to abundant expression of the elongation factor 1α protein required in most eukaryotic cell types (Slobin, 1980). This promoter has also been shown to exhibit higher activity when cloned upstream from cDNA encoding different proteins than a number of commonly used viral promoters in a broad range of cells (Kim *et al.*, 1990; Mizushima and Nagata, 1990). The selection of stable clones that expressed very high levels of recombinant human PC can also be achieved by increasing the concentration of puromycin in the culture medium. This would select for any clones that contain multiple

copies of the transgene integrated into the genome thus also expressing higher levels of puromycin metabolising enzyme (PAC, puromycin acetyl transferase).

The use of 293T cells to produce human PC may have some advantages over heterologous expression system such as *E. coli*, since the 293T cells were derived from human kidney, a cell type in which PC is expressed naturally. Thus 293T cells can provide the folding machinery required to process recombinant PC into its native form. The use of an homologous expression system has also been successfully used to overproduce *S. cerevisiae* PC1 isozyme in either single or null mutants strains (M.G. Nezic, personal communication). The observed 65-70 kDa biotinylated protein which is proportional to the overexpressed PC band in all stable clones suggests that it is a partial breakdown product of PC which is likely to be susceptible to proteases during homogenisation. Similar sized fragments have also been observed when purified yeast (Lim *et al.*, 1988), chicken (Khew-Goodall, 1985), and rat (Booker, 1990; Jitrapakdee *et al.*, 1996) were proteolysed with chymotrypsin or other proteases.

Recombinant human PC can be isolated by one step chromatography i.e. monomeric-avidin affinity chromatography after the selective precipitation of mitochondrial extract with 40% saturated ammonium sulfate solution. However, the recovery of PC obtained from this chromatography is relatively low as a result of poor binding of PC to the column. Similar problems have also been reported in many attempts to purify PC from other sources including bacteria (Modak and Kelly, 1995), insect (Tu and Hagedorn, 1997), chicken (Khew-Goodall, 1985). This seems to conflict with the fact that avidin is known to be a strong inhibitor of PC when it is added in the assay mixture, suggesting that the biotin moiety is readily accessible to the free avidin molecules under these conditions (Purcell and Wallace, 1996). It has been proposed that in the absence of the substrate(s), the biotin group which attached near the C-terminal of PC is less accessible to the avidin while PC is being applied to the monomeric avidin affinity chromatography (Purcell and Wallace, 1996). Although it has been showed that adding substrates of PC including $NaHCO_3^-$, pyruvate and ATP in the running buffer significantly improved the yield of PC from *Pseudomonas aeruginosa* (Purcell and Wallace, 1996) and *M. thermoautotrophicum*, (Mulhapayay *et al.*, 1998), the inclusion of these substrates in the running buffer did not improve the yield of

human PC. Perhaps this reflects a property of the structurally distinct *P. aeruginosa* and *M. thermoautotrophicum* PC which possess $(\alpha\beta)_4$ configuration. Indeed, ATP has been shown to both increase and suppress the interaction of avidin with chicken PC (Scrutton and Utter, 1965). Several other chromatographic separations may also provide the alternative means to purify human PC. These include an ion exchange chromatography, gel filtration and dye-ligand chromatography. The first two methods have been extensively used to purify yeast (Lim *et al.*, 1987), sheep (Khew-Goodall *et al.*, 1991) and chicken (Attwood *et al.*, 1993; Wernberg and Ash, 1993) PCs. Though some minor bands were also observed on SDS-PAGE, these contaminants did not interfere with the PC assay. The dye-ligand chromatography based on its affinity for acetyl-CoA has recently been described to purify PC from *Rhodobacter capsulatus* (Modak and Kelly, 1995). Since vertebrate PC is known to be acetyl-CoA-dependent, this affinity chromatography should also provide an alternative means to avidin affinity chromatography.

The expression system described here has provided for the first time a readily available and reliable source of human PC in native form and will be extremely useful for the production of mutant enzymes, including those reported clinically (Wexler *et al.*, 1998; Carbone *et al.*, 1998) in quantities appropriate for structure and function analysis. These mutations includes substitutions of Val 146->Ala, Arg 451-> Cys, Ala 610 ->Thr and Met 743->Ile. Further work towards this goal will involve site-specific mutagenesis of wild type human PC cDNA to the above substitutions. The most interesting mutants to be created would be Arg 451->Cys, Ala 610-> and Met 743-> Ile as these mutations have been found to alter highly conserved residues and result in reduced catalytic activity of the enzyme. Val 146->Ala substitution mutant has been suggested to affect structural integrity of the enzyme (Wexler *et al.*, 1998) and hence this may not be readily expressed and purified. The mutagenised cassettes could then be used to generate stable clones expressing these mutant enzymes as with wild type enzyme. Mutant enzymes can be purified and their kinetic analysis can be performed in more detail.

It is noteworthy that the 293T cells also produce significant amount of endogenous PC (3.3%) which is likely to contaminate with recombinant PC. This may cause a problem when this cell line is used as a host to produce mutant forms of human PC as the purified

mutant enzymes will also be contaminated with wild type enzyme produced from host cells. Two alternate methods could be used to deal with this problem. First, the level of endogenous PC can be suppressed by adding insulin in the culture media. As shown in Chapter 6 with rat, insulin down-regulated PC gene transcription. This may also occur with the human PC gene as it is closely related organism to rat. Alternatively, a genetic approach can be used to disrupt the endogenous PC gene by homologous recombination. This technique has recently been applied with different cell types in addition to the embryonic stem cells (Wang *et al.*, 1996; Grawunder *et al.*, 1998). There is evidence to suggest that several human cell lines that are defective in PC expression can be grown i.e. the skin fibroblasts derived from PC deficiency patients with indistinguishable phenotype of that wild type (cell lines GM00444 and GM06056 distributed by The Coriell Institute of Medical Research, U.S.A.).

The second approach would be to generate recombinant human PC with histidine tag at the C-terminal end of the enzyme rather than at the N-terminal. Putting the histidine tagged at the N-terminal is likely to be cleaved off since newly synthesised mammalian PCs undergo N-terminal cleavage while targetting to the mitochondria. The few extra histidines at the C-terminal end of the enzyme could interfere with the biotinylation reaction by biotin ligase. A recent report of the three dimensional structure of the C-terminal 87 amino acid residues of the biotin carboxyl carrier subunit of *E. coli* ACC (homologue of biotinyl domain of PC) has shown that the last two residues located at the end of β-strand constributed the hydrophobic core (Roberts *et al.*, 1999). However, the insertion of extra glycine residues (providing flexibility) after the last residue (glutamic acid) of human PC and a C-terminal histidine tag should not interfere with the biotinylation reaction. Indeed, PC from other organisms including *A. terreus* (Li *et al.*, 1998), *M. thermoautotrophicum* (Mukhopadhyay *et al.*, 1998) and *S. cerevisiae* [PC2 isozyme] (Stucka *et al.*, 1991) also contain additional amino acid residues C-terminal of the biotinyl lysine compared wtih mammalian PCs. *An in vitro* study has also been demonstrated that a recombinant PC2 biotinyl domain with C-terminal extension acts as a better substrate for the biotinylation reaction than the shorter version (Val *et al.*, 1995). Generation of histidine-tagged human PC is not only useful for removing endogenous PC away from recombinant PC but also facilities purification

procedures by avoiding inefficient monomeric-avidin Sepharose chromatography. This histidine-tagged protein purification system is commercially available from different sources.

CHAPTER 9

FINAL DISCUSSION, CONCLUSIONS AND

FUTURE DIRECTIONS

PC has been of particular interest to our research laboratory over the past 4 decades since it was first discovered. This enzyme catalyses the first-regulated step in the conversion of pyruvate to oxaloacetate, a Krebs cycle intermediate that is utilised directly and indirectly as the substrate for many biosynthetic purposes. The most important pathways are gluconeogenesis and lipogenesis. Early work carried out in the past focused on the characterisation of the physical properties and kinetics of the enzyme. Since the development of recombinant DNA technology, information on the primary structure of the enzyme has been enhanced enormously by the cloning and sequencing of the genes and cDNA encoding this enzyme. The most significant discovery in our laboratory during such a period is the first report of the complete structure of the gene encoding PC in yeast (Lim *et al.*, 1988). This led to the identification of three functional domains.

Structure

Clearly proteins that play the same role and are crucial for the cells are evolutionarily conserved. Comparison of the primary structures of PC from different sources would therefore allow the identification of structurally or catalytically important residues or motifs that might play common or distinct roles in the reaction. Chapter 3 described the complete sequence of cDNA encoding rat PC. The cDNA contained an open reading frame of 3537 nucleotides encoding a polypeptide of 1178 amino acids and showed high sequence similarity with PC from mammals, insect, yeast and bacteria. Multiple sequence alignment of PC with different portions of other biotin dependent enzymes from different species, together with the recent crystal structure of the biotin carboxylase subunit of *E. coli* acetyl-CoA carboxylase (Waldrop *et al.*, 1994) identified 11 highly conserved residues within the biotin carboxylase domain. Notably a cysteine-lysine pair located in the biotin carboxylation domain of different biotin carboxylases was invariant and has been suggested on the basis of chemical modification studies (Werneberng and Ash, 1993) to form an ion-pair in the enolisation of biotin in the first partial reactions. The functional importance of other conserved motifs found in the transcarboxylation and the biotinyl domains were also discussed within the context of the structural domains of the enzyme (Chapter 3). With the availability of a cDNA clone encoding PC, it is now feasible to investigate the role of particular conserved residues by site-specific mutagenesis. Those candidates includes 11

highly conserved residues located in the biotin carboxylation domain, the putative pyruvate binding site and the metal binding site located in the transcarboxylation domain. A program of site-directed mutagenesis has been targeted to the yeast PC1 isoenzyme followed by expression of these mutant enzymes in the double null mutant strain of yeast lacking both PC1 and PC2 gene expression. The availability of multicopy yeast expression vectors should also facilitate the production of large scale recombinant PC for structure determination by X-ray crystallography. Together with site-directed mutagenesis studies, this should allow us to fully understand the structural basis of the catalytic and allosteric properties of PC.

Gene and evolution

As described in Chapter 5, analysis of the gene structure of rat PC revealed that this gene was located on chromosome 1 and is split into nineteen coding exons and four 5'-untranslated region exons spanning over 40 kb. The organisation of the coding exons and the introns of the rat gene is very similar to the that since described for human gene, suggesting that the PC gene is evolutionarily conserved in mammals. Organisation of exons in the gene was also consistent with the domain structure of the enzyme suggesting a close relationship between exons and protein domains. Of the particular interest was the placement of a very large intron (intron K, ~10 kb) which separated the boundaries of the exons encoding the biotin carboxylase and the transcarboxylase domains, suggesting that this junction might be the region where the fusion of ancestral genes encoding both domains occurred in mammals during evolution. It is very interesting to note that the same region of the gene encoding PC in bacteria, *M. thermoautotrophicum* is also split into two parts i.e. the non-biotinylated subunit (α-subunit) (equivalent to the biotin carboxylase domain of vertebrate enzyme) and the biotinylated subunit (β-subunit) (equivalent to the transcarboxylase and the biotinyl domains) separated by half of a genome.

The availability of embryonic stem cell technology allows one to create a mouse model of PC deficiency mimicking mutations that are naturally occurring in humans or to investigate other physiological roles of PC by disrupting the gene encoding it by homologous recombination. This approach will initially require the generation of a targetting vector containing part of the mouse PC gene flanked by a selectable marker gene. This vector

will be used for two rounds of targetting until homozygous mice lacking PC in both alleles are obtained. With the information on the genomic organisation of both rat and human PC, the way is clear to proceed with the isolation of the mouse gene. This would perhaps reveal new biological role(s) for PC as recently has been shown in insulin secreting cells.

Transcriptional regulation

As described in Chapter 4, RACE-PCR was successfully used to identify and characterise multiple PC mRNA isoforms in rat and human tissues. Five alternative forms of rat PC mRNA isoforms with the same coding sequence but differing in their 5'-UTRs were identified in liver, kidney, brain and adipose tissue and these were expressed in a tissue-specific manner. Likewise, two alternative mRNA isoforms encoding PC were also identified in human liver, and most likely these will be shown to be alternatively transcribed from two promoters as in the rat gene. Comparison of the 5'-UTRs of the different rat PC mRNAs to the 5'-end of the gene clearly showed that differential splicing of two different primary transcripts, alternatively transcribed from two tissue-specific promoters is responsible for the production of multiple mRNA isoforms with 5'-end heterogeneity. Both promoters contained different features of typical eukaryotic promoters: i.e. the proximal promoter contained no canonical TATA or CCAAT boxes but resembled a HIP-1 box, while the distal promoter contained three copies of a CCAAT box and multiple copies of Sp1 transcription factor binding sites. Different putative transcription factor binding sites were found in both promoters suggesting that they are regulated differently. Transient expression studies in COS-1 cells using chimeric gene constructs, created by fusing the proximal and the distal promoters to the luciferase reporter gene, demonstrated that both promoters contained functional elements. Deletion analysis showed that the 153 bp and 187 bp, preceding the transcription start site of the proximal and the distal promoters respectively were required for basal transcription. The distal promoter was found to drive the expression of the luciferase at a 4-8-fold higher level than the proximal promoter when both were transiently expressed in COS-1, CHO-K1 and HepG2 cell lines.

Insulin selectively inhibited the expression of the proximal promoter-luciferase reporter gene by 50% but not the distal promoter, suggesting the presence of an insulin-responsive element in the proximal promoter. However, this potential insulin-responsive

element has not been identified. Further experiments will require transfection of different deletion constructs in COS-1 and HepG2 cells grown in the presence of insulin in the culture medium to see if the deletion causes an increase in the reporter gene activity. Site-directed mutagenesis could then be introduced to confirm the responsiveness of such a motif to insulin. Gel shift analysis could also be performed to identify which protein would bind to an insulin-responsive element. The availability of cloned proximal and distal promoters would also provide an excellent opportunity to investigate the roles of other regulatory proteins that mediate transcriptional regulation of both promoters. The most widely used technique is the DNaseI footprint analysis which involves incubation of the radio-labelled promoter fragment with a nuclear extract obtained from various tissues particularly liver and kidney, followed by digestion with DNaseI. Any regions of the promoter that have transcription factors bound to them will be protected from cleavage by DNaseI. Once the regulatory regions have been identified, gel shift analysis using synthetic sequence(s) corresponding to the DNaseI protected region(s) could then be carried out. Such techniques mentioned above have been used extensively to identify regulatory proteins that play a role in transcriptional regulation of PEPCK and other genes. The anaplerotic role of the distal promoter of rat PC gene in insulin-secreting cells is particularly interesting as the transcripts generated from this promoter can be induced by glucose. The role of PC in insulin signalling has recently been described. It will be interesting to generate a rat insulinoma cell line that has a defect in PC expression and to see how this deficiency can affect the insulin secretion process. A cell line that is defective in PC expression can be generated by disrupting the gene encoding PC using homologous recombination, the same technique that is used with mouse embryonic stem cells.

Physiological role of alternate promoters

The presence of alternate promoters of the rat PC gene has prompted investigation of the physiological roles of these two promoters. In the past, different physiological states, viz. neonatal development, diabetes and genetic obesity have most clearly been shown to affect PC expression in rats. The physiological roles of both promoters were therefore studied in these situations. PC expression is developmentally regulated. Hepatic PC and its activity were low, though readily detectable, immediately before birth but increased markedly

during the suckling period concomitant with an increase in its mRNAs. The proximal promoter has been shown to be responsible for production of the rUTR C transcript which was the major form being accumulated during this period. The level of PC protein and its activity were then decreased concomitant with a decrease in its mRNAs during weaning and approached the level detected in adults. In genetically obese Zucker rats, adipose PC was dramatically increased by 12 weeks of age concomitant with an increase in the rUTR C transcript, indicating that the proximal promoter was trancriptionally active during the hyperlipogenic condition.

In contrast, the proximal promoter was down-regulated in different rat liver cell lines suggesting that this promoter was not functional under cell culture conditions. In pancreatic islets and insulin-secreting cells, only transcripts generated from the distal promoter were expressed and could be induced by glucose. *In vitro* translation using the reticulocyte lysate system and *in vivo* polysome profile analysis clearly showed that different PC mRNA isoforms (rUTR C, rUTR D and rUTR E) were translated with different translational efficiencies, suggesting that translational control plays an important role in regulating PC expression.

As described in Chapter 1, several lines of evidence have implicated other physiological conditions that could also affect PC expression. These include pregnancy, lactation, diabetes, starvation, hypo- and hyperthyroidism. It will therefore be very interesting to see how these physiological conditions affect PC expression.

Short term and long term regulation of PC expression

From the characterisation of multiple PC mRNA isoforms (Chapter 4), gene (Chapter 5) and promoter (Chapter 6) of rat PC gene together with regulation studies in Chapter 7, it can be concluded that transcriptional and post-transcriptional control appear to be the important mechanisms for long term regulation of PC expression. As indicated in Fig. 9.1, the first control step is mediated through a transcriptional mechanism. The presence of two tissue-specific promoters are likely to be an important mechanism to allow transcription of the same gene independently under different circumstances. In gluconeogenic tissues (liver and kidney) and lipogenic tissues (adipose tissues and liver) the proximal promoter is functional. This is perhaps due to the presence of tissue-specific transcription factors that

150

Figure 9.1 A schematic representation of cooperative regulation of rat PC expression by short term and long term mechanisms. Different metabolic signals outside the cells, i.e. glucose-induced insulin release in pancreatic β-cells; gluconeogenesis, lipogenesis, mediated through the hormonal changes which can affect PC expression are shown to alter activity of two alternate promoters (*P1* and *P2*) of the rat PC gene at the transcriptional level (+, stimulate; -, inhibit). This results in the generation of alternative transcripts with 5'-end heterogeneity i.e. liver/adipose-specific PC transcript C (*C* in figure) and housekeeping-specific PC transcripts D and E (*D* and *E*) which can be controlled at the translational level through the formation of a stable secondary structure (stem loop) in the 5'-untranslated region. Short term regulation of PC activity is achieved post-translationally by allosteric activation by acetyl-CoA upon the enzyme targetted to the mitochondria. Boxes shown in the nucleus represent the different 5'-untranslated regions of different PC transcripts derived by alternative splicing.
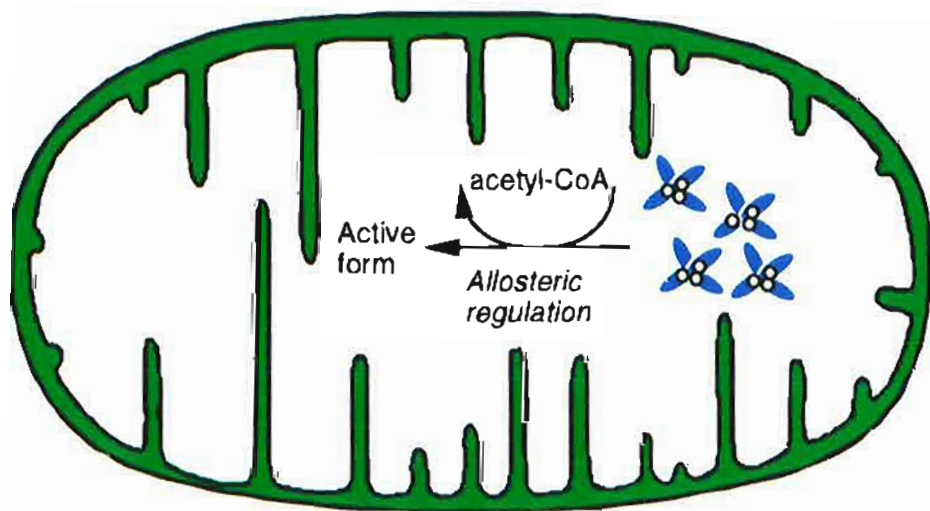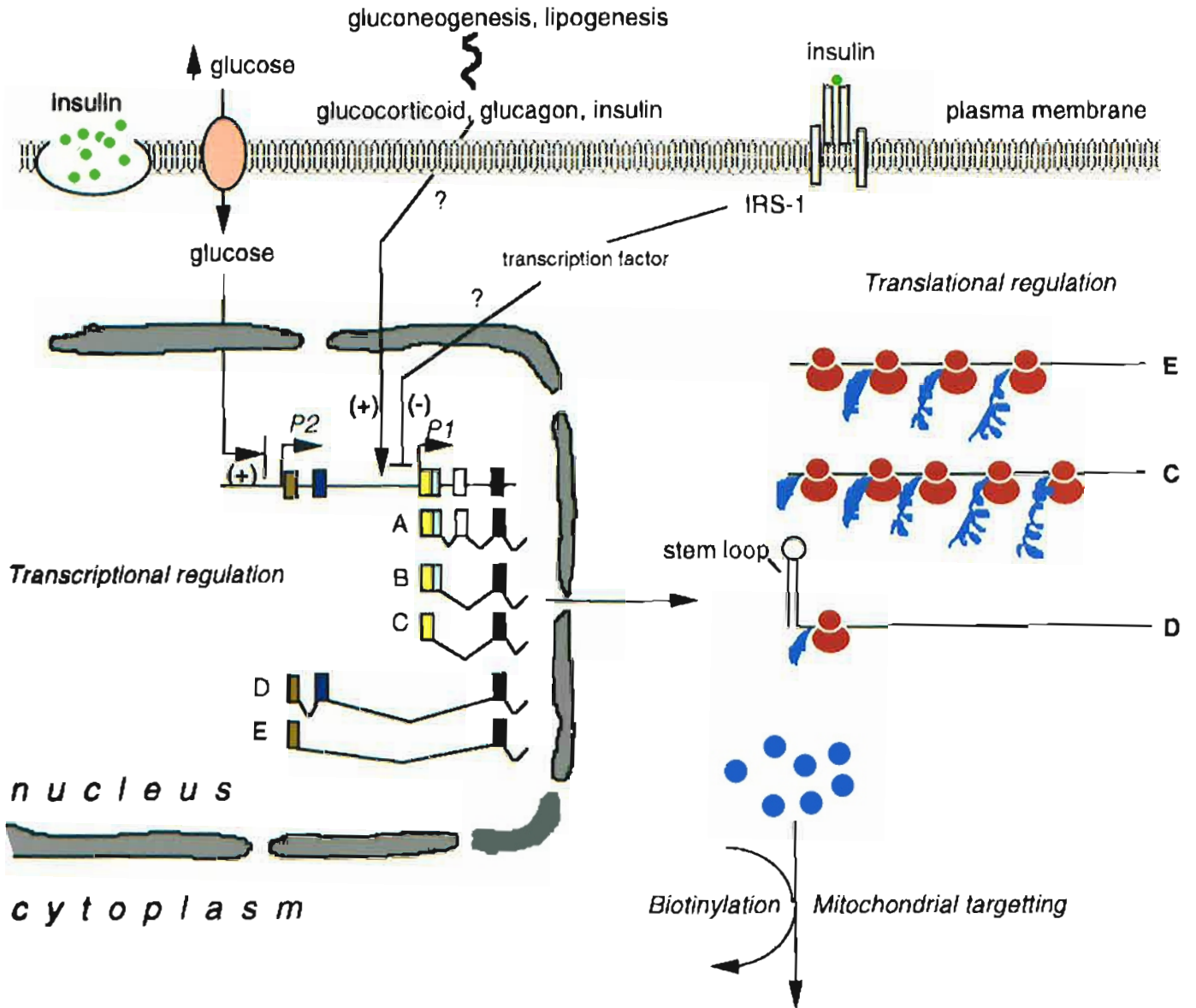
interact with *cis*-acting element(s) in the proximal promoter (presumably hepatic nuclear factor, HNF-4 in liver and fat-specific element, FSE1 in adipose tissue). Alterations of plasma insulin, glucagon and glucocorticoids during the postnatal gluconeogenic period (Girard *et al.*, 1992) and in lipogenesis have been shown to affect PC expression and these hormonal changes may involve up-regulation of transcription from the proximal promoter to supply the demands of the cells under these conditions. An *in vivo* study has shown that insulin down-regulated PC expression in diabetic rats (Weinberg and Utter, 1981), but the mechanism by which insulin works remains unclear. Indirect evidence, obtained from the reporter gene study in chapter 6 has shown that insulin inhibited transcription from the proximal promoter, most likely through an insulin-responsive element binding protein which is currently unknown. On the other hand, the distal promoter appears have a housekeeping function in other tissue types. Of particular interest, this promoter seems to play an anaplerotic role in insulin-secreting cells, i.e. pancreatic islets and insulinoma cells. This promoter is induced if these cell types are grown in higher than physiological concentrations of glucose. The second control step of PC expression is achieved at the translation step i.e. different PC transcripts produced from two tissue-specific promoters exhibit different translation efficiencies. This appears to be mediated through a sequence at the 5'-untranslated regions of certain PC mRNAs which have the tendency to form a secondary structure that could block ribosomal access to the cap site. Therefore, the rate of enzyme synthesis would depend on which mRNA is being produced at the time. The newly synthesised PC then undergoes post-translational modification by biotinylation followed by translocation to the mitochondrial matrix. However, there is no evidence yet to suggest that the biotinylation is the regulatory step. The third control step which has already been known to be short term control is the allosteric regulation by acetyl-CoA (Barritt *et al.*, 1976). The beta oxidation of fatty acids is known to generate a large amount of acetyl-CoA which acts as a physiological regulator of PC. Allosteric activation of PC by acetyl-CoA enhances the production of oxaloacetate in the short term.

Recombinant enzyme and application

Clearly, obtaining PC from human liver at autopsy is not practical as a reliable and reproducible source for further studies of the enzyme. Although several human liver and

kidney cell lines are available, the expression level of PC in these cell lines is extremely low. The availability of cloned human PC cDNA has made it possible to express this recombinant material and replace native material. Full length cDNA encoding human PC was generated and cloned in a mammalian expression vector that allows the production of large amounts of this recombinant enzyme in mammalian cells. So far this is far superior to an *E. coli* expression sytem. With a one step purification using monomeric avidin chromatography, it is possible to obtain a reasonable yield of the recombinant enzyme with sufficient purity to be further characterised. This expression/purification system not only allows one to generate mutant enzymes that mimic natural mutations which have recently been reported in humans but would also be useful for structural studies. Carey (1988) has mapped the different regions of tetrameric sheep PC that interact with various monoclonal antibodies using electron microscopic analysis. However this analysis was performed without information of primary structure of sheep PC. Although cDNA encoding sheep PC is not available, strong cross reactivity observed between monoclonal antibodies raised against sheep PC to other mammalian PCs suggests that the quaternary structure of mammalian PCs are very similar. Having the full length human PC cDNA clone, the epitopes of PC can be mapped. This can be achieved by random cleavage of cDNA encoding human PC and ligating to a C-terminal fusion partner e.g. β-galactosidase for detecting expression in *E. coli.*. This would be followed by reacting the PC peptide library with different monoclonal antibodies. Any reactive clones can be isolated and the cDNA encoding a particular motif can be extracted and sequenced. Together with electron microscopic data, the way is clear to relate primary to quaternary structure of PC.

Perhaps the most useful benefit of having a functional human PC cDNA clone would be for a therapeutic purpose. Clearly, as a defect in PC in humans causes early death or delayed development, the application of gene therapy technology would make it possible to transfer functional cDNA using a retrovirus-based expression vector to correct the PC deficiency.

REFERENCES

Abu-Elheiga, L., Almarza-Ortega, D.B., Baldini, A. and Wakil, S.J. (1997) Human acetyl-CoA carboxylase 2. Molecular cloning, characterization, chromosomal mapping, and evidence for two isoforms. *J. Biol. Chem.* **272**, 10669-77.

Adam, P.A.J. and Haynes, R.C., Jr. (1969) Control of hepatic mitochondrial CO_2 fixation by glucagon, epinephrine, and cortisol. *J. Biol. Chem.* **244**, 6444-6450.

Ahmad, P.M., Russell, T.R. and Ahmad, F.(1979) Rat liver pyruvate carboxylase. Purification, detection and quantification of apo and holo forms by immuno-blotting and by an enzyme-linked immunosorbent assay. *Biochem. J.* **182**, 509-514.

Al-Feel, W., Chirala, S.S. and Wakil, S.J. (1992) Cloning of the yeast FAS3 gene and primary structure of yeast acetyl-CoA carboxylase. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 534-538.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Andreone, T.L., Beale, E.G., Bar, R.S. and Granner, D.K. (1982) Insulin decreases phosphoenolpyruvate carboxykinase (GTP) mRNA activity by a receptor-mediated process. *J. Biol. Chem.* **257**, 35-38.

Angus, C.W. and Lane, M.D. (1981) Increase in translatable mRNA for mitochondrial pyruvate carboxylase during differentiation of 3T3 preadipocytes. *Biochem. Biophys. Res. Commun.* **103**, 1216-1222.

Apte, A.N. and Siebert, P.D. (1993) Anchor-ligated cDNA libraries: a technique for generating a cDNA library for the immediate cloning of the 5' ends of mRNAs. *Biotechniques* **15**, 890-893.

Argaud, D., Zhang, Q., Pan, W., Maitra, S., Pilkis, S.J. and Lange, A.J. (1996) Regulation of rat liver glucose-6-phosphatase gene expression in different nutritional and hormonal states. Gene structure and 5'-flanking sequence. *Diabetes* **45**, 1563-1571.

Artymiuk, P.J., Poirrette, A.R., Rice, D.W. and Willett, P. (1996) Biotin carboxylase comes into the fold. *Nature Struct. Biol.* **3**, 128-132.

Ashman, L.K., Keech, D.B., Wallace, J.C. and Nielsen, J. (1972) Sheep kidney pyruvate carboxylase: Studies on its activation by acetyl-CoA and characterisation of its acetyl coenzyme A independent reaction. *J. Biol. Chem.* **247**, 5818-5824.

Ashman, L.K., Wallace, J.C. and Keech, D.B. (1973) Desensitisation of pyruvate carboxylase against acetyl-CoA by chemical modification. *Biochem. Biophys. Res. Commun.* **51**, 924-931.

Ashman, L.K. and Keech, D. B. (1975) Sheep kidney pyruvate carboxylase: studies on the coupling of adenosine triphosphate hydrolysis and CO_2 fixation. *J. Biol. Chem.* **250**, 14-21.

Ashcroft, F.M., Harrison, D.E. and Ashcroft, S.J.H. (1984) Glucose induces closure of single potassium channels in isolated rat pancreatic beta-cells. *Nature* **312**, 446-448.

Ashworth, J.M. and Kornberg, H.L. (1966) The anaplerotic fixation of carbon dioxide by *Escherichia coli. Proc. Royal Soc. Lond. Biol. Sci.* **165**, 179-188.

Athappily, F.K. and Hendrickson, W.A. (1995) Structure of the biotinyl domain of acetyl-CoA carboxylase determined by MAD phasing. *Structure* **3**, 1407-1419.

Atkin, B.M., Utter, M.F. and Weinberg, M.B.(1979) Pyruvate carboxylase and phosphoenolpyruvate carboxykinase activity in leukocytes and fibroblasts from a patient with pyruvate carboxylase deficiency. *Pediatr. Res.* **13**, 38-43.

Attwood, P.V. and Cleland, W.W. (1986) Decarboxylation of oxaloacetate by pyruvate carboxylase. *Biochemistry* **25**, 8191-8196.

Attwood, P.V., Tipton, P.A. and Cleland, W.W. (1986a) Carbon-13 and deuterium isotope effects on oxalacetate decarboxylation by pyruvate carboxylase. *Biochemistry* **25**, 8197-8205.

Attwood, P.V., Mayer, F. and Wallace, J.C. (1986b) Avidin as a probe of the conformational changes induced in pyruvate carboxylase by acetyl-CoA and pyruvate. *FEBS Lett.* **203**, 191-196.

Attwood, P.V. and Graneri, B.D.L.A. (1991) Pyruvate carboxylase catalysis of phosphate transfer between carbamoyl phosphate and ADP. *Biochem. J.* **273**, 443-448.

Attwood, P.V. and Graneri, B.D.L.A. (1992) Bicarbonate-dependent ATP cleavage catalysed by pyruvate carboxylase in the absence of pyruvate. *Biochem. J.* **287**, 1011-1017.

Attwood, P.V. (1993) Locus of action of acetyl-CoA in the biotin-carboxylation reaction of pyruvate carboxylase. *Biochemistry* **32**, 12736-12742.

Attwood, P.V., Johannssen, W., Chapman-Smith, A. and Wallace, J.C. (1993) The existence of multiple tetrameric conformers of chicken liver pyruvate carboxylase and their roles in dilution inactivation. *Biochem. J.* **290**, 583-590.

Attwood, P.V. (1995) The structure and mechanism of action of pyruvate carboxylase. *Int. J. Biochem. Cell Biol.* **27**, 231-249.

Azizkhan, J.C., Jensen, D.E., Pierce, A.J. and Wade, M. (1993) Transcription from TATA-less promoters: Dihydrofolate reductase as a model. *Crit. Rev. Euk. Gene Express.* **3**, 229-254.

Bacon, D.J. and Anderson, W.F. (1988) A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics* **6**, 219-220.

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. (1965) Carbohydrate metabolism in liver from foetal and neonatal sheep. *Biochem. J.* **95**, 191-200.

Ballard, F.J. and Hanson, R.W. (1967) Phosphoenolpyruvate carboxykinase and pyruvate carboxylase in developing liver. *Biochem. J.* **104**, 866-871.

Ballard, F.J., Hanson, R.W. and Kronfeld, D.S. (1968) Factors controlling the concentration of mitochondrial oxaloacetate in liver during spontaneous bovine ketosis. *Biochem. Biophys. Res. Commun.* **30**, 100-104.

Ballard, F.J.and Hanson, R.W. (1970) The citrate cleavage pathway and lipogenesis in rat adipose tissue: replenishment of oxaloacetate. *J. Lipid Res.* **8**, 73-79.

Ballard, F.J., Hanson, R.W. and Reshef, L. (1970) Immunochemical studies with soluble and mitochondrial pyruvate carboxylase activities from rat tissues. *Biochem. J.* **119**, 735-742.

Barber, M.C. and Travers, M.T. (1995) Cloning and characterisation of mutiple acetyl-CoA carboxylase transcripts in ovine adipose tissue. *Gene* **154**, 271-275.

Barden, R.E. and Scrutton, M.C. (1974) Pyruvate carboxylase from chicken liver. Effects of univalent and divalent cations on catalytic activity. *J. Biol. Chem.* **249**, 4829-4838.

Barnett, J.A. and Kornberg, H.L. (1960) *J. Gen. Microbiol.* **23**, 65.

Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2216-2220.

Barritt, G.J. (1976) The regulation of pyruvate carboxylase activity in gluconeogenic tissues, in Gluconeogenesis (Hanson, R.W. and Mehlman, M.A., Eds), John Wiley & Sons, New York, p 3-46.

Barritt , G.J. (1985) Regulation of enzymatic activity, in Pyruvate carboxylase (Keech, D.B. and Wallace, J.C. eds) CRC Series in Enzyme Biology, CRC Press, Boca Raton, FL, p 141-177.

Barzilai, N. and Rossetti, L. (1993) Role of glucokinase and glucose-6-phosphatase in the acute and chronic regulation of hepatic glucose fluxes by insulin. *J. Biol. Chem.* **268**, 25019-25025.

Beale, E.G., Chrapkiewicz, N.B., Scoble, H.A., Metz, R.J., Quick, D.P., Noble, R.L., Donelson, J.E., Biemann, K. and Granner, D.K. (1985) Rat hepatic cytosolic phosphoenolpyruvate carboxykinase (GTP). Structures of the protein, messenger RNA, and gene. *J. Biol. Chem.* **260**, 10748-10760.

Belyavsky, A., Vinogradova, T. and Rajewsky, K. (1989) PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res.* **17**, 2919-2932.

Benjamin, A.M. and Quastel, J.H. (1974) Fate of L-glutamate in the brain. *J. Neurochem.* **23**, 457-464.

Benoist, C. and Chambon, P. (1981) *In vivo* sequence requirements of the SV40 early promoter region. *Nature* **290**, 304-310.

Berry, M.N., Phillips, J.W. and Grivell, A.R. (1993) Interactions between mitochondria and cytoplasm in isolated hepatocytes. *Curr. Top. Cell. Regul.* **33**, 309-328.

Bevenisty, N. and Reshef, L. (1987) Developmental expression and modification of PEPCK genes. *Biol. Neonate* **52**, 61-69.

Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Transcription termination and 3' processing: the end is in site. *Cell* **41**, 349-359.

Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W. and Azizkhan, J.-C. (1990) Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. *Mol. Cell. Biol.* **10**, 6632-6641.

Bloom, S.J. and Johnson, M.J. (1962) The pyruvate carboxylase of *Aspergillus niger*. *J. Biol. Chem.* **237**, 2718-2720.

Board, P.G., Webb, G.C. and Coggan, M. (1989) Isolation of a cDNA and localization of the human gluthathione S-transferase 3 genes to chromosome bands 11q13 and 12q13->q14. *Ann. Hum. Genet.* **53**, 205-213.

Booker, G.W. (1990) Pyruvate carboxylase: relating structure to function. Ph.D. thesis, University of Adelaide, South Australia.

Böttger, I., Wieland, O., Brdiczka, D. and Pette, D.(1969) Intracellular localization of pyruvate carboxylase and phosphoenolpyruvate carboxykinase in rat liver. *Eur. J. Biochem.* **8**, 113-119.

Böttger, I., Kriegel, H. and Wieland, O. (1970) Fluctuation of hepatic enzymes important in glucose metabolism in relation to thyroid function. *Eur. J. Biochem.* **13**, 253-257.

Boxer, L.M., Prywes, R., Roeder, R.G. and Kedes, L. (1989) The sarcomeric actin CArG-binding factor is indistinguishable from the c-fos serum response factor. *Mol. Cell. Biol.* **9**, 515-522.

Brawerman, G. (1987) Determinants of messenger RNA stability. *Cell* **48**, 5-6.

Breathnach, R. and Chambon, P. (1981) Organization and expression of eukaryotic split gene coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383.

Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* **56**, 467-495.

Brewster, N.K. (1994) Studies on the regulation of PC isozyme (PYC1 PYC2) gene expression in *Saccharomyces cerevisiae.*, Ph.D. thesis, University of Adelaide, South Australia.

Brewster, N.K., Val, D.L., Walker, M.E. and Wallace, J.C. (1994) Regulation of pyruvate carboxylase gene expression in *Saccharomyces cerevisiae* during fermentative and nonfermentative growth. *Arch. Biochem. Biophys.* **311**, 62-71.

Brocklehurst, S.M. and Perham, R.N. (1993) Prediction of the three-dimensional structures of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycine cleavage system: a new automated method for the prediction of protein tertiary structure. *Protein Sci.* **2**, 626-639.

Browner, M.F., Taroni, F., Sztul, E. and Rosenberg, L.E. (1989) Sequence analysis, biogenesis, and mitochondrial import of the α-subunit of rat liver propionyl-CoA carboxylase. *J. Biol. Chem.* **264**, 12680-12685.

Brun, T., Roche, E., Assimacopoulos-Jeannet, F., Corkey, B.E., Prentki, M. (1996) Evidence for an anaplerotic malonoyl-CoA pathway in pancreatic β-cell nutrient signaling. *Diabetes* **45**, 190-198.

Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563-578.

Burgess, D.G. and Penhoet, E.E. (1985) Characterization of the chicken aldolase B gene. *J. Biol. Chem.* **260**, 4604-4610.

Carbone, M.A., MacKay, N, Ling, M., Cole, D.E., Douglas, C., Rigat, B., Feigenbaum, A., Clarke, J.T.R., Haworth, J.C., Greenberg, C.R., Seargeant, L. and Robinson, B.H. (1998) Amerindian pyruvate carboxylase deficiency is associated with two distinct missense mutations. *Am. J. Hum. Genet.* **62**, 1312-1319.

Carey, K.R. (1988) Pyruvate carboxylase: an immunological study. Ph.D. thesis, University of Adelaide, South Australia.

Carson, M. (1991) Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958-961.

Carthew, R.W., Chodosh, L.A. and Sharp, P.A. (1985) An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter.*Cell* **43**, 439-448.

Carver, J.A., Baldwin, G.S., Keech, D.B. and Wallace, J.C. (1988) Inactivation of chicken liver pyruvate carboxylase by 1,10-phenanthroline. *Biochem. J.* **252**, 501-507.

Cassady, A.I. (1987) Pyruvate carboxylase: A molecular biology study. Ph.D. thesis, University of Adelaide, South Australia.

Castella-Escola, J., Ojcius, D.M., LeBoulch, P., Joulin, V., Blouquit, Y., Garel, M.C., Valentin, C., Rosa, R., Climent-Romeo, F. and Cohen-Solal, M. (1990) Isolation and characterization of the gene encoding the muscle-specific isozyme of human phosphoglycerate mutase. *Gene* **91**, 225-232.

Cazzulo, J.J. and Stoppani, A.O.M. (1968) The regulation of yeast pyruvate carboxylase by acetyl-coenzyme A and L-aspartate. *Arch. Biochem. Biophys.* **127**, 563-567.

Cazzulo, J. J., Sundaram, T.K., Dilks, S.N. and Kornberg, H.L. (1971) Synthesis of pyruvate carboxylase from its apoenzyme and (+)-biotin in *Bacillus stearothermophilus*. Purification and properties of the apoenzyme and the holoenzyme synthetase. *Biochem. J.* **122**, 653-661.

Cech, T.R.(1983) RNA splicing: three themes with variations. *Cell* **34**, 713-716.

Chandler, C.S. and Ballard, F.J. (1985) Distribution and degradation of biotin-containing carboxylases in human cell lines. *Biochem. J.* **232**, 385-393.

Chandler, C.S. and Ballard, F.J. (1986) Multiple biotin-containing proteins in 3T3-L1 cells. *Biochem J.* **237**, 123-130.

Chang, H.I. and Cohen, N.D. (1983) Regulation and intracellular localization of the biotin holocarboxylase synthetase of 3T3-L1 cells. *Arch. Biochem. Biophys.* **225**, 237-247.

Chapatawa, K.D., Rajanna, B. and Desaiah, D. (1980) Cadmium induced changes in gluconeogenic enzymes in rat kidney and liver. *Drug Chem. Toxicol.* **3**, 407-420.

Chapatawa, K.D., Boykin, M., Butts, A. and Rajanna, B. (1982) Effect of intraperitoneally injected cadmium on renal and hepatic gluconeogenic enzymes in rats. *Drug Chem. Toxicol.* **5**, 305-317.

Chapman-Smith, A., Booker, G.W., Clements, P.R., Wallace, J.C. and Keech, D.B. (1991) Further studies on the localization of the reactive lysyl residue of pyruvate carboxylase. *Biochem. J.* **276**, 759-764.

Chapman-Smith, A., Forbes, B.E., Wallace, J.C. and Cronan, J.E., Jr. (1997) Covalent modification of an exposed surface turn alters the global conformation of the biotin carrrier domain of *Escherichia coli* acetyl-CoA carboxylase. *J. Biol. Chem.* **273**, 26017-26022.

Chapman-Smith, A. and Cronan, J.E. (1999) Molecular biology of biotin attachment to proteins *J. Nutr.* **129**, 477-484.

Chatelain, F., Pegorier, J-P., Minsassian, C., Bruni, N., Tarpin, S., Girard, J. and Mithieux, G. (1998) Development and regulation of glucose-6-phosphatase gene expression in rat liver, intestine and kidney. *Diabetes* **47**, 882-889.

Chen, M.-J., Shimada, T., Moulton, A.D., Cline, A., Humphries, R.K., Maizel, J. and Niehuis, A.W. (1984) The functional human dihydrofolate reductase gene. *J. Biol. Chem.* **259**, 3933-3943.

Cheng, S., Fockler, C., Barnes, W.M. and Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5695-5699.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Human CCAAT-binding proteins have heterologous subunits. *Cell* **53**, 11-24.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Cípres, G., Urcelay, E., Butta, N., Ayuso, M.S., Parrilla, R. and Martin-Requero, A. (1994) Loss of fatty acid control of gluconeogenesis and PDH complex flux in adrenalectomized rats. *Am. J. Physiol.* **267**, E528-E536.

Clark, J.B., Palmer, C.J. and Shaw, W.N. (1983) The diabetic Zucker fatty rat. *Proc. Soc. Exp. Biol. Med.* **173**, 68-75.

Cohen, N.D., Thomas, M., Stack, M. (1985) in Biotin (Dakshinamurti, K. and Bhagavan, H.N., Eds) *Ann. NY Acad. Sci.* **447**, 393-395.

Cooper, A.J. and Plum, F. (1987) Biochemistry and physiology of cerebral ammonia. *Physiol. Rev.* **67**, 440-519.

Cox, T.C., Bawden, M.J., Martin, A. and May, B.K. (1991) Human erythroid 5-aminolevulinate synthase: promoter analysis and identification of an iron-responsive element in the mRNA. *EMBO J.* **10**, 1891-1902.

Cronan, J.E., Jr. and Wallace, J.C. (1995) The gene encoding the biotin-apoprotein ligase of *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett* **130**, 221-229.

Cullingford, T.E., Dolphin, C.T., Bhakoo, K.K., Peuchen, S., Canevari, L. and Clark, J.B. (1998) Molecular cloning of rat mitochondrial 3-hydroxy-3-methylglutaryl-CoA lyase and detection of the corresponding mRNA and of those encoding the remaining enzymes comprising the ketogenic 3-hydroxy-3-methylglutaryl-CoA cycle in central nervous system of suckling rat. *Biochem. J.* **329**, 373-381.

Dardel, F., Davis, A.L., Laue, E.D. and Perham, R.N.(1993) The three dimensional structure of the lipolyl domain from *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex. *J. Mol. Biol.* **229**, 1037-1048.

Darville, M.I., Crepin, K.M., Hue, L. and Rousseau, G.G. (1989) 5' flanking sequence and structure of a gene encoding rat 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6543-6547.

Darville, M.I., Chikri, M., Lebeau, E., Hue, L. and Rousseau, G.G. (1991) A rat gene encoding heart 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. *FEBS Lett.* **288**, 91-94.

Dawson, R.M.C., Elliott, D.C., Elliott, W.H. and Jones, K.M. (1969) Data for Biochemical Research, Ed. 2, Oxford University Press, p192.

De Vos, A., Heimberg, H., Quartier, E., Yuypens, P., Bouwens, L., Pipeleers, D. and Schuit, F. (1995) Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression. *J. Clin. Invest.* **96**, 2489-2495.

Detimary, P., Dejonghe, S., Ling, Z., Pipeleers, D., Schuit, F. and Henquin, J.-C. (1998) The changes in adenine nucleotides measured in glucose-stimulated rodent islets occur in β cells but not in α cells and are also observed in human islets. *J. Biol. Chem.* **273**, 33905-33908.

Dimroth, P. and Thorner, A. (1992) The sodium ion pumping oxaloacetate decarboxylase of *Klebsiella pneumoniae*. *FEBS Letts.* **300**, 67-70.

Doolittle, R.F. (1995) The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **64**, 87-314.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) A multiplicity of CCAAT box-binding proteins. *Cell* **50**, 863-872.

Dunn, M.F., Encarnación, S., Araíza, G., Vargas, M.C., Dávalos, A., Peralta, H., Mora, Y. and Mora, R. (1996) Pyruvate carboxylase from *Rhizobium etli*:: mutant characterization, nucleotide sequence and physiological role. *J. Bacteriol.* **178**, 5960-5970.

Dunn, F., Araíza, G., Cevallos, M.A. and Mora, J. (1997) Regulation of pyruvate carboxylase in *Rhizobium etli*. *FEMS Microbiol. Lett.* **157**, 301-306.

Easterbrook-Smith, S.B., Hudson, P.J., Goss, N.H., Keech, D.B. and Wallace, J.C. (1976) Pyruvate carboxylase : mechanism of the second partial reaction. *Arch. Biochem. Biophys.* **176**, 709-720.

Edwards, J.B.D., Delort, J. and Mallet, J. (1991) Oligodeoxyribonucleotide ligation to single-stranded cDNAs: a new tool for cloning 5' ends of mRNAs and for constructing cDNA libraries by *in vitro* amplification. *Nucleic Acids Res.* **19**, 5227-5232.

El Khadir-Mounier, C., Le Fur, N., Powell, R.S., Diot, C., Langlois, P., Mallard, J. and Douaire, M. (1996) Cloning and characterization of the 5'-end and promoter region of the chicken acetyl-CoA carboxylase gene. *Biochem. J.* **314**, 613-619.

el-Maghrabi, M.R., Pilkis, J., Marker, A.J., Colosia, A.D., D'Angelo, G., Fraser, B.A., Pilkis, S.J. (1988) cDNA sequence of rat liver fructose-1,6-bisphosphatase and evidence for down-regulation of its mRNA by insulin. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8430-8434.

el-Maghrabi, M.R., Lange, A.J., Kummel, L. and Pilkis, S.J. (1991) The rat fructose-1,6-bisphosphatase gene. Structure and regulation of expression. *J. Biol. Chem.* **266**, 2115-2120.

el-Maghrabi, M.R., Lange, A.J., Jiang, W., Yamagata, K. *et al.* (1995) Human fructose-1,6-bisphosphatase gene (FBP1): exon-intron organization, localization to chromosome bands 9q22.2-q22.3, and mutation screening in subjects with fructose-1,6-bisphosphatase deficiency. *Genomics* **10**, 520-525.

Elson, A., Levanon, D., Brandeis, M., Dafni, N., Bernstein, Y., Daniger, E. and Groner, Y. (1990) The structure of the human liver-type phosphofructokinase gene. *Genomics* **7**, 47-56.

Ercolani, L., Florence, B., Denaro, M.and Alexander, M. (1988) Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. *J. Biol. Chem.* **263**, 15335-15341.

Exton, J.H. and Park, C.R. (1967) Control of gluconeogenesis in the liver: Effect of lactate, pyruvate, fructose, glucagon, epinephrine and adenosine 3', 5'-monophosphate on gluconeogenic intermediates in the perfused rat. *J. Biol. Chem.* **244**, 1424-1433.

Faff-Michalak, L. and Albrecht, J. (1991) Aspartate aminotransferase, malate dehydrogenase and pyruvate carboxylase activities in rat cerebral synaptic and nonsynaptic mitochondria: effects of *in vitro* treatment with ammonia, hyperammonemia and hepatic encephalopathy. *Metab. Brain Dis.* **6**, 187-197.

Fahien, L.A., Davis, J.W. and Laboy, J. (1993) Interactions between pyruvate carboxylase and other mitochondrial enzymes. *J. Biol. Chem.* **268**, 17935-17942.

Faisst, S. and Myer, S. (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.* **20**, 3-26.

Fan, C., Moews, P.C., Walsh, C.T. and Knox, J.R. (1994) Vancomycin resistance: Structure of D-alanine: D-alanine ligase at 2.3 Å resolution. *Science* **266**, 439-443.

Filsell, O.H., Jarrett, I.G., Taylor, P.H. and Keech, D.B. (1969) Effects of fasting, diabetes and glucocorticoids on gluconeogenic enzymes in sheep. *Biochim. Biophys. Acta* **184**, 54-63.

Flores, C.-L. and Gancedo, C. (1997) Expression of PEP carboxylase from *Escherichia coli* complements the phenotypic effects of pyruvate carboxylase mutations in *Saccharomyces cerevisiae. FEBS Lett.* **412**, 531-534.

Forest, C.D., O'Brien, R.M., Lucas, P.C., Magnuson, M.A. and Granner, D.K. (1990) Regulation of P-enolpyruvate carboxykinase gene expression by insulin: Use of stable transfection approach to locate an insulin responsive sequence. *Mol. Endocrinol.* **4**, 1302-1310.

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 Utter, M.F. (1980) Induction of pyruvate carboxylase apoenzyme and holoenzyme in 3T3-L1 cells during differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1321-1325.

Freytag. S.O. and Utter, M.F. (1983) Regulation of the synthesis and degradation of pyruvate carboxylase in 3T3-L1 cells. *J. Biol. Chem.* **258**, 6307-6312.

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.

Friedman, N., Exton, J.H. and Park, C.R. (1967) Interaction of adrenal steroids and glucagon on gluconeogenesis in perfused rat liver. *Biochem. Biophys. Res. Commun.* 29, 113-119.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.

Fuchs, J., Johannssen, W., Rohde, M. and Mayer, F. (1988) Pyruvate carboxylase from *Pseudomonas citronellolis*: Shape of the enzyme and localisation of its prosthetic biotin group by electron microscopic affinity labelling. *FEBS Lett.* **231**, 102-106.

Gamberino, W.C., Berkich, D.A., Lynch, C.J., Xu, B. and LaNoue, K.F. (1997) Role of pyruvate carboxylase in facilitating of synthesis the glutamate and glutamine in cultured astrocytes. *J. Neurochem.* **69**, 2312-2325.

Gancedo, J.M. (1992) Carbon metabolite repression in yeast. *Eur. J. Biochem.* **206**, 297-313.

Garcia-Ruiz, J.P., Ingram, R. and Hanson, R.W. (1978) Changes in hepatic mRNA for P-enolpyruvate carboxykinase (GTP) during development. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4189-4193.

Garrison, J.C. and Haynes, R.C., Jr. (1975) The hormonal control of gluconeogenesis by regulation of mitochondrial pyruvate carboxylation in isolated rat liver cells. *J. Biol. Chem.* **250**, 2769-2777.

Garrison, J.C. and Borland, M.K. (1979) Regulation of mitochondrial pyruvate carboxylation and gluconeogenesis in rat hepatocytes via an alpha-adrenergic, adenosine 3':5'-monophosphate-independent mechanism. *J. Biol. Chem.* **254**, 1129-1133.

Ghosp, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P. and Weissman, S.M. (1978) Heterogeneity and 5'-terminal structures of the late RNAs of simian virus 40. *J. Mol. Biol.* **126**, 841-846.

Giallongo, A., Oliva, D., Cali, L., Barba, G., Barbieri, G. and Feo, S. (1990) Structure of the human gene for alpha-enolase. *Eur. J. Biochem.* **190**, 567-573.

Giallongo, A., Venturella, S., Oliva, D., Barbieri, G., Rubino, P. and Feo, S. (1993) Structural features of the human gene for muscle specific enolase: Differential splicing in the 5'-untranslated sequence generates two forms of mRNA. *Eur. J. Biochem.* **214**, 367-374.

Giralt, M., Park, A.E., Liu, J., Gurney, A. and Hanson, R.W. (1991) Identification of a thyroid hormone response element in the P-enolpyruvate carboxykinase (GTP) gene: Evidence for a synergistic interaction between thyroid hormone and cAMP *cis* regulatory elements. *J. Biol. Chem.* **266**, 21991-21996.

Girard, J.R., Cuendet, G.S., Marliss, E.B., Kervran, A., Rietort, M. and Assan, R. (1973) Fuels, hormones and liver metabolism at term and during early postnatal period in the rat. *J. Clin. Investt.* **52**, 3190-3200.

Girard, J.R., Ferre, P., Pegorier, J.P., Duee, P.H. (1992) Adaptations of glucose and fatty acid metabolism during the perinatal period and the suckling-weaning transition. *Physiol. Rev.* **72**, 507-562.

Goodall, G.J., Baldwin, G.S., Wallace, J.C. and Keech, D.B. (1981) Factors that influence the translocation of the N-carboxyl biotin moiety between the two sub-sites of pyruvate carboxylase. *Biochem. J.* **199**, 603-609.

Goss, J.A., Cohen, N.D. and Utter, M.F. (1981) Characterization of the subunit structure of pyruvate carboxylase from *Pseudomonas citronellolis*. *J. Biol. Chem.* **256**, 11819-11825.

Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5547-5551.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. and Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-1769.

Granner, D.K., Andreone, T., Sasaki, K. and Beale, E. (1983) Inhibition of transcription of the P-enolpyruvate carboxykinase gene by insulin. *Nature* **305**, 549-551.

Granner, D.K, and Pilkis, S. (1990) The genes of hepatic glucose metabolism. *J. Biol. Chem.* **265**, 10713-10716.

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K. and Lieber, M.R. (1998) DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Mol. Cell* **2**, 477-484.

Green, J.D.F., Laue, E.D., Perham, R.N., Ali, S.T. and Guest, J.R. (1995) Three-dimensional structure of a lipoyl domain from the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli. J. Mol. Biol.* **248**, 328-343.

Gunnery, S., Maivali, U. and Mathew, M.B. (1997) Translation of an uncapped mRNA involves scanning. *J. Biol. Chem.* **272**, 21642-21646.

Gushima, H., Miya, T., Murata, K. and Kimura, A. (1983) Purification and characterization of glutathione synthetase from *Escherichia coli. J. Appl. Biochem.* **5**, 210-218.

Guthrie, C. and Patterson, B. (1988) Spiceosomal snRNAs. *Annu. Rev. Genet.* **22**, 387-419.

Ha, J., Daniel, S., Kong, I.-S., Park, C.-K., Tae, H.-J. and Kim, K.-H. (1994) Cloning of the human acetyl-CoA carboxylase cDNA. *Eur. J. Biochem.* **219**, 297-306.

Ha, J., Lee, J.-K., Kim, K.-S., Witters, L.A. and Kim, K.-H. (1996) Cloning of human acetyl-CoA carboxylase-β and its unique features. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11466-11470.

Haarasilta, S. and Oura, E. (1975) Effect of aeration on the activity of gluconeogenetic enzymes in *Saccharomyces cerevisiage* growing under glucose limitation. *Arch. Microbiol.* **106**, 271-273.

Halestrap, A.P. (1978) Stimulation of pyruvate transport in metabolizing mitochondria through changes in the transmembrane pH gradient induced by glucagon treatment of rats. *Biochem. J.* **172**, 389-398.

Halestrap, A.P. and Armston, A.E. (1984) A re-evaluation of the role of mitochondrial pyruvate transport in the hormonal control of rat liver mitochondrial pyruvate metabolism. *Biochem. J.* **223**, 677-685.

Hanson, R.W. and Patel, M.S. (1994) Phosphoenolpyruvate carboxykinase (GTP): the gene and the enzyme. *Adv. Enzymol. Relat. Areas. Mol. Biol*. **69**, 203-281.

Hanson, R.W.and Reshef, L. (1997) Regulation of phosphoenolpyruvate carboxykinase (GTP) gene expression. *Annu. Rev. Biochem.* **66**, 581-611.

Hendrick, J.P., Hodges, P.E. and Rosenberg, L.E. (1989) Survey of amino-terminal protcolytic cleavage sites in mitochondrial precursor proteins: leader peptides cleaved by two matrix proteases share a three-amino acid motif. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4056-4060.

Henrikson, K.P., Allen, S.H.G. and Malloy, W.L. (1979) An avidin monomer affinity column for the purification of biotin-containing enzymes. *Anal. Biochem.* **94**, 366-370.

Heimberg, H., De Vos, A., Vandercammen, A., Van Schaftingen, E., Pipeleers, D. and Schuit, F. (1993) Heterogeneity in glucose sensitivity among pancreatic beta-cells is correlated to differences in glucose phosphorylation rather than glucose transport. *EMBO J.* **12**, 2873-2879.

Herbomel, P., Bourachot, B. and Yaniv, M. (1984) Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. *Cell* **39**, 653-662.

Hobbs, S.M., Jitrapakdee, S. and Wallace, J.C. (1998) Development of a bicistronic vector driven by polypeptide chain elongation factor 1α promoter for the creation of stable mammalian cell lines that express very high levels of recombinant proteins. *Biochem. Biophys. Res. Commun.* **252**, 368-372.

Hod, Y. and Hanson, R.W. (1988) Cyclic AMP stabilizes the mRNA for phosphoenolpyruvate carboxykinase (GTP) against degradation. *J. Biol. Chem.* **268**, 7747-7752.

Hoffman, E.K., Trusko, S.P., Freeman, N.A. and George, D.L. (1987) Structural and functional characterization of the promoter region of the mouse c-Ki-ras gene. *Mol. Cell. Biol.* **7**, 2592-2596.

Howard, P.K., Shaw, J. and Otsuka, A.J. (1985) Nucleotide sequence of the birA gene encoding the biotin operon repressor and biotin holoenzyme synthetase functions of *Escherichia coli. Gene* **35**, 321-331.

Hsu, M.H., Chirala, S.S. and Wakil, S.J. (1996) Human fatty-acid synthase gene: evidence for the presence of two promoters and their functional interaction. *J. Biol. Chem.* **271**, 13584-13592.

Huang, M.T.E. and Gorman, C.M. (1990) Intervening sequences increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA. *Nucleic Acids Res.* **18**, 937-947.

Imagawa, M., Chiu, R. and Karin, M. (1987) Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell* **51**, 251-260.

Imai, E., Stromstedt, P., Quinn, P.G., Carlstedt-Duke, J., Gunstaffson, J. and Granner, D.K. (1990) Characterization of a complex glucocorticoid response unit in the P-enolpyruvate carboxykinase gene. *Mol. Cell. Biol.* **10**, 4712-4719.

Irias, J.J., Olmsted, M.R. and Utter, M.F. (1969) Pyruvate carboxylase : irreversible inactivation by cold. *Biochemistry* **8**, 5136-5148.

Jacob, H.J., Brown, D.M., Bunker, R.K., Daly, M.J., Dzau, V.J., Goodman, A., Koike, G., Krem, V., Kurtz, T., Lernmark, T., Levan, G., Moa, Y., Petterson, A., Pravenec, M., Simon, K.S., Szpirer, C., Szpirer, J., Trolliet, M.R., Winer, E.S., Lander, E.S. (1995) A genetic linkage map of the laboratory rat, *Rattus norvegicus. Nature Genet* . **9**, 63-69.

Jackson, R.J., Howell, M.T. and Kaminski, A. (1990) The novel mechanism of initiation of picornavirus RNA translation. *Trends Biochem. Sci.* **15**, 477-483.

Jang, S.K., Pestova, T.V., Hellen, C.U., Witherell, G.W. and Wimmer, E. (1988) Cap-independent translation of picornavirus RNAs: structure and function of the internal ribosomal entry site. *Enzyme* **44**, 292-309.

Javid-Majd, F., Stapleton, M.A., Harmon, M.F., Hanks, B.A., Mullins, L.S. and Raushel, F.M. (1996) Comparison of the functional differences for the homologous residues within the carboxy phosphate and carbamate domains of carbamoyl phosphate synthetase. *Biochemistry* **35**, 14362-14369.

Jiang, W. and Bond, J.S. (1992) Families of metalloendopeptidases and their relationships. *FEBS. Letts.* **312**, 110-114.

Jitrapakdee, S., Booker, G.W., Cassady, A.I. and Wallace, J.C. (1996) Cloning, sequencing and expression of rat liver pyruvate carboxylase. *Biochem. J.* **316**, 631-637.

Jitrapakdee, S., Booker, G.W., Cassady, A.I. and Wallace, J.C.(1997) The rat pyruvate carboxylase gene structure : Alternate promoters generate multiple transcripts with 5'-end heterogeneity. *J. Biol. Chem.* **272**, 20522-20530.

Jitrapakdee, S. and Wallace, J.C. (1999) Structure, function and regulation of pyruvate carboxylase. *Biochem. J.* (In press).

Johannssen, W., Attwood, P.V., Keech, D.B. and Wallace, J.C. (1983) Localisation of the active site of pyruvate carboxylase by electron microscopic examination of avidin-enzyme complexes. *Eur. J. Biochem.* **133**, 201-206.

Johnson, T.M., Rosenberg, M.P. and Meisler, M.H. (1993) An insulin-responsive element in the pancreatic enhancer of the amylase gene. *J. Biol. Chem.* **268**, 464-468.

Jones, D.H. and Winistorfer, S.C. (1993) Genome walking with 2- to 4-kb steps using panhandle PCR. *PCR Method Appl.* **2**, 197-203.

Jones, C.G., Hothi, S.K. and Titheradge, M.A. (1993) Effect of dexamethasone on gluconeogenesis, pyruvate kinase, pyruvate carboxylase and pyruvate dehydrogenase flux in isolated hepatocytes. *Biochem. J.* **289**, 821-828.

Kanno, H., Fujii, H. and Miwa, S. (1992) Structural analysis of human pyruvate kinase L-gene and identification of the promoter activity in erythroid cells. *Biochem. Biophys. Res. Commun.* **188**, 516-523.

Keech, D.B. and Utter, M.F. (1963) Pyruvate carboxylase II. Properties. *J. Biol. Chem.* **238**, 2609-2614.

Keech, D.B. and Barritt, G.J. (1967) Allosteric activation of sheep kidney pyruvate carboxylase by magnesium ion (Mg^{2+}) and the magnesium adenosine triphosphate ion ($MgATP^{2-}$). *J. Biol. Chem.* **242**, 1983-1987.

Keech, D.B. (1985) Historic background, in Pyruvate carboxylase (Keech, D.B. and Wallace, J.C. eds), CRC Series in Enzyme Biology, CRC Press, Boca Raton, FL, p 1-3.

Khew-Goodall, Y.-S. (1985) Pyruvate carboxylase: Its interactions with acetyl-CoA. Ph.D. thesis, University of Adelaide, South Australia.

Khew-Goodall, Y.-S., Johannssen, W., Attwood, P.V., Wallace, J.C. and Keech, D.B. (1991) Studies on dilution inactivation of sheep liver pyruvate carboxylase. *Arch. Biochem. Biophys.* **284**, 98-105.

Kim, D.W., Uetsuki, T., Kaziro, Y., Yamaguchi, N. and Sugano, S. (1990) Use of the human elongation factor 1α promoter as a versatile and efficient expression system. *Gene* **91**, 217-223.

Kitajima, S. and Uyeda, K. (1982) A binding study of the interaction of beta-D-fructose 2,6-bisphosphate with phosphofructokinase and fructose-1,6-bisphosphatase. *J. Biol. Chem.* **258**, 7352-7357.

Knight, G.B., Gudas, J.M. and Pardee, A.B. (1987) Cell-cycle-specific interaction of nuclear DNA-binding proteins with a CCAAT sequence from the human thymidine kinase gene. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8350-8353.

Knowles, J.R. (1989) The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* **58**, 195-221.

Kondo, H., Sriratsuchi, K., Yoshimoto, T., Matsuda, T., Kitazono, A., Tsuru, D., Anai, M., Sekiguchi, M. and Tanabe, T. (1991) Acetyl-CoA carboxylase from *Escherichia coli*: gene organization and nucleotide sequence of the biotin carboxylase subunit. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9730-9733.

Kondo, H., Kazuta, Y., Saito, A and Fuji, K.-I. (1997). Cloning and nucleotide sequence of *Bacillus stearothermophilus* pyruvate carboxylase. *Gene* **191**, 47-50.

Koffas, M.A.G., Ramamoorthi, R., Pine, W.A., Sinskey, A.J. and Stephanopoulos, G. (1998) Sequence of the *Corynebacterium glutamicum* pyruvate carboxylase gene. *Appl. Microbiol. Biotechnol.* **50**, 346-352.

Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acid Res.* **15**, 8125-8148.

Kozak, M. (1988) A profusion of controls. *J. Cell Biol.* **107**, 1-7.

Kozak, M. (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eukaryotic mRNAs. *Mol. Cell. Biol.* **9**, 5134-5142.

Kozak, M. (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**, 887-903.

Kozak, M. (1992) A consideration of alternative models for the initiation of translation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **27**, 385-402.

Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Krause-Friedmann, N. and Feng, L. (1996) The role of intrancellular Ca^{2+} in the regulation of gluconeogenesis. *Metabolism* **45**, 389-403.

Krebs, H.A. (1954) Considerations concerning the pathways of synthesis in living matter. *Bull. John Hopkins Hosp.* **95**, 19.

Kumar, G.K., Haase, F.C., Phillips, N.F.B. and Wood, H.G. (1988) Involvement and identification of a tryptophanyl residue at the pyruvate binding site of transcarboxylase. *Biochemistry*, **27**, 5978-5983.

Kunst, F. *et al.* (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis. Nature* **390**, 249-256.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lagerstrom, M.J., Parik, J., Malmgren, H., Stewart, J,. Pettersson, U. and Landegren, U. (1991) Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *PCR Method Appl.* **1**, 111-119.

Lamhonwah, A.-M., Barankiewicz, T.J., Willard, H.F., Mahuran, D.J., Quan, F. and Grave, R.A. (1986) Isolation of cDNA clones coding for the alpha and beta chains of human propionyl-CoA carboxylase: chromosomal assignments and DNA polymorphisms associated with PCCA and PCCB genes. *Proc. Natl. Acad. Sci. U.S.A.* **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 pyruvate carboxylase. *Arch. Biochem. Biophys.* **254**, 631-636.

Landschulz, W.H., Johnson, P.F., Adashi, E.Y., Graves, B.J. and McKnight, S.L.(1988) Isolation of a recombinant copy of the gene encoding C/EBP. *Genes & Dev.* **2**, 786-792.

Lane, M.D., Rominger, K.L, Young, D.L. and Lynen, F. (1964) The enzymatic synthesis of holotranscarboxylase from apotranscarboxylase and (+)-biotin: II investigation of the reaction mechanism. *J. Biol. Chem.* **239**, 2865-2871.

Lane, M.D., Moss, J.D. and Polakis, S.E. (1974) Acetyl coenzyme A carboxylase. *Curr. Top. Cell Regul.* **8**, 139-195.

Lange, A.J., Argaud, D., el-Mahgrabi, M.R., Pan, W., Maitra, S.R. and Pilkis, S.J. (1994) Isolation of a cDNA for the catalytic subunit of rat liver glucose-6-phosphatase: regulation of gene expression in FAO hepatoma cells by insulin, dexamethasone and cAMP. *Biochem. Biophys. Res. Commun.* **201**, 302-309.

Lau, E.P., Cochran, B.C. and Fall, R.R. (1979) Isolation of 3-methylcrotonyl-coenzyme A carboxylase from bovine kidney. *Arch. Biochem. Biophys.* **205**, 352-359.

Lehn, D.A., Moran, S.M. and MacDonald, M.J. (1995) The sequence of the rat pyruvate carboxylase-encoding cDNA. *Gene* **165**, 331-332.

Leff, S.E., Evans, R.M. and Rosenfeld, M.G. (1987) Splice commitment dictates neuron-specific-alternative RNA processing in calcitonin/CGRP gene expression. *Cell* **45**, 517-524.

Lei, K.J., Shelly, L.L., Pan, C.J., Sidbury, J.B. and Chou, J.Y. (1993) Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a. *Science* **262**, 580-583.

Leiter, A.B., Weinberg, M., Isohashi, F., Utter, M.F. and Linn, L. (1978) Relationship between phosphorylation and activity of pyruvate dehydrogenase in rat liver mitochondria and the absence of such a relationship for pyruvate carboxylase. *J. Biol. Chem.* **253**, 2716-2723.

Lemaigre, F.P. and Rousseau, G.G. (1994) Transcriptional control of genes that regulate glycolysis and gluconeogenesis in adult liver. *Biochem. J.* **303**, 1-14.

Lemons, J.A., Moorehead, H.C. and Hage, G.P. (1986) Effects of fasting on gluconeogenic enzymes in the ovine fetus. *Pediatr. Res.* **20**, 676-679.

Leon-Del-Rio, A. and Gravel, R.A. (1994) Sequence requirements for the biotinylation of carboxyl-terminal fragments of human-propionyl-CoA carboxylase α subunit expressed in *Escherichia coli. J. Biol. Chem.* **269**, 22964-22968.

Leon-Del-Rio, A., Leclerc, D., Akerman, B., Wakamatsu, N. and Gravel, R.A. (1995) Isolation of a cDNA encoding human holocarboxylase synthetase by functional complementation of a biotin auxotroph of *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.* **92**, 4626-4630.

Levan, G., Klinga Levan, K., Szpirer, C., Szpirer, J. (1992) Genetic Maps 4, Nonhuman Vertebrates, 6th edn (O'Brien SJ edit) Cold Spring Harbor Press, NY.

Li, S. and Cronan, J.E., Jr (1992a) The gene encoding the biotin carboxylase subunit of *Escherichia coli* acetyl-CoA carboxylase. *J. Biol. Chem.* **267**, 855-863.

Li, S. and Cronan, J.E., Jr (1992b) The genes encoding the two carboxyltransferase subunits of *Escherichia coli* acetyl-CoA carboxylase. *J. Biol. Chem.* **267**, 16841-16847.

Li, N., Seetharam, S. and Seetharam, B. (1995) Genomic structure of human transcobalamin II: Comparison to human intrinsic factor and transcobalamin I. *Biochem. Biophys. Res. Commun.* **208**, 756-764.

Li, Y.F., Chen, M.C., Lin, Y.H., Hsu, C.C. and Tsai, Y.C. (1998) GenBank accession number AF097728.

Liang, Y., Najafi, H., Smith, R.M., Zimmerman, E.C., Magnuson, M., Tal, M. and Matschinsky, F.M. (1992) Concordant glucose induction of glucokinase, glucose usage, and glucose-stimulated insulin release in pancreatic islets maintained in organ culture. *Diabetes* **41**, 792-806.

Liao, V.H.-C.and Freedman, J.H. (1998) Cadmium-regulated genes from the nematode *Caenorhabditis elegans. J. Biol. Chem.* **48**, 31962-31970.

Libor, S.M., Sundaram, T.K. and Scrutton, M.C. (1978) Pyruvate carboxylase from a thermophilic *Bacillus*: Studies on the specificity of activation by acyl derivatives of coenzyme A and on the properties of catalysis in the absence of activator. *Biochem. J.* **169**, 543-558.

Lim, F., Morris, C.P., Occhiodoro, F. and Wallace, J.C. (1988) Sequence and domain structure of yeast pyruvate carboxylase. *J. Biol. Chem.* **263**, 11493-11497.

Lin, B., Morris, D.W. and Chou, J.Y. (1997) The role of HNF1α, HNF3γ and cyclic AMP in glucose-6-phosphatase gene activation. *Biochemistry* **36**, 14096-14106.

Lin, B., Morris, D.W. and Chou, J.Y. (1998) Hepatocyte nuclear factor 1alpha is an accessory factor required for activation of glucose-6-phosphatase gene transcription by glucocorticoids. *DNA and Cell Biol.* **17**, 967-974.

Liu, F. and Fromm, H.J. (1988) Purification and characterization of fructose-1,6-biphosphatase from bovine brain. *Arch. Biochem. Biophys.* **260**, 609-615.

Liu, J., Park, E.A., Gurney, A.L., Roesler, W.J. and Hanson, R.W. (1991) Cyclic AMP induction of P-enolpyruvate carboxykinase gene transcription is mediated by multiple promoter elements. *J. Biol. Chem.* **266**, 19095-19102.

Loose, D., Shaw, P., Krauter, K.S., Robinson, C., Englard, S., Hanson, R.W. and Gluechsohn-Waelsch, S. (1986) Trans-regulation of the P-enolpyruvate carboxykinase gene identified by deletions in chromosome 7 of the mouse. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5184-5188.

Lopez-Casillas, F., Bai, D.H., Luo, X.C.L., Kong, I.-S., Hermodson, M.A. and Kim, K.-H. (1988) Structure of the coding sequence and primary amino acid sequence of acetyl-CoA carboxylase. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5784-5788.

Lopez-Casillas, F., Luo, X., Kong, I.-S. and Kim, K.-H. (1989) Characterization of different forms of rat mammary gland acetyl-CoA carboxylase mRNA: analysis of heterogeneity in the 5' end. *Gene* **83**, 311-319.

Lopez-Casillas, F. and Kim, K.-H.(1989) Heterogeneity at the 5' end of rat acetyl-CoA carboxylase mRNA. *J. Biol. Chem.* **264**, 7176-7184.

Lopez-Casillas, F., Ponce-Castaneda, M.V. and Kim, K.-H. (1991) *In vivo* regulation of the activity of the two promoters of the rat acetyl coenzyme-A carboxylase gene. *Endocrinol.* **129**, 1049-1058.

Lopez-Casillas, F. and Kim, K.-H. (1991) The 5'-untranslated regions of acetyl-coenzyme A carboxylase mRNA provide specific translational control *in vitro. Eur. J. Biochem.* **201**, 119-127.

Lucas, P.C., O'Brien, R.M., Mitchell, J.A., Davis, C.M., Imai, E., Forman, B.M., Samuels, H.H. and Granner, D.K. (1991) A retinoic acid response element is part of a pleiotropic domain in the phosphoenolpyruvate carboxykinase gene. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2184-2188.

Lukyanov, S.A. *et al.*(1994) *Bioorganic Chem.* **20**, 701-704.

Luo, X.-C., Park, K., Lopez-casillas, F. and Kim, K.-H. (1989) Structural features of the acetyl-CoA carboxylase gene: Mechanisms for generation of mRNA with 5'-end heterogeneity. *Proc. Nat. Acad. Sci. U.S.A.* **86**, 4042-4046.

Luo, X.-C. and Kim, K.-H. (1990) An enhancer element in the house-keeping promoter for acetyl-CoA carboxylase gene. *Nucleic Acids Res.* **18**, 3249-3254.

Lusty, C.J., Widgren, E.E., Brogile, K.E. and Nyunoya, H. (1983) Yeast carbamoylphosphate synthetase: structure of the yeast gene and homology to *Escherichia coli* carbamoylphosphate synthetase. *J. Biol. Chem.* **258**, 14466-14477.

Lynch, C.J., McCall, K.M., Billingsley, M.L., Bohlen, L.M., Hreniuk, S.P., Martin, L.F., Witters, L.A. and Vannucci, S.J. (1992) Pyruvate carboxylase in genetic obesity. *Am. J. Physiol.* **262**, E608-E618.

MacDonald, M.J. (1981) High content of mitochondrial glycerol-3-phosphate dehydrogenase in pancreatic islets and its inhibition by diazoxide. *J. Biol. Chem.* **256**, 8287-8290.

MacDonald, M.J. and Chang, C.-M. (1985) Do pancreatic islets contain significant amounts of phosphoenolpyruvate carboxykinase or ferroactivator activity ? *Diabetes* **34**, 246-250.

MacDonald, M.J. (1990) Elusive proximal signals of beta-cells for insulin secretion. *Diabetes* **39**, 1461-1466.

MacDonald, M.J., Kaysen, J.H., Moran, S.M. and Pomije, C.E. (1991) Pyruvate dehydrogenase and pyruvate carboxylase: sites of pretranslational regulation by glucose of glucose-induced insulin release in pancreatic islets. *J. Biol. Chem.* **266**, 22392-22397.

MacDonald, M.J., McKenzie, D.I., Walker, T.M. and Kaysen, J.H. (1992) Lack of glyconeogenesis in pancreatic islets: expression of gluconcogenic enzyme genes in islets. *Horm. Metab. Res.* **24**, 158-160.

MacDonald, M.J. (1995a) Influence of glucose on pyruvate carboxylase expression in pancreatic islets. *Arch. Biochem. Biophys.* **319**, 128-132.

MacDonald, M.J. (1995b) Feasibility of a mitochondrial pyruvate malate shuttle in pancreatic islets. *J. Biol. Chem.* **270**, 20051-20058.

MacDonald, M.J., Efendic, S. and Otenson, C.-G. (1996) Normalization by insulin treatment of low mitochondrial glycerol phosphate dehydrogenase and pyruvate carboxylase in pancreatic islets of the GK rat. *Diabetes* **45**, 886-890.

Mackall, J.C. and Lane, M.D. (1977) Role of pyruvate carboxylase in fatty acid synthesis: alterations during preadipocyte differentiation. *Biochem. Biophys. Res. Commun.* **79**, 720-725.

MacKay, N., Rigay, B., Douglas, C., Chen, H.-S. and Robinson, B.H. (1994) cDNA cloning of human kidney pyruvate carboxylase. *Biochem. Biophys. Res.Commun.* **202**, 1009-1014.

Magnuson, M.A., Quinn, P.G. and Granner, D.K. (1987) Multihormonal regulation of phosphoenolpyruvate carboxykinase-chloramphenicol acetyltransferase fusion genes. *J. Biol. Chem.* **262**, 14917-14920.

Magnuson, M.A., Andreone, T.L., Printz, R.L., Koch, S. and Granner, D.K. (1989) Rat glucokinase gene: structure and regulation by insulin. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4838-4842.

Magnuson, M.A. and Shelton, K.D. (1989) An alternate promoter in the glucokinase gene is active in the pancreatic beta cell. *J. Biol. Chem.* **264**, 15936-15942.

Maity, S.N., Golumbek, P.T., Karsenty, G. and de Crombrugghe, B. (1988) Selective activation of transcription by a novel CCAAT binding factor. *Science* **241**, 582-585.

Maity, S.N. and de Crombrugghe, B. (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem. Sci.* **23**, 174-178.

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., Goodbourn, S. and Fischer, J.A. (1987) Regulation of inducible and tissue-specific gene expression. *Science*, **236**, 1237-1241.

Maniatis, T., and Reed, R. (1987) The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* **325**, 673-678.

Martin, A.D., Allan, E.H. and Titheradge, M.A. (1984) The stimulation of mitochondrial pyruvate carboxylation after dexamethasone treatment of rats. *Biochem. J.* **219**, 107-115.

Martinez, E., Chiang, C.-M., Ge, H. and Roeder, R.G. (1994) TATA-binding protein-associated factors(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.* **13**, 3115-3126.

Massillon, D., Barzilai, N., Chen, W., Hu, M. and Rossetti, L. (1996) Glucose-6-phosphatase gene expression in the livers of diabetic rats. *J. Biol. Chem.* **271**, 9871-9874.

Matschinsky, F.M. (1996) A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm. *Diabetes* **45**, 223-241.

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 holoenzyme synthetase and the specificity of holocarboxylase formation. *J. Biol. Chem.* **241**, 2855-2861.

McClure, W.R., Lardy, H.A. and Cleland, W.W. (1971) Rat liver pyruvate carboxylase: isotopic exchange studies of the first partial reaction. *J. Biol. Chem.* **246**, 3584-3590.

McCormack, J.G., Halestrap, A.P. and Denton, R.M. (1990) Role of calcium ions in regulation of mammalian intramitochondrial metabolism. *Physiol. Rev.* **70**, 391-425.

McGrane, M.M., Yun, J.S., Moorman, A.F.M., Lamers, W.H., Hendrick, G.K., Arafah, B.M. and Park, E.A. (1990) Metabolic effects of developmental, tissue and cell specific expression of a chimeric PEPCK/bGH gene in transgenic mice. *J. Biol. Chem.* **265**, 22371-22379.

McKnight, S. and Tjian, R. (1986) Transcriptional selectivity of viral genes in mammalian cells. *Cell* **46**, 795-805.

McVey, J.H., Nomura, S., Kelly, P., Mason, I.J. and Hogan, B.L.M. (1988) Characterization of the mouse sparc/osteonectin gene. *J. Biol. Chem.* **263**, 11111-11116.

Means, A.L. and Farnham, P.J. (1990) Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. *Mol. Cell. Biol.* **10**, 653-661.

Melton, D.W., McEwan, C., McKie, A.B. and Reid, A.M. (1986) Expression of the mouse HPRT gene: deletion analysis of the promoter of an X-chromosome linked housekeeping gene. *Cell* **44**, 319-328.

Menéndez, J., Delgado, J. and Gancedo, C. (1998) Isolation of the *Pichia pastoris PYC1* gene encoding pyruvate carboxylase and identification of a suppressor of the *pyc* phenotype. *Yeast* **14**, 647-654.

Menéndez, J. and Gancedo, C. (1998) Regulatory regions in the promoters of the *Saccharomyces cerevisiae PYC1* and *PYC2* genes encoding isoenzymes of pyruvate carboxylase. *FEMS Microbiol. Lett.* **164**, 345-352.

Merrit, E.A. and Murphy, M. (1994) Raster3D version 2.0-a program for photorealistic molecular graphics. *Acta Cryst* . D **50**, 869-873.

Michelson, A.M., Blake, C.C., Evans, S.T. and Orkin, S.H. (1985) Structure of the human phosphoglycerate kinase gene and the intron-mediated evolution and dispersal of the nucleotide-binding domain. *Proc. Nat. Acad. Sci. U.S.A.* **82**, 6965-6969.

Middleditch, C., Clottes, E. and Burchell, A. (1998) A different isoform of the transport protein mutated in the glycogen storage disease 1b is expressed in brain. *FEBS Lett* **433**, 33-36.

Mitchelle, P.J., Carothers, A.M., Han, J.H., Harding, J.D., Kas, E., Venolia, L. and Chasin, L.A. (1986) Multiple transcription start sites, DNase I-hypersensitive sites, and an opposite-strand exon in the 5'-region of the CHO dhfr gene. *Mol. Cell. Biol.* **6**, 425-440.

Mitchelle, P.J., Wang, C. and Tjiang, R. (1987) Positive and negative regulation of transcription *in vitro*: enhancer-binding protein AP-2 is inhibited by SV40T antigen. *Cell* **50**, 847-861.

Mitchell, G.A., Robert, M.-F., Hruz, P.W., Wang, S., Fontaine, G., Behnke, C.E., Mende-Mueller, L.M., Schappert, K., Lee, C., Gibson, K.M. and Miziorko, H.M. (1993) 3-Hydroxy-3-methylglutaryl coenzyme A lyase (HL). Cloning of human and chicken liver HL cDNAs and characterization of a mutation causing human HL deficiency. *J. Biol. Chem.* **268**, 4376-4381.

Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., Sudo, Y., Miyata, T. and Taniguchi, T. (1988) Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN-β gene regulatory elements. *Cell* **54**, 903-913.

Mizushima, S. and Nagata, S. (1990) pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18**, 5322.

Modak, H.V. and Kelly, D.J. (1995) Acetyl-CoA-dependent pyruvate carboxylase from the photosynthetic bacterium *Rhodobacter capsulatus*: rapid and efficient purification using dye-ligand affinity chromatography. *Microbiology* **141**, 2619-2628.

Moss, J. and Lane, M.D. (1972) The biotin-dependent enzymes. *Adv. Enzymol.* **35**, 321-442.

Mozier, N.M., Walsh, M.P. and Pearson, J.D. (1991) Characterization of a novel zinc binding site of protein kinase C inhibitor-1. *FEBS. Lett.* **279**, 14-18.

Mukhopadhyay, B., Stoddard, S.F. and Wolfe, R. (1998) Purification, Regulation and Molecular and biochemical characterization of pyruvate carboxylase from *Methanobacterium thermoautotrophicum* strain ΔH. *J. Biol. Chem.* **273**, 5155-5156.

Mullis, K.B. and Faloona, F.A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. In Methods in Enzymology, (Wu, R. eds.) **155**, 335-350, San Diego, Academic Press.

Nasrin, N., Buggs, C., Kong, X.-F., Carnazza, J., Goebl, M. and Alexander-Bridges, M. (1991) DNA-binding properties of the product of the testis-determining gene and a related protein. *Nature* **354**, 317-320.

Newgard, C.B. and McGarry, J.D. (1995) Metabolic coupling factors in pancreatic beta-cell. *Annu. Rev. Biochem.* **64**, 689-719.

Nezic, N.G. and Wallace, J.C. (1999) Pyruvate carboxylase: Identification of the functional role of conserved residues in yeast by site-directed mutagenesis. *Proceedings of the 24th Lorne Protein Conference*, C-26.

Noguchi, T., Yamada, K., Inoue, H., Matsuda, T. and Tanaka, T. (1987) The L- and R-type isozymes of rat pyruvate kinase are produced from a single gene by use of different promoters. *J. Biol. Chem.* **262**, 14366-14371.

Nordlie, R.C., Arion, W.J., Hanson, T.L., Gelsdorf, J.R. and Horne, R.N. (1968) Biological regulation of liver microsomal inorganic pyrophosphate-glucose transferase, glucose-6-phosphatase, and inorganic pyrophosphatase. *J. Biol. Chem.* **243**, 1140-1146.

Nordie, R. and Sukalski, K. (1985) Multifunctional glucose-6-phosphatase: a critical review. *In Enzymes of Biological Membranes.* Martonosi, A. Ed., New York, Plenum, p349-398.

Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983) The nucleotide sequence of the rat cytoplasmic beta-actin gene. *Nucleic Acid. Res.* **11**, 1759-1771.

Nyunoya, H. and Lusty, C.J. (1983) The carB gene of *Escherichia coli*: a duplicated gene coding for the large subunit of carbamoyl-phosphate synthetase. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4629-4633.

Obermayer, M. and Lynen, F. (1976) Structure of biotin enzymes. *Trends Biochem. Sci.* **1**, 169-171.

O'Brien, R.M. and Granner, D.K. (1990) PEPCK gene as model of inhibitory effects of insulin on gene transcription. *Diabetes Care* **13**, 327-339.

O'Brien, R.M., Lucas, P.C., Forest, C.D., Magnuson, M.A. and Granner, D.A. (1990) Identification of a sequence in the P-enolpyruvate carboxykinase gene that mediates the negative effect of insulin on transcription. *Science* **249**, 533-537.

O'Brien, R.M., Bonovitch, M.T., Forest, C.D. and Granner, D.K. (1991) Signal transduction convergence: phorbol esters and insulin inhibit P-enolpyruvate carboxykinase gene transcription through the same 10-bp sequence. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6580-6584.

O'Brien, R.M., Lucas, P.C., Yamasaki, T., Noisin, E.L. and Granner, D.K. (1994) Potential convergence of insulin and cAMP signal transduction systems at the phosphoenolpyruvate carboxykinase (PEPCK) gene promoter through CCAAT/enhancer binding protein (C/EBP) *J. Biol. Chem.* **269**, 30419-30428.

Ochman, H., Gerber, A.S., Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.

Ogita, T. and Knowles, J.R. (1988) On the intermediacy of carboxyphosphate in biotin-dependent carboxylations. *Biochemistry* **27**, 8028-8033.

Ohneda, M., Johnson, J.H., Inman, L.R., Chen, L., Suzuki, K.-I., Goto, Y., Alam, T., Ravazzola, M., Orci, L. and Unger, R.H. (1993) GLUT2 expression and function in β-cells of GK rats with NIDDM: dissociation between reductions in glucose transport and glucose-stimulated insulin secretion. *Diabetes* **42**, 1065-1072.

Oliva, D., Cali, L., Feo, S. and Giallongo, A. (1991) Complete structure of the human gene encoding neuron-specific enolase. *Genomics* **10**, 157-165.

Oliva, D., Venturella, S., Passantino, R., Feo, S. and Giallongo, A. (1995) Conserved alternative splicing in the 5'-untranslated region of the muscle-specific enolase gene. Primary structure of mRNAs, expression and influence of secondary structure on the translation efficiency. *Eur. J. Biochem.* **232**, 141-149.

Osmani, S.A., Mayer, F., Marston, F.A.O., Selmes, I.P. and Scrutton, M.C. (1984) Pyruvate carboxylse from *Aspergillus nidulans*: effects of regulatory modifiers on the structure of the enzyme. *Eur. J. Biochem.* **139**, 509-518.

Osmani, S.A., Scrutton, M.C. and Mayer, F. (1985) in Biotin (Dakshinamurti, K. and Bhagavan, H.N., Eds), Ann. N.Y. Acad. Sci. Series, vol. 447, pp. 56-71, New York

Owen, O.E, Patel, M.S., Block, B.S.B., Kreulen, T.H., Reichle, F.A. and Mozzoli, M.A., in Gluconeogenesis (1976), p 533, (Hanson, R.W. and Mehlman, M.A., Eds), John Wiley & Sons, New York.

Owen, M.R. and Halestrap, A.P. (1993) The mechanisms by which mild respiratory chain inhibitors inhibit hepatic gluconeogenesis. *Biochim. Biophys. Acta* **1142**, 11-22.

Palmer, J.D. and Logsdon, J.M. (1991) The recent origins of introns. *Curr. Opin. Gene. Dev.* **1**, 470-477.

Parker, J.D., Rabinovitch, P.S., Burmer, G.C. (1991) Targeted gene walking polymerase chain reaction. *Nucleic Acids Res.* **19**, 3055-3060.

Peshavaria, M. and Day, I.N. (1991) Molecular structure of the human muscle-specific enolase gene (ENO3). *Biochem J.* **275**, 427-433.

Petersen, D.D., Koch, S.R. and Granner, D.K. (1989) 3' noncoding region of phosphoenolpyruvate carboxykinase mRNA contains a glucocorticoid-responsive mRNA-stabilizing element. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7800-7804.

Philippe, J. (1991) Insulin regulation of the glucagon gene is mediated by an insulin-responsive DNA element. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7224-7227.

Phillips, N.F.B., Snoswell, M.A., Chapman-Smith, A., Keech, D.B. and Wallace, J.C. (1992) Isolation of a carboxylphosphate intermediate and the locus of acetyl-CoA action in the pyruvate carboxylase reaction. *Biochemistry* **31**, 9445-9450.

Piechaczyk, M., Blanchard, J.M., Riaad-El Sabouty, S., Dani, C., Marty, L. and Jeanteur, P. (1984) Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes. *Nature* **312**, 469-471.

Pilkis, S.J., el-Maghrabi, M.R. and Claus, T.H. (1988) Hormonal regulation of hepatic gluconeogenesis and glycolysis. *Annu. Rev. Biochem.* **57**, 755-783.

Pilkis, S.J. and Granner, D.K. (1992) Molecular physiology of the regulation of hepatic gluconeogenesis and glycolysis. *Annu. Rev. Physiol.* **54**, 885-909.

Poirer, C. and Guenet, J.-L. (1996) Encyclopedia of the mouse genome V. Mouse chromosome 19. *Mammal Genome* **6**: S309-S316.

Polakis, S.E., Guchhait, R.B., Zwergel, E.E., Lane, M.D. and Cooper, T.G. (1974) Acetyl coenzyme A carboxylase system of *Escherichia coli*: studies on the mechanisms of the biotin carboxylase and carboxy-transferase-catalysed reactions. *J. Biol. Chem.* **249**, 6657-6667.

Polonsky, K.S. (1995) The beta-cell in diabetes: from molecular genetics to clinical research. *Diabetes* **44**, 705-717.

Pryor, H.J., Smyth, J.E., Quinlan, P.T. and Halestrap, A.P. (1987) Evidence that the flux control coefficient of the respiratory chain is high during gluconeogenesis from lactate in hepatocytes from starved rats. Implications for the hormonal control of gluconeogenesis and action of hypoglycaemic agents. *Biochem. J.* **247**, 449-457.

Pugh, B.F. and Tjian, R. (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61**, 1187-1197.

Pugh, B.F. and Tjian, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5**, 1935-1945.

Purcell, A.W. and Wallace, J.C. (1996) The accessibility of the prosthetic group biotin during monomeric avidin affinity chromatography. *Anal. Biochem.* **238**, 213-216.

Quinn, P.G., Wong, T.W., Magnuson, M.A., Shabb, J.B., Granner, D.K. (1988) Identification of basal and cyclic AMP regulatory elements in the promoter of the phosphoenolpyruvate carboxykinase gene. *Mol. Cell Biol.* **8**, 3467-3475.

Radford, S.E., Laue, E.D., Perham, R.N., Martin, S.R. and Apella, E. (1989) Conformational flexibility and folding of synthetic peptides representing an inter-domain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli. J. Biol. Chem.* **264**, 767-775.

Raghow, R. (1987) Regulation of messenger RNA turnover in eukaryotes. *Trends Biochem. Sci.* **12**, 122-166.

Reche, P., Li, Y.-L., Y., Fuller, C., Eichhorn, K. and Perham, R.N. (1998) Selectivity of post-translational modification in biotinylated proteins: the carboxyl carrier protein of the acetyl-CoA carboxylase of *Escherichia coli. Biochem. J.* **329**, 589-596.

Reddy, D.V., Rothemund, S., Shenoy, B.C., Carey, P.R. and Sönnichsen, F.D. (1998) Structural characterization of the entire 1.3S subunit of transcarboxylase from *Propiobacterium shermanii. Protein Sci.* **7**, 2156-2163.

Reed, G.H. and Scrutton, M.C. (1974) Pyruvate carboxylase from chicken liver. Magnetic resonance studies of the effect of substrates and inhibitors on the environment of the bound manganese. *J. Biol. Chem.* **249**, 6156-6162.

Rees, S., Coote, J., Stables, J., Goodson, S., Harris, S. and Lee, M.G. (1996) Bicistronic vector for the creation of stable mammalian cell lines that predisposes all antibiotic-resistant cells to express recombinant protein. *BioTechniques* **20**, 102-110.

Reynolds, G.A., Goldstein, J.L.and Brown, M.S. (1985) Multiple mRNAs for 3-hydroxy-3-methylglutaryl coenzyme A reductase determined by multiple transcription initiation sites and intron splicing sites in the 5'-untranslated region. *J. Biol. Chem.* **260**, 10369-10377.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C. and Markham, A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887-2890.

Rinard, G.A., Okuno, G. and Haynes, R.C. Jr. (1969) Stimulation of gluconeogenesis in rat liver slices by epinephrine and glucocorticoids. *Endocrinol* . **84**, 622-631.

Roberts, E.L., Shu, N., Howard, M.J., Broadhurst, R.W., Chapman-Smith, A., Wallace, J.C., Morris, T., Cronan, J.E.Jr. and Perham, R.N. (1999) Solution structures of Apo- and Holo-biotinyl domains from acetyl-Coenzyme A carboxylase of Escherichia coli determined by triple-resonance NMR spectroscopy. *Biochemistry* (In press).

Robinson, B.H. (1982) Inborn errors of pyruvate metabolism. *Trends Biochem. Sci* 7, 151.

Robinson, B.H., Oei, J., Saunders, M. and Gravel, R. (1983) [^{3}H]Biotin-labeled proteins in cultured human skin fibroblasts from patients with pyruvate carboxylase deficiency. *J. Biol. Chem.* **258**, 6660-6664.

Robinson, B.H., Oei, J., Saudubray, J.M., Marsac, C., Bartlett, K., Quan, F. and Gravel, R. (1987) The French and North American phenotyopes of pyruvate carboxylase deficiency, correlation with biotin containing protein by ^{3}H-biotin incorporation, ^{35}S-streptavidin labeling and Northern blotting with a cloned cDNA probe. *Am. J. Hum. Genet.* **40**, 50-59.

Robinson, B.H., MacKay, N., Chun, K. and Ling, M. (1996) Disorders of pyruvate carboxylase and the pyruvate dehydrogenase complex. *J. Inher. Metab. Dis.* **19**, 452-462.

Roeder, R.G. (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem. Sci.* **16**, 402-408.

Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327-335.

Roesler, W.J., McFie, P.J. and Puttick, D.M. (1993) Evidence for the involvement of at least two distinct transcription factors, one of which is liver-enriched, for the activation of P-enolpyruvate carboxykinase gene promoter by cAMP. *J. Biol. Chem.* **268**, 3791-3796.

Roesler, W.J., Simard, J., Graham, J.G. and McFie, P.J. (1994) Characterization of the liver-specific component of the cAMP response unit in the phosphoenolpyruvate carboxykinase (GTP) gene promoter. *J. Biol. Chem.* **269**, 14276-14283.

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.

Rohde, M., Lim, F. and Wallace, J.C. (1991) Electron microscopic localisation of pyruvate carboxylase in rat liver and *Saccharomyces cerevisiae* by immunogold procedures. *Arch. Biochem. Biphys.* **290**, 197-201.

Rongnoparut, P., Verdon, C.P., Gehnrich, S.C. and Sul, H.C. (1991) Isolation and characterization of the transcriptionally regulated mouse liver (B-type) phosphofructokinase gene and its promoter. *J. Biol. Chem.* **266**, 8086-8091.

Rønne, M., Shibasaki, Y., Poulsen, B.S. and Andersen, O. (1987) The high resolution R-banded karyotype of *Rattus norvegicus*. *Cytogenet. Cell Genet.* **45**, 113-117.

Rosenthal, A. and Jones, D.S.C. (1990) Genomic walking and sequencing by oligocassette mediated polymerase chain reaction. *Nucleic Acids Res.* **18**, 3095-3096.

Rothman, D.L., Magnusson, I., Katz, L.D., Shulman, R.G. and Schulman, G.I. (1991) Quantitation of hepatic glycogenolysis and gluconeogenesis in fasting humans with ^{13}C NMR. *Science* **254**, 573-576.

Ruiz-Lozano, P., de Lecea, L., Buesa, C., Perez de la Osa, P., LePage, D., Gualberto, A. Walsh, K. and Pons, G. (1994) The gene encoding rat phosphoglycerate mutase subunit M: cloning and promoter analysis in skeletal muscle cells. *Gene* **147**, 243-248.

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 sequence. *Arch. Biochem. Biophys.* **183**, 113-122.

Sakimura, K., Kushiya, E., Takahashi, Y. and Suzuki, Y. (1987) The structure and expression of neuron-specific enolase gene. *Gene* **60**, 103-113.

Salditt-Georgieff, M., Harpold, M. and Chen-Kiang, S. (1980) The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell* **19**, 69-78.

Salto, R., Sola, M., Oliver, F.J. and Vargas, A.M. (1996) Effects of starvation, diabetes and carbon tetrachloride intoxication on rat kidney cortex and liver pyruvate carboxylase levels. *Arch. Physiol.Biochem.* **104**, 845-850.

Samols, D., Thornton, C.G., Murtif, V.L., Kumar, G.K., Haase, F.C. and Wood, H.G. (1988) Evolutionary conservation among biotin enzymes. *J. Biol. Chem.* **263**, 6461-6464.

Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature* **334**, 218-222.

Saraste, M., Sibbald, P.R. and Wittinhofer, A. (1990) The P-loop a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430-434.

Sasaki, K., Cripe, T.P., Koch, S.R., Andreone, T.L., Peterson, D.D., Beale, E.G. and Granner, D.K. (1984) Multihormonal regulation of P-enolpyruvate carboxykinase gene transcription: The dominant role of insulin. *J. Biol. Chem.* **259**, 15242-15251.

Sasaki, K. and Granner, D.K. (1988) Regulation of phosphoenolpyruvate carboxykinase gene transcription by insulin and cAMP: reciprocal actions on initiation and elongation. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2954-2958.

Schalinske, K.L., Chen, O.S. and Eisenstein, R.S. (1998) Iron differentially stimulates translation of mitochondrial aconitase and ferritin mRNAs in mammalian cells: implications for iron regulatory proteins as regulators of mitochondrial citrate utilization. *J. Biol. Chem.* **273**, 3740-3746.

Schmoll, D., Führmann, E., Gebhardt, R. and Hamprecht, B. (1995) Significant amounts of glycogen are synthesized from 3-carbon compounds in astroglial primary cultures from mice with participation of the mitochondrial phosphoenolpyruvate carboxykinase isoenzyme. *Eur. J. Biochem.* **227**, 308-315.

Schmoll, D., Allan, B.B. and Burchell, A. (1996) Cloning and sequencing of the 5' region of the human glucose-6-phosphatase gene: transcriptional regulation by cAMP, insulin and glucocorticoids in H4IIE hepatoma cells. *FEBS Letts* **383**, 63-66.

Schuit, F., De Vos, A., Farfari, S., Moens, K., Pipeleers, D., Brun, T. and Prentki, M. (1997) Metabolic fate of glucose in purified islets cells. *J. Biol. Chem.* **272**, 18572-18579.

Schwarz, E., Oesterheld, D., Reinke, H., Beyreuther, K. and Dimroth, P. (1988) The sodium ion translocating oxaloacetate decarboxylase of *Klebsiella pneumoniae. J. Biol. Chem.* **242**, 9640-9645.

Scrutton, M.C. and Utter, M.F. (1965) Pyruvate carboxylase. V. Interaction of the enzyme with adenosine triphosphate. *J. Biol. Chem.* **240**, 3714-3723.

Scrutton, M.C., Young, M.R. and Utter, M.F. (1970) Pyruvate carboxylase from bakers yeast: the presence of bound zinc. *J. Biol. Chem.* **245**, 6220-6227.

Scrutton, M.C., Griminger, P. and Wallace, J.C. (1973) Pyruvate carboxylase: Bound metal content of the vertebrate liver enzyme as a function of diet and species. *J. Biol. Chem.* **247**, 3305-3313.

Scrutton, M. C. and White, M.D. (1973) Pyruvate carboxylase: Specific inactivation of acetyl CoenzymeA-dependent oxaloacetate synthesis during modification of the enzyme by trinitrobenzene sulphonic acid. *J. Biol. Chem.* **248**, 5541-5544.

Scrutton, M.C. and White, M.D. (1974) Purification and properties of human liver pyruvate carboxylase. *Biochem. Med.* **9**, 271-292.

Scrutton, M.C. and Taylor, B.L. (1974) Isolation and characterization of pyruvate carboxylase from *Azotobacter vinelandii* OP. *Arch. Biochem. Biophys.* **164**, 641-654.

Seitz, H.J., Kaiser, M., Krone, W and Tarnowski, W. (1976) Physiological significance of glucocorticoids and insulin in the regulation of hepatic gluconeogenesis during starvation in rats. *Mebabolism* **25**, 1545-1555.

Sekine, N., Cirulli, V., Regazzi, R., Brown, L.J., Gine, E., Tamarit-Rodriguez, J., Girotti, M., Marie, S., MacDonald, M.J., Wollheim, C.B. and Rutter, G.A. (1994) Low lactate dehydrogenase and high mitochondrial glycerol phosphate dehydrogenase in pancreatic beta-cells. Potential role in nutrient sensing. *J. Biol. Chem.* **269**, 4895-4902.

Seubert, W. and Remberger, U. (1961) Reinigung und Wirkungsweise der Pyruvatcarboxylase aus *Pseudomonas citronellolis. Biochem. Z.* **334**, 401-414.

Shank, R.P., Bennett, G.S., Freytag, S.O. and Campbell G.L. (1985) Pyruvate carboxylase: An astrocyte-specific enzyme implicated in the replenishment of amino-acid transmitter pools. *Brain Res.* **329**, 364-367.

Shank, R.P., Leo, G.C. and Zielke, H.R. (1993) Cerebral metabolic compartmentation as revealed by nuclear magnetic resonance analysis of D-[1[^{13}C]glucose metabolism. *J. Neurochem.* **61**, 315-323.

Shelly, L.L., Lei, K.J., Pan, C.J., Sakata, S.F., Ruppert, S., Schutz, G. and Chou, J.Y. (1993) Isolation of the gene for murine glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1A. *J. Biol. Chem.* **268**, 21482-21485.

Shenoy, B.C., Paranjape, S., Murtif, V.L., Kumar, G.K., Samols, D. and Wood, H.G. (1988) Effect of mutations at Met-88 and Met-90 on the biotinylation of Lys-89 of the apo 1.3S subunit of transcarboxylase. *FASEB J.* **2**, 2505-2511.

Shimura, Y., Shimura H., Ohmori, M., Ikuyama, S. and Kohn, L.D. (1994) Identification of a novel insulin-responsive element in the rat thyrotropin receptor promoter. *J. Biol. Chem.* **269**, 31908-31914.

Shows, T.B., Alders, M., Bennett, D., Burbee, P., Cartwright, S.,Chandrasekharapappa, S., Cooper, P., Courseaux, A., Davies, C., Devignes, M.D., Devilee, R., Elliot, R., Evans, G., Fantes, J., Garner, F.H., Gaudray, P., Gerhard, D.S., Gessler, M., Higgins, M., Hummerich, H., James, M., Lagercrantz, J., Litt, M., Little, P., Mannens, M., Munroe, D., Nowak, N., O'Brien, S., Parker, N., Perlin, M., Reid, L., Richard, C., Sawicki, M., Swallow, D., Thakker, R., van Heyninger, V., van Schothorst, E., Vorechovsky, I., Wadelius, C., Weber, B. and Zabel, B. (1996) The report of the fifth international workshop on human chromosome 11 mapping. *Cytogenet. Genet* .**74**: 1-56.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**, 1087-1088.

Singer-Sam, J.D., Keith, D.H., Tani, K., Simmer, R.L., Shively, L., Lindsay, S., Toshida, A.and Riggs, A.D. (1984) Sequence of the promoter region of the gene for human X-linked 3-phosphoglycerate kinase, *Gene* **32**, 409-417.

Sladek, F.M., Zhong, W., Lai, E. and Darnell, J.E.Jr. (1990) Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes & Dev.* **4**, 2353-2365.

Smale, S.T., Schmidt, M.C., Berk, A.J. and Baltimore, D. (1990) Transcriptional activation by Sp1 as directed throught TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4509-4513.

Smale, S.T. and Baltimore, D. (1989) The "initiator" as a transcription control element. *Cell* **57**, 103-113.

Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1989) Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* **23**, 527-577.

Söling, H.D., Willms, B., Klenineke, J. and Gdhlhoff, M. (1970) Regulation of gluconeogenesis in the guinea pig liver. *Eur. J. Biochem.* **16**, 289-302.

Srere, P.A. (1987) Complexes of sequential metabolic enzymes. *Annu. Rev. Biochem.* **56**, 89-124.

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. Commun.* **117**, 344-349.

Stapleton, M.A., Javid-Majd, F, Harmon, M.F., Hanks, B.A., Grahmann, L., Mullins, L.S. and Raushel, F.M. (1996) Role of conserved residues within the carboxy phosphate domain of carbamoyl phosphate synthetase. *Biochemistry* **35**, 14352-14361.

Stoffel, M., Froguel, P., Takeda, J., Zouali, H., Vionnet, N., Nishi, S., Weber, I.T., Harrison, R.W., Pilkis, S.J., Lesage, S. *et al.* (1992) Human glucokinase gene: isolation, characterization and identification of two missense mutations linked to early-onset non-insulin-dependent (type 2) diabetes mellitus. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7698-7702.

Streeper, R.S., Svitek, C.A., Chapman, S., Greenbaum, L.E., Taub, R. and O'Brien, R.M. (1997) A multicomponent insulin response sequence mediates a strong repression of mouse glucose-6-phosphatase gene transcription by insulin. *J. Biol. Chem.* **272**, 11698-11701.

Stucka, R., Dequin, S., Salmon, J. and Gancedo, C. (1991) DNA sequences in chromosomes II and VII code for pyruvate carboxylase isoenzymes in *Saccharomyces cerevisiae*: analysis of pyruvate carboxylase deficient strains. *Mol. Gen. Genet.* **229**, 307-315.

Student, A.K., Hsu, R.Y. and Lane, M.D. (1980) Induction of fatty acid synthetase synthesis in differentiating 3T3-L1 preadipocytes. *J. Biol. Chem.* **255**, 4745-4750.

Sundaram, T.K., Cazzulo, J.J. and Kornberg, H.L. (1971) Pyruvate holocarboxylase formation from the apoenzyme and d-biotin in *Saccharomyces cerevisiae. Arch. Biochem. Biophys.* **143**, 609-616.

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.

Suwanichkul, A., Morris, S.L. and Powell, D.R. (1993) Identification of an insulin-responsive element in the promoter of the human gene for insulin-like growth factor binding protein-1. *J. Biol. Chem.* **268**, 17063-17068.

Suzuki, Y., Aoki, Y., Ishida, Y., Chiba, Y., Iwamatsu, A., Kishino, T., Niikawa, N., Matsubara, Y. and Narisawa K (1994) Isolation and characterization of mutations in the human holocarboxylase synthetase cDNA. *Nature Genet.* **8**, 122-128.

Szepetowski, P., Simon, M.P., Grosgeorge, J., Huebner, K., Bastard, C., Evans, G.A., Tsujimoto, Y., Birnbaum, D., Theillet, C.and Gaudray, P. (1992) Localization of 11q13 loci with respect to regional chromosomal breakpoints. *Genomics* **12**, 738-744.

Tanizawa, Y., Matsutani, A., Chiu, K.C. and Permutt, M.A. (1992) Human glucokinase gene: isolation, structural characterization and identification of a microsatellite repeat polymorphism. *Mol. Endocrinol.* **6**, 1070-1081.

Taylor, P.H., Wallace, J.C. and Keech, D.B. (1971) Gluconeogenic enzymes in sheep liver: Intracellular localization of pyruvate carboxylase and phosphoenolpyruvate carboxykinase in normal, fasted and diabetic sheep. *Biochim. Biophys. Acta* **237**, 179-191.

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.

Thampy, K.G., Huang, W.-Y. and Wakil, S.J. (1987) A rapid purification method for rat liver pyruvate carboxylase and amino acid sequence analyses of NH_2-terminal and biotin peptide. *Arch. Biochem. Biophys.* **266**, 270-276.

Theil, E.C. (1990) Regulation of ferritin and transferrin receptor mRNAs. *J. Biol. Chem.* **265**, 4771-4774.

Thoden, J.B., Holden, H.M., Wesenberg, G., Raushel, F.M. and Rayment, I. (1997) Structure of carbamoyl phosphate synthetase: A journey of 96Å from substrate to product. *Biochemistry* **36**, 6305-6316.

Thomas, A.P. and Halestrap, A.P. (1981) Identification of the protein responsible for pyruvate transport into rat liver and heart mitochondria by specific labelling with [^{3}H]*N*-phenylmaleimide. *Biochem. J.* **198**, 551-564.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Thornton, C.G., Kumar, G.K., Shenoy, B.H., Haase, F.C., Phillips, N.F.B., Park, V.M, Magner, W.J., Hejjlik, D.P., Wood, H.G. and Samols, D. (1993) Primary structure of the 5S subunit of transcarboxylase as deduced from the genomic DNA sequence. *FEBS Lett.* **330**, 191-196.

Tipton, P.A. and Cleland, W.W. (1988) Catalytic mechanism of biotin carboxylase: steady state kinetic investigations. *Biochemistry* **27**, 4317-4325.

Toh, H., Kondo, H. and Tanabe, T. (1993) Molecular evolution of biotin-dependent carboxylases. *Eur. J. Biochem.* **215**, 687-696.

Tsujino, S., Sakoda, S., Mizuno, R., Kobayashi, T., Suzuki, T., Kishimoto, S., Shanske S., DiMauro, S. and Schon, E.A. (1989) Structure of the gene encoding the muscle-specific subunit of human phosphoglycerate mutase. *J. Biol. Chem.* **264**, 15334-15337.

Tsurushita, N., Ho, L. and Korn, L.J. (1988) Nuclear factors in B-lymphoma enhance splicing of mouse membrane bound μ-mRNA in Xenopus oocytes. *Science* **239**, 494-497.

Tsutsumi, K., Mukai, T., Tsutsumi, R., Hidaka, S., Arai, Y., Hori, K. and Ishikawa, K. (1985) Structure and genomic organization of the rat aldolase B gene. *J. Mol. Biol.* **181**, 153-160.

Tu, Z. and Hagedorn, H.H. (1997) Biochemical, Molecular and Phylogenetic analysis of pyruvate carboxylase in the yellow fever mosquito, *Aedes aegypti. Insect Biochem. Mol. Biol.* **27**, 133-147.

Utter, M.F. and Keech, D.B. (1960) Formation of oxaloacetate from pyruvate and CO_2. *J. Biol. Chem.* **235**, 17-18.

Utter, M.F. and Keech, D.B. (1963) Pyruvate carboxylase. I Nature of reaction. *J. Biol. Chem.* **238**, 2602-2608.

Val, D.L., Chapman-Smith, A., Walker, M.E., Cronan, J.E. Jr. and Wallace, J.C. (1995) Polymorphism of the yeast pyruvate carboxylase 2 gene and protein: effects on protein biotinylation. *Biochem. J.* **312**, 817-825.

Vallee, B.L. and Auld, D.S. (1990) Zinc function and structure of zinc enzymes and other proteins. *Biochemistry* **29**, 5647-5659.

van Huijsduijnen, R.-H., Li, X.-Y., Black, D., Matthes, H., Benoist, C. and Mathis, D. (1990) Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-1/CBF) subunits. *EMBO J.* **9**, 3119-3127.

van Roon, M.A., Zonneveld, D., Charles, R. and Lamers, W.H. (1988) Accumulation of carbamoylphosphate-synthase and P-enolpyruvate carboxykinase mRNA in embryonic rat hepatocytes. *Eur. J. Biochem.* **111**, 89-98.

Waldrop, G.L., Rayment, H.M. and Holden, H.M. (1994) Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* **33**, 10249-10256.

Walker, M.E., Val, D.L., Rohde, M., Devenish, R.J. and Wallace, J.C. (1991) Yeast pyruvate carboxylase: Identification of two genes encoding isoenzymes. *Biochem. Biophys. Res. Commun.* **176**, 1210-1217.

Walker, M.E. and Wallace, J.C. (1991) Isolation of a yeast mutant deficient in pyruvate carboxylase activity. *Biochem. Int.* **23**, 697-705.

Walker, M.E., Baker, E., Wallace, J.C. and Sutherland, G.R. (1995a) Assignment of the human pyruvate carboxylase gene (PC) to 11q13.4 by fluorescence *in situ* hybridisation. *Cytogenet. Cell Genet.* **69**, 187-189.

Walker, M.E., Jitrapakdee, S., Val, D.L and Wallace, J.C. (1995b) GenBank Accession number U30891.

Wallace, J.C. (1985) Distribution and biological functions of pyruvate carboxylase in nature, in Pyruvate carboxylase (Keech, D.B. and Wallace, J.C. eds), CRC Series in Enzyme Biology, CRC Press, Boca Raton, FL, p 5-63.

Wallace, J.C. and Easterbrook-Smith, S.B. (1985) The structure of pyruvate carboxylase, in Pyruvate carboxylase (Keech, D.B. and Wallace, J.C. eds), CRC Series in Enzyme Biology, CRC Press, Boca Raton, FL, p 66-108.

Wallace, J.C., Jitrapakdee, S. and Chapman-Smith, A. (1998) Pyruvate carboxylase. *Int. Biochem. Cell Biol.* **30**, 1-5.

Wang, J., Takagaki, Y. and Manley, J.L. (1996) Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes & Dev.* **10**, 2588-2599.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987) Molecular biology of the gene, 4th Edition, Benjamin/Cummings Publishing Co. Inc, p 621-675,

Webb, G.C., Jitrapakdee, S., Bottema, C.D.K. and Wallace, J.C. (1997) Assignment of the pyruvate carboxylase gene to rat chromosome band 1q43 by *in situ* hybridization. *Cytogenet. Cell Genet.* **79**, 151-152.

Weinberg, M.B. and Utter, M.F. (1979) Effect of thyroid hormone on the turnover of rat liver pyruvate carboxylase and pyruvate dehydrogenase. *J. Biol. Chem.* **254**, 9492-9499.

Weinberg, M.B. and Utter, M.F. (1980) Effect of streptozotocin-induced diabetes mellitus on the turnover of rat liver pyruvate carboxylase and pyruvate dehydrogenase. *Biochem. J.* **188**, 601-608.

Werneberg, B.G. and Ash, D.E. (1993) Chemical modifications of chicken pyruvate carboxylase: evidence for essential cysteine-lysine pairs and a reactive sulfhydryl group. *Arch. Biochem. Biophys.* **303**, 214-221.

Werneburg, B.G. and Ash, D.E. (1997) VO^{2+} (IV) complexes with pyruvate carboxylase: activation of oxaloacetate decarboxylation and EPR properties of enzyme-VO^{2+} complexes. *Biochemistry* **36**, 14392-14402.

Wexler, I.D., Du, Y., Ligaris, M.V., Mandal, S.K., Freytag, S.O., Yang, B., Liu, T., Hwon, M., Patel, M.S. and Kerr, D.S. (1994) Primary amino acid sequence and structure of human pyruvate carboxylase. *Biochim. Biophys. Acta* **1227**, 46-52.

Wexler, I.D., Kerr, D.S; Du, Y., Kaung, M.M., Stephenson, W. Lusk, M.M., Wappner, R.S. and Higgins, J.J. (1998) Molecular characterization of pyruvate carboxylase deficiency in two consanguineous families. *Pediatr. Res.* **43** 57 584.

White, M.F. and Hahn, C.R. (1994) The insulin signaling system. *J. Biol. Chem.* **269**, 1-4.

Wicksire, K. and Berdanier, C.D. (1997) Pyruvate metabolism in BHE/cdb rats is altered by the type of dietary fat. *Nutr. Biochem.* **8**, 275-278.

Wiesinger, H., Hamprecht, B. and Dringen, R. (1997) Metabolic pathways for glucose in astrocytes. *GLIA* **21**, 22-34.

Williams, T.J. and Fried, M. (1986) The MES-1 murine enhancer elements is closely associated with the heterogeneous 5' ends of two divergent transcription units. *Mol. Cell. Biol.* **6**, 4558-4569.

Wimhurst, J.M. and Manchester, K.L. (1970) A comparison of the effects of diabetes induced with either alloxan or streptozotocin and of starvation on the activities in rat liver of the key enzymes of gluconeogenesis. *Biochem. J.* **120** 95-103.

Woehlke, G., Wifling, K. and Dimroth, P. (1992) Sequence of the sodium ion pump oxaloacetate decarboxylase from *Salmonella typhimurium. J. Biol. Chem.* **267**, 22798-22803.

Wolodko, W.T., Fraser, M.E., James, M.N.G. and Bridger, W.A. (1994) The crystal structure of succinyl-CoA synthetase from *Escherichia coli* at 2.5Å resolution. *J. Biol. Chem.* **269**, 10883-10890.

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

Wurtele, E.-S.and Nikolau, B.J. (1990) Plants contain multiple biotin enzymes: discovery of 3-methylcrotonyl-CoA carboxylase, propionyl-CoA carboxylase and pyruvate carboxylase. *Arch. Biochem. Biophys.* **278**, 179-186.

Yakunin, A. F. and Hallenbeck, P.C. Regulation of synthesis of pyruvate carboxylase in the photosynthetic bacterium *Rhodobacter capsulatus.* (1997) *J. Bacteriol.* **179**, 1460-1468.

Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J. and Katsube, Y. (1993) Three-dimensional structure of the glutathione synthetase from *Escherichia coli* at 2.0Å resolution. *J. Mol. Biol.* **229**, 1083-1100.

Yamada, S., Otto, P.S., Kennedy, D.L. and Whayne, T.F. (1980) The effects of dexamethasone on metabolic activity of hepatocytes in primary monolayer culture. *In Vitro* **16**, 559-570.

Yamasaki, T., Nakajima, H., Kono, N., Hotta, K., Yamada, K., Imai, E., Kuwajima, M., Noguchi, T., Tanaka, T. and Tarui, S. (1991) Structure of the entire human muscle phosphofructokinase-encoding gene: a two-promoter system. *Gene* **104**, 277-282.

Yao, X., Wei, D., Soden, C.Jr., Summers, M.F. and Beckett, D. (1997) Structure of the carboxyl-terminal fragment of the apo-biotin carboxyl carrier subunit of *Escherichia coli* acetyl-CoA carboxylase. *Biochemistry* **36**, 15089-15100.

Yeung, D., Stanley, R.S. and Oliver, I.T. (1967) Development of gluconeogenesis in neonatal rat liver: effect of triamcinolone. *Biochem. J.* **105**, 1219-1227.

Zempleni, J., Trusty, T.A. and Mock, D.M. (1997) Lipoic acid reduces the activities of biotin-dependent carboxylases in rat liver. *J. Nutr.* **127**, 1776-1781.

Zhang, J., Xia, W. and Ahmad, F. (1993) Adipose pyruvate carboxylase: amino acid sequence and domain structure deduced from cDNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1766-1770.

Zhang, J., Xia, W.-L. and Ahmad, F.(1995) Regulation of pyruvate carboxylase in 3T3-L1 cells. *Biochem. J.* **306**, 205-210.

Zhang, S. and Kim, K.-H. (1995) Glucose activation of acetyl-CoA carboxylase in association with insulin secretion in a pancreatic β-cell line. *J. Endocrinol.* **147**, 33-41.

Zimmer, D.B. and Magnuson, M.A. (1990) Immunohistochemical localization of phosphoenolpyruvate carboxykinase in adult and developing mouse tissues. *J. Histochem. Cytochem.* **38**, 171-178.

Zucker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequence using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133-148.