STUDIES OF COMPLEMENTARY DNAs CORRESPONDING TO SKELETAL MUSCLE PROTEINS

Colin J. McInnes

Submitted for the degree of Doctor of Philosophy in the University of Glasgow

Department of Biochemistry

May 1988

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

CONTENTS

List of Figures
(short title)

ACKNOWLEDGEMENTS

I would like to express my thanks to the following people for their help throughout the course of this work:
The late Professor R.M.S. Smellie for making available the facilities of the Department of Biochemistry. Dr D.P. Leader for his help, encouragement and supervision. All my friends, past and present, in lab C36, in particular Irene, Ying, Caroline and Beate for passing on their technical expertise. And everyone else in the Biochemistry Department for making the last three years bearable. I would also like to thank Patricia Rose for her skill and patience in typing this thesis and to medical illustration for the production of artwork.

Finally thanks must also go to my parents for their endless support as well as board and lodgings, but most of all my thanks must go to Rachel, for everything.

ABBREVIATIONS

The abbreviations used in this thesis are as set out in "Instructions to Authors", Biochemical Journal (1985) 225: 1-26, except for the following

Amp	Ampicillin
bp	base pairs
BSA	Bovine Serum Albumin
DNAse	deoxyribonuclease
dNTP	deoxynucleotide-5'-triphosphate
EMBL	European Molecular Biology Laboratory
kb	kilobase (1000 base pairs)
PEG	polythylene glycol
RNAse	ribonuclease
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
tet	tetracycline

SUMMARY

The initial objective of the work described in this thesis was to isolate complementary DNAs (cDNAs) corresponding to mRNAs encoding proteins expressed solely in skeletal muscle. To this end a mouse skeletal muscle cDNA library was differentially screened with radioactively-labelled single-stranded cDNA probes derived from skeletal and cardiac muscle. Of 15,000 individual colonies subjected to the mass screening procedure of colony hybridisation, 247 were selected on the basis that they gave a hybridisation signal only with the skeletal muscle probe. Southern blot analysis of plasmid DNA isolated from each of these revealed eight which continued to produce a differential hybridisation signal with the two single-stranded cDNA probes. However subsequent hybridisation of these eight radioactively-labelled cDNA clones with poly(A)$^+$ RNA isolated from skeletal muscle, cardiac muscle, liver and brain, revealed only a single clone which appeared to represent an mRNA specific to skeletal muscle.

The nucleotide sequence of the cDNA insert of this clone, which was approximately 500 base pairs (bp) in length, was determined. It was found to contain an open reading-frame, the predicted product of which corresponded to part of the sequence of the β-isoform of rabbit tropomyosin. This isoform has only been found in skeletal muscle, although other isoforms are found in cardiac muscle, smooth muscle, and non-muscle tissues. No previous mouse β-tropomyosin cDNA clones have been described.

The skeletal muscle cDNA library was rescreened in an attempt to obtain a full-length β-tropomyosin cDNA clone. Three clones were selected and sequenced. One of these was found to be identical to the original clone selected from the initial screening. The second, however, contained the entire β-tropomyosin amino-acid coding region together with 72 nucleotides of the 5' non-coding region and 151 nucleotides of the 3' non-coding region. The remaining clone lacked sequence from the 3' half of the mRNA, but extended the 5' non-coding region to 95 nucleotides. From comparison with the size of the mRNA (1.3kb), revealed by Northern blotting, it is estimated that the sequence of the mouse β-tropomyosin determined represents approximately 90% of the mRNA.

The expression of tropomyosin mRNAs was examined during the differentiation of a cultured mouse cell line from the individual myoblast stage to the multinucleate cell stage. Northern blot analysis of the mRNA isolated from the cells at different stages of their development revealed that upon fusion of the cells, the expression of several different tropomyosin mRNAs is induced. In particular two species, with lengths of 1.2kb and 2.6kb, which are not found in the RNA isolated from 11-day old mouse skeletal muscle, are expressed during the differentiation of skeletal muscle cells. The two remaining species expressed upon fusion of the cells, a 1.3kb species and a 2.4kb species, correspond to the species identified in the RNA from 11-day old mice. The identities of all four tropomyosin species are unknown, except for the 1.3kb β-tropomyosin mRNA. However comparison with published work suggests that the 1.2kb and 2.4kb species may correspond respectively to the

alternatively spliced products from the primary transcripts which also give rise to the β-tropomyosin and the α_2-tropomyosin species. No previous work has reported the possible coexpression of alternatively spliced isoforms of tropomyosin.

Analysis of the nucleotide sequence corresponding to mouse β-tropomyosin mRNA revealed some unusual features. There was a deficit of the dinucleotide CpG in the codon position [2,3], but not in the [3,1] position, throughout the amino-acid coding region of β-tropomyosin. One possible explanation of this is that there is strand-specific hemi-methylation of the corresponding germ line DNA.

Previously published comparisons of a partial human β-tropomyosin cDNA sequence with a human non-muscle tropomyosin (TM_{36}) cDNA sequence had indicated the likelihood that β-tropomyosin and TM_{36}-tropomyosin are encoded by the same gene, with each of the isoforms arising from the alternative splicing of two pairs of mutually exclusive exons. Comparison of these sequences with the mouse β-tropomyosin cDNA allowed the conclusion that none of the 5' exons not represented in the human β-tropomyosin cDNA clone are involved in alternative splicing.

Nucleotide comparisons with other β-tropomyosin cDNA sequences reported during the completion of this work indicated that the 3' and 5' non-coding regions of the β-tropomyosin mRNA had been subject to some selective evolutionary pressure for conservation, although not to the extent found for certain other mRNAs encoding muscle proteins (e.g. actin mRNAs).

CHAPTER 1

Introduction

Skeletal muscle is the most abundant tissue in mammals, constituting up to 40% of the total body mass. Historically it has been used as the major source of enzymes and proteins for the study of intermediary metabolism and protein structure. However, more recently, it has been used increasingly as a model system for investigating the molecular events occurring during terminal differentiation.

There are obvious physiological and histological differences between skeletal and cardiac muscle. However, little is known about these differences, in terms of the protein species expressed in the two tissues, other than that different isoforms of several ubiquitously expressed proteins are specific to the two tissues and are expressed at particular stages of muscle development. The investigation of the genes encoding such proteins is important for understanding the mechanisms controlling differential gene expression. The identification of genes encoding unique protein species expressed in skeletal muscle (as opposed to isotypic variants) will also help to elucidate these mechanisms, but, in addition, may help to identify major differences between skeletal and cardiac muscle which in turn may lead to the identification of the primary biochemical lesions in those skeletal muscle myopathies in which cardiac muscle does not appear to be affected.

This thesis will present work, the objective of which was to identify complementary DNAs (cDNAs) corresponding to proteins expressed solely in skeletal muscle. This introduction will first review the general biochemistry of striated muscle, emphasising differences that exist between skeletal and cardiac muscle. It will then consider in more detail those muscle proteins already shown to exist in multiple isoforms, and will examine the tissue-specific and developmentally-specific manner in which they are expressed.

1.1 General Features of Striated Muscle Tissues

The general biochemistry of muscle tissue is such a wide ranging topic that no attempt has been made to provide references to the orignal work, which in some instances dates back over 100 years. Instead the reader is referred to those books by Darnell et al (1983), Bourne (1973) and Passmore and Robson (1968), and the references therein.

Histology and Physiology

Despite their common striated appearance, the general morphology of skeletal muscle cells is different from that of cardiac muscle cells.

Skeletal muscle cells are elongated and cylindrical, frequently measuring up to 10cm in length, with a diameter of approximately 30-60μm. Each cell is multinucleate and has its own nerve sypply.

Skeletal muscle connects the bones of the skeleton so that movement is made possible. Contraction is under voluntary control, and therefore, skeletal muscle cells are able to remain inactive, or at rest, for considerable periods of time in addition to generating movement. In response to such stimuli as hormones, neural input and the changing energy requirement of the muscle, skeletal muscle cells can change their metabolism, composition and bulk.

Cardiac muscle cells are much shorter than skeletal muscle cells, never exceeding 100μm in length. They are irregular in shape, branched and interconnected through intercalated discs, thus forming a meshwork of closely coupled cells. Cardiac muscle cells are mononucleate, with a high proportion of their volume (up to 40%) being taken up by mitochondria, reflecting their aerobic metabolism. The heart as a whole has a nerve supply, but because the cells are coupled each individual cell need not have its own nervous input. Contraction is under involuntary control, with each cell having its own inherent self-propagating rhythmic pattern of contraction. However again as a result of the cells being coupled, the cell with the fastest rhythm dictates the contraction of all the others. Cardiac muscle cells have the ability to change their metabolism composition and bulk in response to stimuli similar to those which effect skeletal muscle cells.

General Metabolism of Striated Muscle

The primary role of all striated muscle is the conversion of chemical energy, in the form of ATP, into mechanical work. However,

even when no work is being performed, ATP is still required by the muscle cells, not only for the normal cellular biosynthetic processes, but also for the active transport of ions, which maintains the membrane potential so essential to the contraction process itself. In resting skeletal muscle the energy for these processes is provided by the β-oxidation of fatty acids, whereas in the working muscle ATP is supplied by other means. Cardiac muscle, however, is only ever transiently at rest and therefore it is difficult to determine whether separate mechanisms are used for producing ATP for each of the different processes.

During contraction the type of work performed by each muscle cell type largely dictates the metabolic pathways utilised for producing ATP. For the purpose of this review striated muscle will be divided into two general classifications, fast twitch cells and slow twitch/cardiac cells. Fast twitch skeletal muscle cells, those which contract and fatigue rapidly, obtain the greatest proportion of their energy from the anaerobic glycolysis of glycogen and glucose, thus producing lactic acid as the end product. The lactic acid cannot be metabolised by these cells during contraction, and although much of it is passed to the liver via the blood, the reduction in pH caused by its production, results in the rapid fatiguing of fast skeletal muscle cells. In cardiac muscle cells and slow twitch skeletal muscle cells it is imperative that no lactic acid is produced during contraction, although, paradoxically cardiac muscle cells are usually net consumers of the lactic acid produced by fast skeletal muscle cell contraction. Nevertheless, however, the energy in these cell types is provided by aerobic

metabolism, a fact reflected in the large number of mitochondria present in both cell types. The majority of ATP is produced from the aerobic metabolism of fatty acids, but some energy is also provided by the aerobic glycolysis of glycogen and glucose, where the end product, pyruvic acid, is further metabolised with the muscle itself, via the reactions of the Tricarboxylic Acid (TCA) cycle.

The other major difference in ATP production is found between skeletal muscle and cardiac muscle cells in general. As discussed previously, cardiac muscle is only ever transiently at rest and as a consequence the reactions responsible for producing ATP are occurring at all times. In resting skeletal muscle the ATP produced by the β-oxidation of fatty acids is not sufficient to allow contraction to occur, but the reactions responsible for producing ATP during contraction, generally do not occur or are occurring only at a basal level in the resting cells. Therefore at the onset of contraction, when the demand for ATP increases by several factors, skeletal muscle cells require another mechanism, for producing ATP, which bridges the time gap from the onset of contraction to the sufficient provision of ATP by the glycolytic pathway. This alternative mechanism involves the high energy compound phosphocreatine which is stored in large quantities in striated muscle. In a reaction catalysed by the enzyme creatine kinase, the phosphate group of phosphocreatine is transferred to ADP resulting in the synthesis of ATP. The reaction is fully reversible under slightly alkaline conditions and thus in the resting muscle cell the phosphocreatine pool is regenerated. The

phosphocreatine pool, although limited, can produce sufficient ATP to allow contraction to proceed until the glycolytic reactions provide enough ATP for continued contraction. The creatine kinase reaction does occur in cardiac cells, but to a much lesser extent and certainly does not have such an important role as in skeletal muscle cells.

In addition to the catabolic pathways already discussed, striated muscle cells are capable of producing ATP from the oxidisation of glutamate and other amino acids. However the amount of ATP generated by such reactions is so small that it is not considered important in the overall provision of energy within the cell. Instead it is thought that amino acid oxidation probably serves to replenish some of the TCA cycle intermediates.

In addition to the general biosynthetic pathways required for homeostasis, the principle anabolic reactions of striated muscle cells involve the synthesis of glycogen from glucose. The other major energy sources, fatty acids, can be produced in the mitochondria via chain elongation reactions but it is not thought that this is important in maintaining the fatty acid pool in muscle cells.

The Contractile Apparatus of Striated Muscle

The contractile apparatus constitutes up to two thirds of the dry weight of a striated muscle cell. The basic unit of contraction is the sarcomere, a highly organised array of several different

proteins, which in the resting state is 2.5 μm in length (Figure 1.1). The serial arrangement of these sarcomeres form what is called the myofibril, a cylindrical structure which can vary in length from approximately 100μm in cardiac cells to as long as 50cm in some skeletal muscle cells. It is the parallel array of hundreds of these myofibrils, with sarcomeres aligned in register, which give the muscle its characteristic striated pattern under the light microscope. Table 1.1 lists those proteins so far identified as being associated with the sarcomere.

Figure 1.1 The structure of the sarcomere

A diagrammatic representation of the sarcomere is shown, indicating the major protein species associated with it (Novikoff and Holtzman, 1970).

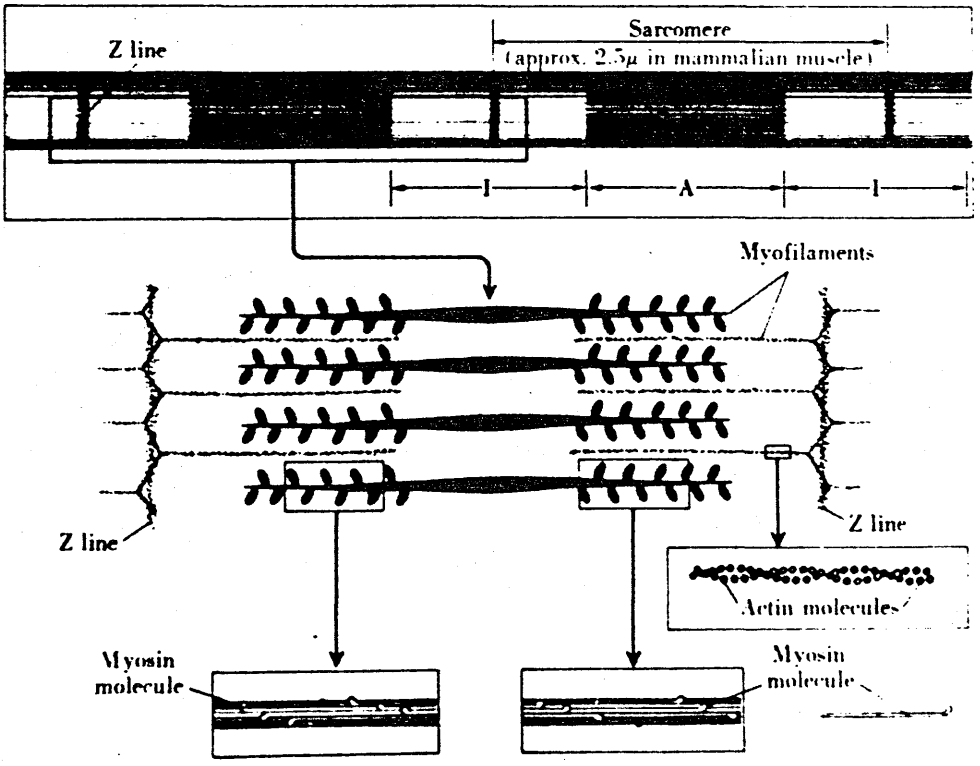


Table 1.1 Proteins of the sarcomere

I THICK FILAMENTS

- Myosin heavy chain
- Myosin light chain
- C protein
- I protein
- Creatine kinase

II THIN FILAMENTS

- Actin
- Tropomyosin
- Troponins C, I and T
- α-actinin
- Amorphin

III Z-line

- α-actinin
- Amorphin
- Z-protein
- Eu-actinin

The sliding filament theory of muscle contraction proposes that upon contraction the thick and thin filaments (Figure 1.1) slide past each other producing a subsequent shortening of the sarcomere to 2.25μm. This motion requires energy, which is provided by the hydrolysis of ATP, a reaction catalysed by the actomyosin ATPase. When myosin interacts with actin in the presence of ATP, the ATP is hydrolysed with the concomitant production of movement. The movement, probably due to a conformational change in the myosin molecule, is very slight and consequently many molecules of ATP have to be hydrolysed to produce the full 10% shortening of the sarcomere.

The other proteins of the sarcomere are either involved with the maintenance of its structural integrity, or in the regulation of the contraction process, with the possible exception of one or two proteins. These latter include creatine kinase, the importance of which has already been discused, and the majority of which is found sequestered in the M-line of the sarcomere.

The major regulatory proteins, tropomyosin, troponin-I, troponin-C and troponin-T all mediate their effect in the presence of calcium ions, without which contraction could not occur. Therefore, in addition, striated muscle cells contain not only an active transport system for calcium ions, but also a large number of proteins which bind calcium ions.

The Excitability of Striated Muscle Tissue

The stimulus for contraction of striated muscle is depolarisation of the cell. Thus striated muscle cells contain many of the enzymes and proteins associated with excitable membrances. These include the acetylcholine receptor, acetylcholine esterase and the sodium and potassium channels required to maintain the membrane potential.

The Differentiation of Skeletal Muscle

The mature skeletal muscle cell arises from mononucleate precursors called myoblasts, which do not possess contractile activity. In vivo these myoblasts proliferate until, at a certain stage of development, they fuse end to end to form post-mitotic multinucleate myotubes, which at a later stage, further develop into the mature multinucleate muscle cell with its characteristic striated appearance and contractile activity. These three definite stages of development can also be followed in vitro by culturing a myogenic cell line. The myoblasts are grown to confluence after which they can be induced to fuse and thus form the multinucleate myotubes and subsequently the mature muscle cells.

1.2 Different Isoforms of Muscle Proteins

The similar sarcomeric structure and metabolic pathways present in skeletal and cardiac muscle, might at first suggest that identical proteins are involved in these tissues, albeit in

different amounts. However, structural studies, enzyme kinetics, immunochemistry and the isolation of specific cDNAs have, in fact, demonstrated that many of the proteins thought to be common to the two tissues, are structurally distinct. Such proteins are referred to as isoforms and can be defined as individual proteins which have very similar though not identical primary structures, but which serve a similar if not identical role within the cell. Many muscle proteins, both sarcomeric and metabolic, have been shown to exist in multiple isoforms, and a list of these appears in Table 1.2. In the following section these muscle proteins are reviewed and the significance of their various different isoforms is discussed.

Table 1.2 Muscle proteins known to exist in more than one isotypic form

I SARCOMERIC	II METABOLIC
Myosin heavy chain	Creatine kinase
Myosin light chain	Phosphorylase
C-protein	Phosphofructokinase
Actin	Aldolase
Tropomyosin	Pyruvate kinase
Troponins C, I and T	and many more
Creatine kinase	

Myosin

Myosin is the major protein of the thick filament (Figure 1.1). It is composed of two heavy chains, each having a molecular weight of 200000, and four light chains each having a molecular weight of 16000-30000. In vivo myosin interacts with actin resulting in the hydrolysis of ATP, the rate of which is dependent on the isoform of the myosin heavy chain (MHC) present (Lowey and Risby 1971; Sarkar et al 1971; Close 1972, Hoh et al 1978). Thus by combining different isoforms of MHCs a fine tuning of the rate of ATP hydrolysis, and therefore the speed of contraction, can be achieved in different muscle cells. Myosin light chains (MLCs) fall into two categories, the "alkali" light chains and the phosphorylatable, or "regulatory", light chains. Two of each type of myosin light chain is associated with each native myosin molecule. The MLCs were also thought to affect the rate of ATP hydrolysis since different MLCs are found associated with MHCs of different ATPase activity. Indeed early research workers classified muscle as either "slow twitch" or "fast twitch" (with slow or fast hydrolysation of ATP), according to the isoforms of the MLCs present within the muscle (Sarkar et al 1971; Lowey and Risby 1971; Holt and Lowey 1977, Trayer et al 1977). However, more recently, evidence has been presented which shows that MLCs do not affect the ATPase activity of isolated myosin heavy chains (Wagner and Giniger 1981). The exact role of the MLCs remains unclear.

The actual number of MHC isoforms is not known, but at least eleven have been found in the rat: one adult superfast

skeletal, two adult fast skeletal, one adult slow skeletal, one perinatal, one embryonic/foetal, two adult cardiac, one smooth muscle and two non-muscle isoforms (Buckingham and Minty, 1983). As these classifications suggest, the MHC isoforms are expressed both in a tissue-specific and developmentally-specific manner, although one of the cardiac isoforms has also been found in skeletal muscle (Lompre et al, 1984; Saez and Leinwand 1986). A sequential transition from the expression of the embryonic isoform first to the perinatal isoform, and then to one of the four adult skeletal isoforms during terminal differentiation of skeletal muscle has been demonstrated in vitro and in vivo in the mouse (Weydert et al 1987).

The exact number of MLCs is also not known, but several developmentally regulated isoforms have been reported. These include two fast skeletal isoforms (Robert et al 1982), a cardiac ventricular isoform, which is identical to the slow skeletal isoform, (Barton et al 1985) and the cardiac atrial isoform which is identical to the embryonic ventricular and embryonic skeletal isoforms (Whalen et al 1982). The two fast skeletal isoforms reported by Robert et al (1982) were shown to be identical in part of their sequence, but had different amino-termini. Therefore it was suggested that these isoforms would be able to interact with myosin heavy chains in exactly the same way, but with the different termini providing distinct effector activities (Robert et al 1984).

C-Protein

C-Protein is the other major protein of the thick filament. It appears to be involved in maintaining the structural integrity of the myosin filaments. At least two distinct isoforms have been isolated, one from fast skeletal muscle and the other from slow skeletal muscle (Starr and Offer 1983), although Reinach *et al* (1983) have shown, using monoclonal antibodies, that in some chick fast skeletal muscle cells, two c-protein isoforms co-exist. Until the exact role of the c-protein has been defined more precisely it will not be possible to deduce the significance of there being separate isoforms.

Actin

Actin is the major constituent of the thin filaments. It is a globular protein of molecular weight 46000, but has the ability to form long polymers. It is the polymeric form which is found *in vivo* in the sarcomere. Amino acid sequencing studies have demonstrated the existence of at least six different isoforms of actin in mammals, each of which tends to predominate in a particular tissue type (Vandekerckhove and Weber 1979), although coexpression of some of the isoforms is also found in certain tissues (Gunning *et al* 1983, Vandekerckhove *et al* 1981). The six isoforms that have been isolated are the α-skeletal isoform, the α-cardiac isoform, and α and γ-smooth muscle isoforms and the β- and γ- non-muscle isoforms. Like myosin heavy chain isoforms, the expression of actin isoforms is coordinately regulated throughout muscle

development (Storti et al 1986; Schwartz and Rothblum 1981; Caravatti et al 1982; Minty et al 1982; Bains et al 1984). In particular Bains et al (1984) demonstrated in a cultured skeletal muscle cell line that only β- and γ-non muscle actin could be found in the mononucleate myoblasts. However after fusion the amount of these two isoforms fell, to be replaced, firstly, by a massive transient increase in the α-cardiac isoform and then by a much slower steady accumulation of the expected α-skeletal isoform. In the mononucleate myoblasts actin is primarily associated with the cytoskeleton, whereas in the fused cells it forms part of the sarcomere, and this alone may provide a functional rationale for the existence of different isoforms at these two stages of development. However the functional significance of a switch from α-cardiac to α-skeletal expression remains obscure.

Tropomyosin

Tropomyosin is a dimeric filamentous protein with a subunit molecular weight of 33000. In striated muscle tropomyosin forms head to tail aggregates which are found lying in the groove of the actin polymer. It has fourteen binding sites for actin per molecule and is thought to help to stabilise the actin polymer and thus the thin filament (Stone and Smillie 1978). Each molecule has at least one binding site for tropinin-T and, along with the troponin complex, it regulates the activity of the actomyosin ATPase. Tropomyosin is also found in non-muscle cells, most probably associated with the cytoskeleton (Moosekar 1983).

Several isoforms of tropomyosin have been reported (Cummins and Perry 1973; Mak et al 1980; Romero-Herrera et al 1982; Heeley et al 1984). In man, Talbot and McLeod (1983) and McLeod (1986) have described the isolation, from fibroblasts, of at least five polypeptides with properties expected of tropomyosin, and have further described cDNA clones representing the mRNAs for three of these non-muscle isoforms (McLeod et al 1985; McLeod et al 1986; McLeod et al 1987). Complementary DNA clones have also been described for the human α- and β-striated muscle isoforms, and a smooth muscle isoform isolated from chicken gizard (Helfman et al 1984; Sanders and Smillie 1985). The amino acid sequence of the smooth muscle isoform was reported to be similar to one of the five non-muscle isoforms reported by McLeod et al (1985), but as yet no mammalian smooth muscle isoform has been isolated.

The non-muscle tropomyosins are 247 amino acids in length, compared with the 284 amino acids of the striated muscle isoforms. This results in them lacking one of the binding sites for actin and also the binding sites for troponin-T, which are not required in non-muscle tissues which do not contain the troponin complex. Although the α- and β- striated muscle isoforms differ at 39 out of 284 amino acids (Mak et al 1980), it is not clear how (or indeed if) they differ in function, and indeed, it has not yet been determined whether hetero-dimers of α and β-tropomyosin can form in vivo.

Troponin

The troponin complex of striated muscle is composed of three separate proteins which interact with each other, and with tropomyosin, actin, and calcium ions, to regulate muscle contraction. Troponin-T binds to tropomyosin, troponin-c binds to calcium ions, and troponin-I binds to actin and troponin-c. The interaction between troponin I and actin has been shown to inhibit the ATPase activity of actomyosin.

Two separate troponin-Cs have been isolated, one from cardiac/slow skeletal muscle, the other from fast skeletal muscle, which differ in their stoichiometry of binding calcium ions (Potter and Johnson 1982). These forms differed only in a single amino acid, which would make them the most closely related isoforms yet described.

Different troponin-I isoforms have been isolated from rabbit slow skeletal, fast skeletal and cardiac muscle cells. They all show a high amino acid homology in the actin-binding region, but are heterologous in the N-terminal region where the binding site for tropinin-C is found (Wilkinson and Grand 1978).

Peptide mapping, amino acid sequencing and two-dimensional gel electrophoresis have shown the existence of several troponin-T isoforms. Briggs et al (1984) have isolated two rabbit fast skeletal isoforms, and Wilkinson et al (1984), three chicken fast skeletal isoforms. However Imai et al (1986) using two-

dimensional gel electrophoresis and immunoblotting have reported the existence of at least forty distinct troponin-T isoforms in chicken leg muscle. The existence of so many isoforms was reinforced by the isolation of four cDNAs and the gene, encoding rat fast skeletal troponin-T isoforms, and will be discussed in more detail later (Breitbart et al 1987, Nadal-Ginard et al 1987).

Creatine Kinase

Creatine kinase is required for the synthesis of ATP both in muscle and non-muscle cells. It is a dimeric enzyme with a subunit molecular weight of approximately 40,000, and is composed of two polypeptides, the M- or the B-isoforms. The three possible isozymes MM, BB and MB seem to have both a tissue-specific and developmentally specific pattern of expression. A sequential transition from the BB isoform to the MM isoform has been demonstrated in differentiating embryonic muscle tissue (Ziter 1974) and in myogenic cell lines (Turner et al 1974; Lebherz and Rutter 1969). In adult tissue the BB isoform occurs predominantly in smooth muscle and non-muscle tissue, whereas the MM isozyme seems confirmed to skeletal muscle, with the mixed dimer MB being predominant in mammalian heart tissue (Eppenberger et al, 1961). The separate cytoplasmic isoform of creatine kinase may be explained by it having no requirement to associate with the sarcomere. However it is not clear why two different sarcomeric isoforms are required, one for skeletal muscle and the other for cardiac muscle.

Enzymes of Carbohydrate Metabolism

Many metabolic enzymes have been shown to exist in multiple isoforms, some of the best studied being Phosphorylase (Sato et al 1972, Delain et al 1973), Phosphofructokinase (Layzer and Epstein 1972, Korata et al 1972, Delain et al 1973), Aldolase (Horecker et al 1972, Toutsumi et al 1983) and Pyruvate Kinase (Ibsen 1977, Peters et al 1981, Moore and Bulfield 1981). For each of the enzymes cited there is a clear tissue-specific distribution of isozymes in adult tissues and the same developmentally regulated sequential pattern of expression of isozymes during the terminal differentiation of skeletal muscle, as was found for actin and myosin isoforms.

Each of these enzymes is multimeric in structure. The principles involved in the formation of isotypic variants are similar for each enzyme and therefore only the case of Pyruvate Kinase will be discussed.

Pryuvate Kinase, an enzyme of the glycolytic pathway, is a tetrameric protein. At least four different subunit types have been found which result in four different homo-tetrameric isozymes, although many hetero-tetramers may also exist (Ibsen 1977). The four different homologous isozymes exhibit a very definite tissue-specific distribution, with one being found in muscle tissue, one in liver, one in kidney and one in erythrocytes. Each isozyme is distinguishable kinetically and therefore different isozymes may be adapted to the environment of the tissue in which they are found.

Many isoforms of muscle proteins are classified by the adult tissue in which they are found. However, it is quite clear, from the examples already discussed, that isoforms, other than those classified as being specific to skeletal muscle, are expressed during the normal development of a skeletal muscle cell. Consequently it has been suggested that, rather than being tissue-specific many isoforms are expressed in response to the physiological and metabolic demands of the individual muscle cell at the various stages of its development. This suggestion is consistent with the fact that different isoforms of the same protein do possess different properties in many of the cases already discussed. However, for some proteins, eg actin, functional advantages conferred by particular isoforms, in specific tissues, have not yet been discovered.

1.3 Organisation and Expression of Genes for Muscle Proteins

Analysis of the genes encoding several muscle proteins has suggested that there are two ways in which the differential expression of isoforms may be controlled. The first involves simple transcriptional units; that is those in which one gene encodes one protein. Expression of these proteins is, therefore, most likely to be controlled through the transcription of their genes. The second involves what may be termed complex transcriptional units; those in which one gene encodes two or more proteins. There is no evidence that DNA rearrangements, similar to those occurring in immunoglobulin gene expression, occur here, but rather it appears that in this case expression of a particular protein species requires control of pre-mRNA processing. Table 1.3 lists the proteins discussed in the preceding sections, according to the way in which the expression of their isoforms is thought to be controlled.

Table 1.3 Muscle proteins and the method by which the expression of their isoforms is controlled.

I Transcriptional Control

Myosin heavy chains

Actin

Creatine kinase

Metabolic enzymes?

II Alternative processing of pre-mRNA

Troponin T

Tropomyosin

Myosin light chains

Of the proteins discussed previously, only the actins and myosin heavy chains have been shown to have simple transcription units. The exact number of genes encoding MHC isoforms is not known, but has been estimated as being between seven and twenty-two, with the genes coding for the skeletal muscle isoforms being clustered on chromosome 11 in rats (Czosneck et al 1982, Weydert et al 1985, Gelbeiter et al 1986) and chromosome 17 in humans (Saez et al 1987). The cardiac isoforms, too, are also linked being on chromosome 14 in both species. Weydert et al (1985) have proposed that the clustering of skeletal MHC genes is essential for their sequential expression throughout development. However, actin isoforms are also expressed sequentially during differentiation of skeletal muscle cells, and although the chromosomal locations of the genes for all six isoforms have not yet been determined, the genes for α-skeletal, α-cardiac and β-cytophasmic actins in the rat are situated on separate chromosomes - 3, 17 and 5 respectively (Czosneck et al 1983, Robert et al 1985). Therefore the coordinate expression of actin isoforms, at least, does not require close chromosomal linkage.

Studies of the control of the expression of specific genes in other cases have revealed the presence of regions of DNA sequence 5' to the structural gene (cis-acting elements) which bind protein factors (trans-acting elements), present in the nucleus, in such a way as to stimulate transcription of the gene (Dynan and Tjian 1985). Furthermore, several research workers have shown that certain regions of DNA 5' to the gene are essential for the tissue-specific expression of the α1-antitrypsin and albumin genes in the liver (Ciliberto et al 1985, Ott et al 1984) the insulin gene in the pancreas (Walker et al 1983), the α-A-crystallin

gene in the lens (Chepelinsky et al 1985) and actin genes in muscle (Melloul et al 1984).

Recently Minty and Kedes (1986) and Miwa et al (1987) have identified two short regions of DNA upstream of the human cardiac actin gene which modulate its transcription when transfected into a myogenic cell line. These same two regions did not affect transcription when the gene was transfected into either a fibroblast or pheochromocytoma cell line and this led to the proposal that these sequences must interact with a tissue-specific transcriptional factor(s) present within muscle cells. Their results, however, also suggest that the postulated tissue-specific factor(s), although necessary, is not sufficient for expression. Thus, myoblasts of the C2 skeletal muscle cell line show an abundant expression of their cardiac actin gene (albeit transiently) at the onset of fusion (Bains et al 1984), but do not express the (endogenous) gene before differentiation. However if a human cardiac actin gene is transfected into the undifferentiated myoblasts it is transcribed (Minty et al 1986). Minty et al (1986) thus proposed that the tissue specific transcription of the cardiac actin gene must be, at least, a two stage process in which the gene had first to be activated in some way before the DNA regulating regions could respond to the controlling transcriptional factor(s).

It is not known why the cardiac actin gene should remain innactive until the onset of differentiation, but several theoretical possibilities exist. The presence of protein factors which bind to DNA in such a way as to prevent gene transcription, as is the case in both adenovirus gene transcription (Borelli et al 1984) and gene transcription in mammalian embryonic stem cells (Gorman et al 1985); or protein factors which

prevent post-transcriptional processing of the mRNA, are both possibilities. However Minty et al (1986) demonstrated that the transcription of a human actin gene, containing 5kb of upstream sequence and 2kb of downstream sequence, transfected into a C2 myogenic cell line, is not repressed in any way during the time that the endogenous actin gene transcription is being repressed.

It has been found that regions of DNA close to several genes contain a large proportion of methylated cytosine residues in the dinucleotide CpG, but that these regions become selectively demethylated when the genes are being transcribed (Chiu and Blau 1985, Keshet et al 1985). However when Shani et al (1984) investigated the methylation state of the skeletal muscle actin and myosin light chain -2 genes, they did not detect any change in the methylation state during the terminal differentiation of a rat L_8 myogenic cell line, and therefore concluded that methylation was not involved in the inactivation in this case. However, since the methylation state of DNA in vivo can only be apraised by using certain restriction enzymes which can discriminate between methylated and unmethylated DNA, but which will only act on a small proportion of the possible methylation sites, one cannot rule out some involvement for DNA methylation.

Chromatin structures within the chromosome, can innactivate genes by blocking access to them by the transcriptional factors. Conformational changes in chromatin structure are then needed to activate the gene (Weintraub 1985). Such a change of chromatin structure has been shown for the skeletal muscle actin gene in differentiating L_8 cells, as assayed by DNase I sensitivity (Carmon et al 1982). Therefore an

analogous situation has been proposed for the cardiac actin gene.

Whatever the gene activation or deactivation mechanism is, a primary stimulus is required to initiate either the conformational change in the chromatin structure, the selective (de)methylation of DNA, or the expression of a repressor protein. Candidates for such a stimulus include the release of a neural or hormonal factor. Evidence for a neural factor being involved was supplied when it was shown that the full maturation process of myogenic cell lines, including the sequential expression of MHC genes, is obtained only when the culture medium is supplemented with nerve extracts (Toyota and Shimada 1983; Gonoi et al 1983; Periasamy et al 1984; Weydert et al 1985). However other research groups have disputed this fact by demonstrating that the same sequential patterns of isoform expression are obtained, albeit at a slower rate, both in vivo and in vitro in the absence of any nervous input (Phillips and Bennet 1984; Kano et al 1983; Weydert et al 1984), and so the involvement of a neural factor may still be questionable.

Thyroid hormone, on the other hand, has been shown to be involved in MHC gene expression (Lampre et al 1984, Izumo et al 1986). It was demonstrated that α-cardiac MHC expression increases in hyperthyroid rats, but decreases, with a subsequent increase in β-cardiac MHC expression, in hypothyroid animals. The mechanisms by which thyroid hormone might initiate gene (de)activation has yet to be elucidated. However the identification of its receptor and the demonstration that this has DNA binding properties reveals a possible mechanism by which thyroid hormone may mediate its effect.

Before discussing the question of the control of the alternative processing of gene transcripts, some specific examples of differential splicing will be described.

The expression of troponin-T isoforms has been shown to be subject to alternative splicing. Comparison of four cDNAs with each other and the gene encoding them (Breitbart et al 1984, Breitbart et al 1987, Nadal-Ginard et al 1987) demonstrated that, theoretically, as many as sixty-four distinct isoforms of fast skeletal troponin-T could be produced by the alternative splicing of exons within one gene. Furthermore, analysis of nuclear RNA showed that at least two, and possibly many more, of the isoforms arise from identical primary transcripts, and therefore alternative splicing mechanisms alone are sufficient for producing protein diversity. Imai et al (1986) had previously shown by 2D-gel electrophoresis and immunoblotting that at least forty troponin-T isoforms were coexpressed in checken leg muscle, but whether or not the expression of each was tightly controlled, by specific alternative splicing mechanisms, still has to be determined.

Alternative splicing is also used to generate two distinct isoforms of skeletal muscle myosin light chains (Robert et al 1982). These isoforms are first expressed in late embryonic development and then accumulate at different rates in skeletal muscles of different metabolic types. Comparison of the DNAs of these, with the gene encoding them, indicated that the majority of the coding region and all of the 3' non-coding region was shared by the two isoforms, but that they differed in the first exon and the 5' non-coding region. Thus it would appear that two distinct transcriptional promoters had been used to produce two

distinct primary transcripts, which differed in length by some 10 kb. However since the longer transcript also includes the 5' non-coding region and first exon of the smaller transcript, alternative splicing mechanisms must be used to produce a specific protein isoform (Periasamy et al 1984). In addition since the isoforms accumulate in different muscles at different rates, then it is suggested that two distinct transcriptional factors (or sets of factors) would be required to produce the separate primary transcripts.

McLeod et al (1986) have described the isolation of two tropomyosin cDNAs; one encoding the striated nuscle α-tropomyosin and the other encoding TM_{30}, a non-muscle isoform. Comparison of the nucleotide sequences of these two cDNAs suggests that they were both transcribed from a single gene, and arose via the alternative splicing of mutually exclusive exons. A similar proposal was made about the origin of the β-tropomyosin isoform and TM_{36}, another non-muscle isoform (McLeod et al 1985). However in this case only part of the β-tropomyosin sequence was known. There is no evidence to suggest that the α-tropomyosin and TM_{30} isoforms are coexpressed in either muscle or non-muscle tissues and therefore it would appear that in this particular case one gene encodes two protein isoforms in separate tissues. However as in the case of MLC isoforms, the different 5' termini of the cDNAs suggest that two separate transcriptional promoters may have been used to produce separate pre-mRNAs.

Little is known about any of the mechanisms mediating alternative splicing of exons. However the theoretical possibilities are strikingly similar for each of the three examples outlined above. The first model

involves the primary transcripts forming specific secondary structures which juxtapose exons that have to be spliced together. Computer analysis of the troponin-T gene sequence has shown that the formation of stem and loop structures, which juxtapose specific exons, is possible, and, furthermore, that these structures theoretically could be stabilised by interaction with certain small nuclear RNAs (Breitbart and Nadal-Ginard 1986). In the case of the MLC isoforms, however, since the primary transcripts differ in length by about 10 kb it is thought that each should have a different "folding" capability, and therefore the interaction with additional small nuclear RNAs may not be required to stabilise specific secondary structures.

It has also been suggested that the use of alternative promoters and hence alternative transcription initiation sites, as is the case for MLC transcripts, may predetermine the splicing pattern to be used. Whether or not sequence elements within the primary transcript, acting in-cis, would be sufficient to direct alternative splicing, is not yet known. Following on from this suggestion, however, is one in which the choice of a particular pair of exons for the first splicing event, is thought to predetermine the subsequent splicing pattern. This is because it has been shown that certain "donor" and "acceptor" splice junctions have a different affinity for each other and therefore some pairs of exons may be more compatible for splicing together than others. In this model the control would be exercised by a specific factor(s).

It is not clear why some isoforms of muscle proteins, the actins and the myosin heavy chains, are subject to transcriptional control alone, while others are also subject to alternative splicing of pre mRNA. Any

evolutionary advantage gained by being able to produce greater protein diversity from a relatively small section of the genome, by using alternative splicing mechanisms, must surely be lost by the need for at least one more tier of control. Only a fuller investigation of the genes encoding isoforms of muscle proteins may help to solve this paradox, but more importantly will provide valuable evidence about the control of expression of eukaryotic genes throughout development.

CHAPTER 2

Materials and Methods

A. Materials

2.1 Chemicals

All chemicals were AnalaR Grade supplied by BDH Chemicals Ltd or Formachem (Research International) Ltd, unless otherwise stated in the text. A list of all chemical and equipment suppliers is given in Table 2.1

2.2 Media and Antibiotics

All culture media, and other solutions used in the cloning work, were sterilised by autoclaving at 15 p.s.i. for 15 mins. Solutions containing glucose, sucrose or glycerol were autoclaved at 5 p.s.i. for 30 mins. Heat labile compounds such as antibiotics and thymine were sterilised by passage through a 0.2 μm filter, (Flow Research), and added aseptically to the cooled culture media.

2.2.1 Antibiotics

Antibiotics were supplied by Sigma and used at the following concentrations;

TABLE 2.1 Chemical and Equipment Suppliers

Aldrich Chemical Company, Gillingham, Dorset, England

Anglian Biotechnology Limited, Colchester, Essex, England

Amersham International plc, Amersham, Bucks, England

BDH Chemicals Ltd, Poole, Dorset, England

Beckman Instruments Inc, Edinburgh, Scotland

B.R.L. UK Ltd (GIBCO), Paisley, Scotland

The Boehringer Corporation (London) Ltd, Lewes, E. Sussex, England

James Burroughs Ltd, Fine Alcohols Division, London, England

Difco Laboratories, West Molesey, Surrey, England

Fisons Scientific Apparatus, Loughborough, Leicestershire, England

Flow Research Labs Ltd, Irvine, Scotland

Fluka A.G., Fluorochem Ltd (UK distr.), Glossop, Derbyshire, England

Koch-Light Laboratories Ltd, Colnbrook, Bucks, England

Kodak Ltd, Kirby, Liverpool, England

LKB Instruments Ltd, South Croydon, Surey, England

New England Biolabs, CP Labs Ltd, (UK distr.), Bishops Stortford, Herts, England

Pharmacia Ltd, Milton Keynes, England

Schleicher and Schuell, Anderman & Co. (UK distr.), East Molesey, Surrey, England

Sigma London Chemical Co Ltd, Poole, Dorset, England

Whatman Lab Sale Ltd, Maidstone, Kent, England

UV Products Ltd, Winchester, Hants, England

Ampicillin: stored at -20°C as a 5mg/ml solution in water, used at a concentration of 30 µg/ml in culture media.

Tetracycline: stored at -20°C as a 20mg/ml solution in 50% ethanol and used at a concentration of 20 µg/ml in culture media for routine growth, and 13 µg/ml for transformations.

2.2.2. Liquid Media

L-Broth: 10g Bacto Tryptone (Difco 0123-01)
5g Yeast Extract (Difco 1880-17)
5g NaCl

per litre of distilled water. pH was adjusted to 7.2 with NaOH and the media autoclaved. Antibiotics were added as appropriate.

2xYT Broth; 16g Bacto Tryptone
10g Yeast Extract
5g NaCl

per litre of distilled water.

2.2.3 Medium Containing Agar

LB Agar: 15g Bacto Agar (Difco 0140-01)

per litre of L-Broth with or without antibiotics as appropriate.

2xYT Agar: 15g Bacto Agar

per litre of 2xYT Broth

Hammersmith Stabs: 0.9g Nutrient Broth (Difco 0003-02)

0.75g Bacto Agar

0.5g NaCl

per 100ml of distilled water. pH was adjusted to 7.2 and the solution autoclaved. Once cool 1ml Thymine (10mg/ml stock in water; Sigma) was added.

Glucose/Minimal Media Agar: M9 Salts: 6g Na_2HPO_4

3g KH_2PO_4

1g NH_4Cl

0.5g NaCl

per litre of distilled water.

For minimal media each reagent was sterilised separately and then mixed asceptically;

1 litre	M9 salts
1 ml	1M $MgSO_4$
1 ml	0.1M $CaCl_2$
1 ml	1M Thiamine - HCl
10 ml	20% Glucose
15 g	Bacto Agar

H-Agar:

10 g	Bacto Tryptone
8 g	NaCl
12 g	Bacto Agar

per litre of distilled water.

H-Top Agar: 10 g Bacto Tryptone

8 g NaCl

8 g Bacto Agar

per litre of distilled water. H-top agar was stored as 50-100ml aliquots.

2.2.4 Buffers

T.E. Buffer: 10 mM Tris-HCl pH 8.0

1 mM EDTA

B. METHODS

2.3 General Procedures

The following section describes the maintenance of bacteria, DNA and RNA stocks, as well as those procedures which were used frequently throughout the course of this work.

2.3.1 Bacterial strains

Two strains of bacteria were used and their genotypes appear in Table 2.2.

E.Coli DH1 was used as the host strain for the growth of the two plasmid DNA libraries.

E.coli JM109 was used as the host strain for the growth of the pUC 18 and pUC 19 and related recombinants, as well as the growth of both RF (replicative form) and single-stranged DNA of bacteriophage M13 and M13 derived recombinants.

TABLE 2.2 The genotypes of the E.Coli strains used in this work.

Organism	Strain	Genotype	Reference
E.coli	DH1	F^-, rec A1, end A1 gyr A96, thi, hsdR17 sub E44, relA1, λ^-	Law, 1968 Meselson and Yuan, 1968
E.coli	JM109	rec A1, end A1, gyr A96, thi, hsd R17, SupE44, relA1 λ^-, Δ(lac-pro AB), [F^-, tra D36, pro AB, lac IqZ ΔM15]	Yanish-Perron et al 1985

2.3.2 Storage of Bacteria

Stock cultures of JM109 and DH1 transformed with plasmid DNA, were stored at -70°C under glycerol. Sterile glycerol was added, to 50%, to an exponentially growing culture and vortexed to ensure a homogenous solution was obtained. The bacteria were then stored as 1ml aliquots at -70°C where they remained viable for the period of this work, at least.

Working stocks of DH1 were stored at room temperature for up

to one year in the form of Hammersmith stabs.

Working stocks of JM109 were stored at 4°C on minimal agar plates to ensure the retention of the F episome required for the detection of recombinants.

2.3.3. Storage of bacteriophage M13

M13 virus particles which cannot be frozen were stored in sterile 2xYT medium at 4°C.

2.3.4 Storage of DNA and RNA

Replicative form (RF) and single stranded DNA from M13, and plasmid DNA, was stored at -20°C in T.E. buffer, in tightly fitting capped tubes (Eppendorf).

Genomic DNA was stored in T.E. at 4°C to prevent shearing of the DNA caused by freezing.

RNA, both poly A^+ and total RNA, was stored in sterile tubes at a concentration of 1mg/ml, in sterile T.E., at -70°C.

2.3.5 Digestion of DNA with Restriction Endonucleases

Restriction endonucleases were purchased from the following companies; Bethesda Research Laboratories (BRL), New England Biolabs, The Boehringer Corporation (London) Ltd., Pharmacia, and

Digestion of DNA using restriction endonucleases from BRL were performed using the "REact" system of buffers supplied by B.R.L. Other digestions were performed using High, Medium or Low Salt buffers as indicated by the manufacturer. Restriction endonucleases, Sma I, Bgl I, and Bgl II required more specialised buffers. A list of the 10X stock concentration of all buffers appears in Table 2.3.

One unit of enzyme activity is defined as the amount of enzyme required to digest 1 μg of DNA in 1 hour. In a typical reaction 1 μg of DNA was incubated with a 5-10 fold excess of enzyme to ensure complete digestion. Reaction volumes were kept to a minimum (usually 20 μl), but with the volume of enzyme added never exceeding 10% of the total reaction volume. The reaction mixture was incubated at the appropriate temperature for at least two hours after which the digestion was stopped by the heat denaturation of the enzyme or by the addition of EDTA to 10mM. Digested DNA samples could then be analysed by gel electrophoresis.

2.3.6 Gel electrophoresis of DNA/RNA.

DNA, DNA fragments and RNA were separated by electrophoresis through agarose or polyacrylamide gels as follows:

size range	% gel concentration
> 5kb	0.7% agarose
0.5-5kb	1% agarose
0.1-2kb	2% agarose or
	4% polyacrylamide

In general polyacrylamide gel electrophoresis was used for separating radiolabelled fragments of DNA in preparation for Maxam and Gilbert sequence analysis, and agarose gel electrophoresis used at all other times, unless greater resolving power was required.

Table 2.4 lists the electrophoresis buffers used as well as the sample loading buffers.

TABLE 2.3 Contents of the buffers used for restriction endonuclease digestion of DNA

Buffer	Tris-HCl	NaCl	KCl	$MgSO_4$	DTT
Low Salt	10mM pH7.4	-	-	10mM	1mM
Med Salt	10mM pH7.4	50mM	-	10mM	1mM
High Salt	10mM pH7.4	100mM	-	10mM	-
Sma I	10mM pH8.0	-	20mM	10mM	1mM
Bg1I/Bg1II	10mM pH9.4	-	-	10mM	1mM

The above concentration of ingredients are for the 1x reaction butter. 10x stock solutions of each buffer were prepared and stored at -20°C.

2.3.6.1 Agarose Gel Electrophoresis

Agarose (Type II; BRL) gels were prepared by adding the appropriate amount of agarose to 80ml of 1 x electrophoresis buffer (normally TBE) and heating the mixture to boiling point. The solution was allowed to cool to approximately 50°C and the gel cast in a 12cm x 12cm gel apparatus. Gels were then immersed in the same 1 x running buffer in an electrophoresis tank and the samples loaded using one third volume of "loading buffer". The DNA samples were then subjected to electrophoresis at 50mA for approximately one hour or until the desired resolution was obtained.

After electrophoresis DNA bands could be visualised by illumination under long wave u.v. light using a u.v. transilluminator (U.V. Products), and photographed if desired, using a Polaroid Cu-5 land camera and Type 665 Positive/Negative film.

TABLE 2.4 Content of the buffers required for electropheris with agarose and polyacrylamide gels

Buffer		Loading buffer
10 x T.B.E.:	1M Tris-HCl 1M Boric Acid 10mM EDTA pH is 8.3	0.02% bromo- phenol blue,50% glycerol in 1 x TBE
20 x Tris-Acetate Buffer:		
	800mM Tris 100mM Sodium Acetate 20mM EDTA pH to 7.4 with Acetic Acid	0.02% bromo- phenol blue, 50% glycerol in 1 x Tris-Acetate

Ethidium Bromide was added to 5ng/ml in the 1 x buffer except for polyacrylamide sequencing gels, when it was omitted altogether.

2.3.6.2 Low Melting Point Agarose Gel Electrophoresis

LMP agarose (BRL) gels were prepared as above, but were allowed to set at 4°C. The current was not allowed to exceed 40mA during electrophoresis.

2.3.6.3 Polyacrylamide Gel Electrophoresis

Polyacrylamide gel electrophoresis was performed using 16cm x 16cm glass plates, 1.5mm spacers and the V161, BRL vertical gel apparatus. A typical gel contained 4% acrylamide (19% acrylamide, 1% N,N-methylene bisacrylamide stock; BRL), 10% glycerol and 0.04% ammonium persulphate in a total volume of 50ml 1 x TBE (Table 2.4). The gel plug was made by mixing 5ml of the above gel mixture with 40 µl TEMED (N,N,N'N'-tetramethylethylene-diamine) and pouring. Once the gel plug had set 30 µl TEMED was added to the remaining gel mixture and the main gel cast.

DNA samples were prepared by the addition of 0.3 volumes of loading buffer (50% glycerol, 0.05% Xylene Cyanol, 0.05% Bromophenolblue, in 10xTBE). They were then loaded onto the gel and electrophoresed, in 1x TBE, at 200V for 2-3 hours depending on the resolution required. In a typical 4% polyacrylamide gel the xylene cyanol migrated with a mobility equivalent to a DNA fragment of 350bp in length, and the bromophenol blue, one of 70bp in length.

After electrophoresis the gel was left on one glass plate and stained for approximately 10 minutes in a 40µg/ml solution of

Ethidium Bromide. The DNA was then visualised as before, under UV illumination, and photographed if required.

2.3.7 Extraction of DNA from Agarose/Polyacrylamide Gels

2.3.7.1 Elution of DNA from Agarose Gels

Electrophoresis of digested DNA was performed in Tris-acetate buffer at 60V for as long as was necessary to separate the required fragment(s).

The DNA band(s) was then excised from the gel with a scalpel and placed in 9mm dialysis tubing, (The Scientific Instrument Co.), with 200 µl of 1 x Tris-acetate buffer, and sealed at both ends. The tubing was then placed in the electrophoresis tank, in known orientation, and electrophoresis performed at 60V for one hour, after which the solution was removed from the dialysis bag and retained. A further 200 µl of 1x Tris-acetate buffer was then added to the dialysis bag and electrophoresis continued in the same orientation for a further 15 mins, after which the orientation was reversed and back-electrophoresis performed for 1-2 minutes. The solution was again removed from the bag and pooled with the earlier aliquot. This was then centrifuged to remove agarose. The supernatant was then transferred to a fresh Eppendorff tube and the DNA precipitated with ethanol (see Section 2.3.8).

2.3.7.2 Extraction of DNA from Low Melting Point Agarose Gels.

Electrophoresis was performed for as long as was necessary to separate the required fragment(s), which was then excised from the gel and placed in a 1.5ml Eppendorff tube with 100 μl TE buffer. The gel was then melted by heating to 65^{o}C for 15 minutes, and the agarose removed with phenol. The DNA was then recovered from the supernatant by ethanol precipitation (both Section 2.3.8).

2.3.7.3. Elution of DNA from Polyacrylamide Gels

The appropriate DNA band(s) was cut from the gel and placed in a siliconised 1ml Eppendorff tip heat-sealed at one end, and containing siliconised glass wool. The gel was ground up using a glass rod, and then washed to the bottom of the tip using 600 μl of Eluting buffer (500mM ammonium acetate, 10mM magnesium acetate 1mM EDTA, 0.1% SDS). The tip was then sealed at the other end using parafilm and incubated at 37^{o}C overnight.

After incubation the sealed end of the tip was cut off with scissors, the parafilm removed, and the buffer allowed to drain into a siliconised glass centrifuge tube (e.g. 15ml Corex). The tip was then rinsed, four times, with 200 μl of Elution Buffer and the DNA precipitated from the pooled eluate with ethanol (Section 2.3.8). The DNA was sedimented by centrifugation at 2000g and resuspended in 400 μl 3M sodium acetate pH 6.0. Pieces of acrylamide were then removed by centrifugation in a table-top Eppendorff centrifuge, after which the DNA was recovered from the supernatant by ethanol

precipitation.

2.3.8 Phenol Extraction and Ethanol Precipitation of DNA/RNA.

DNA was routinely purified free of protein by denaturation, and subsequent removal of the protein, by phenol, after which it was recovered from the aqueous supernatant by ethanol precipitation.

2.3.8.1 Phenol Extraction

Typically a mixture of phenol:chloroform:isoamyl alcohol (25:24:1) was used. Generally the aqueous phase was extracted three times by vortexing it with an equal volume of the phenol mixture, with the phases being separated each time, by centrifugation. When gentle handling was required, the aqueous and phenol phases were mixed by hand, since vortexing would have sheared the high molecular weight DNA.

Any residual phenol remaining in the aqueous phase after extraction, was removed by extracting twice with water saturated diethylether, after which the nucleic acids were recovered from the aqueous phase by ethanol precipitation.

2.3.8.2 Ethanol Precipitation

Nucleic acids were recovered from aqueous solutions by the addition of 0.1 volume 3M sodium acetate pH 6.0 and 2.5 volumes ethanol, and incubating overnight at -20°C or at -70°C for

15 minutes.

After incubation the nucleic acids were sedimented by centrifugation at ca 10,000 rpm (17,000g) and then washed with 80% ethanol to remove excess salt. After recentrifutation, the nucleic acids were dried by lyophilisation, and finally redissolved, to the required concentration in T.E. buffer.

2.4 Isolation of Nucleic Acids

The following section describes the various methods used for isolating, plasmid DNA, mRNA and bacteriophage M13 replicative form (RF) and single-stranded DNA.

2.4.1 Isolation of Plasmid DNA

2.4.1.1 Large Scale Isolation of Plasmid DNA (Birnboim & Doly 1976)

Growth and Amplification of Plasmid DNA

25ml of L-Broth containing the appropriate antibiotic was inoculated with a single colony of the bacterial clone in question and incubated overnight at 37°C in an orbital shaker. Two 5ml aliquots of the overnight culture was then taken to seed two flasks, each containing 800ml of L-Broth (without antibiotic) and the bacteria grown until the turbity of the broth reached an $O.D._{600} = 0.8$. Bacterial growth was then halted and plasmid copy number amplified by the addition of chloramphenicol (Sigma) to 165 μg/ml and further

incubation at 37°C on the orbital shaker overnight.

Lysis of Bacteria and Extraction of Nucleic Acid

Bacteria were sedimented by centrifugation at 5000 rpm (4000_g) for 10 minutes, in a suitable large capacity rotor, the supernatant discarded, and the cells resuspended in 9.5ml total of 0.05M Glucose, 0.01M EDTA, 0.025M Tris-HCl pH8.0. The cells were then lysed by the addition of lysozyme (Sigma) to 2mg/ml and incubation at 0°C for 30 minutes. A further incubation at 0°C for five minutes with 20ml 0.2M NaOH, 0.1% w/v SDS, served to irreversably denature the host cell protein and chromosomal DNA, after which the plasmid DNA was selectively renatured by neutralising the solution with the addition of 15ml 3M sodium acetate pH4.8 and a final incubation at 0°C for 1 hour.

Cellular debris was removed by centrifugation at 30,000 rpm (64,000g) for 30 minutes in a Beckman Ti60 rotor, the supernatant collected and the nucleic acids precipitated by the addition of 0.6 volume isopropanol.

Purification of Plasmid DNA

The nucleic acids were sedimented by centrifugation at 10000 rpm (17,000g) for 15 minutes in a suitable rotor, such as the Sorvall SS-34, and then redissolved in 30ml T.E.

28.9g caesium chloride (BCL) was dissolved in this to give a

solution of density = 1.59 to which ethidium bromide was added to a concentration of 0.4mg/ml. Protein and other debris precipitated by the addition of caesium chloride was removed by subjecting the solution to a clarifying centrifugation at 1500 rpm (500g) for 30 minutes in a Beckman TJ-6 'table top' centrifuge.

The clarified solution was then transfered to sealable Beckman VTi50 centrifuge tubes which were filled to capacity, if necessary, by the addition of ceasium chloride solution of the same density. The tubes were sealed and the solution subjected to ultracentrifugation at 50,000 rpm (240,000g) and 20°C for at least 16 hours in order to equilibrate the DNA.

After untracentrifugation the plasmid DNA could be visualised, under long wave ultra violet illumination, as the lowest discrete band in the tube. The DNA was then collected by piercing the tube with a 21g hypodermic needle just below the band and allowing it to drain into a sterile centrifuge tube, being careful not to contaminate the collected band with any other band appearing above it in the tube. At this stage, however, the collected plasmid DNA was usually still contaminated by tRNA, but which was removed if necessary by ultracentrifugation, at 65,000 rpm (400,000g) and 20°C for a further 16 hours, in a Beckman VTi65 rotor. The band of plasmid DNA was subsequently collected as outlined previously.

Ethidium bromide was removed from the pooled solution by extracting it four times with 3-methyl-1-butanol (Koch-Light). The volume of solution was then measured and four times this volume of

T.E. was added to it. The DNA was recovered by ethanol precipitation overnight.

The precipitated DNA was sedimented by centrifugation at 10000 rpm (17,000g) and 0°C for 30 minutes in a suitable rotor (Sorvall HB-4), thereafter resuspended in a small volume (eg 100 µl) T.E. and transfered to a 1.5ml Eppendorff tube where it was again precipitated with ethanol. The final DNA pellet obtained from this stage was lyophilised and then dissolved in a suitable volume of T.E. for storage. The yield was usually of the order of 1mg.

2.4.1.2 Small Scale Isolation of Plasmid DNA (Holmes & Quigley 1981)

Growth of Bacteria

A single colony of the bacterial clone in question was spread onto the appropriate antibiotic selection plate in such a way as to produce a lawn of cells covering half of the plate. The bacteria were allowed to grow overnight.

Lysis of Bacteria and Extraction of Plasmid DNA

The bacteria were harvested by gently scraping the cells off the agar plate, and then resuspending them in 1ml of a solution containing 8% w/v Sucrose, 0.04M Tris-HCl pH8.0, 0.05M EDTA pH7.5, 5% v/v Triton X-100, in a 1.5ml Eppendorff tube. Lysis and denaturation of protein and genomic DNA was achieved by the addition of Lysozyme (Sigma) to 200 µg/ml and subsequent boiling of the

solution for 7 minutes.

Cellular debris was then sedimented by centrifugation in an Eppendorff centrifuge, and 0.6ml of the remaining supernatant taken for isolating the plasmid DNA.

The supernatant, at this stage, was usually contaminated with RNA and so boiled RNAase A (Sigma - 1mg/ml stock in H_2O) was added to approximately 4 μg/ml and the solution incubated at 37°C for 15 minutes. The RNAase A was then inactivated by the addition of 1 μl diethylpyrocarbonate (Sigma) and a further incubation at 65°C for 10 minutes.

Plasmid DNA was precipitated by the addition of 2/5 volume, 5M ammonium acetate (unbuffered) and 0.9 volume isopropanol with an incubation on dry ice for five minutes. DNA was sedimented by centrifugation in an Eppendorff centrifuge for 10 minutes and then resuspended in 0.3M ammonium acetate/70% isopropanol. The DNA was again sedimented and then washed with 80% ethanol, after which the final pellet was lyophilised, dissolved in 30 μl T.E. and stored at -20°C.

The DNA obtained was occasionally still contamined with tRNA and protein, but was usually suitable for performing restriction enzyme analysis. 1-2μl of this prep was sufficient for visualisation on an agarose gel.

2.4.2 Isolation of RNA and Purification of Poly A^+ RNA

2.4.2.1 Isolation from mouse tissue (Kirby 1956)

All solutions and items of equipment, including dissecting instruments and glassware, were sterilised, by autoclaving, before use. Disposable gloves were worn at all times to prevent degradation of the RNA by contaminating ribonucleases.

Preparation of Tissue

RNA was prepared from 11 day old mice (Porton), and although different tissues were used (brain, liver, skeletal muscle and cardiac muscle) the technique for isolating the RNA was identical in each case.

The tissue was excised rapidly from the dead animal and placed immediately into liquid nitrogen. It was ground in a mortar until it was completely powdered and then added to 20ml Calvaria Buffer (0.05M NaCl, 0.005M sodium acetate pH5.0, 0.001M EDTA, 1% w/v SDS).

Extraction of Protein

Protein was denatured and extracted using phenol (Section 2.3.8). The extraction was performed, at 65°C for 3 minutes, in 30ml glass tubes (Corex), with the phases being separated each time by centrifugation; the aqueous phase was drawn off with a pasteur pipette and stored until later. The phenol phase was reextracted

with a further 20ml Calvaria Buffer which was then pooled with the first aqueous phase and reextracted with phenol as often as was required to produce a clear solution. Nucleic acids were then recovered from this aqueous phase by precipitation with ethanol (Section 2.3.8).

Isolation of RNA

The nucleic acids were sedimented by centrifugation and then redissolved in 4ml 1% w/v N-lauroyl sarcosine (Sigma L5125), 0.01M EDTA pH 7.0. This slow process was accelerated by passing successively through sterile 19g and 23g hypodermic needles which also served to sheer the genomic DNA, which helped to reduce losses of RNA.

Once dissolved, 4g of caesium chloride was added and allowed to dissolve. The sample was then halved and each was layered over a ceasium chloride cushion (5g CsCl in 4ml 0.01M EDTA pH 7.0) in Beckman SW40 rotor centrifuge tubes. Approximately 0.5ml of the sarcosyl buffer was then laid over this and the tubes then filled to capacity with parafin oil.

RNA was separated from the DNA, by ultracentrifugation at 17000 rpm ($36,000_g$) and 20°C overnight. After centrifugation the supernatant was removed to below the level of the caesium chloride interface, the tubes drained, and cut down to prevent the RNA precipitate being contaminated by DNA or protein higher up the tube. The RNA precipitate was then dissolved in 2ml sterile water

and the yield and purity checked by measuring the absorbance at 260nm and 280nm (1 A_{260} unit=40μg RNA). Essentially pure RNA should give an A_{260}: A_{280} ratio of two, but in practice was sometimes slightly less. Finally the RNA was recovered by ethanol precipitation and the resulting precipitate redissolved in a suitable volume of T.E. buffer.

Isolation of Poly $(A)^+$ RNA (Edmonds et al 1971)

Since poly $(A)^+$ RNA represents only 5% of total RNA a minimum of 1mg of total RNA must be obtained from the previous stage before it is beneficial to proceed with poly$(A)^+$ RNA isolation.

Isolation was carried out by means of oligo dT cellulose chromatography through a small column. 100mg of oligo dT cellulose (BRL) was applied to the column in water/diethylpyrocarbonate. The column was then washed with water/diethylpyrocarbonate followed by a large volume of loading buffer (0.5M NaCl, 0.01M Tris-HCl pH 7.5, 0.001M EDTA). Total RNA, in 2ml loading buffer, was heated to 65°C for two minutes, cooled rapidly, and then applied to the column. The eluate was collected and the procedure repeated twice. After the third passage through the column unbound RNA was eluted from it by washing the column with loading buffer plus 0.5% w/v SDS. At this stage 1ml fractions were collected and the absorbance at 260nm monitored until it fell below 0.05. The fractions collected up to this point contain the poly $(A)^-$ RNA. Poly $(A)^+$ RNA was eluted from the column using elution buffer, (0.01M Tris-HCl pH 7.5, 0.001M EDTA, 0.2% w/v SDS), with the absorbance at 260nm being measured to

calculate the yield.

Once the appropriate fractions were pooled the poly $(A)^+$ RNA was precipitated with ethanol overnight and the resultant precipitate dissolved and stored in sterile water at -70°C.

2.4.2.2 Isolation of RNA from C2C12 cells, a mouse myogenic cell line (Blau et al 1985)

C2C12 cells are derived from the satellite cells isolated from crushed thigh muscle (Yaffe and Saxel 1977). At high concentrations of serum they can be cultured as mononucleate myoblasts, but at low serum concentrations they withdraw from the cell cycle and form multinucleated myotubes. The C2C12 cell line was a gift from R. Akhurst, St Mary's Hospital, London.

C2C12 cells were cultured in 80cm^2 flasks in 1 x Delbeco's Modification of Eagle's Medium (DMEM; - NBL) containing 0.375% sodium bicarbonate (GIBCO), 0.05 mM glutamine (GIBCO), 20% foetal bovine serum (Flow) and 0.5% chicken embryo extract (Flow). The cells were grown at low density, with the culture medium being changed every 48 hours, to prevent premature withdrawal from the cell cycle. Differentiation was promoted by allowing the cells to grow to about 70% confluence and then reducing the serum concentration. The fusion media used was 1 x DMEM, 0.375% sodium bicarbonate, 0.05mM glutamine, 2% horse serum (Flow).

Two 80cm^2 flasks were used to prepare RNA. The cells were

harvested by scraping them into the culture medium with a rubber policeman. They were then sedimented by centrifugation at approximately 1500g and then resuspended in 0.5ml 1 x Calvaria buffer (Section 2.4.3.1). The protein was denatured and extracted with phenol (Section 2.3.8) and total nucleic acids recovered from the aqueous phase by precipitation with ethanol. The RNA was separated by ultracentrifugation through a ceasium chloride step gradient, as before (Section 2.4.3.2), but this time the preparation was scaled down for use with the Beckman SW56 rotor, where the tube capacity is only 4ml. Ultracentrifugation was at 19,000 rpm (36,000g) and 20°C, overnight.

The resulting RNA precipitate was treated as before, with the RNA finally being redissolved in an appropriate volume of T.E. buffer and stored at -70°C.

2.4.3 Isolation of Single-Stranded Bacteriophage M13 DNA

The following procedure was used to isolate single-stranded DNA from both wild-type M13 phage and recombinants derived from M13.

Growth and Amplification of Phage

Single plaques, obtained from the transformation of competent JM109 cells, with single stranded (ss) or replicative form (RF) DNA, were picked and inoculated into separate 2ml volumes of a 1/100 dilution of an overnight culture of JM109 cells. The cells were then shaken on an orbital incubator at 37oC for 5-6 hours after

which they were harvested by centrifugation in an Eppendorff centrifuge for five minutes.

At this stage the supernatant contains the phage particles and so should be retained for preparation of single-stranded DNA. However the sedimented cells can also be retained for a small scale isolation of replicative form DNA by the procedure outlined in section 2.4.1.2.

Isolation of single-stranded DNA

The supernatant was poured into a new Eppendorff tube and the phage particles precipitated by the addition of 20% w/v polyethylene glycol 6000, 2.5M NaCl to 20% v/v and incubating at room temperature for fifteen minutes. The phage was then sedimented by centrifugation in an Eppendorff centrifuge for five minutes after which the supernatant was aspirated off. It is very important to remove all traces of the polyethlene glycol 6000, therefore the tubes were recentrifuged and any remaining supernatant was again drawn off with an aspirator.

The phage pellet was resuspended in 100 μl T.E. buffer to which 100 μl phenol was added and the mixture vortexed to destroy the phage coat and other phage proteins. Approximately 500 μl of chloroform was then added and the mixture vortexed to further reduce the risk of contamination by polyethylene glycol 6000. The two phases were then separated by centrifugation and the aqueous phase drawn off. Single- stranded DNA was then recovered from the aqueous

phase by precipitation with ethanol (Section 2.3.8) and the final DNA precipitate redissolved and stored in 25 µl T.E. Buffer.

2.4.4 Preparation of M13 Replicative Form DNA

The following procedure was used to isolate replicative form (RF) DNA from wild type M13 mp 18 and mp 19 virus and that of recombinants derived from M13.

Growth of Bacteria and Amplification of Replicative Form DNA

Competent JM 109 cells were transformed with the single-stranded DNA template of interest and plated out on H-Agar plates. A single plaque was picked and inoculated into 1.5ml of a 1/100 dilution of an overnight culture of JM109 cells and incubated on an orbital shaker at 37°C for 4-5 hours. The cells were then sedimented and the supernatant shored at +4°C overnight. 100ml of fresh 2 x YT medium was inoculated with 1ml of an overnight culture of JM109 cells, 1ml of the stored supernatant, and then incubated at 37°C for five hours.

The cells were harvested, in a suitable large capacity centrifuge, and the RF DNA isolated in the same manner as for the large scale isolation of plasmid DNA outlined in Section 2.4.2.1

2.5 Synthesis of single-stranded complementary DNA

(Wickens et al 1978)

Single stranded complementary DNA, (cDNA), was synthesised from poly $(A)^+$ RNA isolated from both skeletal and cardiac muscle tissue. A minimum of 1 μg of poly $(A)^+$ RNA was required for each batch of cDNA synthesis.

The reaction buffer was assembled fresh on each occasion and contained 50mM Tris-HCl pH 8.3, 10mM $MgCl_2$, 100 mM KCl, 50 μM of non-radiolabelled dATP, 1mM of each of the other three dNTPs, 10mM dithiothreitol and 0.1 mg/ml oligo-dT (12-18; DCL). Preliminary experiments had been performed to determine the optimum concentration of the unlabelled radioactive nucleotide (dATP) which would give the best combination of length and specific activity of the cDNA synthesised.

The reaction mixture, containing 60 μl Reaction buffer, 10μg poly $(A)^+$ RNA, 25 units of AMV Super Reverse Transcriptase (Anglian) and 20 μl of α-^{32}P dATP, in a total volume of 100 μl, was assembled on ice and then incubated at 42°C for one and half hours. Poly $(A)^+$ RNA was removed from the reaction mixture by boiling for three minutes, followed by rapid chilling of the solution. The cDNA was then recovered from the reaction mixture by size-exclusion chromotography, using the same method as outlined in section 2.6.2. The cDNA could then be checked for length by electrophoresing small sample of it on an agarose gel, drying the gel down and subjecting it to autoradiography.

2.6 Radioactive Labelling of DNA

Fragments of DNA were labelled with ^{32}P or ^{35}S prior to Maxam and Gilbert sequencing, for use as radioactive probes in Southern and Northern transfer, and as radioactive molecular weight markers for use in Southern and Northern Blots. Three methods were used.

2.6.1 End labelling of DNA fragments with the Klenow Fragment of *E.Coli* DNA Polymerase I.

This method was used to prepare DNA for subsequent sequencing by the method of Maxam and Gilbert and for the preparation of radioactive molecular weight markers. The method relies primarily on the initial restriction of the DNA with a restriction enzyme which leaves 5' protruding ends. The DNA is then labelled by filling in these "sticky" ends, with complementary radioactive bases, using the Klenow enzyme.

After restricting the DNA with an appropriate enzyme it was ethanol precipitated (Section 2.3.8) and the following added to the dried DNA precipitate; 3.75 µl of the Medium Salt Restriction Buffer (Table 2.3), 30 µl ^{32}P dNTP (an appropriate radioactive dNTP complementary to one of the nucleotides in the "sticky" end), 1 µl of a mixture of the other three non-radioactive dNTPs each at a concentration of 0.1mM, and one unit of Klenow fragment (BRL). The reaction volume was made up to 25 µl with sterile distilled water and the mixture incubated at room temperature for 30 minutes.

The radioactively labelled DNA fragment was then removed from the majority of unreacted mucleotides by precipitating it with 90 µl 2.5M ammonium acetate (unbuffered) and 360 µl ethanol for five minutes in dry ice (solid CO_2; Distillers). The DNA was sedimented by centrifugation, with the supernatant containing the unincorporated radioactive nucleotides being discarded. The DNA precipitate was then resuspended in 100 µl 0.3M sodium acetate pH 6.0, and precipitated with 300 µl ethanol and incubation in dry ice for 15 minutes. The DNA was then sedimented, lyophilised, and stored at -20°C.

The two separate precipitation steps were usually sufficient to remove the unreacted radioactive dNTP, however since the DNA is subjected to electrophoresis, at a later stage, slight contamination did not matter.

When labelling DNA for use as radioactive molecular weight markers ^{35}S dNTP was substituted for the ^{32}P dNTP.

2.6.2 Nick-Translation of DNA Fragments (Rigby et al 1977)

The following method was used to radioactively label DNA fragments for use as probes for Northern and Southern transfer.

The DNA fragment to be labelled (for optimum labelling 0.3-1 µg were used) was dissolved in a total of 2.5 µl water. To this was added 1.8 µl of 10x Medium Salt Restriction Buffer (Table 2.3), 50 µCi of

any α-^{32}P dNTP, 1.5 µl of a mixture of the other three non-radioactively labelled dNTPs, each at a concentration of 0.1mM, 1 µl of a fresh 1 in 10,000 dilution of DNAase I (5 mg/ml stock in distilled water) and one unit of E. Coli DNA Polymerase I (BRL). The DNAase I was used to introduce nicks into the DNA fragment which were subsequently repaired by DNA polymerase using the mixture of radioactive and non-radioactive dNTPs. The reaction mixture was incubated at 15°C for at least four hours.

Once the reaction was complete the unreacted radioactive nucleotides were separated from the labelled DNA fragment by size-exclusion chromotography. A 1ml plastic Eppendorff tip, or a pasteur pipette, was packed with siliconised glass fibre and a column of Biogel P60 (Biorad) was poured. The Biogel P60 is stored preswollen in N.E. buffer (50mM NaCl, 0.5mM EDTA pH 7.0) with 0.02% w/v sodium azide as preservative.

The volume of the sample was increased to 100 µl with NE Buffer, and then passed through the column which had been previously preequillibrated with several volumes of NE buffer. The eluate was collected and the column washed with further 100 µl volumes of N.E. Each 100 µl fraction was collected in separate Eppendorff tubes, with ten fractions in total being collected.

The labelled DNA fragment eluted first, appearing as a peak of radioactivity centered around fraction 5 or 6, with the unreacted radioactive nucleotides appearing mainly in fractions 9 and 10. The radioactivity in each 100 µl fraction was measured either in a

scintillation spectrometer using Cherenkoff counting or with a Type 5.10E hand held Geiger counter.

The fractions containing the labelled DNA fragment were pooled and stored in a lead pot at -20°C.

2.6.3 Randomly Primed Labelling of DNA Fragments by E.Coli DNA Polymerase I (Feinberg and Vogelstein 1983)

The following method was used for labelling of DNA fragments used as probes in Northern and Southern transfer. It was found to be a particularly useful method of labelling DNA if the quantity of DNA available was limited, since as little as 30ng of DNA could be labelled to high specific activity. However problems of probe specificity may arise if the length of the DNA fragment to be labelled is less than 1500 b.p.

Solutions required for the Labelling Reaction

Solution A: Each of the three non-radioactively labelled dNTPs each at a concentration of 100mM in 1 ml 1.25 M Tris-HCl pH 8.0, 0.125 M $MgCl_2$, 18 μl β-mercaptoethanol. This solution is stored at -20°C.

Solution B: 2M Hepes (Sigma) pH 6.6, stored at +4°C.

Solution C: Hexadesoxyribonucleotides (Pharmacia Ph No 2166) evenly suspended in T.E. buffer at 90 A_{260} units per ml, and stored at -20°C.

Stop Solution: 0.02M NaCl, 0.02M Tris HCl pH 7.5, 0.002M, EDTA, 0.25% w/v SDS, 1 μM dNTP (the non-radioactive version of the labelled dNTP used).

Bovine Serum Albumin (BSA) as a stock solution of 10 mg/ml, ^{32}P dNTP, sterile distilled water and E. Coli DNA polymerase I were also needed for the reaction.

The DNA was added to 10 μl reaction mixture, 2 μl BSA, 20 μCi ^{32}P dNTP, 2-3 units Pol 1 and the volume increased to 50 μl with sterile distilled water. The solution was then incubated at room temperature overnight, after which 50 μl of Stop Solution was added. The unreacted radioactive nucleotides were then separated from the labelled DNA fragments by size-exclusion chromatography as outlined for nick-translation (Section 4.2).

2.7 Southern and Northern Transfer

Southern and Northern transfer procedures were used to transfer DNA and RNA samples respectively from agarose gels on to nitrocellulose (Schlier & Schull) or nylon (NEN or PALL) membranes for subsequent hybridisation to radioactive probes. DNA and RNA samples were also spotted directly onto the nitrocellulose or nylon membranes for "dot-blots".

2.7.1 Southern Transfer (Southern 1975)

DNA was separated on an agarose gel (Section 2.3.6) with T.B.E. as the electrophoresis buffer. The gel was photographed before proceeding. It was then bathed in approximately 2.5 volumes 0.5M NaOH, 1.5M NaCl for 30 mins in order to denature the DNA. The denaturing solution was poured off, the gel rinsed twice in distilled water and then neutralised in approximately 5 volumes of 0.5M Tris-HCl pH 7.6, 1.5M NaCl for 30 minutes. It was important to neutralise the gel, since alkaline conditions have a deleterious effect on the nitrocellulose membranes.

While the gel was soaking, the blotting apparatus was assembled, comprising two reservoirs, each containing approximately 300ml 20 x SSC (3M NaCl, 0.3M sodium citrate), connected by a solid support. A double layer of 3MM Chromatography paper (Whatman) was placed on the solid support with each end extending into the two reservoirs to act as wicks. The 3MM paper was allowed to soak completely before proceeding. Once neutralisation of the gel was complete it was placed, in known orientation, on top of the solid support. The nitrocellulose membrane, appropriately marked for easy identification of the gel tracks, and four layers of 3MM paper, were then placed on top of the gel, ensuring that no air bubbles were trapped between any of the layers, especially between the gel and the nitrocellulose membrane. The nitrocellulose membrane and the four layers of blotting paper were presoaked in 2 x SSC (0.3M NaCl, 0.03M sodium citrate), to promote transfer of the DNA from the gel

to the nitrocellulose by the salt gradient. Finally approximately 6cm of dry "tissues", weighted down by a 1kg weight, were stacked on top of the blotting paper to draw the buffer from the reservoirs up through the gel, thus carrying the DNA with it onto the nitrocellulose membrane where it binds.

Transfer was allowed to proceed overnight, and once complete the nitrocellulose membrane was carefully washed in 2 x SSC for 5 minutes to remove any gel still adhering to it. The nitrocellulose was then allowed to air dry at room temperature, and the DNA immobilised covalently, by baking the filter in a vacuum oven at 80°C for two hours. The filter was then ready for hybridisation to a radioactive probe.

2.7.2 Northern Transfer (Thomas 1980 , Ellis et al 1983)

The procedure for transferring RNA from agarose gels to nitrocellulose or nylon membrances was identical to that for Southern transfer (Section 2.7.1). However since RNA is hydrolysed by strong alkali, it had to be denatured in a manner other than that of soaking the gel in sodium hydroxide. Therefore RNA was separated on a denaturing agarose gel.

A 1% agarose acetate formaldehyde gel was prepared by dissolving 0.6g agarose (BRL) in 60ml 1 x RB (0.02M MOPS, 0.005M sodium acetate, 0.001M EDTA pH 7.0). The solution was allowed to cool below 60°C after which 10ml 35-37% formaldehyde (Koch-Light) was added to a final concentration of 2.2M and the gel cast.

RNA samples were prepared by adding three volumes of a solution containing 66% v/v Formamid; FLUKA, 13% v/v 10x RB, 20% v/v 35-37% formaldehyde and heating to 60°C for 5 minutes. 0.5 volumes of loading buffer (10% w/v Ficol; Pharmacia, 0.1% w/v Bromophenol blue; Sigma; in water) was added to the samples and electrophoresis performed at 60V in 1 x RB containing 2.2M formaldehyde for the length of time it took for the dye to migrate 80% of the length of the gel.

The gel was not stained with ethidium bromide since the formaldehyde prevents fluorescence of the RNA and DNA size markers. Therefore radiolabelled DNA size markers (Section 2.6.1) were generally used instead.

2.7.3 Southern and Northern 'Dot Blots'

This procedure involved spotting the DNA and RNA directly onto the membrane, thus eliminating the gel eletrophoresis stage. Before loading the nucleic acid samples, the nitrocellulose filter was soaked in 20 x SSC and then allowed to air dry at room temperature.

DNA samples, usually about 2μl of the plasmid DNA from the small scale isolation outlined in Section 2.4.1.2, were denatured by heating to 90-100°C for 5 minutes followed by rapid cooling on ice. Samples were then spotted, 1 μl at a time, directly onto the nitrocellulose filter. Once all the samples had been loaded the filter was baked in a vacuum over at 80°C for two hours.

The procedure for spotting RNA samples onto the membrane was identical, but for the fact that the RNA was denatured by heating to 65°C for 15 minutes, before rapid cooling and loading.

2.7.4 Hybridisation of Immobilised DNA to Radioactive Probes

Detection of RNA or DNA immobilised on nitrocelulose or nylon membranes is dependent on hybrids forming between the immobilised nucleic acids and the radiolabelled probe. Consequently the conditions under which stable hybrids can form have to be considered.

The use of Formamid (Fluka) in the hybridisation medium lowers the temperature at which stable hybrids can form, thus eliminating the problems caused by higher temperatures needed in its absence. Similarly salt concentration in the hybridisation medium is kept relatively high to promote the formation of stable hybrids. Thus by manipulating temperature and salt concentration a variety of hybrids can be detected dependent on their relative stability.

The baked nitrocellulose filter was taken and presoaked in 5 x SSPE (0.9M NaCl, 0.01M NaH_2PO_4, 0.05M NaOH, 0.001M EDTA) and then prehybridised, at 42°C for at least two hours, in the hybridisation solution (50% v/v deionised formamid, 5 x SSPE, 0.1% w/v SDS, 10 x Denhardt's solution) 50 x Denhardt's solution is 1% w/v Ficoll, 1% w/v Polyvinylpyrolidine, 1% w/v B.S.A. - all Sigma - filtered through Whatman No.1 filter paper. This is to block non-

specific binding of the probe to the nitrocellulose filter.

After prehybridisation was complete the radiolabelled probe was denatured by heating to approximately 80°C for 10 minutes or by the adddition of 0.1 volume 1M NaOH, followed 10 minutes later by 0.1 volume 1M Tris-HCl pH 7.5 and 0.1 vol 1M HCl. The denatured probe was added to fresh hybridisation solution (as above) and sealed in a polythene bag with the prehybridised filter, ensuring that no air bubbles were present. Hybridisation was then performed at 42°C for at least 16 hours.

Once hybridisation was complete the radioactive solution was carefully removed and the filter washed to remove non specifically bound probe. The conditions under which the filters were washed were altered to suit the stringency required for detecting the sequences of interest.

The stability of the hybrids is decreased by raising the temperature and lowering the salt concentration of the washing solution. Thus conditions were obtimised for detecting non-identical hybrids as well as identical hybrids.

Filters were washed initially in 2 x SSC, 0.1% w/v SDS (2 x SSC is 0.3M NaCl, 0.03M Sodium Citrate) at room temperature for 50 minutes with the washing solution being changed every 10 minutes. This was sufficient for detecting partially homologous sequences, however if highly homologous sequences were to be detected then the stringency was increased by raising the washing temperature to

CONDITIONS OF STRINGENCY USED FOR WASHING HYBRIDISED FILTERS

Filters washed at high stringency were washed in 2 x SSC at room temperature for 50 minutes and then for a further hour in 0.1 x SSC at a temperature of 60^{o}C.

Filters washed at medium stringency were also washed for 50 minutes at room temperature in 2 x SSC and then in 0.1 x SSC for 1 hour at a temperature in the range 42^{o}-50^{o}.

Filters washed at low stringency were subjected only to the 50 minute wash in 2 x SSC at room temperature.

45°C and above, whilst also lowering the salt concentration to 0.1 x SSC, 0.1%, w/v SDS. The filter was then washed for a further 60 minutes with one change of washing solution.
– see opposite for detailed washing conditions.

When washing was complete, the filers were allowed to air dry and were then subjected to autoradiography using X-OMAT film (Kodak).

2.8 Screening of a Plasmid Vector Complementary DNA Library (Grunstein and Hogness 1975)

The plasmid library was stored as frozen cells under glycerol at -70°C. When it was to be screened a small piece of the frozen stock was chipped off and into 1ml sterile L-Broth (Section 2.2) containing 30% v/v glycerol. Serial dilutions were made of this, and the library titred by growth of the bacteria overnight or L- plates containing tetracycline. The original 1ml of L-Broth with glycerol was stored overnight at -20°C.

Once titred the appropriate dilutions were made, and the plasmid library grown on L-plates plus tetracycline (20 μg/ml) at a density of approximately 500 colonies per 9cm petri dish.

When the colonies had grown, but before they became too large and diffuse, a nitrocellulose filter was placed directly on top of the colonies, for approximately 5 minutes, and its orientation marked precisely both on the filter and the master plate of bacteria. The filter was then taken and placed 'other side up' on a

nutrient agar plate (L-broth plus tetracycline). The colonies were then allowed to grow directly on the nitrocellulose filter. In this way several duplicate filters of bacteria could be prepared for use with different probes. Once grown the colonies could be lysed and the plasmid DNA, contained within them, baked onto the nitrocellulose filters.

The colonies were lysed, and the proteins and genomic DNA denatured, by placing the filters colony side up, successively on top of a small volume (1ml) of several different solutions which are listed in Table 2.5. Bacterial debris was then removed from the filters before proceeding with covalent attachment of the plasmid DNA and subsequent hybridisation steps. Figure 2.1 presents a flow chart indicating this procedure.

Table 2.5 The Solutions requried for Colony Lysis

Solution A:- 25% Sucrose, 50mM Tris-HCl pH 7.6, 0.5mg/ml Lysozyme.

Solution B:- 0.5M NaOH, 0.2% v/v Triton X-100

Solution C:- 0.5M NaOH

Solution D:- 1M Tris-HCl pH 7.6

Solution E:- 1.5M NaCl, 0.1M Tris-HCl pH 7.6

Solution F:- 0.3M NaCl.

All solutions can be prepared in advance and stored at room temperature, except Solution A. Solution A is prepared in advance minus the lysozyme, which is added to the appropriate concentration immediately prior to use. None of the solutions needed to be sterilised.

Figure 2.1 Flow Chart indicating the Protocol for Colony Lysis

Step 1	3 x 1 minute	solution A
Step 2	1 x 1 minute	solution B
Step 3	1 x 3 minutes	solution C
Step 4	2 x 1 minute	solution D
Step 5	1 x 5 minutes	solution E
Step 6	Filters are allowed to air dry	
Step 7	Immersion in solution F	
Step 8	Filters rolled between Whatman 54 chromatography paper	
Step 9	Filters are allowed to air dry	
Step 10	Filters baked in 80°C vacuum oven.	

The introcellulose filters were laid successively on top of 1ml of solutions A to E for varying lengths of time (steps 1 to 5), allowed air dry (step 6), and then immersed in solution F (step 7). The filters were then rolled between Whatman 54 chromatography paper (Ashless), using considerable weight, to remove bacterial debris (step 8), and were once again allowed to air dry (step 9). Steps 7 to 9 were then repeated as often as was necessary to remove all visible traces of bacterial debris before the plasmid DNA was immobilised on the filter by baking in a vacuum oven for two hours (step 10).

2.9 Subcloning of DNA into Plasmid and Viral Vectors

DNA fragments were subcloned into the appropriate vectors either to facilitate direct sequencing of the fragments or to provide material for the preparation of radiolabelled probes.

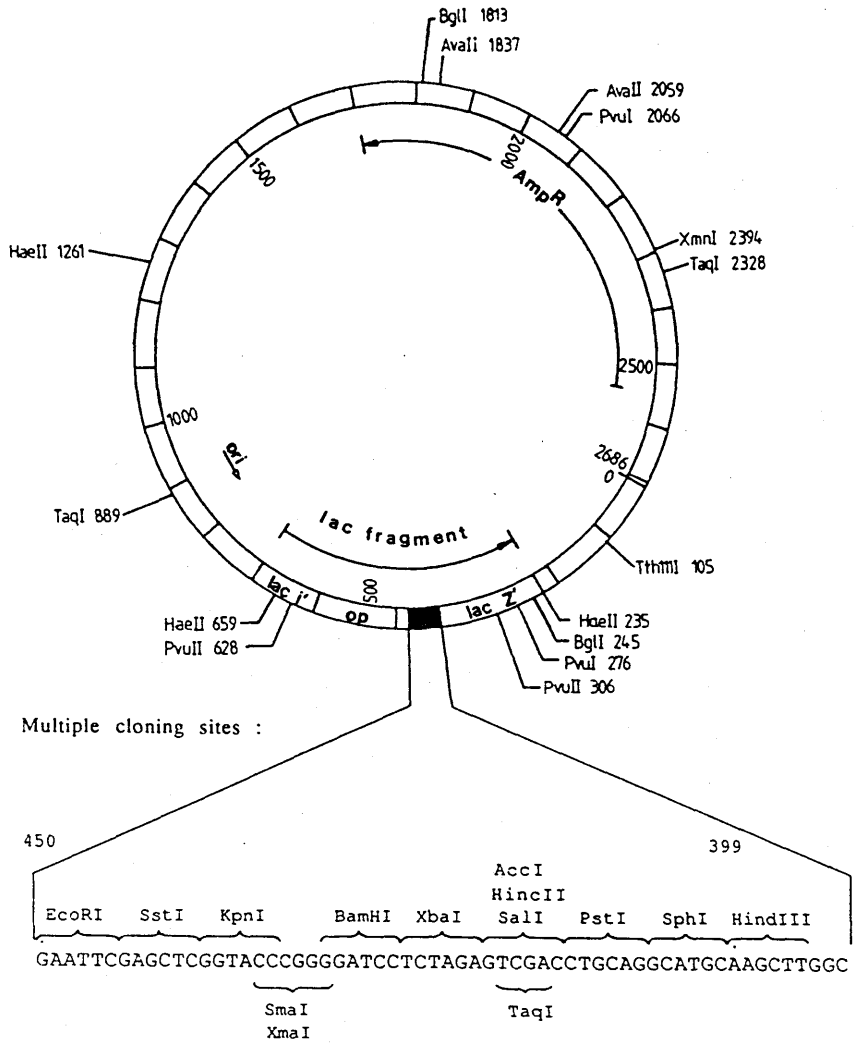
The plasmid vector pUC18 (Figure 2.2) and indeed the M13 vectors, use a simple chromogenic method for selecting recombinants which requires the host strain of bacteria, JM109, to be totally deficient in β-galactosidase activity. When the host is transformed with a non-recombinant plasmid, or M13 phage particle, β-galactosidase activity is restored and can be detected by the blue colour of colonies or plaques.

β-galactosidase activity is restored by non recombinant pUC18 because of what is known as α-complementation. The host strain, JM109, carries the F-episome which contains a mutant β-galactosidase gene (lac Z MI5) which is capable of complementation with the N-terminal fragment of the β-galactosidase gene carried by pUC18. The restored galactosidase activity of the transformed cell is used to convert the colourless substrate x-gal (5-bromo-4- chloro-3-indolyl-β-D-galactoside) into a blue chromophor which subsequently gives a blue colour to bacterial colonies.

The insertion of a piece of foreign DNA into the 54bp multiple cloning site of pUC18 prevents α-complementation and therefore β-galactosidase activity is not restored by such recombinant plasmids. Thus when grown in the presence of X-gal, no blue colour

is formed, and the resulting colonies have a white appearance. In order to distinguish these white colonies from colonies resulting from untransformed host, the transformation mixture is grown on LB plates containing Ampicillin, thus making use of the Ampicillin resistance gene in pUC18. Therefore, when the transformation mixture is gown on LB plates plus Ampicillin only colonies transformed with pUC18 grow, and in the presence of X-gal non-recombinant plasmids give rise to blue colonies whereas recombinant colonies are white.

Figure 2.2 Plasmid Vector pUC18



pUC18 is a small (2826 bp) E. Coli plasmid, constructed by Messing and coworkers, out of pBR322 and M13 mp18. It carries a 54bp multiple cloning region which forms the basis of the subcloning strategy and also allows easy detection of recombinants. The above figure illustrates this 54bp region with the 13 unique restriction sites. It also shows the location of the lac gene fragment, the Ampicillin resistance gene and the origin of replication.

2.9.1 Subcloning into Plasmid Vectors pU18 and pU19

The multiple cloning site of pUC18 contains sites for thirteen hexanucleotide-specific restriction enzymes and it is into this region that foreign DNA is inserted.

The vector DNA, pUC18, (5-10 μg) and the cloned foreign DNA (2 μg) were separately digested with the appropriate enzyme(s), and agarose gel electrophoresis performed to check for completeness of digestion or isolation of particular fragments, as appropriate. Both vector and insert DNA were purified by extraction with phenol and then precipitated with ethanol (Section 2.3.8) After lyophilisation the insert DNA was redissolved in T.E. buffer at 0.2 mg/ml, while the vector DNA was kept as a lyophilised pellet.

Alkaline Phosphatase Treatment of Vector DNA

Transformation of JM109 with a ligation mixture in which there are undigested plasmid molecules would result in a high background of blue colonies. Similarly if digested vector DNA is ligated, even in the presence of insert DNA, without first of all treating with alkaline phosphatase, then the greater efficiency of intramolecular ligation will lead to a large number of religated native pUC18 molecules, thus also giving a high background of blue colonies upon transformation. Therefore it is important to check for completion of digestion and also to treat the vector DNA with alkaline phosphatase to prevent self religation, by removing the 5' phosphate groups at the ends of the DNA molecule. However, the need for

alkaline phosphatase treatment was reduced when the vector DNA was cut with two different enzymes and was not be performed at all when subsequent blunt end (into the Sma I site) ligation was attempted.

The lyophilised vector DNA was redissolved in 20 µl 50mM Tris-HCl pH9.5, 0.1mM EDTA, 1mM spermidine, to which 1 unit of Calf Intestinal Alkaline Phosphatase (Boehringer grade 1) was added. The mixture was incubated at 37°C for 30 minutes after which the volume was increased to 100 µl with TE and the DNA purified by phenol extraction and ethanol precipitation. The vector DNA was finally redissolved in TE at a concentration of 0.2mg/ml.

Ligation of Insert and Vector DNA

The ratio of insert DNA to vector DNA was carefully assembled so as to avoid either too little or too much insert DNA being incorporated thus leading respectively to low efficiency of incorporation or to the formation of insert polymers. Consequently ratios of 1:1 and 3:1, vector:insert, were generally prepared. The ratio is dependent on the number of restriction sites present in the foreign DNA for the particular enzyme(s) used.

The ligation mixture was assembled on ice by combining the appropriate amount of insert DNA, vector DNA (usually 0.5 µg), 4 µl 5mM ATP, 1 µl 100mM DTT, 3 µl 10 x Ligase Buffer (400mM Tris-HCl pH 7.6, 100mM $MgCl_2$), 1 unit T4 DNA Ligase (BRL) and increasing the volume to a total of 30µl with sterile distilled water. Incubation was at 15°C overnight after which the ligation mixture

was stored at -20°C ready for transformation.

Transformation of Competent JM109 Cells

Frozen competent JM109 cells (Section 2.9.3) were allowed to thaw on ice and then aliquoted into 100μl volumes in Eppendorff tubes. 10 μl of the ligation mixtures were then added to separate 100 μl aliquots of competent cells and these mixtures incubated on ice for 30 minutes. During this incubation LB plates containing ampicillin were spread with 120μl 2% X-gal: 2% IPTG: H_2O (1:1:2). IPTG (isopropylthio-β-galactoside) is a gratuitous inducer of the lac operon and thus induces β- galactosidase expression by the plasmid and the F-episome. It is stored at +4°C. 2% X-gal in dimethyformamide is stored at -20°C. This mixture should be allowed to absorb onto the plates before the transformation mixtures are spread on them.

After the 30 minute incubation the transformation mixtures were taken and incubated at 42°C for 2 minutes and then transferred back to ice until they could be spread on the Amp plates.

Once the transformation mixtures had been spread the plates were inverted and incubated at 37°C overnight.

Any resulting white colonies were picked and plasmid DNA isolated by the small scale isolation method (Section 2.4.1.2). The presence of the desired insert was then checked by restriction with the appropriate enzyme(s) and subsequent agarose gel electrophoresis.

2.9.2 Subcloning into Bacteriophage M13 Vectors

DNA was subcloned into M13mp18 and M13mp19 routinely to provide single stranded DNA for sequencing by the method of Sanger.

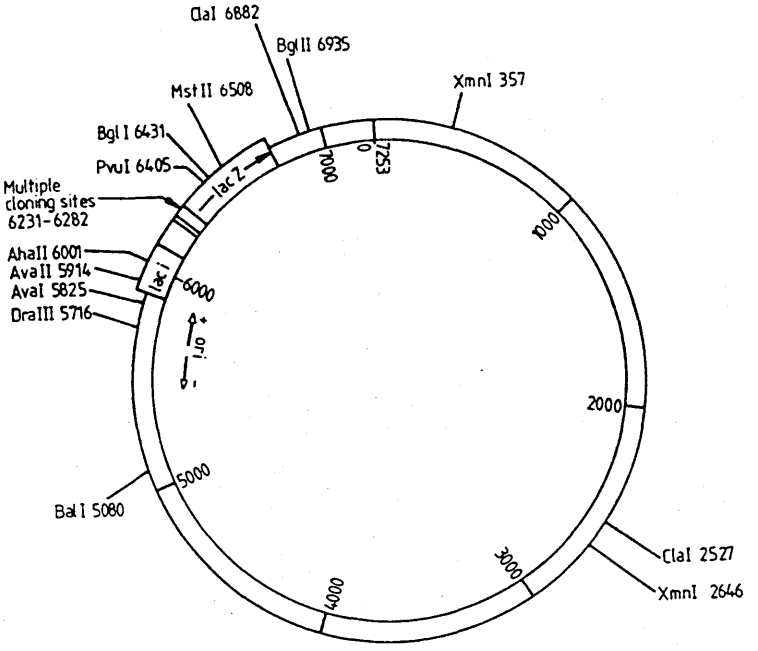
Subcloning of Foreign DNA into M13mp Vectors

The general features of subcloning into pUC18 are also applicable to subcloning into the M13mp vectors. The same 54 bp multiple cloning region is present and recombinants are also detected by growth on X-gal and IPTG. However in M13 it is "plaques" that are formed instead of colonies of bacteria. Recombinant "plaques" are clear whilst non-recombinant "plaques" are blue. In fact, however, since M13 is not a lysogenic bacteriophage it is not true phage plaques that are formed but instead areas of restricted bacterial growth with the appearance of plaques.

Ligation of Insert and Vector DNA

The insert and vector DNA were preparated as for pUC18 (Section 2.9.1) with the vector DNA being treated with phosphatase as appropriate. The ligation protocol was also identical but for the fact that smaller quantities (20-100ng) of DNA were used.

Figure 2.3 Bacteriophage M13mp18



Multiple cloning sites in M13 mp18:

6230 6289

AccI
HincII
EcoRI SstI KpnI BamHI XbaI SalI PstI SphI HindIII

GAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGC

SmaI
XmaI

TaqI

M13mp18 (Yanish-Perron et al 1985) is a filamentious bacteriophage of E. Coli. The phage particle contains a single stranded DNA molecule. However, upon infection, this strand of DNA serves as a template for the synthesis of the complementary strand thus creating a double-stranded replicative form of the viral DNA which is subsequently amplified to about 200 copies per cell. It is this RF DNA which is used for subcloning since it can be manipulated as if it were a large plasmid (7250 bp).

M13mp18 and M13mp19 contain the same 54 bp multiple cloning region as pUC18 and pUC19 respectively, and only differ from each other in the orientation of this region thus making sequencing possible from either direction. The above figure shows that 54 bp region of M13mp18, the position where the 17bp universal primer anneals and the direction of sequence initiation. It also shows the position of the lac gene fragment and the origins of plus and minus strand replication.

Transformation of Competent JM109 Cells

The transformation procedure was identical to that for pUC18 however the plating procedure was different.

H-plates (Section 2.2.3) were used and were preincubated at 37°C for about two hours. Meanwhile H-top agar was melted, aliquoted into 2.5ml volumes, and kept liquified at 42°C. 10 µl 10% X-gal, 10µl 2% IPTG and 100µl of a saturated suspension of fresh JM109 cells were added to the H-top agar immediately prior to plating. The fresh JM109 cells were required to produce a lawn of bacteria on the H-plates to enable easy identification of plaques.

After raising the temperature of the transformation mixture to 42°C for two mintues the 100 µl of transformed cells were added to the 2.5ml H- top agar (with additions) mixed gently, and then poured onto an H- plate. The top agar was allowed to solidify and plates incubated overnight at 37°C. Clear plaques were picked and single-stranded DNA prepared by the method outlined in Section 2.4.5.

2.9.3. Preparation of Competent JM109 Cells (Cohen et al 1972)

A single colony of E. Coli JM109, stored on minimal agar, was taken, inoculated into 25ml L-Broth, and grown to saturation at 37° overnight on an orbital incubator. 2.5ml of this overnight culture was then used to inoculate 500ml of L-Broth in a 2 litre flask and the cells grown until they reached a turbility of O.D.

= 0.2. A convenient quantity of cells was produced from four such 500ml cultures.

The cells were harvested in a suitable large capacity centrifuge by centrifugation at 5000 rpm (4000g) for 15 minutes after which the cells were carefully resuspended in 250ml cold sterile 100mM $CaCl_2$ (i.e. half volume of 500ml culture). After incubating the cells on ice for 20 minutes they were resedimented at 5000 rpm (4000g) and 4°C for 15 minutes. All subsequent manipulations were performed in a cold room (temperature 4°C). The cells were resuspended in 5ml ice cold 100mM $CaCl_2$ and sterile glycerol added to 10%. They were then aliquoted into suitable volumes (1ml and 0.5ml) snap frozen in liquid nitrogen, and then stored at -70°C until required.

2.10 Sequencing of DNA

2.10.1 Chemical Cleavage Method of Maxam and Gilbert (1977,1980)

DNA sequencing by the method of Maxam and Gilbert requires a double stranded fragment of DNA labelled at one end. The labelling reaction was performed using the Klenow fragment of E. Coli DNA Polymerase 1 (Section 2.6.1) which requires an initial restriction enzyme cut to be made leaving a linear piece of DNA with 5' protruding ends. All 5' protruding ends are labelled by the Klenow "filling in" reaction and therefore a second restriction enzyme cut must be made. The second restriction enzyme is chosen so that the labelled DNA is cleaved asymmetrically, preferably into two

fragments, which can subsequently be separated by polyacrylamide gel electrophoresis (Section 2.3.6.3). The appropriate band was excised from the gel, the DNA eluted from the polyacrylamide and temporarily stored as a lyophilised precipitate.

Chemical Modification and Strand Scission

Sequencing by this method is dependent on base-specific modification and cleavage of the DNA. Table 2.6 lists the solutions required.

The lyophilised precipitate of labelled DNA was dissolved in 11 µl water with 4 µg (4mg/ml) calf thymus DNA being added as carrier. The mixture was then aliquoted into four equal parts in siliconised Eppendorff tubes, in preparation for the four sequencing reactions, the procedure of which is outlined in Table 2.7.

Table 2.6 The solutions required for sequencing by the chemical cleavage method of Maxam and Gilbert

Dimethylsulphate	- DMS (Aldrich Chemical Co)
Hydrazine	- HZ (Kodak) stored at -70°C
Piperidine	- (Koch-Light) stored at 4°C
Pyridine Formate	- 4% v/v formic acid adjusted to pH2.0 with pyridine (BDH)
DMS Buffer	- 50mM sodium cacodylate; 10mM $MgCl_2$; 0.1mM EDTA adjusted to pH8.0 and stored at 4°C
DMS Stop	- 1.5M sodium acetate pH7.0, 1M β-mercaptoethanol (Koch-Light), 200 g/ml yeast tRNA. Stored at -20°C.
HZ Stop	- 0.3M sodium acetate, 0.1mM EDTA 50 μg/ml yeast tRNA. Stored at 4°C
NaCl	- As 5M solution in H_2O

TABLE 2.7 Protocol for base specific reactions for Maxam and Gilbert sequencing

	G	G + A	T + C	C
Mix on ice	200 µl DMS buffer 4µl ^{32}p DNA	10 µl H_2O 4µl ^{32}p DNA	10µl H_2O 4µl ^{32}p DNA	15 µl 5M NcCl 4µl ^{32}p DNA
Add	1µl DMS	50µl Pyridine Formate	30µl HZ	30µl NZ
Incubate at 20°C	3 mins	5 mins	5 mins	5 mins
Add	50µl DMS stop 750µl ethanol	180µl HZ stop 750µl ethanol	200µl HZ stop 750µl ethanol	200µl HZ stop 750µl ethanol
Store	-70°C, 10 min	-70°C, 10 min	-70°C, 10 min	-70°C, 10 min
Centrifuge	10 mins	10 mins	10 mins	10 mins
Add to Pellet	250µl 0.3M sodium acetate 750µl ethanol	250µl 0.3M sodium acetate 750µl ethanol	250µl 0.3M sodium acetate 750µl ethanol	250µl 0.3M sodium acetate 750µl ethanol
Store	-70°C, 10 min	-70°C, 10 min	-70°c, 10 min	-70°C, 10 min
Centrifuge	10 mins	10 mins	10 mins	10 mins
Rinse Pellet	70% ethanol lyophilise	70% ethanol lyophilise	70% ethanol lyophilise	70% ethanol lyophilise
Add to Pellet	100µl 1M Piperidine	100µl 1M Piperidine	100µl 1M Piperidine	100µl 1M Piperidine
Heat to	90°C, 30 min lyophilise	90°C, 30 min lyophilise	90°C, 30 min lyophilise	90°C, 30 min lyophilise
Add	10µl H_2O lyophilise	10µl H_2O lyophilise	10µl H_2O lyophilise	10µl H_2O lyophilise
Add	10µl H_2O lyophilise	10µl H_2O lyophilise	10µl H_2O lyophilise	10µl H_2O lyophilise

All reactions were stored as a lyophilised pellet until they were ready to be loaded on to the gel.

Once strand scission was complete the four sequencing reactions were electrophoresed through a denaturing polyacrylamide gel with the DNA fragments being separated according to molecular weight.

A typical gel (40cm x 20cm x 0.4mm) contained 6% acrylamide (from 19% acrylamide, 1% bis-acrylamide stock), 7M Urea and 0.01% ammonium persulphate in a total volume of 100ml 1 x TBE buffer. The gel plates were siliconised before use with Repelcote (2% Dimethyldichlorosilane in 1,1,1-trichloroethane; BDH) and assembled with 0.4mm spacers. 40 µl TEMED (NNN'N' - tetramethylethylenediamine) was added to the gel mixture immediately prior to pouring. Once poured a comb (14 teeth; 7mm x 0.4mm) was inserted and the gel allowed to set.

Although DNA samples were denatured before loading, and the gel itself is denaturing, the temperature of the gel was also kept high during electrophoresis to help prevent secondary structure forming. To this end the gel was pre-electrophoresed at 25mA for 1 to 2 hours during which the four sequencing reactions were dissolved in 5 µl of loading buffer (99% deionised formamide, 0.05% xylene cyanol).

Immediately prior to loading, the DNA samples were denatured by boiling for 2 minutes and then rapidly chilled on ice. 1.5 µl of each of the four samples was then loaded into separate adjacent wells and electrophoresis performed at 25 mA until the dye front had travelled 20cm. A further 1.0µl of each sample was then loaded into

the next set of wells and electrophoresis continued until the dye front of these samples had travelled 15cm. The remainder of each sample was then loaded into the final set of wells and electrophoresis continued for a further 13cm after which the gel was ready for autoradiography. The DNA samples were also denatured by boiling immediately prior to the 2nd and 3rd loadings.

After electrophoresis, one gel plate was carefully removed, the gel covered with clingfilm, and autoradiography performed at -70°C overnight. Autoradiography was performed using Kodak X-Omat H film and an intensifying screen (Cronex Lighting Plus, Dupont).

The above procedure, with the three separate loadings, was sufficient for reading between 150 to 200 bases per gel.

2.10.2 Dideoxynucleotide Method of Sanger (Sanger 1981)

DNA sequencing by this method requires a single-stranded DNA template for synthesis of the complementary stand. To achieve this, DNA is cloned into the bacteriophage, M13 (Section 2.9.2), and single-stranded DNA prepared from recombinants (Section 2.4.5). The complementary strand is synthesised by the Klenow fragment of E. Coli DNA polymerase 1.

Strand-Elongation

The complementary strand of DNA is synthesised by the Klenow enzyme using the single-stranded recombinant DNA as template and a short region of double stranded DNA, 3' to the multiple cloning site (Figure 2.3), as primer to direct synthesis in the 5' to 3' direction. Dideoxy sequence analysis is dependent upon the random incorportion of dideoxribonucleotides into the growing DNA strand thus causing chain termination. To determine the DNA sequence of a template, four separate elongation reactions are performed, each containing only one of the ddNTPs, ^{35}S dATP and the three remaining dNTPs. The ratio of dNTP:ddNTP in each raction is controlled so that chain termination occurs randomly over the complete sequence. Table 2.8 outlines the dNTP:ddNTP content of each of the four base specific reactions.

Table 2.8 The dNTP/ddNTP content of the four base specific working solutions used in Sanger dideoxy sequencing.

	A	G	C	T
10 x Klenow buffer	20µl	20µl	20µl	20µl
0.5mM dATP	-	-	-	-
0.5mM dGTP	20µl		1µl	20µl
0.5mM dCTP	20µl	20µl	1µl	20µl
0.5mM dTTP	20µl	20µl	20µl	1µl
0.1mM ddATP	80µl	-	-	-
0.1mM ddCTP	-	-	61µl	-
0.3mM ddGTP	-	61µl	-	-
0.5mM ddTTP	-	-	-	61µl

The working solutions were stored at -20°C.2 1 of each working solution was used for each base specific sequencing reaction. The above table represents working solutions when the label being used was α-^{35}S dATP. If another α-^{35}S labelled dNTP was to be used the working solution were altered appropriately.

Annealing of Primer and Sequence Reaction

Before performing the sequencing reactions, the M13 17-base primer (BRL), which is complementary to a region of DNA immediately 3' to the multiple cloning site of M13, has to be annealed to produce the double-stranded primer region. This was achieved by assembling 7.5 μl of the single-stranded template (Section 2.4.5) 1.5 μl 10 x reaction buffer (100mM Tris HCl pH8.0, 100mM $MgCl_2$, 300mM NaCl) and 1.0 μl of the M13 universal 17-base primer. The mixture was incubated at 56°C for two hours and then allowed to cool to room temperature in preparation for the sequencing reaction. Once cooled, 1.2 μl of the solution was aliquoted into each of four reaction tubes. 2 μl of the appropriate dNTP;ddNTP mixture was then added to each tube and the reaction started by centrifugation of the two solutions to the bottom of the reaction tube. Incubation was at room temperature for 20 minutes after which 2 μl of chase mix (all four dNTPs at 0.5 mM) was added, and incubation continued, at room temperature, for a further 15 minutes.

After incubation the reaction mixes were ready to be separated on a sequencing gel, but could also be stored at -20°C for up to two weeks beforehand. Also only half of the reaction mix was required for visualisation by autoradiography and thus the other half could also be stored for performing a second longer sequencing gel. If the separation of the samples was to be performed immediately, 2 μl loading dye (0.03% xylene cyanol; 0.03% bromophenol blue; 20mM EDTA in formamide) were added to 3 μl of each sample to stop the reaction. The DNA was denatured at 100°C for three

minutes and then loaded onto the sequencing gel.

DNA Sequencing Gels

Denaturing polyacrylamide gels were prepared as outlined in Section 2.10.1 but using a 32 tooth (2.5mm x 0.4mm) comb to prepare the wells. DNA samples were denatured by boiling for three minutes before loading and electrophoresis performed at 25mA until the bromophenol blue (Sigma) was 0.5cm above the bottom of the gel (approximately two hours). Such electrophoresis was sufficient for resolving 180-200 bases, however further resolution could be gained by electrophoresing a second loading of the samples for five and a half hours.

2.11 Deletion Subcloning (Labeit et al 1986)

Sequence determination by either the method of Maxam and Gilbert or that of Sanger, requires that there be appropriate restriction sites, either for direct labelling of the DNA, or for subcloning into the multiple cloning region of M13. Occasionaly, however, some stretches of DNA are inaccessible to sequence determination because they lack suitable restriction enzyme sites. The following procedure can be used to generate a library of M13 derived recombinants containing serially random deletions of insert DNA, thus enabling sequence determination of regions lacking suitable restriction sites.

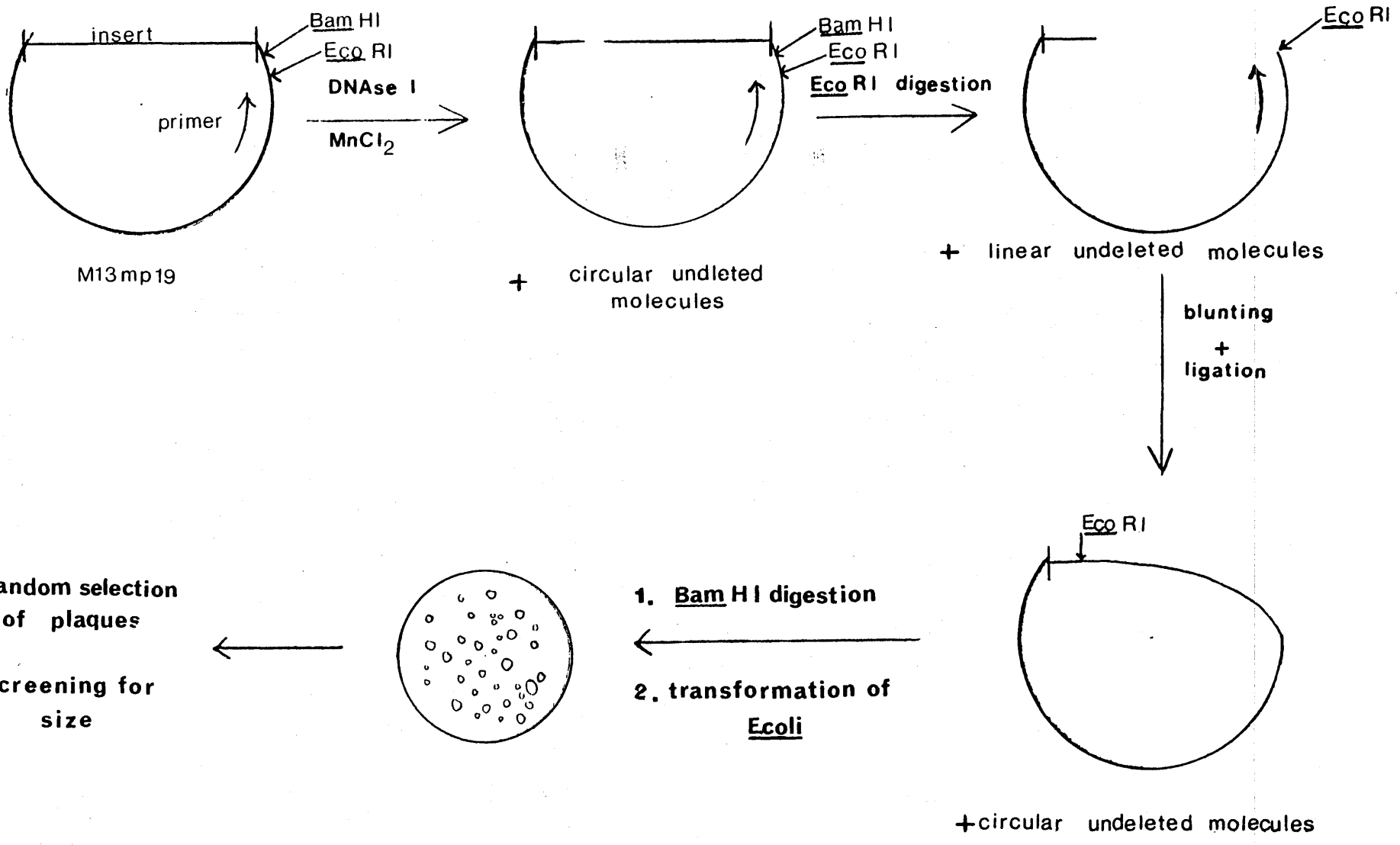
The fragment of DNA to be sequenced was subcloned into M13 in

the normal manner (Section 2.9.2), but RF DNA was isolated by the procedure outlined in Section 2.4.6. Figure 2.4 outlines the principles involved in generating the deletion library. Briefly, the RF DNA of the recombinant M13 was subjected to limited digestion by DNase I, in the presence of Mn^{2+} ions, to ensure that only a small fraction of the DNA molecules were linearised. This resulted in breaks which were nearly random in their distribution. Subsequent digestion with the unique restriction enzyme A (i.e. enzyme A has only one restriction site in the original M13 recombinant, occuring in the multiple cloning region between the primer region and the inserted foreign DNA), introduced deletions extending from the priming site into the insert DNA. The resulting staggered ends were "filled in" using Klenow enzyme, and a blunt end ligation is performed to recircularise the deleted molecules.

Digestion with a second unique restriction enzyme B, the site for which was situated between that for enzyme A and the inserted foreign DNA in the original recombinant, should linearise any non-deleted DNA molecules, since only deleted molecules will have lost restriction site B. Since transformation of linear DNA is inefficient, subsequent transformation of competent JM109 cells will produce plaques representing deleted molecules which can be selected for the preparation of single-stranded template DNA. The single-stranded DNA is then run on an agarose gel and size selected recombinants chosen for subsequent sequencing.

Figure 2.4 Generation of random deletions in M13 replicative form DNA

The 2.5kb Pst I fragment of clone pCM212 was cloned into M13mp19 and replicative form DNA isolated. This was digested for 2 minutes with DNAse I in the presence Mn^{2+} ions (Figure 2.5a) and then with the restriction endonuclease Eco RI. The ends of the DNA were made flush using the Klenow reaction outlined in section 2.6.1 and then ligated. Undeleted molecules were then linearised with the restriction endonuclease Bam HI and the whole reaction mixture used to transform competent JM109 cells. Since intact circular DNA molecules transform at a higher efficiency than linear molecules the resulting plaques represent deletion subclones of the original M13mp19 clone containing the 2.5kb Pst I fragment. Single-stranded DNA was prepared from randomly selected plaques and screened for size (Figure 2.5b) before sequencing by the method of Sanger (1977).


insert
Bam HI
Eco RI
primer
DNAse I
MnCl2
M13mp19
+ circular undleted molecules
Bam HI
Eco RI
Eco RI digestion
+ linear undeleted molecules
Eco RI
blunting
+
ligation
Eco RI
+circular undeleted molecules
1. Bam HI digestion
2. transformation of Ecoli
1. random selection of plaques
2. screening for size

DNase I Digestion

A 10^6-fold dilution of a 5mg/ml DNAse I stock was prepared in 2mM $MnCl_2$/80mM Tris-HCl pH 7.4. 10 µg of M13 RF DNA was diluted in 50µl H_2O and then added to 50µl of the DNAse I solution which was then incubated at room temperature. 20µl aliquots were withdrawn at two minute time intervals and the reaction in the aliquots immediately halted, by the addition of 2 µl 0.1M EDTA.

The extent of DNAse I digestion in each of the aliquots were visualised by subjecting 3µl of each to electrophoresis in a 0.7% agarose gel. Meanwhile the aliquots were stored frozen. A typical result is shown in Figure 2.5a.

The appropriate aliquot which showed digestion to approximately a fifth to a third of the circular form of the RF DNA (lane 2 Fig 2.5a) was chosen and the DNAse I extracted with phenol. After precipitation with ethanol, the DNA was then digested with the appropriate enzyme, blunt ended and religated, and then redigested, with the second unique enzyme. Figure 2.5b shows a typical size distribution of recombinants obtained from the subsequent transformation of JM109 with the deleted DNA molecules.

Figure 2.5 Digestion of M13 Replicative Form DNA with DNAse I in the presence of Mn^{2+} ions and the size selection of the resultant deletion subclones

The 2.5kb Pst I fragment of clone pCM212 was cloned into M13mp19 and replicative form (RF) DNA isolated. (a) The RF DNA was digested with DNAse I in the presence of Mn^{2+} ions and the reaction stopped by the addition of EDTA to 10mM at 2, 4, 6, 8 and 10 minutes. A sample from each reaction mixtures was electrophoresed through a 0.7% agarose gel. Lanes 1-6 contain, from left to right, the samples taken after 10 minutes, 8 minutes, 6 minutes, 4 minutes, 2 minutes and a zero time control (undigested DNA). The sample which was digested for 2 minutes was selected for performing the subsequent steps to produce the deletion library (Figure 2.4). (b) The single-stranded M13 DNA isolated from the deletion subclones was electrophoresed in a 0.7% agarose gel, overnight at 12 mA. A large difference in the sizes of molecules generated by the DNAse I delection method can be seen. Appropriate clones, spanning all sizes of molecules were chosen and sequenced by the method of Sanger (1977).

minutes
minutes
minutes
minutes
minutes

(a)

(b)

CHAPTER 3

Results

3.1 Differential Screening of a Mouse cDNA Library

The initial aim of this work was to isolate complementary DNA (cDNA) clones corresponding to proteins expressed solely in skeletal muscle. The approach adopted was to screen a mouse skeletal muscle cDNA library with two radiolabelled single-stranded cDNA probes, one derived from skeletal muscle mRNA, the other derived from what may be regarded as the most closely related tissue, cardiac muscle.

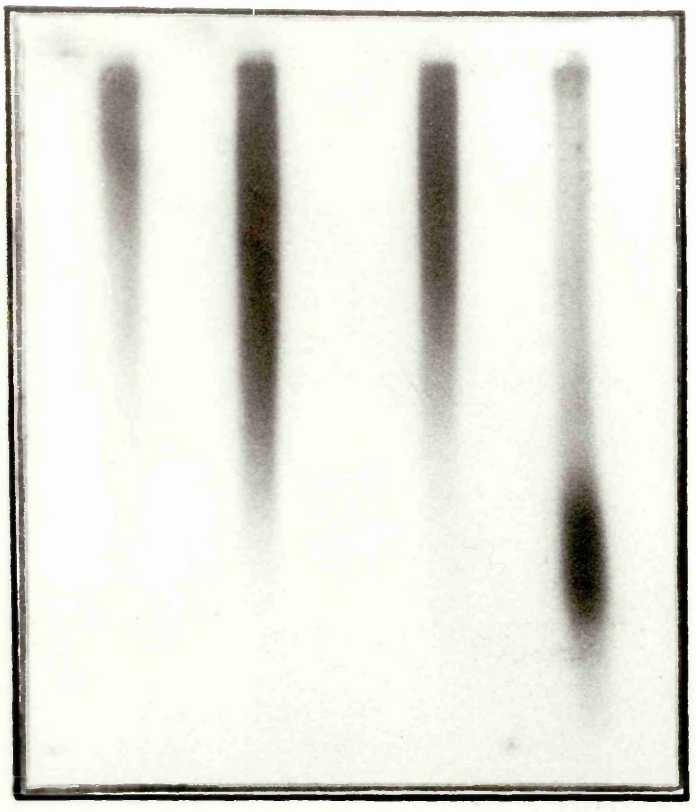
First, experiments were undertaken to optimise the conditions for producing the ^{32}P-labelled single-stranded cDNA probes. It was found that the lengths and specific radioactivity of the probes were dependent on the chemical concentration of the radioactively-labelled dNTP used in their synthesis. Thus, when a fixed radiochemical concentration of (^{32}P) dATP was used to label the cDNAs, increasing the chemical concentration of dATP in the reaction mixture resulted in the synthesis of cDNAs of greater length (Figure 3.1), but, of course, also decreased their specific radioactivity. As the cDNAs were to be used as probes, it was necessary to make the best compromise between length and specific radioactivity. A concentration of 50μM dATP was adopted on the basis of the results in Figure 3.1.

The mouse skeletal muscle cDNA library was a gift from Dr D.P.

Figure 3.1 The variation in length and specific readioactivity of single-stranded cDNA as a function of the chemical concentration of the radioactively-labelled dNTP.

Single-stranded cDNA was prepared from 1 μg poly(A)$^+$ from skeletal muscle using various chemical concentrations of dATP, but using a fixed concentration of α-(^{32}P) dATP. Aliquots from each reaction mixture, containing approximately 20 x 10^4cpm, were taken and electrophoresed in a denaturing 1% agarose gel. The gel was dried and autoradiographed for approximately four hours. Each track of the dried gel was then sliced into 1cm portions (from top to bottom) and the radioactivity present in each portion was measured as Cherenkhov radiation in a scintillation spectrometer. (a) shows the autoradiograph of the initial gel. The concentrations of dATP were (4) 20nM, (2) 15μM, (3) 50μM and (1) 150 μM. (b) The graph shows for different concentrations of dATP the yield of cDNA in terms of the percentage of the ^{32}P-label present in each gel slice in relation to the length of DNA in that slice. (c) The table summarises the yield, length and specific radioactivity of the cDNA obtained using the different chemical concentrations of dATP.

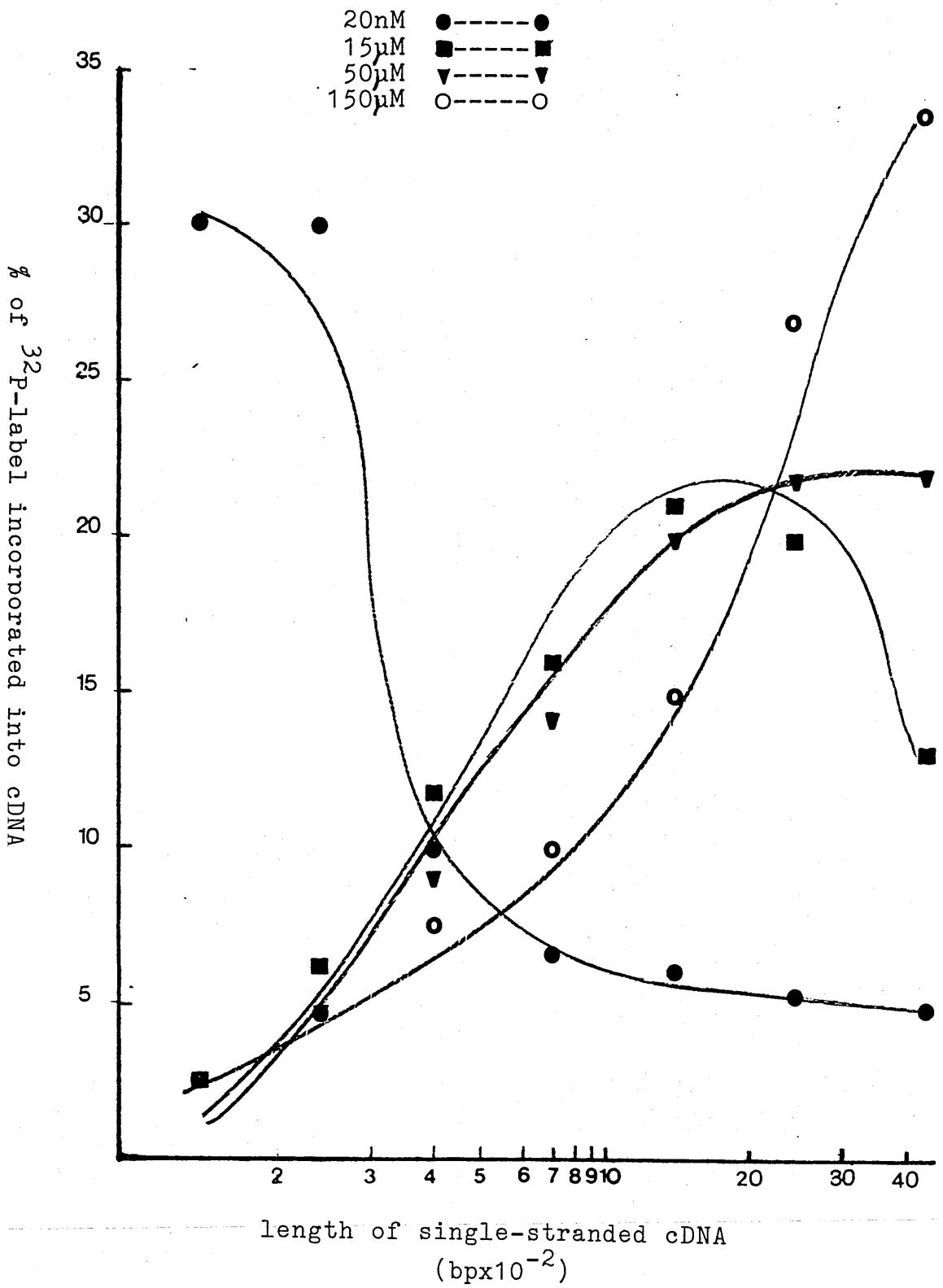
(a)



1 2 3 4

(c)

dATP	peak length of cDNA	specific radioactivity	yield of cDNA
20nM	300 bases	2.6×10^{10} cpm/μg of cDNA	5.7×10^{-5} μg
15μM	1600 bases	3.6×10^{7} cpm/μg of cDNA	1.1×10^{-2} μg
50μM	2-3000 bases	1.1×10^{7} cpm/μg of cDNA	1.8×10^{-2} μg
150μM	3000 bases	3.6×10^{6} cpm/μg of cDNA	3.9×10^{-2} μg


20nM
15μM
50μM
150μM
35
30
25
20
15
10
5
% of ^{32}P-label incorporated into cDNA
2
3
4
5
6
7
8
9
10
20
30
40
length of single-stranded cDNA
(bpx10^{-2})

Leader, University of Glasgow, who had constructed it in the plasmid vector pBR322, using cDNA reverse transcribed from poly $(A)^+$ RNA isolated from eleven day old mice (Leader et al, 1986). The library was plated as described (section 2.8.1) with two nitrocellulose filter replicas being taken of each master plate. These were then hybridised to ^{32}P-labelled cDNA probes derived from skeletal and cardiac muscle, washed at medium stringency and then subjected to autoradiography.

Figure 3.2 presents the results of hybridisation to two such replica filters, from which several points can be made. First, only approximately 20% of the bacterial colonies actually gave a hybridisation signal on the autoradiograph of the filter probed with the skeletal muscle cDNA (Figure 3.2(a)). Those colonies not giving rise to a hybridisation signal presumably contain cloned cDNAs corresponding to low abundance mRNAs.

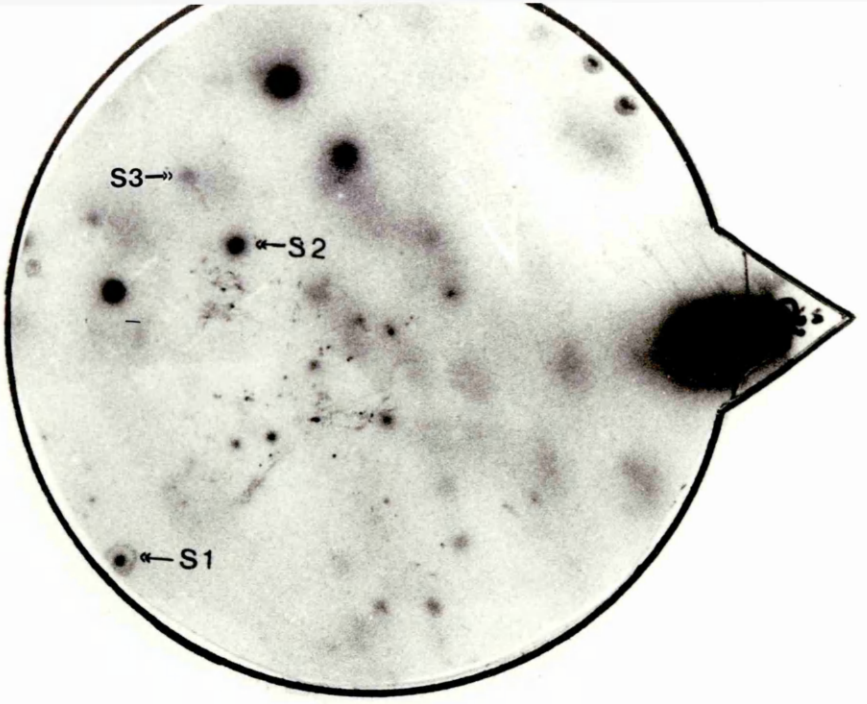
A second point to note is that there were clones in the skeletal muscle cDNA library which gave a hybridisation signal with the cardiac muscle probe, but not with the skeletal muscle probe (Figure 3.2(b)). Examples of these are clones C1 and C2. Such colonies presumably contain cloned cDNAs derived from mRNAs which are more abundant in cardiac muscle than in skeletal muscle.

Figure 3.2 shows that there were also colonies of the type sought; those which gave a hybridisation signal with the skeletal muscle probe, but not with the cardiac muscle probe. Examples of these are clones S1, S2 and S3. 247 colonies were selected for

Figure 3.2 Examples of the results of the initial screening of the mouse skeletal muscle cDNA library.

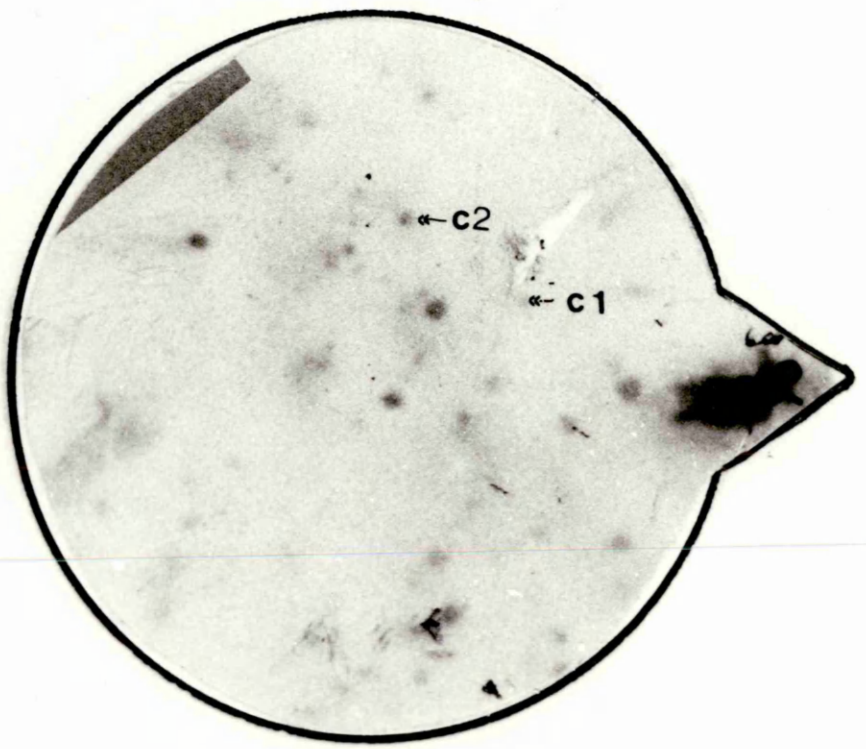
The mouse skeletal muscle cDNA library was screened with ^{32}P single-stranded cDNA reverse-transcribed from (a) skeletal muscle poly $(A)^+$ RNA and (b) cardiac muscle poly $(A)^+$ RNA. The filters were hydridised overnight, washed five times for ten mintues in 2 x SSC at room temperature and then twice for 30 minutes in 0.1 x SSC and 42°C. S1, S2 and S3 are examples of colonies giving hybridisation signals only with the skeletal muscle probe. C1 and C2 are colonies which give hybridisation signals only with the cardiac muscle probe.

(a)



^{32}P-labelled skeletal
muscle cDNA probe

(b)



^{32}P-labelled cardiac
muscle cDNA probe

further investigation, out of the approximately 1.5×10^4 colonies screened.

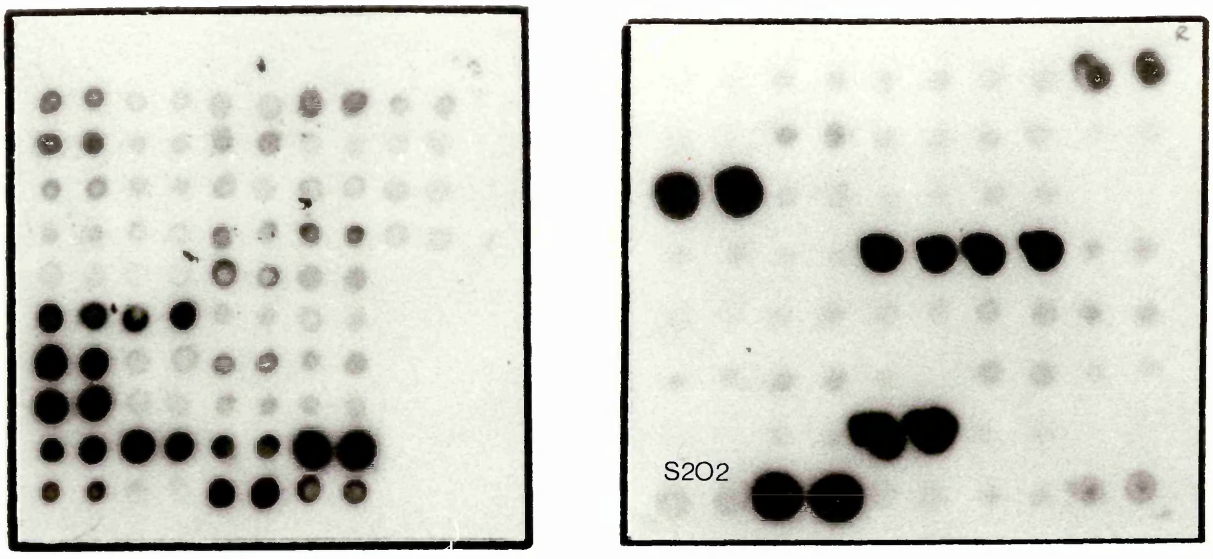
The initial screening of the cDNA library by the replica plating and colony hybridisation procedure had several inherent technical difficulties associated with it. There was a considerable chance of unequal transfer of colonies to each of the replica filters, and also of contamination of the filters with cell debris, both of which could have distorted the hybridisation signal. This in turn may have resulted in the selection of colonies regarded as being specific to skeletal muscle, but which were not in fact so. As the first screening had reduced the number of clones to more manageable proportions, it was possible to circumvent these problems by small scale preparation of plasmid DNA from each clone and application of approximately equal amounts of the DNA directly onto duplicate nitrocellulose filters (dot-blotting) for subsequent hybridisation to the two ^{32}P-labelled single-stranded cDNA probes.

Figure 3.3 presents the autoradiographs of two such duplicate filters of dot blotted DNA. It is apparent, from Figure 3.3, that the majority of the clones selected from the initial screening as being specific to skeletal muscle, were not in fact so, since in the rescreen they gave a hybridisation signal with the cardiac muscle probe. Indeed there were even a number (clones C156 and C152 in Figure 3.3) which appeared to specify proteins more abundant in cardiac muscle than in skeletal muscle. Nevertheless, there were several clones which did appear to represent mRNAs more abundant in skeletal muscle than in cardiac muscle, eight of which gave a

Figure 3.3 Examples of the results of the secondary screening of the 247 clones selected from the initial screening of the mouse skeletal muscle cDNA library.

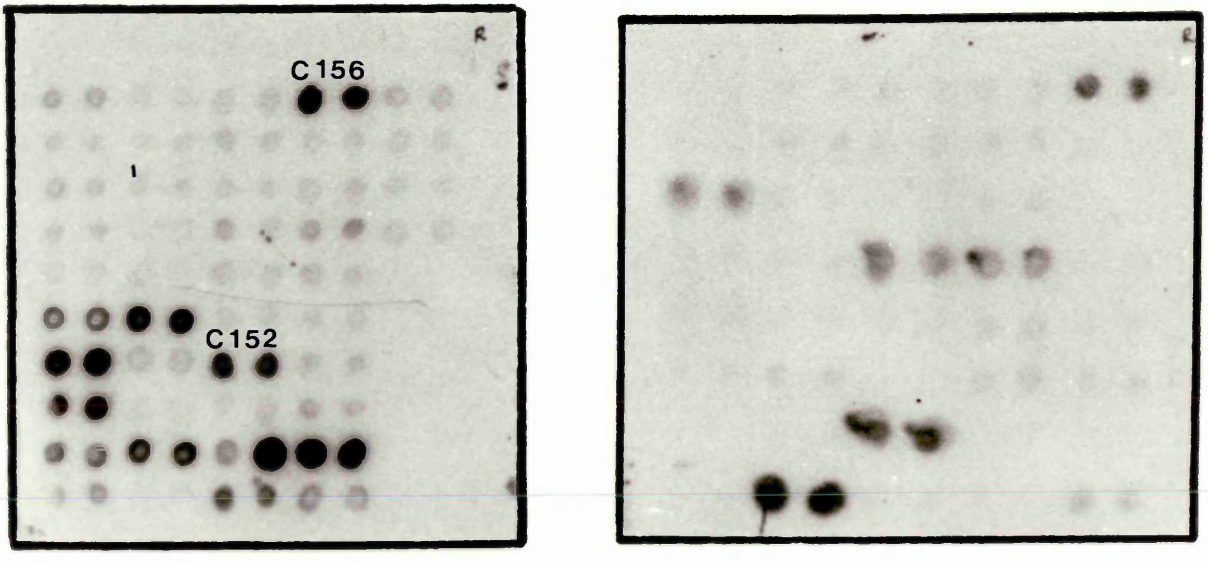
Plasmid DNA was prepared from the 247 clones selected from the initial screening of the mouse cDNA library (cf. Fig 3.2). 2 μl of each preparation was spotted directly in duplicate onto separate duplicate nitrocellulose filters and hybridised to (a) (^{32}P) single-stranded skeletal muscle cDNA and (b) (^{32}P) single-stranded cardiac muscle cDNA. The filters were bydridised overnight, washed five times for ten minutes in 2 x SSC, 0.1% SDS at room temperature, twice for 30 minutes in 0.1x SSC, 0.1% SDS at 45^{o}C and then autoradiographed. S202 is a clone which gave a hybridisation signal only with the skeletal muscle probe. C152 and C156 are clones which are apparently more abundant in cardiac muscle than in skeletal muscle.

(a)



^{32}P-labelled skeletal
muscle cDNA probe

(b)



^{32}P-labelled cardiac
muscle cDNA probe

hybridisation signal only with the skeletal muscle cDNA probe (e.g. clone S202).

To characterise these eight clones further, a different approach was adopted. Instead of screening the cloned cDNAs with single-stranded cDNA probes derived from reverse-transcribed mRNA, the mRNAs themselves were screened with the nick-translated plasmid DNA isolated from each of the clones. This approach was adopted to overcome problems in the previous screenings caused by the underrepresentation of a particular cDNA within the total cDNA probe, as it allowed hybridisation of a high specific activity probe to a full length mRNA. In addition, poly $(A)^+$ RNAs isolated from liver and brain were included as further controls for expression of the proteins, represented by the clones, in non-muscle tissue.

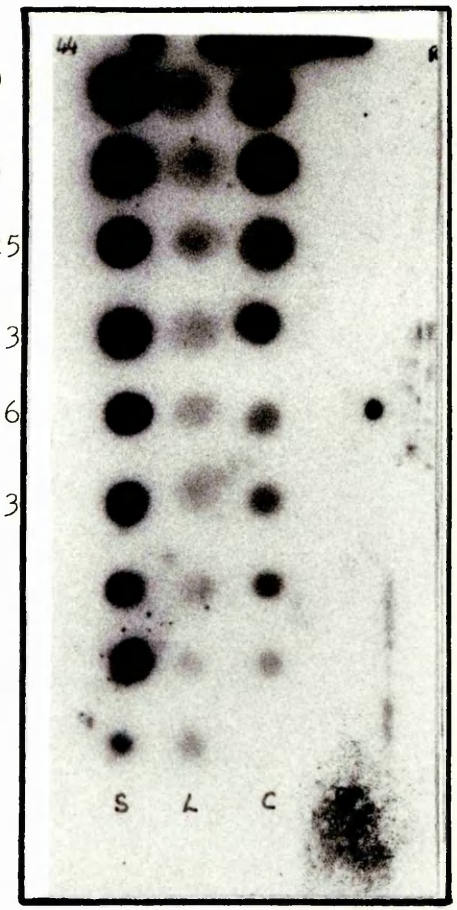
Figure 3.4 presents the autoradiographs of the RNA blots obtained for each of the eight clones. A variety of hybridisation patterns were obtained from this third screening. All but one of the clones hybridised with skeletal muscle poly $(A)^+$ RNA and at least one of the other poly $(A)^+$ RNAs. Four clones, numbers 21, 33, 215 and 239 gave a hybridisation signal with all four of the poly $(A)^+$ RNAs, suggesting that the mRNA represented by each of these clones was expressed ubiquitously. Clones 40 and 44 hybridised to the skeletal muscle poly $(A)^+$ RNA, and to a more or lesser extent, to poly A+ RNA from two of the other sources, whereas clone 9 appeared to be specific to muscle tissue. Clone 202 was the only one which gave a hybridisation signal solely with skeletal muscle poly $(A)^+$ RNA, and so it was selected for further characterisation. Clone 202 will

Figure 3.4 Results of hybridising the eight clones from the second screening of the skeletal muscle cDNA library to RNA from various sources.

Serial dilutions of poly $(A)^+$ RNA were prepared from skeletal muscle, cardiac muscle, liver and brain and spotted directly onto nitrocellulose filters. Eight such filters were probed with nick-translated plasmid DNA prepared from clones 9, 21, 33, 40, 44, 202, 215 and 239. The filters were hybridised overnight, washed five times for ten minutes in 2 x SSC, 0.1% SDS at room temperature and twice for 30 minutes in 0.1 x SSC, 0.1% SDS and 50°C and then autoradiographed.
Amounts of RNA indicated are in µg.

skeletal
liver
cardiac
brain
clone 9
1.0
0.5
0.25
0.13
0.06
0.03
skeletal
liver
cardiac
brain
clone 21
1.0
0.5
0.25
0.13
0.06
0.03
skeletal
liver
cardiac
brain
clone 33
1.0
0.5
0.25
0.13
0.06
0.03
skeletal
liver
cardiac
brain
clone 40

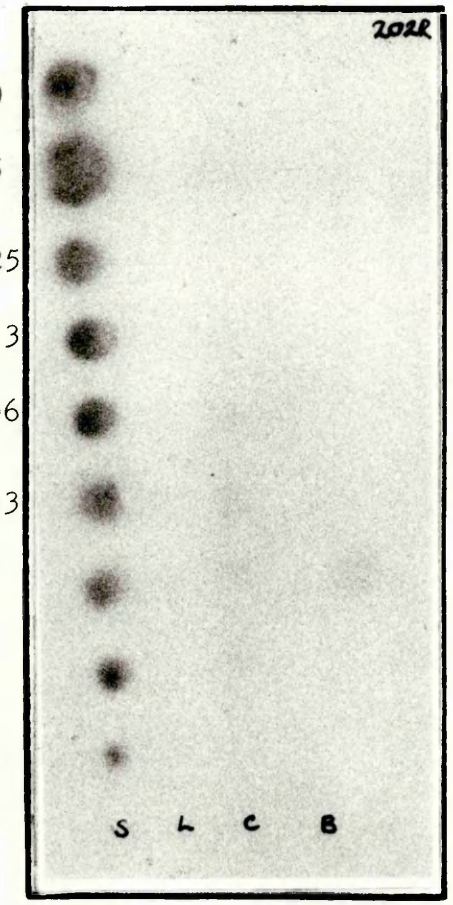
clone 44



0
5
25
13
06
03

skeletal
liver
cardiac
brain

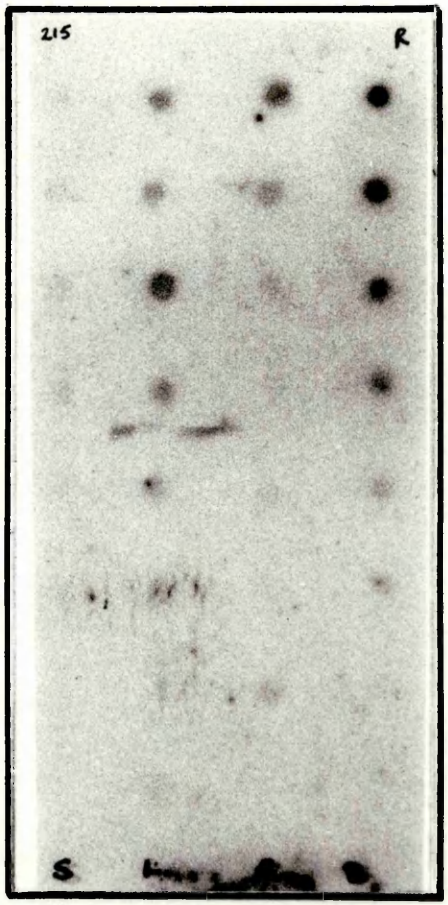
clone 202



1.0
0.5
0.25
0.13
0.06
0.03

skeletal
liver
cardiac
brain

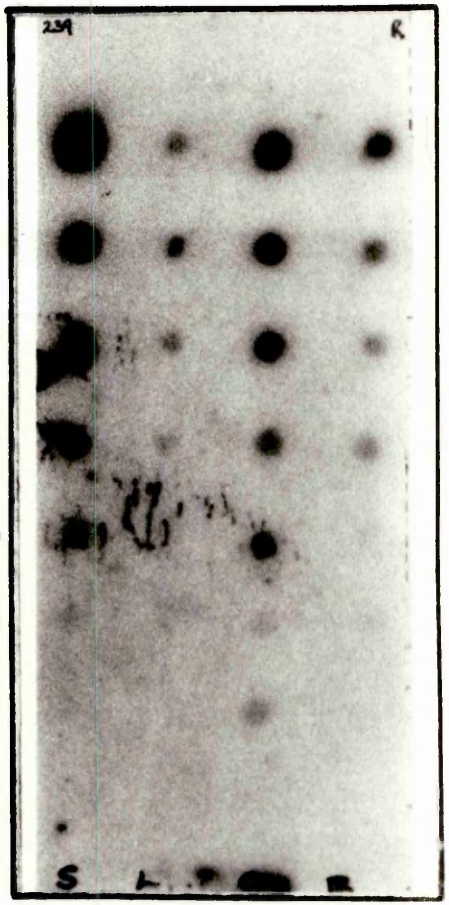
clone 215



1.0
0.5
0.25
0.13
0.06
0.03

skeletal
liver
cardiac
brain

clone 239



1.0
0.5
0.25
0.13
0.06
0.03

skeletal
liver
cardiac
brain

subsequently be referred to as pCM202.

3.2 Characterisation and Identification of pCM202

Construction of a cDNA library by G/C tailing and cloning into the unique Pst I restriction site of pBR322 should result in the regeneration of a Pst I restriction site at both ends of the inserted foreign DNA (Figure 3.5a). Therefore pCM202 was digested with Pst I and the resulting fragments separated by agarose gel electrophoresis. Two bands resulted, with apparent lengths of 4.4 kbp and 450 bp (Figure 3.6). The 4.4 kbp fragment comigrated with the linearised vector DNA and therefore it was concluded that pCM202 contained an insert of approximately 450 bp in length.

In view of the relatively small size of the cDNA insert in pCM202, it was decided to transfer the entire 450 bp Pst I fragment into the plasmid vector pUC 18 (Figure 2.2) and to determine its nucleotide sequence using the chemical method of Maxam and Gilbert (1980). By labelling the insert DNA at either the flanking Hind III or Eco RI site, in the multiple cloning region, using the Klenow reaction (section 2.6.1), and sequencing in opposite directions, the entire nucleotide sequence of the 450bp cDNA fragment of pCM202 was obtained. This revealed a G/C tail (16-20 G or C residues in series) at only one end of the fragment, indicating that one of the Pst I sites lay within the cDNA insert. As digestion of pCM202 with Pst I had resulted in only two detectable fragments (Figure 3.6) it was assumed that one of the Pst I sites had not been regenerated by the G/C tailing.

Figure 3.5 Regeneration of PstI restriction sites by the G/C-tailing cloning procedure.

(a) Construction of a cDNA library by the G/C tailing procedure and cloning into the unique Pst I restriction site of pBR322 should result in the regeneration of a Pst I restriction site at both ends of the inserted foreign DNA (Villakomeroff et al 1978). (b) All the β-tropomyosin clones found in this study lacked at least one regenerated Pst I site. In most cases it was found that a C- T mutation in the first base of the recognition site had resulted in the Pst I site not being regenerated. In pCM202 and pCM212, however, misalignment of the G/C tails resulted in only 3 out of the 6 bases of the Pst I recognition site being regenerated.

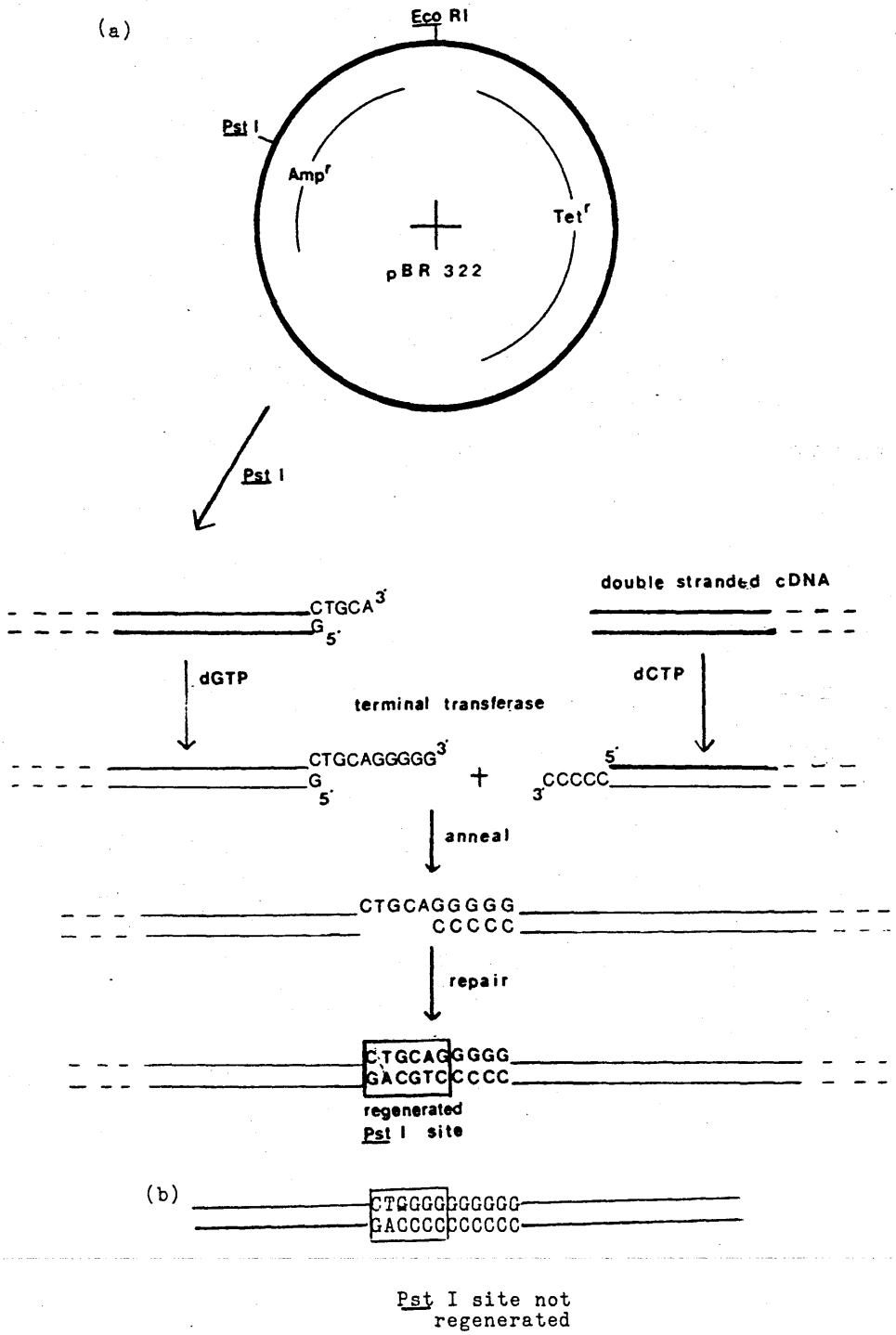
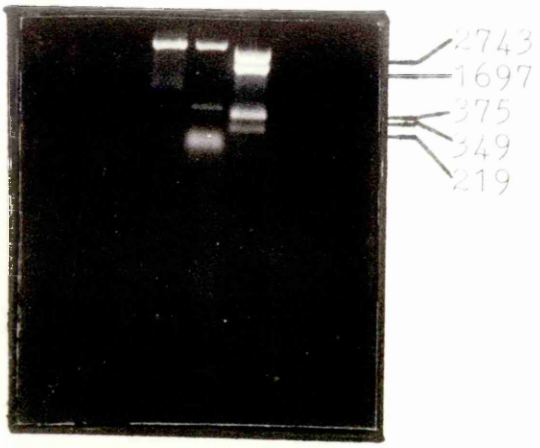

(a)
Eco RI
Pst I
Amp^r
Tet^r
pBR 322
Pst I
double stranded cDNA
CTGCA 3'
G 5'
dGTP
terminal transferase
dCTP
CTGCAGGGGG 3'
G 5'
+
5'
3' CCCCC
anneal
CTGCAGGGGG
CCCCC
repair
CTGCAGGGGG
GACGTCCCCC
regenerated
Pst I site
(b)
CTGGGGGGGGGG
GACCCCCCCCCC
Pst I site not
regenerated

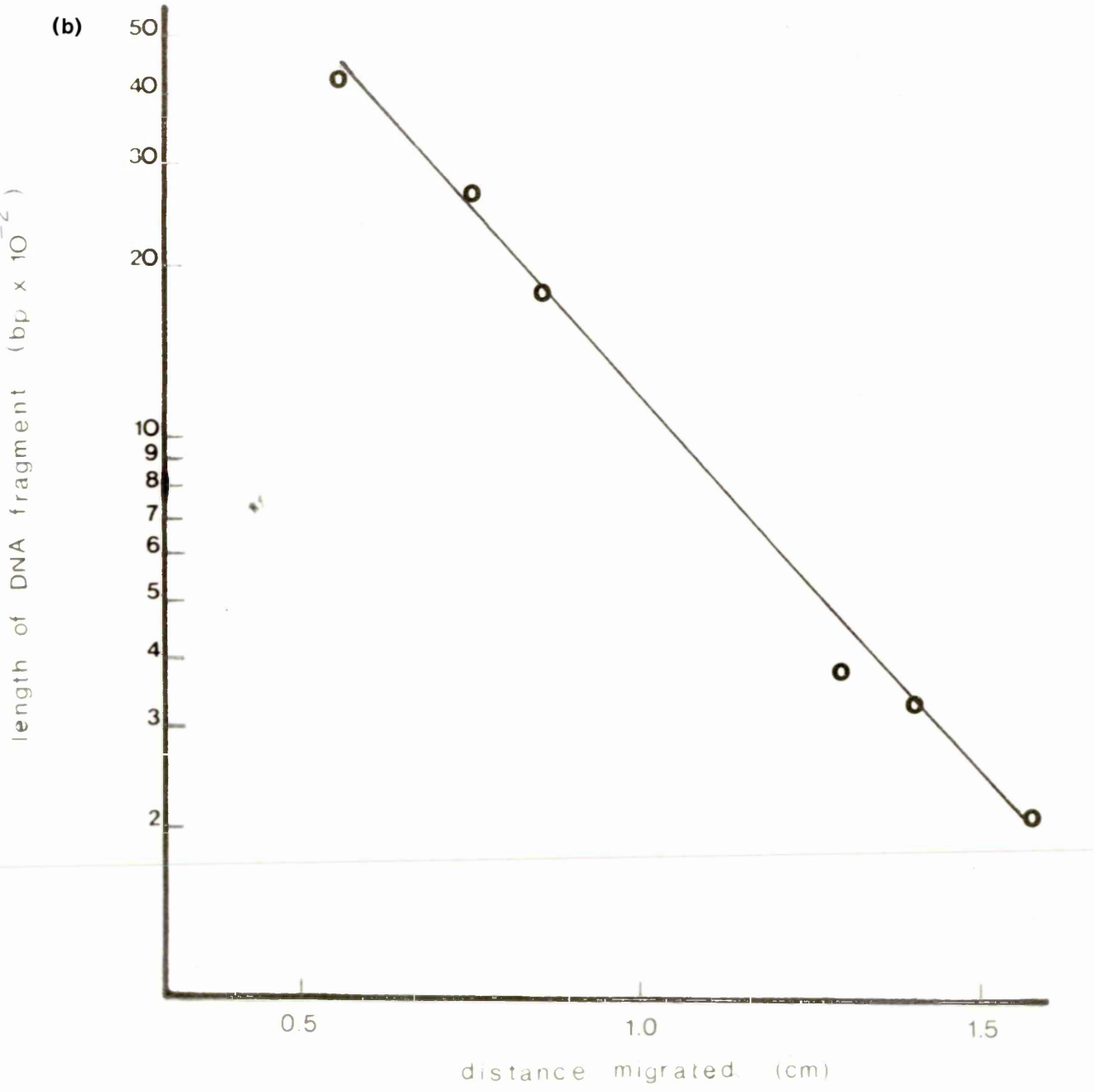
Figure 3.6 Analysis of pCM202 with restriction endonuclease Pst I.

pCM202 was digested with Pst I at 37°C for two hours and the resulting fragments separated by electrophoresis in a 1% agarose gel (a) Lanes 1-3 contain (from left to right) pBR322 digested with Pst I, pCM202 digested with Pst I and Øx 174 digested with Hpa II (DNA size markers). The size of the small Pst I fragment of pCM202 was estimated at approximately 450 bp from the graph (b).

(a)


1 2 3
2743
1697
375
349
219

(b)


50
40
30
20
10
9
8
7
6
5
4
3
2
length of DNA fragment (bp x 10^{-2})
0.5
1.0
1.5
distance migrated (cm)

Therefore pCM202 was digested with Pst I/Hind III and Pst I/Bgl I to determine in which orientation the cDNA insert lay (Figure 3.7) and its sequence determined from the internal Pst I site to the G/C cloning tail (which lacked a regenerated Pst I site - see Figure 3.5b). This completed the nucleotide sequence of the cDNA insert of pCM202, which is presented in Figure 3.8 along with the translation of the six possible reading frames.

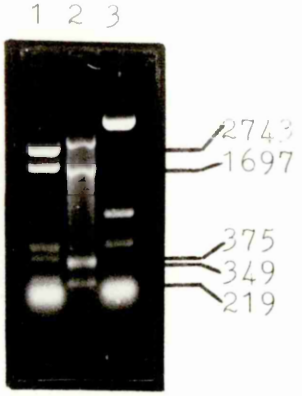
Although the nucleotide sequence of pCM202 had not been determined on both strands (Figure 3.7), the presence of an open reading frame (i.e. one which was not interrupted by stop codons), extending over the entire 489bp of sequence (Figure 3.8), suggested that it would be worthwhile, at this stage, to subject the sequence to analysis on the VAX cluster at EMBL, Heidelberg, to see if it was similar to any of the sequences contained in the Genbank and EMBL nucleotide sequence databases. This was done using the program WORDSEARCH, of the UWGCG software package (Devereux et al, 1984). The result of the computer search indicated that pCM202 represented the mRNA for an isoform of tropomyosin. It was found to share the greatest sequence homology with the major chicken α-tropomyosin (McLeod 1982; Figure 3.9), but because there were no mouse tropomyosin sequences in the databases the exact identity of the pCM202 insert could not be determined.

When the amino-acid sequence, predicted from the open reading frame in Figure 3.8, was compared with the protein sequences in the NBRF amino-acid databases on the VAX cluster at EMBL, Heidelberg, an almost perfect match was revealed with rabbit β-tropomyosin (Mak et

Figure 3.7 Partial restriction map of pCM202 showing sequencing strategy

(a) The orientation of the cDNA insert of pCM202, with respect to the vector DNA, was determined by performing "double digests" of the DNA and separating the resulting fragments on a 1% agarose gel. Lanes 1-3 contain (from left to right) pCM202 digested with Pst I/ Bgl I, Øx 174 digested with Hpa II (DNA size markers) and pCM202 digested with Pst I/Hind III. This allows one to deduce the orientation indicated in (b)(ii). (b)(i) scale indicating the size of the cDNA insert of pCM202. (ii) The partial restriction map of pCM202 is presented, indicating the vector DNA (dotted line), and the orientation of thepCM202 cDNA insert. 516 refers to nucleotide 516 in the β-tropomyosin sequence (the A of the initiator methionine codon AUG being numbered 1), as predicted from Figure 3.10, the rabbit β-tropomyosin amino-acid sequence. Plasmids pCM1 and pCM2 were constructed in pUC18 as indicated, and the nucleotide sequence of the cDNA determined using the sequencing strategy outlined in (iii). Arrows (→) denote the direction and approximate length of the sequence determined from individual subclones. Arrows above the central line indicate sequence determined on the coding strand of the DNA. Those below the central line indicate sequence determined on the non-coding strand of DNA.

(a)


1 2 3
2743
1697
375
349
219

(b)

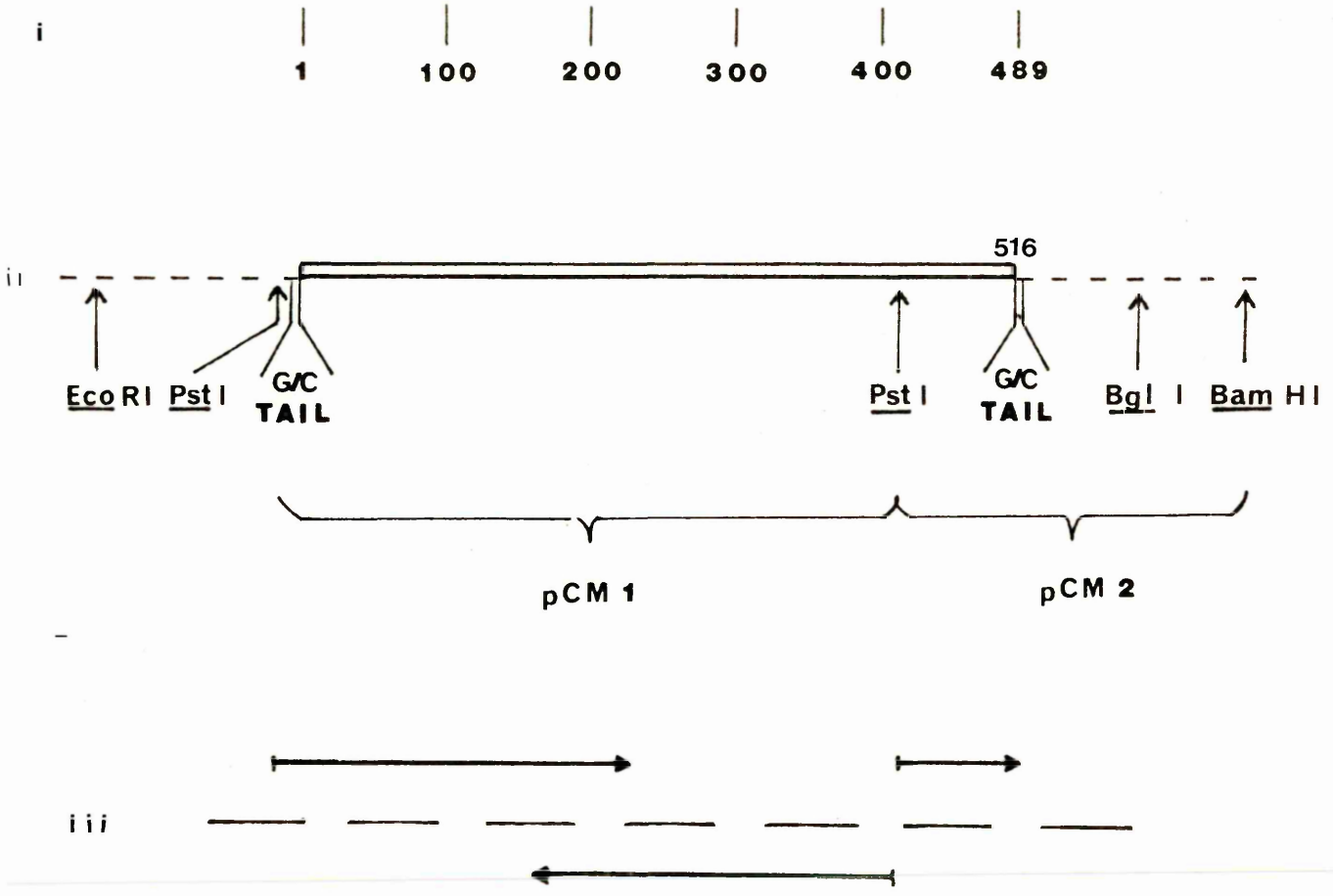

i
1
100
200
300
400
489
ii
516
Eco RI
Pst I
G/C TAIL
Pst I
G/C TAIL
Bgl I
Bam HI
pCM 1
pCM 2
iii

Figure 3.8 Conceptual translation of reading frames in pCM202.

The nucleotide sequence of pCM202 is presented opposite together with a translation of all six possible reading frames (A1- A3, B1-B3). The one letter amino-acid code is used. Stop codons are indicated by an asterix (*). A3 was the only reading frame found to be open, i.e. not interrupted by stop codons.

```
 L  V  R  N  V  G  T  *  R  M  P  S  T  A  R  S  R  P  K  P  ....A1
  *  E  *  M  W  G  P  R  E  C  H  R  P  R  G  A  G  R  S  R  ....A2
   E  S  K  C  G  D  L  E  N  A  I  D  R  A  E  Q  A  E  A   ....A3
CTGAGAGTAAATGTGGGGACCTAGAGAATGCCATCGACCGCGCGGAGCAGGCCGAAGCCG

 T  K  S  K  L  K  T  D  A  S  S  W  R  K  S  S  R  P  S  R  ....A1
  Q  K  A  S  *  R  P  M  Q  A  A  G  G  R  A  A  G  P  P  E  ....A2
D  K  K  Q  A  E  D  R  C  K  Q  L  E  E  E  Q  Q  A  L  Q   ....A3
ACAAAAAGCAAGCTGAAGACCGATGCAAGCAGCTGGAGGAAGAGCAGCAGGCCCTCCAGA

 R  S  *  R  G  Q  R  T  R  W  K  S  I  P  S  P  *  R  M  P  ....A1
  E  A  E  G  D  R  G  R  G  G  K  V  F  R  V  R  E  G  C  P  ....A2
K  K  L  K  G  T  E  D  E  V  E  K  Y  S  E  S  V  K  D  A   ....A3
AGAAGCTGAAGGGGACAGAGGACGAGGTGGAAAAGTATTCCGAGTCCGTGAAGGATGCCC

 R  R  N  W  S  R  L  R  R  R  P  P  T  L  K  Q  M  W  P  L  ....A1
  G  E  T  G  A  G  *  E  E  G  H  R  R  *  S  R  C  G  L  S  ....A2
Q  E  K  L  E  Q  A  E  K  K  A  T  D  A  E  A  D  V  A  S   ....A3
AGGAGAAACTGGAGCAGGCTGAGAAGAAGGCCACCGACGCTGAAGCAGATGTGGCCTCTC

 *  T  A  A  F  S  S  *  R  R  S  W  I  G  H  R  S  A  W  L  ....A1
  E  P  P  H  S  A  R  R  G  G  V  G  S  G  T  G  A  P  G  Y  ....A2
L  N  R  R  I  Q  L  V  E  E  E  L  D  R  A  Q  E  R  L  A   ....A3
TGAACCGCCGCATTCAGCTCGTAGAGGAGGAGTTGGATCGGGCACAGGAGCGCCTGGCTA

 Q  P  C  K  S  W  R  R  L  R  K  P  R  M  R  A  R  E  E  *  ....A1
  S  L  A  K  A  G  G  G  *  E  S  R  G  *  E  R  E  R  N  E  ....A2
T  A  L  Q  K  L  E  E  A  E  K  A  A  D  E  S  E  R  G  M   ....A3
CAGCCTTGCAAAAGCTGGAGGAGGCTGAGAAAGCCGCGGATGAGAGCGAGAGAGGAATGA

 R  S  L  K  T  G  P  *  R  M  R  K  R  W  S  C  R  R  C  S  ....A1
  G  H  *  K  P  G  H  E  G  *  G  K  D  G  A  A  G  D  A  A  ....A2
K  V  I  E  N  R  A  M  K  D  E  E  K  M  E  L  Q  E  M  Q   ....A3
AGGTCATTGAAAACCGGGCCATGAAGGATGAGGAAAAGATGGAGCTGCAGGAGATGCAGC

 *  R  K  P  S  T  S  L  R  T  Q  T  A  N  M  R  R  W  P  G  ....A1
  E  G  S  Q  A  H  R  *  G  L  R  P  Q  I  *  G  G  G  Q  E  ....A2
L  K  E  A  K  H  I  A  E  D  S  D  R  K  Y  E  E  V  A  R   ....A3
TGAAGGAAGCCAAGCACATCGCTGAGGACTCAGACCGCAAATATGAGGAGGTGGCCAGGA

S  W  *   ....A1
 A  G  D  ....A2
K  L  V  T ....A3
AGCTGGTGAC
```

```
 V  T  S  F  L  A  T  S  S  Y  L  R  S  E  S  S  A  M  C  L  ....B1
  S  P  A  S  W  P  P  P  H  I  C  G  L  S  P  Q  R  C  A  W  ....B2
   H  Q  L  P  G  H  L  L  I  F  A  V  *  V  L  S  D  V  L   ....B3
GTCACCAGCTTCCTGGCCACCTCCTCATATTTGCGGTCTGAGTCCTCAGCGATGTGCTTG

 A  S  F  S  C  I  S  C  S  S  I  F  S  S  S  F  M  A  R  F  ....B1
  L  P  S  A  A  S  P  A  A  P  S  F  P  H  P  S  W  P  G  F  ....B2
G  F  L  Q  L  H  L  L  Q  L  H  L  F  L  I  L  H  G  P  V   ....B3
GCTTCCTTCAGCTGCATCTCCTGCAGCTCCATCTTTTCCTCATCCTTCATGGCCCGGTTT

 S  M  T  F  I  P  L  S  L  S  S  A  A  F  S  A  S  S  S  F  ....B1
  Q  *  P  S  F  L  S  R  S  H  P  R  L  S  Q  P  P  P  A  F  ....B2
F  N  D  L  H  S  S  L  A  L  I  G  R  F  L  S  L  L  Q  L   ....B3
TCAATGACCTTCATTCCTCTCTCGCTCTCATCCGCGGCTTTCTCAGCCTCCTCCAGCTTT

 C  K  A  V  A  R  R  S  C  A  R  S  N  S  S  S  T  S  *  M  ....B1
  A  R  L  *  P  G  A  P  V  P  D  P  T  P  P  L  R  A  E  C  ....B2
L  Q  G  C  S  Q  A  L  L  C  P  I  Q  L  L  L  Y  E  L  N   ....B3
TGCAAGGCTGTAGCCAGGCGCTCCTGTGCCCGATCCAACTCCTCCTCTACGAGCTGAATG

 R  R  F  R  E  A  T  S  A  S  A  S  V  A  F  F  S  A  C  S  ....B1
  G  G  S  E  R  P  H  L  L  Q  R  R  W  P  S  S  Q  P  A  P  ....B2
A  A  V  Q  R  G  H  I  C  F  S  V  G  G  L  L  L  S  L  L   ....B3
CGGCGGTTCAGAGAGGCCACATCTGCTTCAGCGTCGGTGGCCTTCTTCTCAGCCTGCTCC

 S  F  S  W  A  S  F  T  D  S  E  Y  F  S  T  S  S  S  V  P  ....B1
  V  S  P  G  H  P  S  R  T  R  N  T  F  P  P  R  P  L  S  P  ....B2
Q  F  L  L  G  I  L  H  G  L  G  I  L  F  H  L  V  L  C  P   ....B3
AGTTTCTCCTGGGCATCCTTCACGGACTCGGAATACTTTTCCACCTCGTCCTCTGTCCCC

 F  S  F  F  W  R  A  C  C  S  S  S  S  C  L  H  R  S  S  A  ....B1
  S  A  S  S  G  G  P  A  A  L  P  P  A  A  C  I  G  L  Q  L  ....B2
L  Q  L  L  L  E  G  L  L  L  F  L  Q  L  L  A  S  V  F  S   ....B3
TTCAGCTTCTTCTGGAGGGCCTGCTGCTCTTCCTCCAGCTGCTTGCATCGGTCTTCAGCT

 C  F  L  S  A  S  A  C  S  A  R  S  M  A  F  S  R  S  P  H  ....B1
  A  F  C  R  L  R  P  A  P  R  G  R  W  H  S  L  G  P  H  I  ....B2
L  L  F  V  G  F  G  L  L  R  A  V  D  G  I  L  *  V  P  T   ....B3
TGCTTTTTGTCGGCTTCGGCCTGCTCCGCGCGGTCGATGGCATTCTCTAGGTCCCCACAT

 L  L  S   ....B1
  Y  S  Q  ....B2
F  T  L  R ....B3
TTACTCTCAG
```

Figure 3.9 Result of the search of the nucleotide databases, Genbank and EMBL.

The nucleotide sequence of pCM202 was used to search the nucleotide sequences stored in the Genbank and EMBL databases. The "Quality" of the match indicates the extent of homology between the nucleotide sequence of pCM202 with a given sequence in the nucleotide database. The highest quality of match (156.2) was found to be with the major chicken α-tropomyosin (McLeod 1982).

```
      cm:202.rft                check: 6087  from: 37  to: 400  /reverse
     genbank.sst  entry: 1818  check: 7681  from: 1  to: 989
ENTRY: 1818 SEARCHSET of: chktropal.  check: 7681  from: 1  to: 989
disk$users:[pubdata.genbank_40.uwgcg.]]othervert]chktropal.
 Gaps: 3  Quality: 156.2  Ratio: 0.442  Words: 51  Width: 3  Limits: +/-.

     359 GCGCAGCAGGCCGAAGCCGACAAAAAGCAAGCTGAAGACCGATGCAAG.. 312
         !!!!!  .!   ! !!! !!       !  !!  ! !  ! ! !! ! !
      10 gcggacaagaaggcagcgga.....ggagaggagcaagcagctggaggac 54

     311 CAGCTGGAGGAAGAGCAGCAGGCCCTCCAGAAGAAGCTGAAGGGGACAGA 262
          !!!! ! !!    !!!  !!   !!  !!  !   !!!! !  !!
      55 gagctggtggctctgcaaaagaacctgaag..ggcactgaggatgagctg 102

Press RETURN to continue

     261 GGACGAGGTGGAAAGTATTCCGGTCGTGAGGTGCCAGGACTGGAGCAGGC 212
         !      :   !:!           !!   :  !   !!!! ! !!!
     103 gacaaatattccgagtcccttaaagatgcacaggaaaagttggaactggc 152

     211 TGAGAAGAAGGCCACCGACGCTGAAGCAGATGTGACCTCTCTGAACCGCC 162
         !!! !  !!!!!!! !! !!!!!   !! !!  . !! !!!!! ! !
     153 tgacaaaaaggccacagatgctgagagtgaagtagctttccctgaacagac 202

     161 GCATTCAGCTCGTAGAGGAGGAGTTGGATCGCGCACAGGAGCGCCTGGCT 112
         !!!.    ! !! !!!!! !!!!!.!!! .!!  !! ! !!!.  !!!!!
     203 gcatccaactggttgaggaagagttggatcgggctcaggagcgcttggct 252

     111 ACAGCCTTGCAAAAGCTGGAGGAGGCTGAGAAAGCGCGCGATGAGAGCGA 62
         .  !    ! . !!!.!!!!!!!!!! !!!!.!  ! .    !!!  .  .
     253 actgcccctgcagaagctggaggaggctgagaaggctgcagatgaggagtga 302

      61 GAGAGGAATGAAGGTCATTGAAAACCGGGGCATCAAGGATCAGGAAAAGA 12
         !    !.!!. !!.!!!!!!!!!  ! !     !.! ...! !! !!!!
     303 aagaggaatgaaggtcattgaaaatagagcccagaaggatgaagagaaga 352
```

al, 1980) as illustrated in Figure 3.10. There were only eight mismatched amino acids, all of which occurred at the N-terminus of the amino acid sequence, predicted for pCM202. However, these corresponded exactly to eight amino acids nearer the C-terminus of the rabbit β-tropomyosin in a region not covered by pCM202. Thus it was possible that this region of mismatched amino-acids represented a cloning artefact rather than a real difference between the amino-acid sequences of mouse and rabbit β-tropomyosin.

The skeletal muscle cDNA library was rescreened using the nick-translated Pst I fragment of pCM202 as probe, in order to try to obtain clones containing more of the mouse β-tropomyosin cDNA. Approximately 10,000 colonies were screened, from which twenty-two putative positives were selected. Plasmid DNA was prepared from each, subjected to electrophoresis on an agarose gel, transferred to nitrocellulose and then hybridised to the pCM202 probe. The resulting autoradiograph is presented in Figure 3.11, from which it can be seen that only three clones, pCM212, pCM222 and pCM232 gave positive hybridisation signals.

The DNA from each clone was digested with Pst I in order to determine the size of their inserts. However, as with pCM202, it was possible that one or both Pst I sites may not have been regenerated by the G/C tailing procedure and, therefore, each of the clones was also linearised with the restriction enzyme Hind III and their sizes determined. It can be seen from Figure 3.12 that pCM232 is similar in size to pCM202, containing two Pst I fragments of 4.4 kbp and 450 bp with an overall size of

Figure 3.10 Comparison of the predicted amino-acid sequence of pCM202 with that of rabbit β-tropomyosin

The amino-acid sequence of pCM202, predicted from A3 the only open reading frame in Figure 3.8 was compared with tropomyosin sequences in the NBRF amino-acid database. An almost perfect match was found with that of rabbit β-tropomyosin (Mak et al 1980). The figure opposite shows a comparison of the sequence with identities being indicated by vertical lines. Notice the identity between the boxed regions, which is discussed in the text.

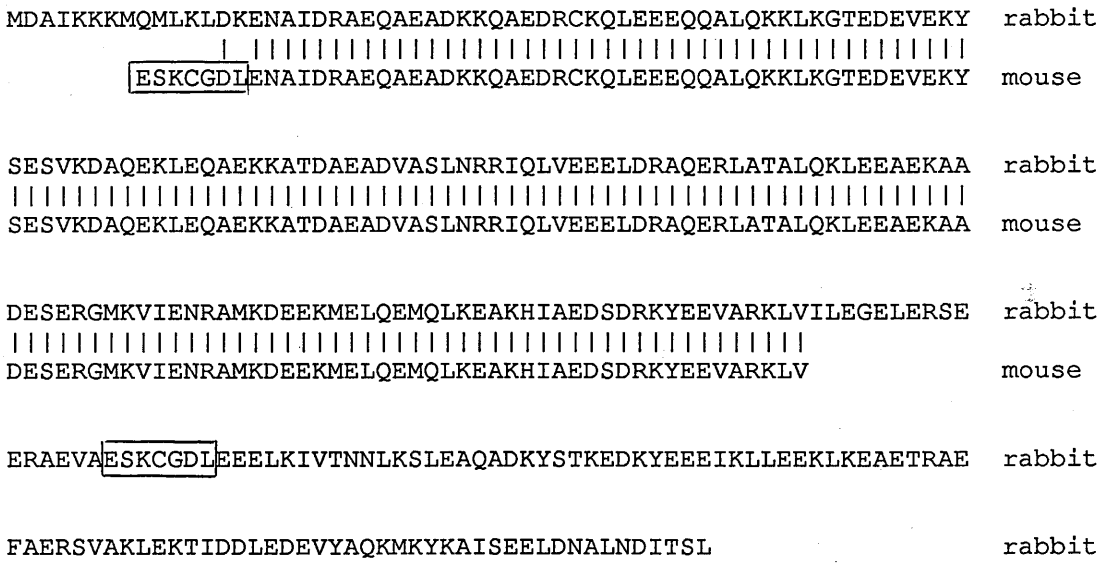
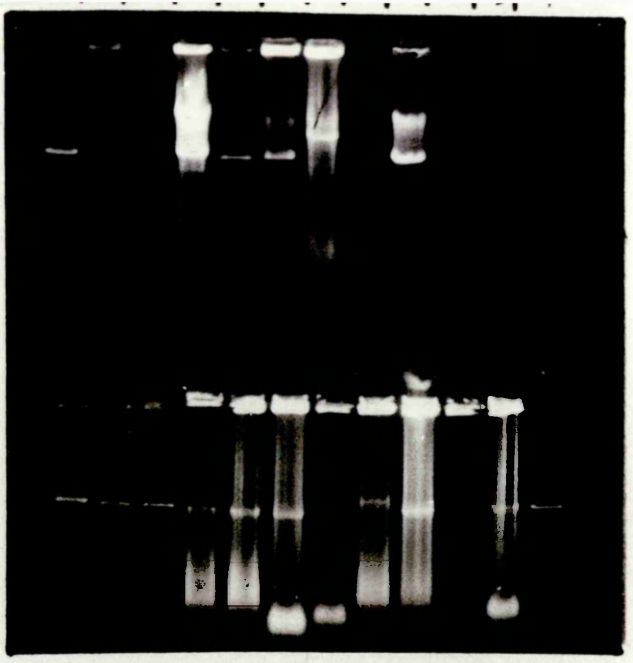

MDAIKKKMQMLKLDKENAIDRAEQAEADKKQAEDRCKQLEEEQQALQKKLKGTEDEVEKY rabbit
ESKCGDLENAIDRAEQAEADKKQAEDRCKQLEEEQQALQKKLKGTEDEVEKY mouse
SESVKDAQEKLEQAEKKATDAEADVASLNRRIQLVEEELDRAQERLATALQKLEEAEKAA rabbit
SESVKDAQEKLEQAEKKATDAEADVASLNRRIQLVEEELDRAQERLATALQKLEEAEKAA mouse
DESERGMKVIENRAMKDEEKMELQEMQLKEAKHIAEDSDRKYEEVARKLVILEGELERSE rabbit
DESERGMKVIENRAMKDEEKMELQEMQLKEAKHIAEDSDRKYEEVARKLV mouse
ERAEVAESKCGDLEEELKIVTNNLKSLEAQADKYSTKEDKYEEEIKLLEEKLKEAETRAE rabbit
FAERSVAKLEKTIDDLEDEVYAQKMKYKAISEELDNALNDITSL rabbit

Figure 3.11 Rescreen of the mouse skeletal muscle cDNA library using pCM202 as probe

The Pst I fragment of pCM202 was nick-translated and used to rescreen the mouse skeletal muscle cDNA library. Twenty-two puta tive positives were selected and used for small-scale preparations of plasmid DNA. 2 µl of each plasmid was then electro- phoresed through a 1% agarose gel, blotted onto nitrocellulose and probed with pCM202. Hybridisation was overnight at 42°C followed by washing five times for ten minutes in 2xSSC, 0.1% SDS, twice for 30 minutes in 0.1xSSC at 45°C and then autoradiographed. (a) shows the gel stained with ethidium bromide. (b) shows an autoradiograph of the nitrocellulose filter after hybridisation. Three clones, numbers 9, 27 and 28 gave a positive hybridisation signal and were designated pCM212, pCM232 and pCM222 respectively.

(a)



15 16 17 18 20 21 24 25 26 27 28 29

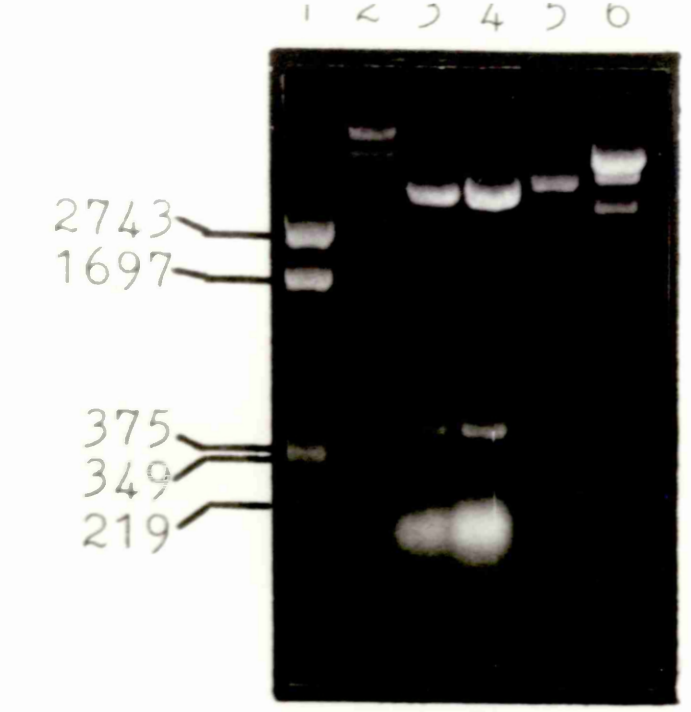
9

(b)

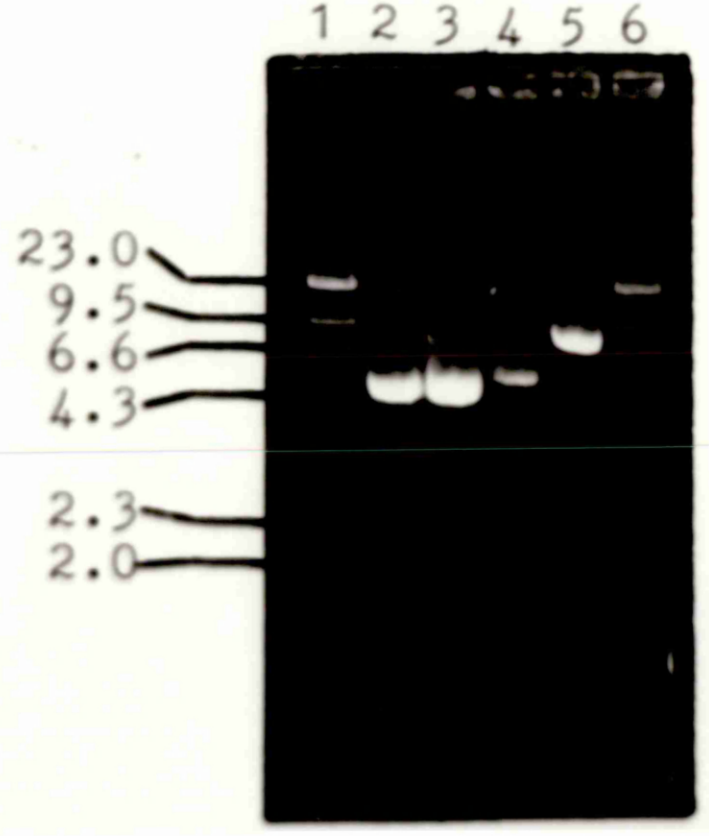
27 28

Figure 3.12 Estimation of the sizes of the inserts in the tropomyosin clones pCM212, pCM222 and pCM232.

Each of the tropomyosin clones were digested with Pst I and Hind III in order to determine the size of its insert. (a) shows a 1% agarose gel. Lanes 1-6 contained (from left to right) Øx 174 digested with Hpa II, λDNA digested with Hind III, pCM202 digested with Pst I, pCM232 digested with Pst I, pCM212 digested with Pst I (NB In lane 6 the relative intensites of the two upper bands indicate partial digestion.) (b) shows a 0.7% agarose gel. Lanes 1-6 contain (from left to right) λDNA digested with Hind III, pCM202 digested with Hind III, pCM232 digested with Hind III, pCM222 digested with Hind III, pCM212 digested with Hind III and λDNA digested with Hind III.



(b)



approximately 4.8 kbp. Digestion of pCM222 produced two Pst I fragments of approximately 4.6 kbp and 700 bp whereas pCM212 produced two Pst I fragments of approximately 4.9 kbp and 3.0 kbp.

It was decided to determine the sequence of each of cDNA inserts of pCM212, pCM222 and pCM232 by subcloning fragments of each into the bacteriophage vectors M13mp18 and M13mp19 and using the chain termination sequencing reactions of Sanger (1977). Figure 3.13 outlines the sequencing strategy adopted for pCM232. It also presents a portion of the sequence to demonstrate that pCM232 is, in fact, identical to pCM202. Thus it contains the region at the 5' end containing the misplaced amino acids (cf Figure 3.10) and the G/C tails which are of identical length to those found in pCM202.

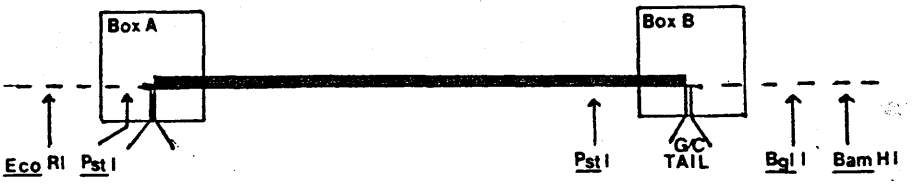
pCM222 was found to contain the coding region of β-tropomyosin from the initiator methionine to amino acid 172 (516 nucleotides), and thus was able to confirm that the first eight amino acids at the 5' end of pCM202 were indeed misplaced (Figure 3.14) although the region of tropomyosin from which they were considered to be derived was also not covered by pCM222. In addition to the coding region of β-tropomyosin pCM222 also contained at least 88 nucleotides (probably 95 - see Discussion) of the 5' non-coding region of the mRNA. However preceding these was a region of DNA corresponding to nucleotides 287-505 of the coding region of β-tropomyosin (numbering from the A of the initiator methionine codon AUG), but in an inverted orientation, as illustrated in Figure 3.15. The nucleotide sequence at the junction between this inverted DNA and the 5' non-coding region of β-tropomyosin is shown in Figure 3.16. It should

Figure 3.13 Strategy adopted for sequencing pCM232

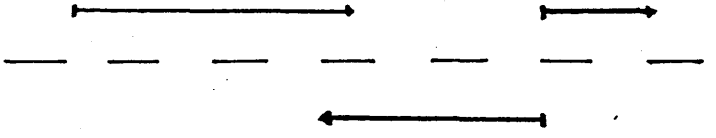
(i) Scale indicating the size of the cDNA insert of pCM232. (ii) The partial restriction map of pCM232 is shown indicating the restriction enzyme sites used for determining the nucleotide sequence. The vector DNA is indicated by the dotted line. The nucleotide sequence was determined by subcloning fragments of the plasmid into M13pm18 and M13mp19 and using the dideoxy chain termination method. (iii) The sequencing strategy is indicated by the arrows (→). Arrows above the central line indicate sequence determined on the coding strand of DNA, those below the line indicate sequence on the non-coding strand of DNA. The arrows also denote the direction and approximate length of the sequence determined for each individual subclone. (iv) A comparison between the extreme 5' and 3' boundaries of pCM202 and pCM232 is shown. Identities are indicated by a vertical line.

(i)

(ii)



(iii)



(iv)

BOX A

$(G)_{16}$CTGAGAGTAAATGTGGGGAC-----pCM202

$(G)_{16}$CTGAGAGTAAATGTGCGGAC-----pCM232

BOX B

pCM202-----AGGTGGCCAGGAAGCTGGTGA$(C)_{15}$

pCM232-----AGGTGGCCAGGAAGCTGGTGA$(C)_{15}$

Figure 3.14 Partial restriction map and strategy adopted for sequencing pCM222

(i) Scale indicating the size of the cDNA insert of pCM222. (ii) The partial restriction map of pCM222 is shown indicating the restriction enzyme sites used in determining the nucleotide sequence. The nucleotide sequence was determined by the chain termination method. Numbering is with respect to the β-tropomyosin sequence with the A of the initiator methionine codon AUG counted as 1. The vector DNA is indicated by the dotted line. The open box (95 nucleotides) indicates the 5' non-coding region of the mRNA, and the shaded box (b) the main tropomyosin coding sequence. The second shaded box, (a), also contains tropomyosin sequence, but opposite in orientation to the main sequence (b), as indicated by the open arrows (⇨). (iii) The sequenciang strategy adopted for pCM222 is presented. Arrows (→) indicate the direction and approximate length of sequence determined for each individual subclone. Arrows above the central line indicate sequence determined on the coding strand of DNA, while those below the line indicate sequence determined on the non-coding strand of DNA.

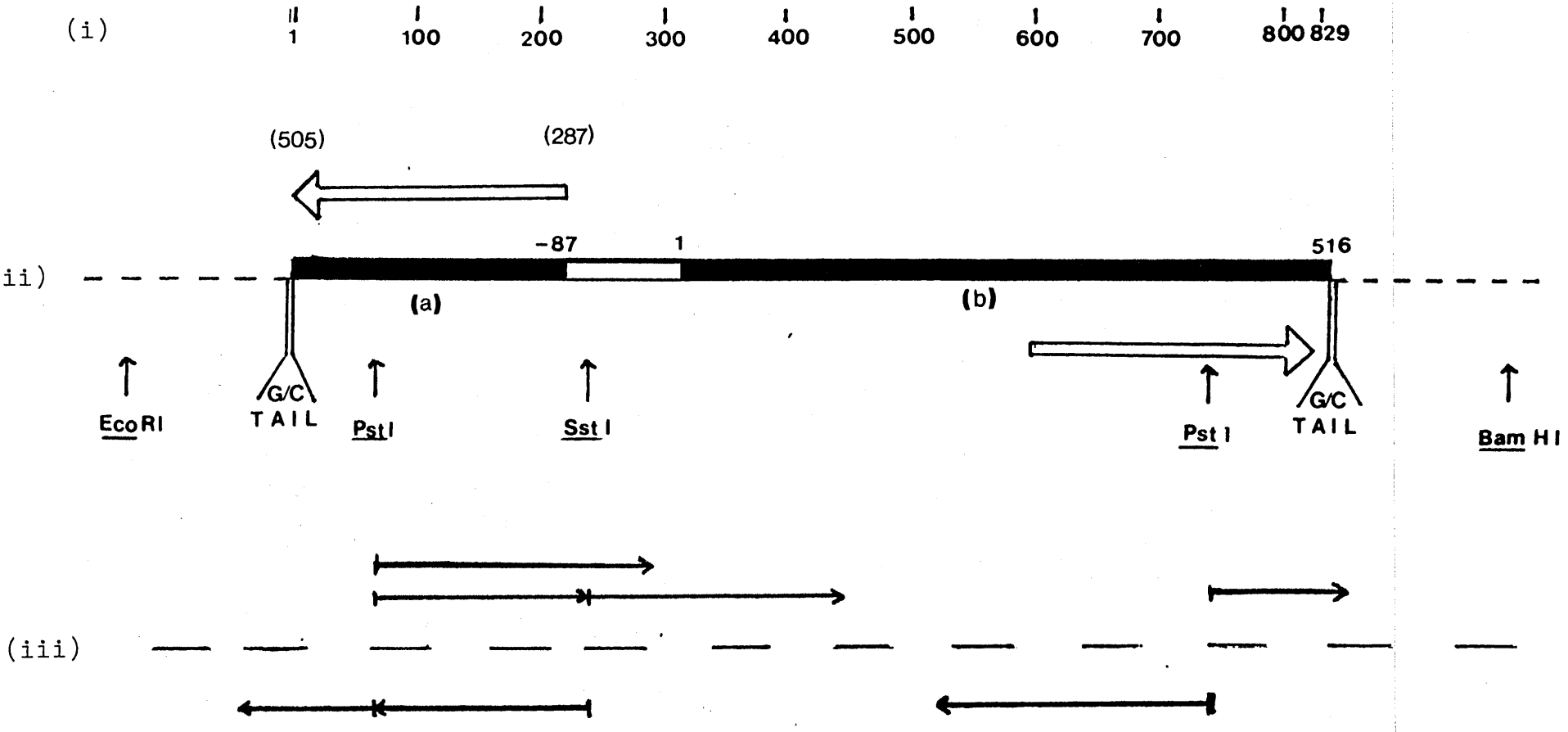

(i)
1
100
200
300
400
500
600
700
800 829
(505)
(287)
-87
1
516
(ii)
(a)
(b)
EcoRI
G/C
TAIL
PstI
SstI
PstI
G/C
TAIL
BamHI
(iii)

Figure 3.15 Partial nucleotide sequence of pCM222

Part of the nucleotide sequence of pCM222 is presented opposite including the region of inverted DNA, the 5' non coding sequence of the mRNA and the first 150 nucleotides of the coding sequence. The amino-acid sequence is also presented, the three letter amino-acid code being used. Numbering above each line is according to the amino-acid sequence of β-tropomyosin starting with the initiator methionine. (NB The first 218 nucleotides contain the inverted repeat DNA and therefore the amino-acids are in reverse orientation. The nucleotide sequence presented in this region is from the non-coding strand of β-tropomyosin, but the same strand of DNA as the rest of the sequence presented, and so the codons and amino-acids presented do not correspond to each other.) Numbering below each line is according to nucleotide position. A comparison between the nucleotide sequences of pCM222 and pCM202 is also presented. Identities are indicated by vertical lines. Note the boxed region containing the first seven codons of pCM202 which is discussed in the text.

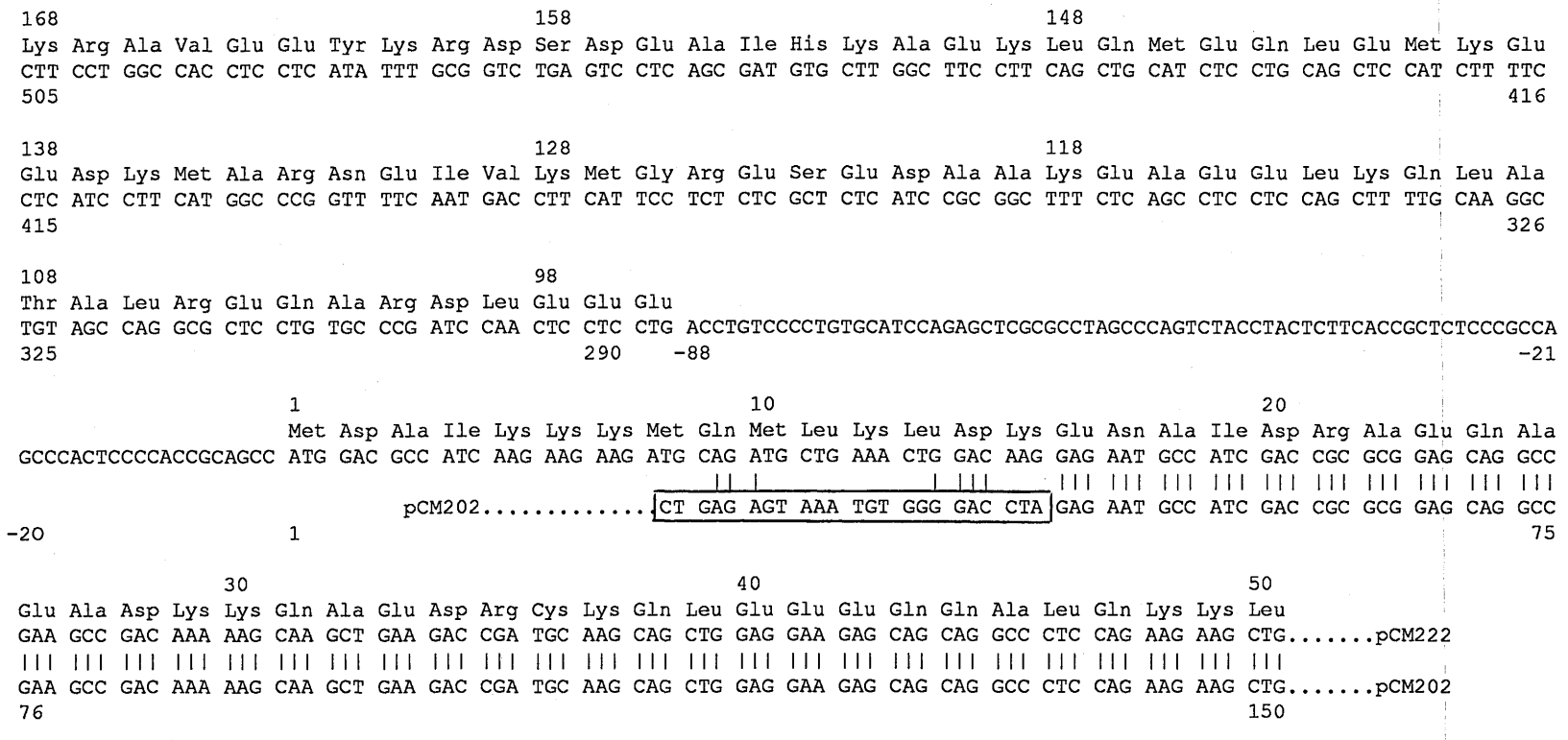
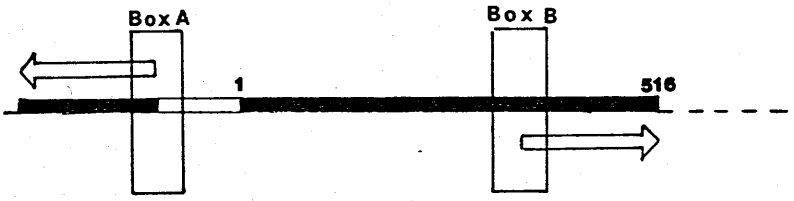

168
158
148
Lys Arg Ala Val Glu Glu Tyr Lys Arg Asp Ser Asp Glu Ala Ile His Lys Ala Glu Lys Leu Gln Met Glu Gln Leu Glu Met Lys Glu
CTT CCT GGC CAC CTC CTC ATA TTT GCG GTC TGA GTC CTC AGC GAT GTG CTT GGC TTC CTT CAG CTG CAT CTC CTG CAG CTC CAT CTT TTC
505
416
138
128
118
Glu Asp Lys Met Ala Arg Asn Glu Ile Val Lys Met Gly Arg Glu Ser Glu Asp Ala Ala Lys Glu Ala Glu Glu Leu Lys Gln Leu Ala
CTC ATC CTT CAT GGC CCG GTT TTC AAT GAC CTT CAT TCC TCT CTC GCT CTC ATC CGC GGC TTT CTC AGC CTC CTC CAG CTT TTG CAA GGC
415
326
108
98
Thr Ala Leu Arg Glu Gln Ala Arg Asp Leu Glu Glu Glu
TGT AGC CAG GCG CTC CTG TGC CCG ATC CAA CTC CTC CTG ACCTGTCCCCTGTGCATCCAGAGCTCGCGCCTAGCCCAGTCTACCTACTCTTCACCGCTCTCCCGCCA
325
290
-88
-21
1
10
20
Met Asp Ala Ile Lys Lys Lys Met Gln Met Leu Lys Leu Asp Lys Glu Asn Ala Ile Asp Arg Ala Glu Gln Ala
GCCCACTCCCCACCGCAGCC ATG GAC GCC ATC AAG AAG AAG ATG CAG ATG CTG AAA CTG GAC AAG GAG AAT GCC ATC GAC CGC GCG GAG CAG GCC
pCM202.....CT GAG AGT AAA TGT GGG GAC CTA GAG AAT GCC ATC GAC CGC GCG GAG CAG GCC
-20
1
75
30
40
50
Glu Ala Asp Lys Lys Gln Ala Glu Asp Arg Cys Lys Gln Leu Glu Glu Glu Gln Gln Ala Leu Gln Lys Lys Leu
GAA GCC GAC AAA AAG CAA GCT GAA GAC CGA TGC AAG CAG CTG GAG GAA GAG CAG CAG GCC CTC CAG AAG AAG CTG.....pCM222
GAA GCC GAC AAA AAG CAA GCT GAA GAC CGA TGC AAG CAG CTG GAG GAA GAG CAG CAG GCC CTC CAG AAG AAG CTG.....pCM202
76
150

Figure 3.16 The nucleotide sequence at the junction between the inverted DNA and the 5' non-coding DNA of pCM222

(i) A schematic representation of pCM222 is presented. The vector DNA is indicated by the dotted line. The β-tropomyosin coding sequence is indicated by the shaded boxes, the 5' non-coding region of the mRNA by the open box. The open arrows (⇨) indicate the regions of inverted DNA. (ii) A comparison is made between the nucleotide sequence on the coding strand of DNA of Boxes A and B. Numbering is with respect to the β-tropomyosin sequence (the A of the initiator methionine codon AUG being counted as 1). Identities are indicated by a vertical line. Note that the nucleotide sequence in Box B is in reverse orientation. Note also that the unique nucleotides of the 5' non-coding region extend only as far as nucleotide -87.

(i)



(ii)

BOX A	-107		-97	-87	-77
5'-----	TGCCCGATCCAACTCCTCCTGACCTGTCCCCT				-----3'
3'-----	AGGACACGGGCTAGGTTGAGGAGGAGATGCTCG				-----5'
	313	303	293	283	BOX B

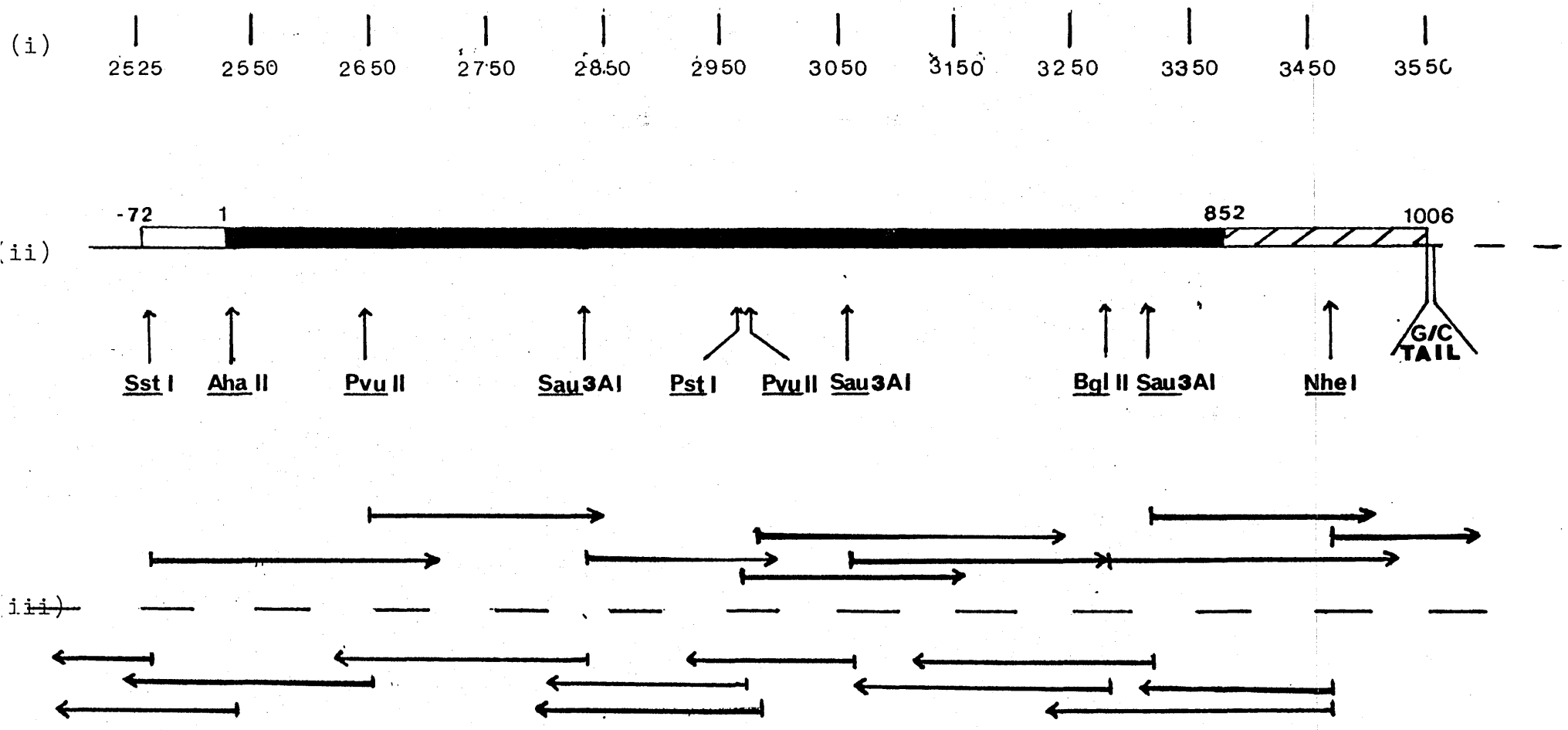
be mentioned here that this artefact is thought to have arisen during the construction of the double stranded cDNA and not after transformation of the cells. However the possible origins of this (and other) artefact(s) will be considered in the Discussion.

Clone pCM212 had been shown by restriction endonuclease digestion (Figure 3.12) to be approximately 3.5kb larger than pBR322. On analysis of this clone it was found that the primary tropomyosin insert accounted for approximately 1.1kb of this difference, the remainder being due to a duplication. The sequence of the tropomyosin will be presented first, followed by a description of how the overall structure of pCM212 was determined.

pCM212 contained the entire coding region of mouse β-tropomyosin (852 nucleotides) together with 72 nucleotides of the 5' non-coding (cf. the 5' non-coding sequence of pCM222), and 151 nucleotides of the 3' non-coding, regions of the mRNA (Figure 3.17). It was from this clone that the majority of the sequence data were obtained, with more than 95% of the β-tropomyosin sequence being determined on both strands of the DNA. The 72 nucleotides of the 5' non-coding region were identical to those found in pCM222 and, furthermore, the extensive sequencing was able to confirm that there was only one Sst I restriction site in this region of the DNA and therefore that the additional 5' non-coding sequence predicted from pCM222 was valid. The presence of the entire coding region also confirmed that the misplaced amino-acids in pCM202 had arisen as a result of a duplication of nucleotides 561-582 of the coding region of β-tropomyosin. However the mechanism by which this could have

Figure 3.17 Partial restriction map and strategy for sequencing the β-tropomyosin region of pCM212

(i) Scale indicating the extent of the tropomyosin sequence of pCM212. Numbering is according to the scale in Figure 3.18. (ii) The partial restriction map of pCM212 is presented indicating restriction enzyme sites which were used in determining the nucleotide sequences. The vector DNA is indicated by the dotted line, the tropomyosin coding sequence by the shaded box, the 3' non-coding sequence by the hatched box and the 5' non-coding sequence by the open box. (iii) The sequencing strategy adopted for pCM212 is presented. The arrows (→) indicate the direction and approximate length of the sequence determined for each individual subclone. Arrows above the central line indicate sequence determined on the coding strand of the DNA, while those below the line indicate sequence on the non-coding strand of DNA.


(i)
2525
2550
2650
2750
2850
2950
3050
3150
3250
3350
3450
3550
(ii)
-72
1
852
1006
G/C TAIL
Sst I
Aha II
Pvu II
Sau3AI
Pst I
Pvu II
Sau3AI
Bgl II
Sau3AI
Nhe I
(iii)

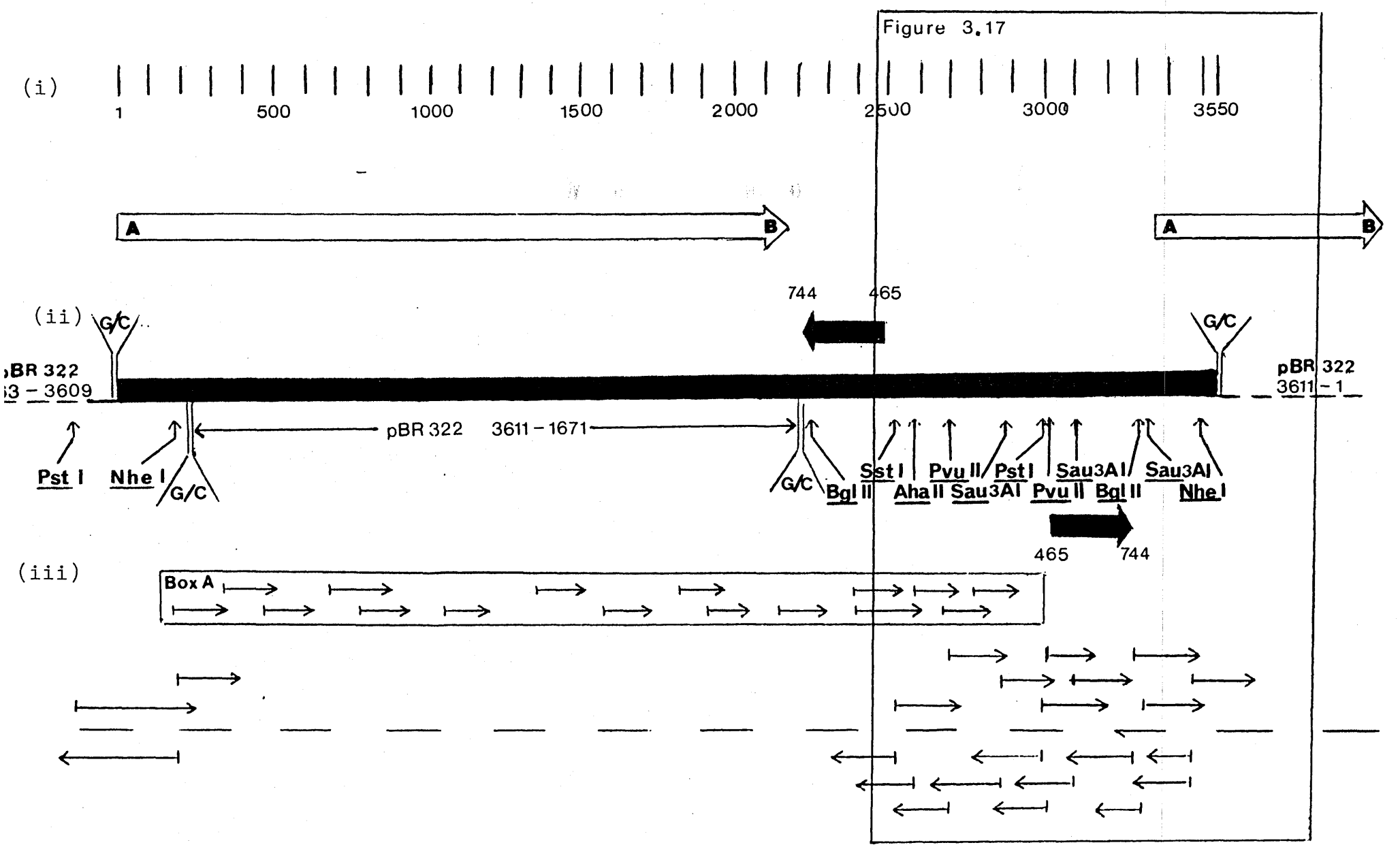
occurred will be considered in the Discussion.

The extra DNA in clone pCM212, other than the primary tropomyosin insert, lay 5' of the latter. Much of this lacked restriction enzyme sites convenient for cloning into the M13 sequencing vectors and so a different strategy had to be adopted for determining its sequence. The entire 3.0kb Pst I fragment of pCM212, containing the region for which no sequence had been determined, was subcloned into M13mp18. A recombinant phage was selected, replicative form DNA prepared from this, and a library of deletion subclones made using the method outlined in section 2.11. Clones carrying deletions of varying lengths were selected, as described, and sequenced (see Box A of Figure 3.18). The overall structure of the DNA bounded by the G/C cloning tails in pCM212 was as follows. Following nucleotide 3615 of the vector there was a Pst I site and a G/C cloning tail. Then there was a region corresponding to nucleotides 778-1008 of the β-tropomyosin sequence in the orientation 5' - 3'. Following this was an "internal" G/C tail, lacking a Pst I site (because of a misalignment of the G/C tails during the ligation of insert to vector) and nucleotides 3608-1671 of the plasmid vector. Following this was a second "internal" G/C tail. Interestingly, nucleotides 1675-1671 contain part (5 out of 6) of a Pst I recognition site which may be significant in explaining the origin of the direct repeat - see Discussion. However, following the second internal G/C tail was a region of DNA corresponding to nucleotides 744-468 (amino-acids 248-156) of the β-tropomyosin sequence in inverted orientation (3' - 5'). Then the β-tropomyosin sequence, described in Figure 3.17, was found

Figure 3.18 Partial restriction map, and strategy adopted for sequencing pCM212

(i) Approximate scale indicating the length of the cDNA insert of pCM212. Bases 2450-3550 are represented in more detail in Figure 3.17. (ii) Partial restriction map of pCM212 indicating the restriction enzyme sites which were used to determine the nucleotide sequence. The sites of the four G/C cloning tails are indicated. The vector DNA is indicated by the dotted line, while the solid shaded box indicates the total insert DNA. pCM212 also contained regions of repeated DNA. Directly repeated DNA is indicated by the open arrows (⇨). Invertedly repeated DNA is indicated by the shaded arrows (➡). (iii) The sequencing strategy adopted for pCM212 is shown. Arrows (→) indicate the direction and approximate length of sequence determined from each individual subclone. Arrows above the central line indicate sequence determined on the coding strand of the DNA, while those below the line indicate sequence on the non-coding strand of DNA. Over 95% of the mouse β-tropomyosin cDNA sequence was determined on both strands of the DNA. Arrows enclosed within Box A indicate sequence determined using the deletion subclone library prepared by the method outlined in Section 2.11.

Figure 3.17


(i)
1
500
1000
1500
2000
2500
3000
3550
A
B
A
B
744
465
(ii)
G/C
G/C
pBR 322
3609 - 3611 - 1
pBR 322 3611 - 1671
Pst I
Nhe I
G/C
G/C
Bgl II
Sst I
Aha II
Pvu II
Sau3A I
Pst I
Pvu II
Sau3A I
Bgl II
Sau3A I
Nhe I
465
744
(iii)
Box A

immediately following this inverted repeat. At the extreme 3' end of the insert there was a G/C cloning tail, lacking a regenerated Pst I site (because of a misalignment of the G/C tails during the ligation of insert to vector) then the sequence of the plasmid vector resuming at nucleotide 3608 and continuing in the direction 3608 - 1 (Figure 3.18). A large part of the DNA sequence determined for pCM212 is presented in Figure 3.19.

The presence of an "internal" G/C tail immediately preceding the inverted repeat suggests that the original cDNA clone contained only the β-tropomyosin sequence with an associated inverted repeat, which was presumed to have arisen in a manner similar to that for the inverted repeat in pCM222 and before the double stranded cDNA was cloned into pBR322 (see Discussion). However, the direct repeat, including both the β-tropomyosin 3' non-coding sequence and the fragment of pBR322, must have arisen in vivo after the ligation of the cDNA with the vector and subsequent to the transformation of E. Coli. This will be considered further in the Discussion. Figure 3.20 shows the junction between the 5' non-coding sequence of the β-tropomyosin cDNA and the region of inverted DNA, as well as comparing the internal G/C tails with those at the extreme ends of the cDNA insert.

Approximately 1.1kb of mouse β-tropomyosin sequence is presented in Figure 3.21 together with the predicted amino acid sequence.

Figure 3.19 The nucleotide sequence of clone pCM202

The partial nucleotide sequence of pCM212 is presented. Numbers above each line refer to nucleotide position within the total insert DNA, whereas numbers below each line refer to nucleotide position with respect to the β-tropomyosin sequence. Note the boxed regions indicating the repeated DNA. The open box indicates the directly repeated DNA whereas the shaded box indicates the invertedly repeated DNA. (NB Only part of the internal pBR322 sequence is presented.) Restriction enzyme sites which were used for sequencing are indicated. G/C cloning tails are indicated by lower case letters.

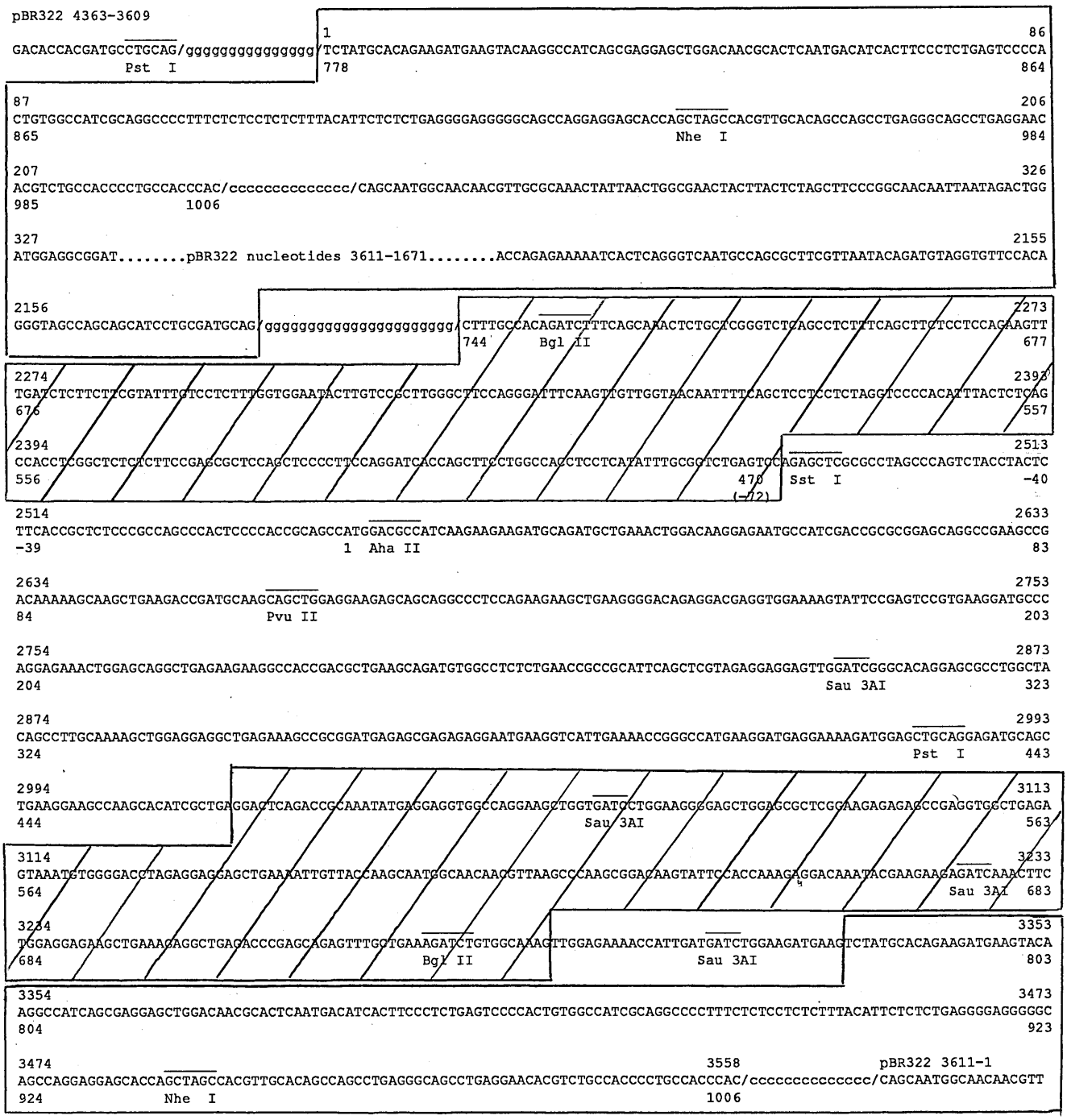
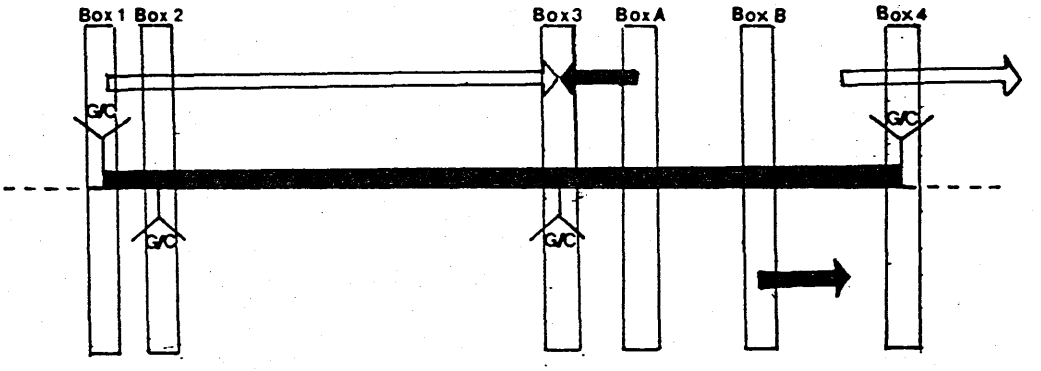

pBR322 4363-3609
1
86
GACACCACGATGCCTGCAG/ggggggggggggggg/TCTATGCACAGAAGATGAAGTACAAGGCCATCAGCGAGGAGCTGGACAACGCACTCAATGACATCACTTCCCTCTGAGTCCCCA
Pst I
778
864
87
206
CTGTGGCCATCGCAGGCCCCTTTCTCTCCTCTCTTTACATTCTCTCTGAGGGGAGGGGGCAGCCAGGAGGAGCACCAGCTAGCCACGTTGCACAGCCAGCCTGAGGGCAGCCTGAGGAAC
865
Nhe I
984
207
326
ACGTCTGCCACCCCTGCCACCCAC/ccccccccccccc/CAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGG
985
1006
327
2155
ATGGAGGCGGAT.....pBR322 nucleotides 3611-1671.....ACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACA
2156
2273
GGGTAGCCAGCAGCATCCTGCGATGCAG/gggggggggggggggggggggg/CTTTGCCACAGATCTTTCAGCAAACTCTGCTCGGGTCTCAGCCTCTTTCAGCTTCTCCTCCAGAAGTT
744
Bgl II
677
2274
2393
TGATCTCTTCTTCGTATTTGTCCTCTTTGGTGGAATACTTGTCCGCTTGGGCTTCCAGGGATTTCAAGTTGTTGGTAACAATTTTCAGCTCCTCCTCTAGGTCCCCACATTTACTCTCAG
676
557
2394
2513
CCACCTCGGCTCTCTCTTCCGAGCGCTCCAGCTCCCCTTCCAGGATCACCAGCTTCCTGGCCAGCTCCTCATATTTGCGGTCTGAGTGCAGAGCTCGCGCCTAGCCCAGTCTACCTACTC
556
470
(-72)
Sst I
-40
2514
2633
TTCACCGCTCTCCCGCCAGCCCACTCCCCACCGCAGCCATGGACGCCATCAAGAAGAAGATGCAGATGCTGAAACTGGACAAGGAGAATGCCATCGACCGCGCGGAGCAGGCCGAAGCCG
-39
1 Aha II
83
2634
2753
ACAAAAAGCAAGCTGAAGACCGATGCAAGCAGCTGGAGGAAGAGCAGCAGGCCCTCCAGAAGAAGCTGAAGGGGACAGAGGACGAGGTGGAAAAGTATTCCGAGTCCGTGAAGGATGCCC
84
Pvu II
203
2754
2873
AGGAGAAACTGGAGCAGGCTGAGAAGAAGGCCACCGACGCTGAAGCAGATGTGGCCTCTCTGAACCGCCGCATTCAGCTCGTAGAGGAGGAGTTGGATCGGGCACAGGAGCGCCTGGCTA
204
Sau 3AI
323
2874
2993
CAGCCTTGCAAAAGCTGGAGGAGGCTGAGAAAGCCGCGGATGAGAGCGAGAGAGGAATGAAGGTCATTGAAAACCGGGCCATGAAGGATGAGGAAAAGATGGAGCTGCAGGAGATGCAGC
324
Pst I
443
2994
3113
TGAAGGAAGCCAAGCACATCGCTGAGGACTCAGACCGCAAATATGAGGAGGTGGCCAGGAAGCTGGTGATCCTGGAAGGGGAGCTGGAGCGCTCGGAAGAGAGAGCCGAGGTGGCTGAGA
444
Sau 3AI
563
3114
3233
GTAAATGTGGGGACCTAGAGGAGGAGCTGAAAATTGTTACCAAGCAATGGCAACAAGGTTAAGCCCAAGCGGACAAGTATTCCACCAAAGAGGACAAATACGAAGAAGAGATCAAACTTC
564
Sau 3AI
683
3234
3353
TGGAGGAGAAGCTGAAAGAGGCTGAGACCCGAGCAGAGTTTGGTGAAAGATCTGTGGCAAAGTTGGAGAAAACCATTGATGATCTGGAAGATGAAGTCTATGCACAGAAGATGAAGTACA
684
Bgl II
Sau 3AI
803
3354
3473
AGGCCATCAGCGAGGAGCTGGACAACGCACTCAATGACATCACTTCCCTCTGAGTCCCCACTGTGGCCATCGCAGGCCCCTTTCTCTCCTCTCTTTACATTCTCTCTGAGGGGAGGGGGC
804
923
3474
3558
pBR322 3611-1
AGCCAGGAGGAGCACCAGCTAGCCACGTTGCACAGCCAGCCTGAGGGCAGCCTGAGGAACACGTCTGCCACCCCTGCCACCCAC/ccccccccccccc/CAGCAATGGCAACAACGTT
924
Nhe I
1006

Figure 3.20 Detailed examination of the direct and inverted repeats of pCM212

(i) A schematic representation of clone pCM212 is presented. The vector DNA is indicated by the dotted line and the solid box indicates the insert DNA. The four G/C cloning tails are also indicated. Directly repeated DNA is indicated by the open arrows (⟹). Invertedly repeated DNA is indicated by the shaded arrows (➡). (ii) A comparison between the nucleotide sequences found on the same strand of DNA in Box A and Box B is shown. Numbering is with respect to the β-tropomyosin cDNA sequence (the A of the initiator methionine codon, AUG, is counted as 1). Note that the sequence in Box B is in reverse orientation. Note also that the unique nucleotides in Box A only extend up to nucleotide -70. (iii) A comparison between the nucleotide sequences surrounding the G/C tails in Box 2 and Box 4 is shown. The nucleotide sequences presented are from the same strand of DNA. Note the complete identity betwen the two sequences as indicated by the vertical lines. (iv) A comparison between the nucleotide sequences surrounding the G/C tails in Box 1 and Box 3 is shown. The nucleotide sequences presented are from the same strand of DNA. Note the partial homology immediately preceding the G tails, as indicated by the vertical lines.

(i)



(ii)

```
BOX A                -84        -74        -64        -54
      5'----CATATTTGCGGTCTGAGTCCAGAGCTCGCGCCTAGCCCA----3'
3'----GGAGGAGTATAAACGCCAGACTCAGGAGTCGCTACACGAA----3'
    493        483        473        463              BOX B
```

(iii)

```
BOX 2
     5'----GCCACCCCTGCCACCCAC(C)15AGCAATGGCAACAACGTT----3'
     5'----GCCACCCCTGCCACCCAC(C)15AGCAATGGCAACAACGTT----3'
BOX 4
```

(iv)

```
BOX 1
     5'----GACACCACGATGCCTGCAG(G)14GTCTATGCACAGAAGAT----3'
     5'----GCAGCATCCTGCGATGCAG(G)22CTTTGCCACAGAGATCT----3'
BOX 3
```

Figure 3.21 The nucleotide sequence of mouse β-tropomyosin cDNA

The nucleotide sequence and deduced amino-acid sequence of mouse β-tropomyosin cDNA is presented. Over 95% of the sequence was determined on both strands of the DNA, with the majority of the sequence derived from clone pCM212. Nucleotides -95 → -73 were obtained from clone pCM222. Numbers above each line refer to amino acid position and numbers below each line to nucleotide position.

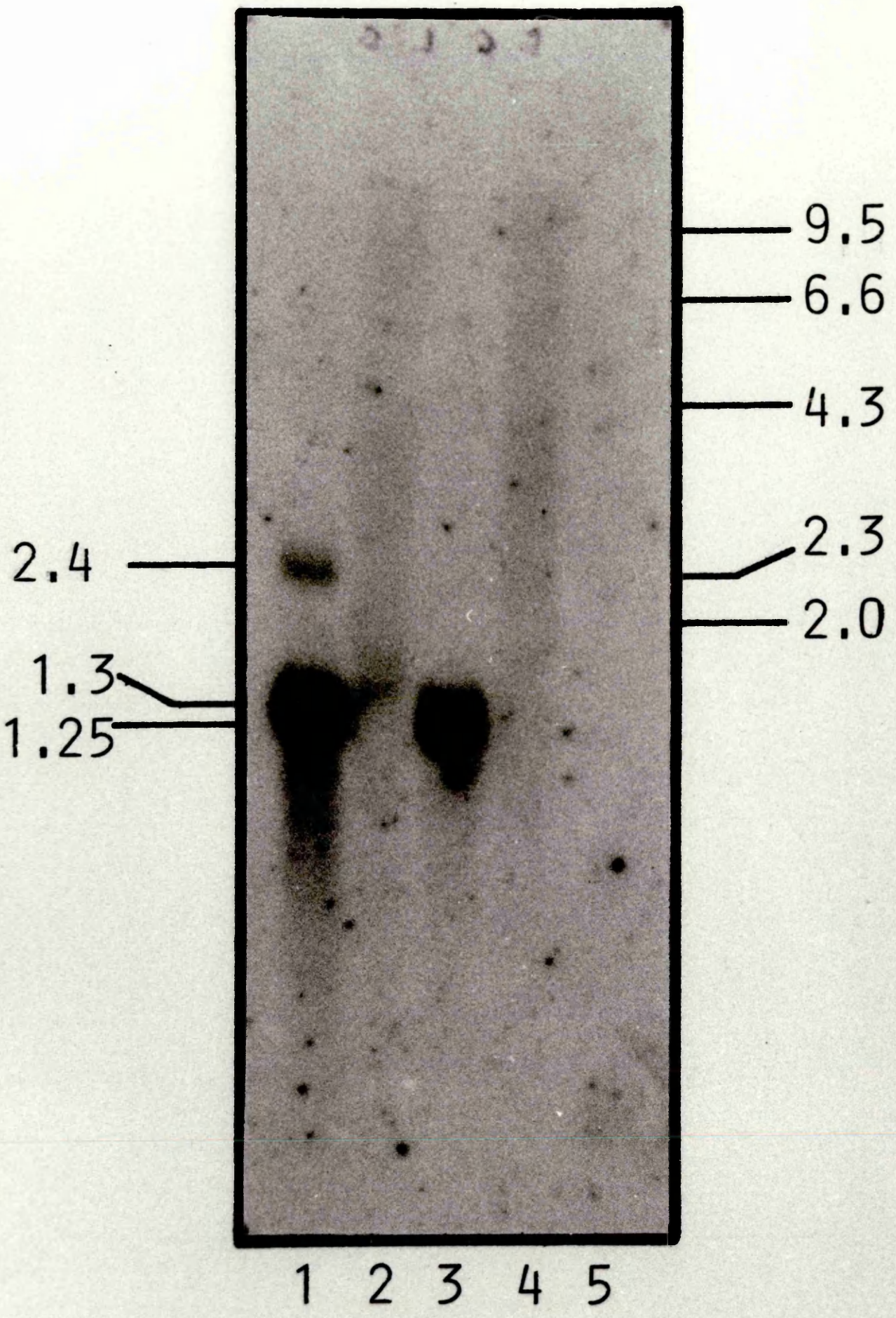
```
                                                                                                   1
                                                                                                   Met Asp Ala Ile Lys Lys
CCTCCTGACCTGTCCCCTGTGCATCCAGAGCTCGCGCCTAGCCCAGTCTACCTACTCTTCACCGCTCTCCCGCCAGCCCACTCCCCACCGCAGCC ATG GAC GCC ATC AAG AAG
-95                                                                                                1                    18

            10                                      20                                      30
Lys Met Gln Met Leu Lys Leu Asp Lys Glu Asn Ala Ile Asp Arg Ala Glu Gln Ala Glu Ala Asp Lys Lys Gln Ala Glu Asp Arg Cys
AAG ATG CAG ATG CTG AAA CTG GAC AAG GAG AAT GCC ATC GAC CGC GCG GAG CAG GCC GAA GCC GAC AAA AAG CAA GCT GAA GAC CGA TGC
19                                                                                                                  108

            40                                      50                                      60
Lys Gln Leu Glu Glu Glu Gln Gln Ala Leu Gln Lys Lys Leu Lys Gly Thr Glu Asp Glu Val Glu Lys Tyr Ser Glu Ser Val Lys Asp
AAG CAG CTG GAG GAA GAG CAG CAG GCC CTC CAG AAG AAG CTG AAG GGG ACA GAG GAC GAG GTG GAA AAG TAT TCC GAG TCC GTG AAG GAT
109                                                                                                                 198

            70                                      80                                      90
Ala Gln Glu Lys Leu Glu Gln Ala Glu Lys Lys Ala Thr Asp Ala Glu Ala Asp Val Ala Ser Leu Asn Arg Arg Ile Gln Leu Val Glu
GCC CAG GAG AAA CTG GAG CAG GCT GAG AAG AAG GCC ACC GAC GCT GAA GCA GAT GTG GCC TCT CTG AAC CGC CGC ATT CAG CTC GTA GAG
199                                                                                                                 288

            100                                     110                                     120
Glu Glu Leu Asp Arg Ala Gln Glu Arg Leu Ala Thr Ala Leu Gln Lys Leu Glu Glu Ala Glu Lys Ala Ala Asp Glu Ser Glu Arg Gly
GAG GAG TTG GAT CGG GCA CAG GAG CGC CTG GCT ACA GCC TTG CAA AAG CTG GAG GAG GCT GAG AAA GCC GCG GAT GAG AGC GAG AGA GGA
289                                                                                                                 378

            130                                     140                                     150
Met Lys Val Ile Glu Asn Arg Ala Met Lys Asp Glu Glu Lys Met Glu Leu Gln Glu Met Gln Leu Lys Glu Ala Lys His Ile Ala Glu
ATG AAG GTC ATT GAA AAC CGG GCC ATG AAG GAT GAG GAA AAG ATG GAG CTG CAG GAG ATG CAG CTG AAG GAA GCC AAG CAC ATC GCT GAG
379                                                                                                                 468

            160                                     170                                     180
Asp Ser Asp Arg Lys Tyr Glu Glu Val Ala Arg Lys Leu Val Ile Leu Glu Gly Glu Leu Glu Arg Ser Glu Glu Arg Ala Glu Val Ala
GAC TCA GAC CGC AAA TAT GAG GAG GTG GCC AGG AAG CTG GTG ATC CTG GAA GGG GAG CTG GAG CGC TCG GAA GAG AGA GCC GAG GTG GCT
469                                                                                                                 558

            190                                     200                                     210
Glu Ser Lys Cys Gly Asp Leu Glu Glu Glu Leu Lys Ile Val Thr Asn Asn Leu Lys Ser Leu Glu Ala Gln Ala Asp Lys Tyr Ser Thr
GAG AGT AAA TGT GGG GAC CTA GAG GAG GAG CTG AAA ATT GTT ACC AAC AAC TTG AAA TCC CTG GAA GCC CAA GCG GAC AAG TAT TCC ACC
559                                                                                                                 648

            220                                     230                                     240
Lys Glu Asp Lys Tyr Glu Glu Glu Ile Lys Leu Leu Glu Glu Lys Leu Lys Glu Ala Glu Thr Arg Ala Glu Phe Ala Glu Arg Ser Val
AAA GAG GAC AAA TAC GAA GAA GAG ATC AAA CTT CTG GAG GAG AAG CTG AAA GAG GCT GAG ACC CGA GCA GAG TTT GCT GAA AGA TCT GTG
649                                                                                                                 738

            250                                     260                                     270
Ala Lys Leu Glu Lys Thr Ile Asp Asp Leu Glu Asp Glu Val Tyr Ala Gln Lys Met Lys Tyr Lys Ala Ile Ser Glu Glu Leu Asp Asn
GCA AAG TTG GAG AAA ACC ATT GAT GAT CTG GAA GAT GAA GTC TAT GCA CAG AAG ATG AAG TAC AAG GCC ATC AGC GAG GAG CTG GAC AAC
739                                                                                                                 828

            280             284
Ala Leu Asn Asp Ile Thr Ser Leu END
GCA CTC AAT GAC ATC ACT TCC CTC TGA GTCCCCACTGTGGCCATCGCAGGCCCCTTTCTCTCCTCTCTTTACATTCTCTCTGAGGGGAGGGGGCAGCCAGGAGGAGCACC
829                         852                                                                                     938

AGCTAGCCACGTTGCACAGCCAGCCTGAGGGCAGCCTGAGGAACACGTCTGCCACCCCTGCCACCCAC
939                                                                1006
```

3.3 Expression of β-tropomyosin *in vivo*

Total RNA was isolated from liver, brain, skeletal muscle and cardiac muscle, subjected to electrophoresis through denaturing formaldehyde agarose gels and then transferred to nitrocellulose or nylon membranes (section 2.7.2). In order to determine the size of the β-tropomyosin mRNA the 450bp *Pst* I fragment of clone pCM202 was nick-translated and hybridised to the immobilised RNA. The resulting autoradiograph is presented in Figure 3.22. Several different species of RNA hybridised to this probe. Two of them occur in RNA isolated from skeletal muscle tissue and have lengths of approximately 1.3kb and 2.4kb, whereas one, possibly two, species are detected in the RNA isolated from cardiac muscle and have lengths of approximately 1.2kb. No tropomyosin mRNA was detected in the RNA isolated from either liver or brain, even when the amount loaded onto the gel was increased to ten times that which was loaded in the skeletal and cardiac tracks. In order to determine which of the mRNA species detected by the *Pst* I fragment of pCM202 corresponded to the β-tropomyosin isoform a probe which was considered to be specific for β-tropomyosin was prepared. This was done by digesting clone pCM212 with the restriction endonucleases Bgl II and Bgl I. This resulted in a fragment of DNA containing nucleotides 732-1008 of the β-tropomyosin sequence together with the G/C cloning tail and approximately 150 nucleotides of the plasmid vector pBR322. This fragment of DNA was nick-translated and hybridised to RNA from the four different tissues. The resulting autoradiograph is presented in Figure 3.23. Only the 1.3kb tropomyosin mRNA species in skeletal muscle RNA was detected by this

Figure 3.22 Hybridisation of the Pst I fragment of clone pCM202 with total RNA

Total RNA was isolated from skeletal muscle, cardiac muscle, liver and brain. 1 μg (10 μg for liver and brain) was electrophoresed through a 0.7% agarose gel containing formaldehyde and the RNA blotted onto nitrocellulose. The Pst I fragment of clone pCM202 was labelled with ^{32}P. Hybridisation was overnight in 50% Formamide, 5 x SSC, 5x Denhardt's, 0.1% SDS at 42°C. The filter was washed 5 times for 10 minutes in 2 x SSC, 0.1% SDS and then autoradiographed. Tracks 1-4 contain, from left to right, skeletal muscle RNA, liver RNA, cardiac muscle RNA, brain RNA. Molecular weight size markers are indicated.


9.5
6.6
4.3
2.3
2.0
2.4
1.3
1.25
1 2 3 4 5

probe, although hybridisation signals were also detected in liver and brain RNA. However these bands correspond to the 18S and 28S ribosomal RNA fraction, and probably arise as a result of the high G/C content of these RNAs and the presence of the G/C cloning tail contained within the probe. The fact that similar bands were not found in the skeletal and cardiac RNA can be explained in terms of the lower concentration of RNA applied to these tracks. The conclusion that β-tropomyosin has an mRNA of length 1.3kb is consistent with the results reported by McLeod et al 1985 for the human β-tropomyosin mRNA.

To determine the pattern of tropomyosin expression throughout the development of skeletal muscle, a skeletal muscle cell line (C2C12) was examined. This cell line can be cultured as individual myoblasts at a high concentration of serum but can be induced to fuse into post-mitotic myotubes if the concentration of serum is decreased. RNA was isolated from these cells at five different stages of their development (section 2.4.3.2). Figure 3.24 presents photographs of the cells at the individual myoblast stage, the final fused myotube stage and an intermediate stage taken 16 hours after reduction of the serum concentration where the cells can be seen to be aligning themselves in preparation for fusion. The RNA isolated from each developmental stage was separated on a denaturing formaldehyde agarose gel and transferred to a nitrocellulose membrane. The 400bp Pst I fragment of clone pCM202 was then hybridised to the immobilised RNA. Figure 3.25 presents the resulting autoradiograph. No tropomyosin mRNA species were detected in either of the samples of RNA isolated from the cells grown in

Figure 3.23 Hybridisation of the β-tropomyosin-specific probe with total RNA

Total RNA was isolated from skeletal muscle, cardiac muscle, liver and brain. 1 μg (10 μg for liver and brain) was electrophoresed in a 0.7% agarose gel containing formaldehyde and the RNA blotted onto nitrocellulose. The filter was hybridised overnight at 42°C in 50% formamide, 5 x SSC, 5 x Denhardt's and 0.1% SDS. After hybridisation the filter was washed in 2 x SSC, 0.1% SDS at room temperature 5 times for 10 minutes and autoradiographed. Tracks 1-5 contain, from left to right ^{35}S-labelled DNA digested with Hind III, skeletal muscle RNA, liver RNA, cardiac muscle RNA and brain RNA. Note the bands corresponding to 18S and 28S rRNA discussed in the text. One single tropomyosin species is seen in skeletal muscle, corresponding to a mRNA of 1.3kb in length.

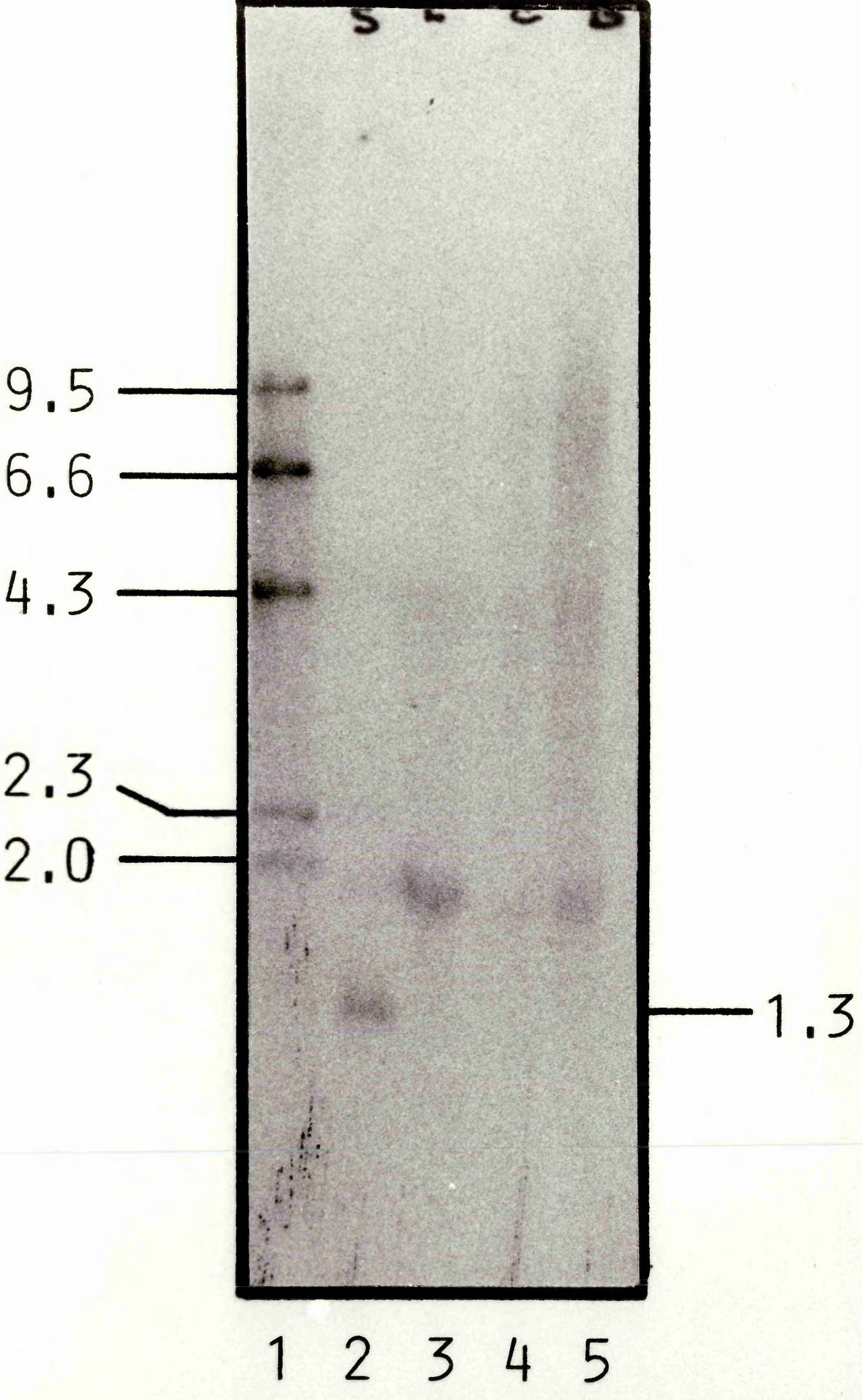
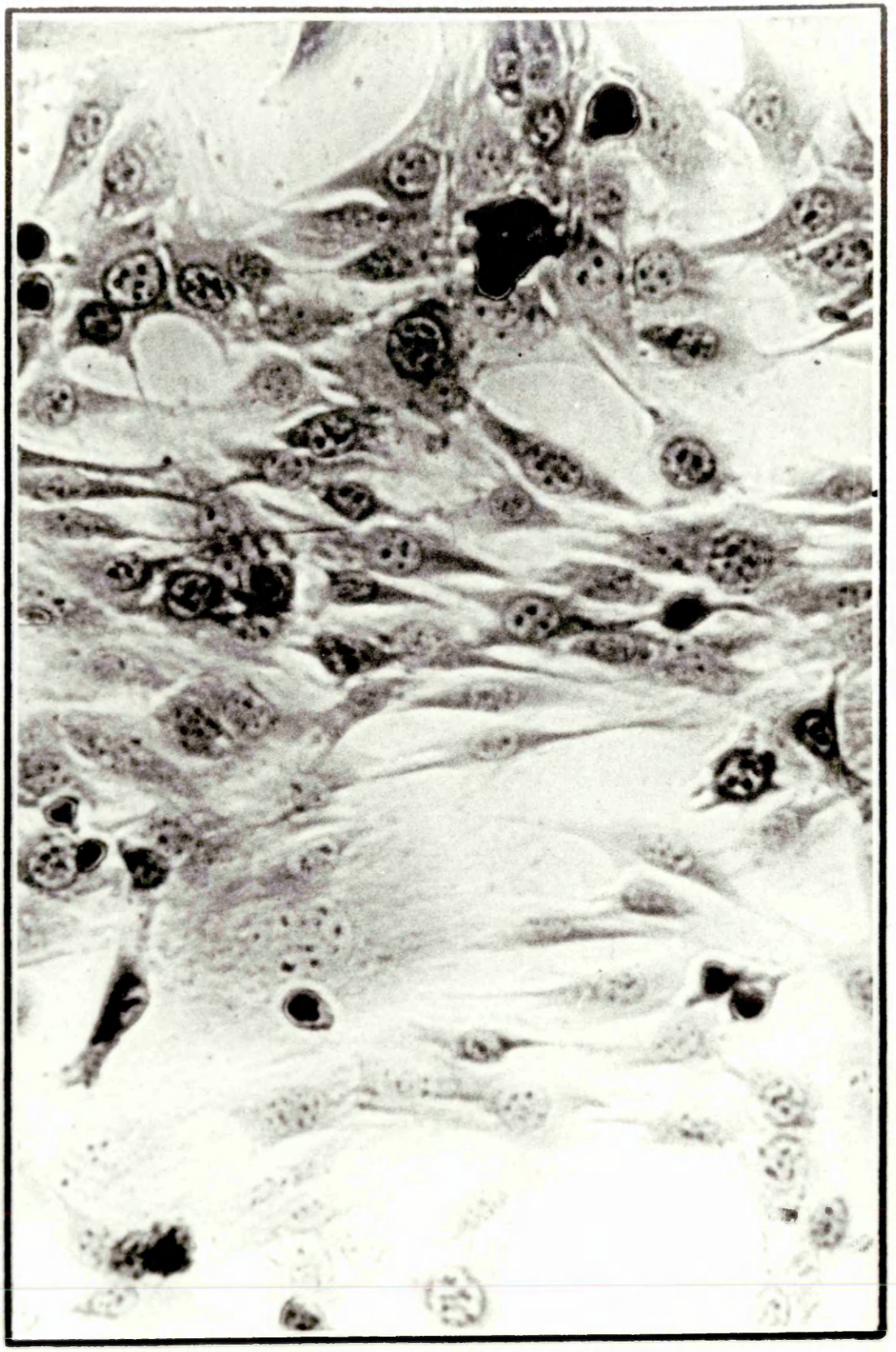

S L C B
9.5
6.6
4.3
2.3
2.0
1.3
1 2 3 4 5

Figure 3.24 Differentiating myoblasts in culture

(a) Individual myoblasts cultured in 20% foetal calf serum.
(b) The morphology of the cells 16 hours after the reduction of the serum concentration to 2%. Note that the cells can be seen to be aligning themselves in preparation for fusion.
(c) The morphology of the cells 64 hours after reduction of the serum concentration. Note the multinucleate nature of the cells and in particular the long muscle fibre in the middle part of the photograph.

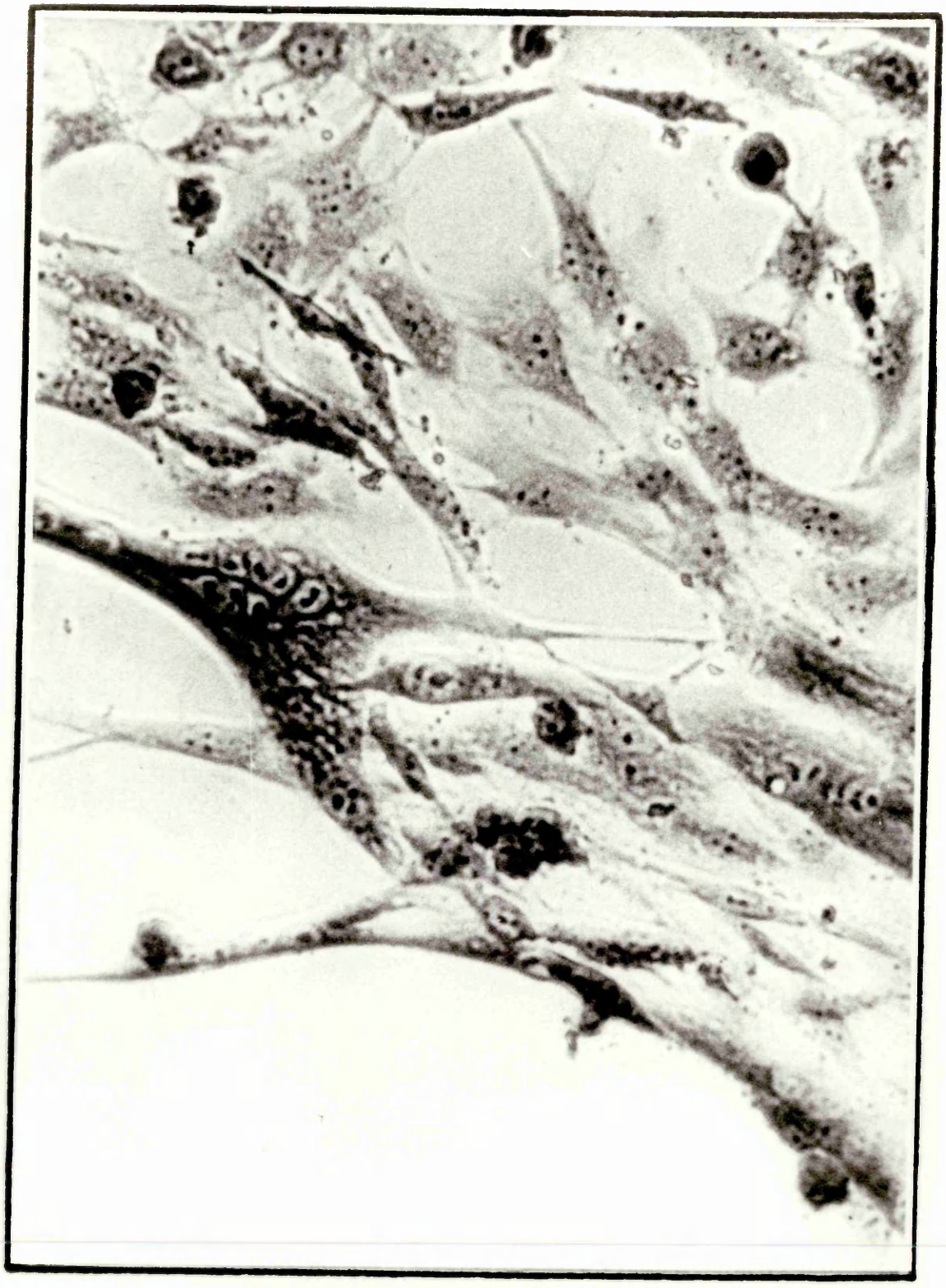
-24 hours

(b)



+16 hours

(c)



+64 hours

high concentrations of serum (tracks 1 and 2). However in the first sample taken after reduction of the serum concentration a 1.2kb species of mRNA is clearly detectable. The 1.3kb species is not present whereas the 2.4kb species is just apparent. RNA samples taken 24 hours later clearly show all three species of mRNA and a further high molecular weight species (approximately 2.6kb in length), as do the samples taken after a further 24 hours. No further samples were taken as the cells at this stage were fully fused and thereafter began to die, as is typical of fused myoblasts in culture (Blau et al 1983).

A control experiment performed with a (^{32}P)-labelled cDNA probe for actin, from D.P. Leader, (Figure 3.26) clearly shows actin mRNA species in the RNA samples taken at all five timepoints and therefore it is evident that induction of the tropomyosin mRNAs occurs only after fusion of the myoblasts.

Figure 3.25 Hybridisation of the β-tropomyosin coding probe with RNA isolated at various stages of muscle cell development

Total RNA was isolated from C2C12 cells, at five stages in their development, separated on a 0.7% agarose gel containing formaldehyde and blotted onto nitrocellulose. Hybridisation with the (^{32}P)-labelled Pst I fragment of clone pCM202 was overnight in 50% formamide. 5 x SSC, 5 x Denhardt's and 0.1% SDS. The filter was washed in 2 x SSC and 0.1% SDS for 5 times 10 minutes at room temperature and then autoradiographed. Tracks 1-7 contain, from left to right, skeletal muscle RNA, cardiac muscle RNA, RNA isolated from C2C12 cells 24 hours before reduction of serum, 1 hour before reduction of serum, 16 hours after reduction of serum, 40 hours after reduction of serum and 64 hours after reduction of serum.

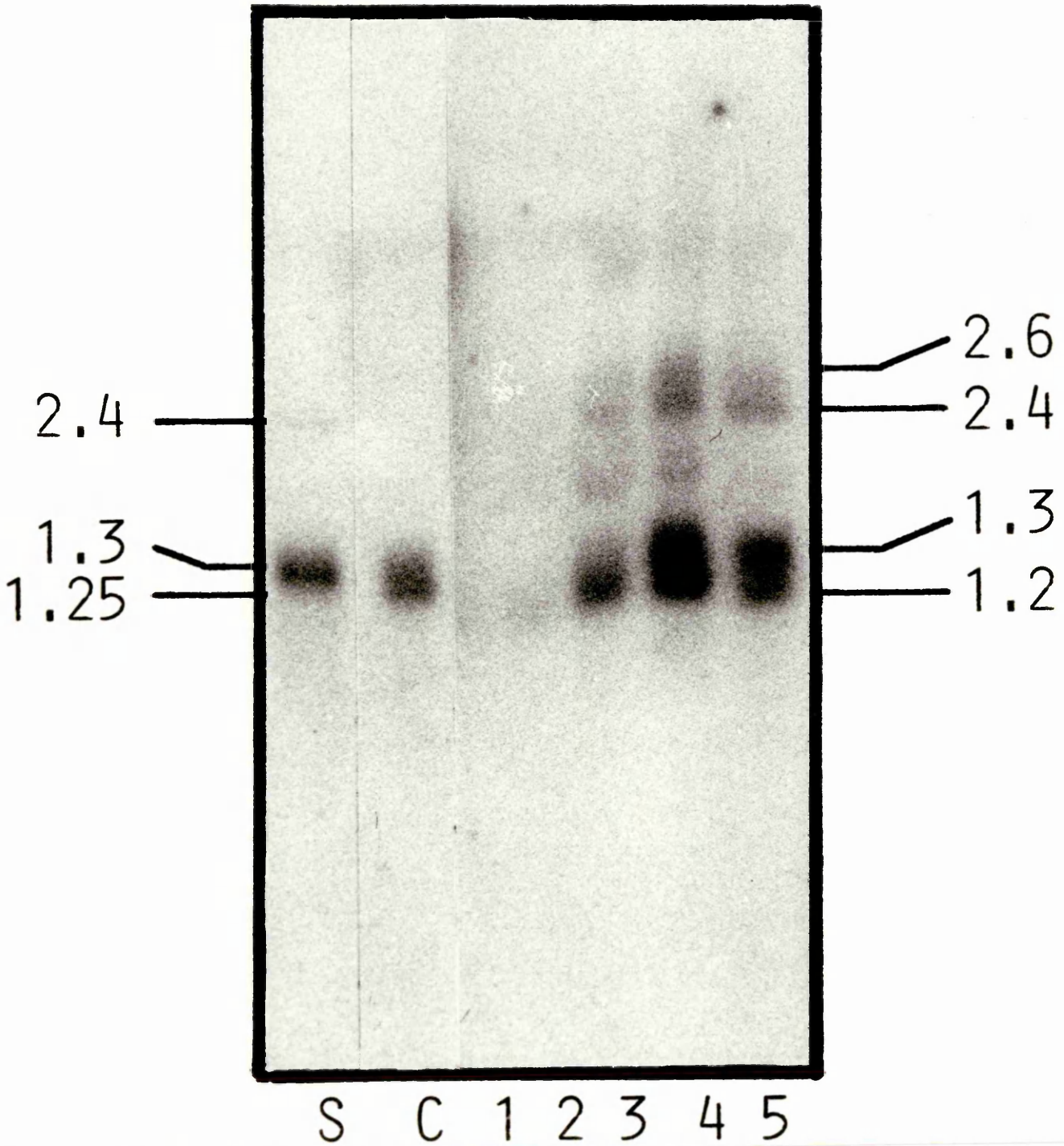
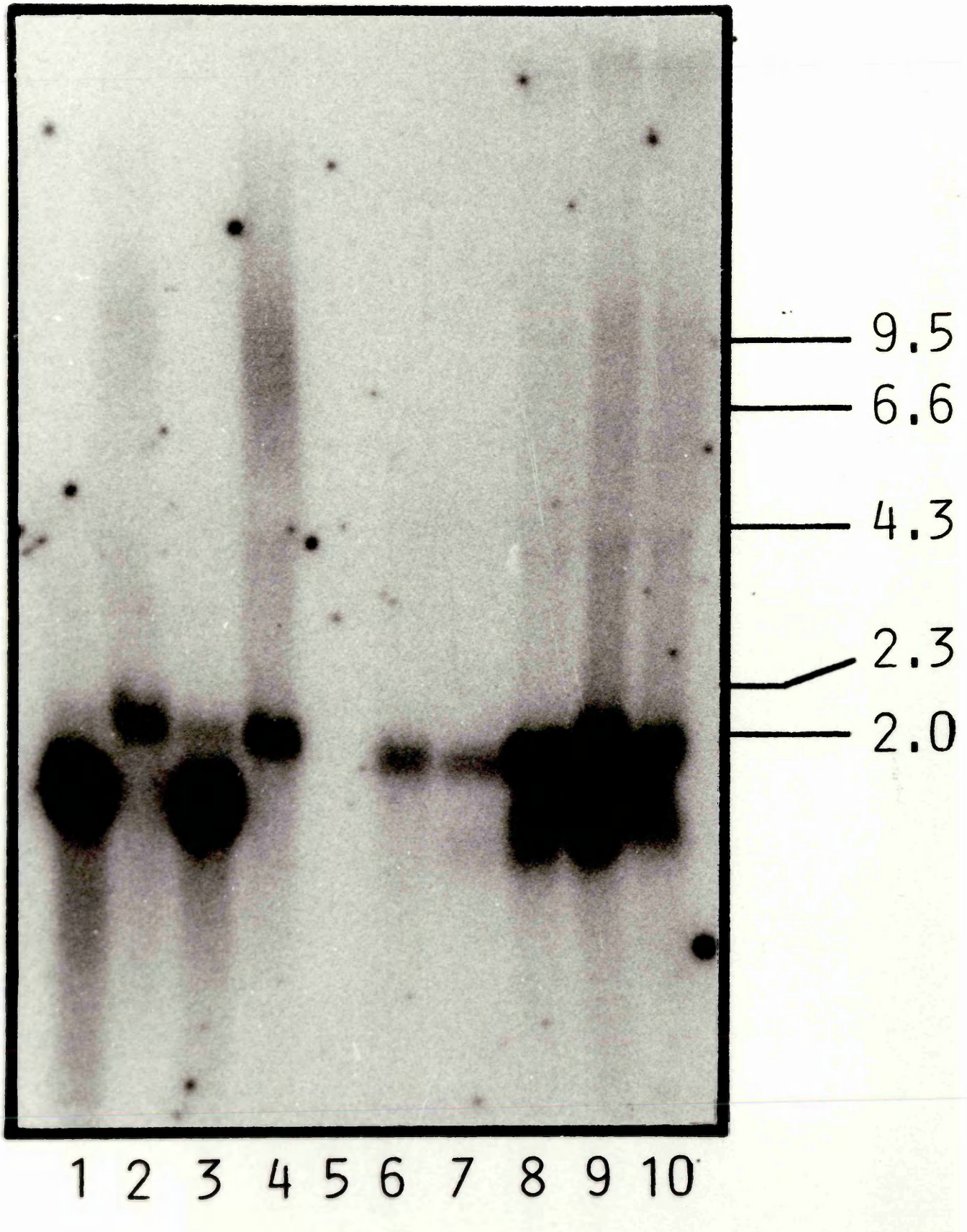

2.4
1.3
1.25
2.6
2.4
1.3
1.2
S C 1 2 3 4 5

Figure 3.26 Hybridisation of an actin cDNA probe with RNA isolated at various stages of muscle cell development

Total RNA was isolated from C2C12 cells at different stages of their development, separated on a 0.7% agarose gel containing formaldehyde and transferred onto a nitrocellulose filter. Hybridisation with the (^{32}P)-labelled actin cDNA probe (from D.P. Leader, University of Glasgow) was overnight in 50% formamide, 5 x SSC, 5 x Denhardt's and 0.1% SDS. After hybridisation the filter was washed in 2 x SSC and 0.1% SDS for 5 times 10 minutes at room temperature and then for 2 times 30 minutes at 42°C in 0.1 x SSC, 0.1% SDS. The filter was allowed to dry and then autoradiographed. Lanes 1-10 contain, from left to right, skeletal muscle RNA, cardiac muscle RNA, liver RNA, brain RNA, λ/Hind III DNA size markers, and C2C12 RNA isolated 24 hours before reduction of serum, 1 hour before reduction of serum, 16 hours after reduction of serum, 40 hours after reduction of serum and 64 hours after reduction of serum.


9.5
6.6
4.3
2.3
2.0
1 2 3 4 5 6 7 8 9 10

CHAPTER 4

Discussion

4.1 Isolation of cDNAs specific for skeletal muscle

The initial objective of this work was to isolate a cDNA clone corresponding to a mRNA expressed in skeletal muscle but not in cardiac muscle. One set of cDNA clones isolated which satisfy this description are those corresponding to the mRNA encoding β-tropomyosin.

4.1.1 β-tropomyosin

Two lines of evidence support the specificity of β-tropomyosin for skeletal muscle. Firstly, although tropomyosin is a protein which is expressed ubiquitously, protein analysis has detected the β-isoform only in skeletal muscle (Cummins and Perry, 1973; Fine and Blitz, 1975; Mak et al 1980). Secondly, numerous studies involving the isolation of tropomyosin cDNAs from various tissues have resulted in the isolation of the β-tropomyosin cDNA only from skeletal muscle (McLeod et al 1985). This is consistent with the results obtained in this thesis, when a (^{32}P)-labelled cDNA probe, representing β-tropomyosin coding sequence, was hybridised to "dot-blotted" mRNA isolated from skeletal muscle, cardiac muscle, liver and brain (Figure 3.4). The stringency under which the above hybridisation was performed did not allow cross-hybridisation with

the isoform of tropomyosin found in cardiac muscle which is clearly visible in track 3 of Figure 3.22, a "Northern" blot of RNA which was hydridised to the same probe, but at a much lower stringency. Furthermore a (^{32}P)-labelled cDNA probe containing the 3' non-coding region of β-tropomyosin (containing sequence which was considered to be unique to the β-isoform) selected a complementary species only in skeletal muscle (Figure 3.23). However at this stage it may be worth noting that band selected by this probe was of a lower intensity than that selected by the coding region probe. One interpretation of this may be that more than one 1.3kb species of tropomyosin mRNA is present in skeletal muscle. This would be consistent with the findings of McLeod and Gooding (1988) who reported that two α-isoforms (α_1 and α_2) both with an mRNA of 1.3kb in length are found in adult human skeletal muscle.

The Northern blot of RNA isolated from skeletal muscle, cardiac muscle, liver and brain (Figure 3.22) which was hybridised to the tropomyosin coding region probe at low stringency also revealed at least two other species of RNA. These presumably correspond to other isoforms of tropomyosin present in cardiac muscle and skeletal muscle but which do not correspond to the 1.3kb mRNA of β-tropomyosin.

In particular, in skeletal muscle there is a second, minor but quite clear, species of approximately 2.4kb in length. This does not correspond to any species of tropomyosin previously reported to be present in adult skeletal muscle. In fact the only tropomyosin mRNA species of approximately this size so far reported is the 2.5kb

fibroblast mRNA formed by the alternative splicing of the transcript which also gives rise to αTM2, one of the α-isoforms previously reported to be present in skeletal muscle (McLeod et al, 1986; McLeod and Gooding 1988). Interestingly, however, McLeod and Gooding (1988) do report that this 2.5kb species is present in mRNA isolated from foetal skeletal muscle, even although it is absent from the mature tissue.

There is also at least one mRNA species in cardiac muscle which cross-hybridises with the β-tropomyosin coding region probe. This diffuse band corresponding to a mRNA of approximately 1.2kb in length is presumed to contain the two α-tropomyosin isoforms reported by McLeod and Gooding (1988). Only isoform specific probes, however, would help to resolve this.

The expression of tropomyosin mRNAs was next examined during the differentiation of the mouse myoblast cell line C2C12. In prefusion myoblasts no tropomyosin mRNA species were detectable, but upon fusion both the 1.3kb species and the 2.4kb species were induced (Figure 3.25). The 2.4kb mRNA appeared to be expressed slightly before the 1.3kb mRNA and its abundance relative to the latter was greater than that found in the mRNA isolated from skeletal muscle of 11 day old mice (Figure 3.22). Thus its presence in the later may be due to its persistence from an earlier developmental stage. Furthermore if its identity is that of the fibroblast mRNA reported by McLeod et al (1986) this poses the question of its function in skeletal muscle. Since it appears only after fusion of the myoblasts and is clearly present in the mature

tissue, the possibility exists that it may also be involved with the sacromere along with the β-isoform. Studies, in vivo, with an isoform-specific antibody, for example, would help to elucidate this.

It would also appear that other tropomyosin mRNA species, not present in 11 day old leg muscle, are also expressed in the fused myoblasts. As equally abundant as, and appearing before the 1.3kb species is one of approximately 1.2kb in length. Comparison of this with the 1.2kb species of cardiac muscle resolved on the same gel indicates clearly that these two species are different. The only other species of this size previously reported is another non-muscle isoform with a mRNA of 1.1kb, the alternative splicing product of the primary transcript which also gives rise to the rat β-tropomyosin mRNA (Helfman et al 1986). In addition a second high molecular weight species (approximately 2.6kb in length) is detected after fusion of the myoblasts. This, however, does not correspond to anything previously described and may perhaps represent a separate isoform or else be a variant of one of the other spec ies poly-adenylated at a different site.

If the identities of the 1.2kb and 2.4kb myoblast species are as suggested it would appear that the initial induction of the mRNAs for the major tropomyosins of skeletal muscle (the β- and α_2-isoforms) is accompanied by expression of the alternatively spliced mRNAs, although this expression is subsequently suppressed. Table 4.1 briefly summarises the origin of each of the tropomyosin species discussed above.

Positive identification of each of the tropomyosin species present in differentiating myoblasts requires the use of isoform-specific probes. Nevertheless it is evident that the programme leading to the expression of tissue-specific isoforms of tropomyosin during muscle differentiation involves the transient expression of several other species. This is consistent with the expression of multiple isoforms of other muscle proteins such as actin and myosin, during differentiation of skeletal muscle (Bains et al, 1984; Meydert et al, 1987). However in the case of actin, at least, differential splicing is not involved.

Table 4.1 The genetic origins of different tropomyosin mRNAs

(McLeod and Gooding, 1988)

Muscle mRNA product (kb)	Gene	Fibroblast mRNA product
1.3	TM β	1.1
1.3 αTM.1	TM α	1.3
		2.0
1.3 αTM.2	TM_{nm}	2.5

The mRNA and protein products of three alternatively spliced tropomyosin genes are shown. This thesis proposes that each of these mRNA species, apart from the 1.3 and 2.0kb species derived from TMα, are expressed in skeletal muscle at some stage of its development.

4.1.2 Other cDNA clones specific for skeletal muscle.

None of the other cDNA clones isolated by diferential screening proved, in the end, to be specific for skeletal muscle. Little was known beforehand about what differences one might expect to find between two such closely related tissues as skeletal and cardiac muscle, other than the existence of isotypic variants of the major structural and functional proteins.

The mammalian genome has been estimated to have the potential of coding for up to 300,000 separate genes (Lewin, 1984), but each individual tissue is considered to express only a fraction of this number. Studies on the mRNA expressed in several terminally differentiated tissues have shown that each tissue expresses approximately 15-20000 genes (Young et al, 1985; Getz et al, 1976; Hastie and Bishop, 1976). Of these only about 20% are thought to be specific to the individual tissue. In addition, the mRNA expressed within a tissue falls into one of three categories, high abundance mRNA, medium abundance mRNA and low abundance mRNA. The high abundance mRNAs of which there are several thousand copies per cell constitute less than 1% of the mRNA species, the medium abundance mRNAs, present at several hundred copies per cell, a further 5-10%, but the great majority of mRNA species are low abundance mRNAs, which are present at only 1-2 copies per cell. It is thought that the majority of mRNA species specific to a particular tissue fall within the category of low abundance mRNAs.

It is quite clear from the initial screening of the cDNA

library by the method of Grunstein and Hogness (section 2.8.1) that the technique was not sensitive enough to detect any but the most abundant mRNA species. Thus any differences detected between skeletal and cardiac muscle cDNA clones would necessarily involve these more abundant mRNAs. Furthermore the similarity between cardiac and skeletal muscle isoforms in many cases would be expected to lead to cross hybridisation. This then probably explains why only one skeletal muscle-specific cDNA was selected. Table 4.2 lists the possible classes of proteins corresponding to the other cDNAs in the clones selected during the screening process which turned out not to be specific for skeletal muscle.

Table 4.2 Possible classes of cDNAs isolated during the screening of the mouse skeletal muscle library

Seven of the eight cDNA clones subjected to the final screening procedure involving hybridisation to poly(A)$^+$ RNA turned out not to be specific for skeletal muscle. The table below indicates the possible classes of proteins which these seven clones may represent, based on the hybridisation pattern produced when each was hybridised to skeletal muscle, cardiac muscle, liver and brain poly(A)$^+$ RNA.

Clone	Skeletal muscle	Cardiac muscle	Liver	Brain	Possible protein classification
9	++	++	-	-	Muscle eg troponin
21	++	++	++	++	Metabolic
33	+++	+++	+++	+++	Metabolic
40	+	+	-	+++	Neuromuscular eg acetylcholine receptor
44	+++	++	+	-	Metabolic
215	+	+	++	++	Metabolic
239	+++	+++	+	+	Muscle eg actin

4.2 Significance of the mouse β-tropomyosin cDNA sequence.

The mouse β-tropomyosin cDNA sequence isolated during the course of this work is almost full length. 1.1kb of sequence was determined from a total of 1.3kb as estimated from Northern blot analysis (Figure 3.22). The sequence, obtained from overlapping cDNA clones, contained all 852 nucleotides of the coding region together with 87 (probably 95) nucleotides of the 5' non-coding region and 151 nucleotides of the 3' non-coding region. The only β-tropomyosin cDNA described before this work was that of human, which, although complete at the 3' end lacked the 5' non-coding sequence together with the sequence encoding the first 152 amino-acids (McLeod et al, 1985). Comparison of the 3' non-coding region of the mouse β-tropomyosin cDNA with that of the human (Figure 4.1), suggests that the mouse cDNA is approximately 80 nucleotides short at the 3' end.

When McLeod et al (1985) compared the partial human β-tropomyosin cDNA sequence with that of TM_{36}, a 284 amino-acid human fibroblast tropomyosin they found that the regions encoding amino-acids 153 to 187 and 214 to 257 shared complete (100%) nucleotide identity and thus proposed that the β- and TM_{36}-isoforms arose via the differential splicing of the same primary transcript. Thus two regions encoding amino-acids 188 to 213 and 258 to 284 must be contained within alternatively spliced exons. This is analogous to the differential splicing found for the α_2- and the $TM30_{nm}$- isoforms of tropomyosin, but in that particular case a further region involving the 5' end of the tropomyosin was

Figure 4.1 Comparison of mouse β-tropomyosin cDNA with human β- and TM_{36}-tropomyosin cDNAs

A comparison of the mouse β-tropomyosin cDNA sequence with that of human β-tropomyosin and the human TM_{36}-tropomyosin cDNA sequences is shown. The initiator methionine and termination codons are indicated. Identities with the mouse sequence are indicated by dashes. Where the sequences differ, the replacement nucleotide is given. Note the gaps introduced into the 5' and 3' non-coding regions to give the best match. Note also that the 3' non-coding region of the TM_{36} isoform bears little relation to those of both β-isoforms and therefore little matching has been performed. Two regions corresponding to amino-acids 188 to 213 and 258 to 284 show the greatest sequence divergence between the β-isoforms and the TM_{36}-isoform. These regions are referred to in the text.

```
                                CTCCTGACCTGTCCCCTGTGCATCCAGA GCTCGCG...Mouse Beta
                                -----CG----C-A--G-----C----TC----A-C...Human Cytoplasmic
                                                                  Met
CCTAGCCCAGTCTACCTACTCTTCACCGCTCTCCCGCCAGCCCACTCCCCACCGCAGCC ATG...Mouse Beta
-  ---------CGTCCGG--C--------CTG--G---G------C-----------GCC---...Human Cytoplasmic

GAC GCC ATC AAG AAG AAG ATG CAG ATG CTG AAA CTG GAC AAG GAG AAT...Mouse Beta
--- --- --- --- --- --- --- --- --- --- --G --- --- --- --- --C...Human Cytoplasmic

GCC ATC GAC CGC GCG GAG CAG GCC GAA GCC GAC AAA AAG CAA GCT GAA...Mouse Beta
--- --- --- --- --C --- --- --- --- --- --- --G --- --- --- --G...Human Cytoplasmic

GAC CGA TGC AAG CAG CTG GAG GAA GAG CAG CAG GCC CTC CAG AAG AAG...Mouse Beta
--- --C --- --- --- --- --- --G --- --- --- --- --- --- --- ---...Human Cytoplasmic

CTG AAG GGG ACA GAG GAC GAG GTG GAA AAG TAT TCC GAG TCC GTG AAG...Mouse Beta
--- --- --- --- --- --T --- --- --- --- --- --T --A --- --- ---...Human Cytoplasmic

GAT GCC CAG GAG AAA CTG GAG CAG GCT GAG AAG AAG GCC ACC GAC GCT...Mouse Beta
--G --- --- --- --- --- --- --- --C --- --- --- --- --T --T ---...Human Cytoplasmic

GAA GCA GAT GTG GCC TCT CTG AAC CGC CGC ATT CAG CTC GTA GAG GAG...Mouse Beta
--G --- --- --- --- --C --- --- --- --- --- --- --G --T --- ---...Human Cytoplasmic

GAG TTG GAT CGG GCA CAG GAG CGC CTG GCT ACA GCC TTG CAA AAG CTG...Mouse Beta
--- C-- --C --- --C --- --- --- --- --- --- --- C-- --G --- ---...Human Cytoplasmic

GAG GAG GCT GAG AAA GCC GCG GAT GAG AGC GAG AGA GGA ATG AAG GTC...Mouse Beta
--- --- --C --- --G --G --T --- --- --- --- --- --- --- --- ---...Human Cytoplasmic

ATT GAA AAC CGG GCC ATG AAG GAT GAG GAA AAG ATG GAG CTG CAG GAG...Mouse Beta
--C --- --- --- --- --- --- --- --- --G --- --- --A --- --- ---...Human Cytoplasmic

ATG CAG CTG AAG GAA GCC AAG CAC ATC GCT GAG GAC TCA GAC CGC AAA...Mouse Beta
--- --- --- --- --G --- --- --- --- --- --- --T --- --- --- ---...Human Cytoplasmic
                          -- --- --- --- --T --- --- --- ---...Human Beta

TAT GAG GAG GTG GCC AGG AAG CTG GTG ATC CTG GAA GGG GAG CTG GAG...Mouse Beta
--- --A --- --- --- -A- --- --- --- --- --- --- --A --- --- ---...Human Cytoplasmic
--- --A --- --- --- -A- --- --- --- --- --- --- --A --- --- ---...Human Beta

CGC TCG GAA GAG AGA GCC GAG GTG GCT GAG AGT AAA TGT GGG GAC CTA...Mouse Beta
--- --- --G --- --G --T --- --- --C --- --- --C CG- GCC A-A C-G --G...Human Cytoplasmic
--- --- --G --- --G --T --- --- --C --- --- --- --- --- --- ---...Human Beta

GAG GAG GAG CTG AAA ATT GTT ACC AAC AAC TTG AAA TCC CTG GAA GCC...Mouse Beta
--- --- --A --T CG- -CC A-G GA- C-G GC- C-C --G T-- --- ATG ---...Human Cytoplasmic
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --G ---...Human Beta

CAA GCG GAC AAG TAT TCC ACC AAA GAG GAC AAA TAC GAA GAA GAG ATC...Mouse Beta
TCA -A- --G G-- --- --- --- --- --A --T --- --T --- --G --- ---...Human Cytoplasmic
--G --- --- --- --- --- --- --- --A --T --- --T --- --G --- ---...Human Beta

AAA CTT CTG GAG GAG AAG CTG AAA GAG GCT GAG ACC CGA GCA GAG TTT...Mouse Beta
--- --G T-- --- --- --- --- --G --- --- --- --- --- --- --- ---...Human Cytoplasmic
--- --G T-- --- --- --- --- --G --- --- --- --- --- --- --- ---...Human Beta

GCT GAA AGA TCT GTG GCA AAG TTG GAG AAA ACC ATT GAT GAT CTG GAA...Mouse Beta
--C --G --G --- --- --- --- --- --- --- --- --C --- --C --A ---...Human Cytoplasmic
--C --G --G --- --- --- --- --- --- --- --- --C --- --C --A ---...Human Beta

GAT GAA GTC TAT GCA CAG AAG ATG AAG TAC AAG GCC ATC AGC GAG GAG...Mouse Beta
--G ACC T-G GCC AGT GCC --- GA- G-- A-- GTC -AG --T CA- C-- ACC...Human Cytoplasmic
--- --- --- --- --C --- --- --- --- --- --- --- --T --- --- --A...Human Beta
                                          End
CTG GAC AAC GCA CTC AAT GAC ATC ACT TCC CTC TGA GTCCCCAC     TG...Mouse Beta
T-- --- C-G A-C --G CTG --A C-- -AC AA- --G TGA -GG ---GCCCCACC...Human Cytoplasmic
--- --- --- --- --- --- --- --- --C --- --- --- - ------GCCAGC-...Human Beta

TGGCCATCGCAGGCCCCTTTCTCTCCTCTC TTTACATTCTCTCTGAGGGGAGGGG GCAGCC...Mouse Beta
CCCAGCCA-GCTATGGT-GC-A-C--AAC-CCCA-T-AAACTGA--TTACT--CCTCTCpolyA..Human Cytoplasmic
-- ---C-T--- -T-TC----------C---C----------AT---------A----G-...Human Beta

AGGAGGAGCACCAGCTAGCCACGTTGCACAGCCAGCCTGAGGGCAGCCTGAGGAACACGTCTG...Mouse Beta
----------GA-AT-GC-A--A----------G---G-A-------AG---GAG-CC-CA...Human Beta

CCACCCCTGCCACCCAC                                              ...Mouse Beta
T--TG--CA--------TCTGGCACTGGCTTCATCCTTTACCTATCCCCTTCCACCCTCCTTT...Human Beta

GGTTGCTTAATAAATTCTGAACTTGGpolyA                                ...Human Beta
```

also found to be subject to alternative splicing (McLeod et al, 1986).

The mouse β-tropomyosin cDNA sequence thus made it possible to compare the 5' regions of the β- and TM_{36}-isoforms to see if the identity between them extends over the remaining sequence (encoding amino-acids 1-152) not covered by McLeod et al (1985).

The result of the comparison (Figure 4.1) indicated that the nucleotide sequence corresponding to the first 152 amino-acids showed no less homology between the mouse β-tropomyosin and human TM_{36}-isoform (93.5% identity) than did the mouse and human β-tropomyosin isoforms over the region corresponding to amino-acids 153-284 (93% identity). Furthermore those differences which did exist in the 5' region (encoding amino-acids 1-152) between the mouse β- and human TM_{36}-isoforms were scattered randomly and not clustered like those in the regions encoding amino-acids 188 to 213 and 258 to 284. Thus it would appear that the β- and TM_{36}-isoforms do share common exons in the 5' region of the gene and that the only exons subject to alternative splicing are those previously reported by McLeod et al (1985).

While the final part of the mouse β-tropomyosin clone was being determined Helfman et al (1986) published the approximately full length cDNA sequences of the rat β-tropomyosin and TM_{36}-tropomyosin isoforms. Comparison of these two isoforms also accords with the view that both isoforms arise via the alternative splicing of the same primary transcript, with only the two exons, encoding amino-

acids 188 to 213 and 258 to 284 being subject to alternative splicing.

Although Helfman et al (1986) did not have the complete 5' non-coding region of the β-tropomyosin cDNA, primer extension and S1 nuclease experiments allowed its deduction. Comparison of this deduced sequence with the 5' non-coding sequence of the mouse β-tropomyosin cDNA indicated that 95 nucleotides of the mouse 5' non-coding region had been cloned (Figure 4.3) and that it lacked only the terminal 16 bases of the 5' end.

It is clear from the comparison of β-tropomyosin amino-acid sequences, presented in Figure 4.2, that β-tropomyosin is very highly conserved between species. The only differences to be found between the amino-acid sequences of mouse, rat, human and rabbit β-tropomyosin is that the human sequence (as predicted from the human TM_{36} sequence) has a Glu residue at position 66 whereas the other species have an Asp residue at this position. This difference, however, is a conservative substitution and as such should not dramatically affect the properties of tropomyosin.

The availability of the human and rat β-tropomyosin cDNA sequences made it possible to compare them with the mouse sequence (Figure 4.3) to see to what extent the 5' and 3' non-coding sequences of β-tropomyosin have been conserved between species. In order to do this, the most useful frames of reference are the neutral positions in the coding regions of the cDNAs. To obtain a value for the number of neutral mutations which have occurred since

Figure 4.2 Comparison of the deduced amino-acid sequences of mouse, human, rat and rabbit β-tropomyosin

A comparison of the deduced mouse β-tropomyosin amino-acid sequence with that of human, rat and rabbit is shown. Identities with the mouse sequence are indicated by dashes. Note the only difference between the four species is that human β-tropomyosin has a glutamic acid residue at position 66.

```
        10         20         30         40         50
MDAIKKKMQM LKLDKENAID RAEQAEADKK QAEDRCKQLE EEQQALQKKL....Mouse
---------- ---------- ---------- ---------- ----------....Human
---------- ---------- ---------- ---------- ----------....Rat
---------- ---------- ---------- ---------- ----------....Rabbit

        60         70         80         90        100
KGTEDEVEKY SESVKDAQEK LEQAEKKATD AEADVASLNR RIQLVEEELD....Mouse
---------- -----E---- ---------- ---------- ----------....Human
---------- ---------- ---------- ---------- ----------....Rat
---------- ---------- ---------- ---------- ----------....Rabbit

       110        120        130        140        150
RAQERLATAL QKLEEAEKAA DESERGMKVI ENRAMKDEEK MELQEMQLKE....Mouse
---------- ---------- ---------- ---------- ----------....Human
---------- ---------- ---------- ---------- ----------....Rat
---------- ---------- ---------- ---------- ----------....Rabbit

       160        170        180        190        200
AKHIAEDSDR KYEEVARKLV ILEGELERSE ERAEVAESKC GDLEEELKIV....Mouse
---------- ---------- ---------- ---------- ----------....Human
---------- ---------- ---------- ---------- ----------....Rat
---------- ---------- ---------- ---------- ----------....Rabbit

       210        220        230        240        250
TNNLKSLEAQ ADKYSTKEDK YEEEIKLLEE KLKEAETRAE FAERSVAKLE....Mouse
---------- ---------- ---------- ---------- ----------....Human
---------- ---------- ---------- ---------- ----------....Rat
---------- ---------- ---------- ---------- ----------....Rabbit

       260        270        280  284
KTIDDLEDEV YAQKMKYKAI SEELDNALND ITSL....Mouse
---------- ---------- ---------- ----....Human
---------- ---------- ---------- ----....Rat
---------- ---------- ---------- ----....Rabbit
```

Figure 4.3 Comparison of the mouse, human and rat β-tropomyosin cDNA sequences

A comparison between the mouse, human and rat β-tropomyosin cDNA sequences is shown. The human sequence was partially predicted from the human TM_{36}-tropomyosin cDNA sequence as is referred to in the text. The initiator methionine and termination codons are indicated. Identities with the mouse sequence are indicated by dashes. Where the human and rat sequences differ from the mouse the replacement nucleotide is indicated. Note the gaps introduced into the 5' and 3' non-coding regions to produce the best match.

```
                                CTCCTGACCTGTCCCCTGTGCATCCAGA GCTCGCG...Mouse Beta
                                -----CG----C-A--G-----C----TC----A-C...Human Beta
                                -----A--------- G-GTGCAT---T ----T--...Rat Beta
                                                                 Met
CCTAGCCCAGTCTACCTACTCTTCACCGCTCTCCCGCCAGCCCACTCCCCACCGCAGCC ATG...Mouse Beta
-  ---------CGTCCGG--C--------CTG--G---G------C-----------GCC---...Human Beta
G-------G---------------------G------G-------------------- ---...Rat Beta

GAC GCC ATC AAG AAG AAG ATG CAG ATG CTG AAA CTG GAC AAG GAG AAT...Mouse Beta
--- --- --- --- --- --- --- --- --- --- --G --- --- --- --- --C...Human Beta
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---...Rat Beta

GCC ATC GAC CGC GCG GAG CAG GCC GAA GCC GAC AAA AAG CAA GCT GAA...Mouse Beta
--- --- --- --- --C --- --- --- --- --- --- --G --- --- --- --G...Human Beta
--- --- --- --- --A --- --- --- --- --- --- --G --- --- --- ---...Rat Beta

GAC CGA TGC AAG CAG CTG GAG GAA GAG CAG CAG GCC CTC CAG AAG AAG...Mouse Beta
--- --C --- --- --- --- --- --G --- --- --- --- --- --- --- ---...Human Beta
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --A...Rat Beta

CTG AAG GGG ACA GAG GAC GAG GTG GAA AAG TAT TCC GAG TCC GTG AAG...Mouse Beta
--- --- --- --- --- --T --- --- --- --- --- --T --A --- --- ---...Human Beta
--- --- --- --- --- --T --- --- --- --- --- --- --- --T --- ---...Rat Beta

GAT GCC CAG GAG AAA CTG GAG CAG GCT GAG AAG AAG GCC ACC GAC GCT...Mouse Beta
--G --- --- --- --- --- --- --- --C --- --- --- --- --T --T ---...Human Beta
--- --- --- --- --- --- --- --- --C --- --- --- --- --- --T ---...Rat Beta

GAA GCA GAT GTG GCC TCT CTG AAC CGC CGC ATT CAG CTC GTA GAG GAG...Mouse Beta
--G --- --- --- --- --C --- --- --- --- --- --- --G --T --- ---...Human Beta
--- --- --- --- --T --- --- --T --A --- --- --- --- --- --- ---...Rat Beta

GAG TTG GAT CGG GCA CAG GAG CGC CTG GCT ACA GCC TTG CAA AAG CTG...Mouse Beta
--- C-- --C --- --C --- --- --- --- --- --- --- C-- --G --- ---...Human Beta
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---...Rat Beta

GAG GAG GCT GAG AAA GCC GCG GAT GAG AGC GAG AGA GGA ATG AAG GTC...Mouse Beta
--- --- --C --- --G --G --T --- --- --- --- --- --- --- --- ---...Human Beta
--- --- --- --- --- --- --T --- --- --- --- --- --- --- --- ---...Rat Beta

ATT GAA AAC CGG GCC ATG AAG GAT GAG GAA AAG ATG GAG CTG CAG GAG...Mouse Beta
--C --- --- --- --- --- --- --- --- --- --G --- --- --A --- ---...Human Beta
--- --G --- --- --- --- --- --- --- --- --G --- --- --- --- ---...Rat Beta

ATG CAG CTG AAG GAA GCC AAG CAC ATC GCT GAG GAC TCA GAC CGC AAA...Mouse Beta
--- --- --- --- --G --- --- --- --- --- --- --T --- --- --- ---...Human Beta
--- --- --- --A --- --- --- --- --- --- --- --- --- --- --- ---...Rat Beta

TAT GAG GAG GTG GCC AGG AAG CTG GTG ATC CTG GAA GGG GAG CTG GAG...Mouse Beta
--- --A --- --- --- --- --- --- --- --- --- --- --- --A --- ---...Human Beta
--C --- --- --- --- --- --- --- --- --- --- --- --- --A --- ---...Rat Beta

CGC TCG GAA GAG AGA GCC GAG GTG GCT GAG AGT AAA TGT GGG GAC CTA...Mouse Beta
--- --- --G --- --G --T --- --- --C --- --- --- --- --- --- ---...Human Beta
--- --T --- --- --- --- --- --- --- --- --- --- --- --- --- ---...Rat Beta

GAG GAG GAG CTG AAA ATT GTT ACC AAC AAC TTG AAA TCC CTG GAA GCC...Mouse Beta
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --G ---...Human Beta
--- --- --- --- --- --- --- --- --- --- --- --- --T --- --- ---...Rat Beta

CAA GCG GAC AAG TAT TCC ACC AAA GAG GAC AAA TAC GAA GAA GAG ATC...Mouse Beta
--G --- --- --- --- --- --- --- --A --T --- --T --- --G --- ---...Human Beta
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---...Rat Beta

AAA CTT CTG GAG GAG AAG CTG AAA GAG GCT GAG ACC CGA GCA GAG TTT...Mouse Beta
--- --G T-- --- --- --- --- --G --- --- --- --- --- --- --- ---...Human Beta
--- --- --- --- --- --- --- --G --- --- --- --- --- --- --- ---...Rat Beta

GCT GAA AGA TCT GTG GCA AAG TTG GAG AAA ACC ATT GAT GAT CTG GAA...Mouse Beta
--C --G --G --- --- --- --- --- --- --- --- --C --- --C --A ---...Human Beta
--- --- --- --- --- --- --- --- --- --- --- --C --- --C --- ---...Rat Beta

GAT GAA GTC TAT GCA CAG AAG ATG AAG TAC AAG GCC ATC AGC GAG GAG...Mouse Beta
--- --- --- --- --C --- --- --- --- --- --- --- --T --- --- --A...Human Beta
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---...Rat Beta
                                                 END
CTG GAC AAC GCA CTC AAT GAC ATC ACT TCC CTC TGA GTCCCCAC      T...Mouse Beta
--- --- --- --- --- --- --- --- --C --- --- --- - ------GCCCAGC...Human Beta
--- --- --- --G --- --- --- --- --- --- --- --- --T-----      -...Rat Beta

GTGGCCATCGCAGGCCCCTTTCTCTCCTCTC TTTACATTCTCTCTGAGGGGAGGGG GCAGC...Mouse Beta
--- ---C-T--- -T-TC------------C---C----------AT---------A----G...Human Beta
------------------------    ---T ---C---------------------- -----...Rat Beta

CAGGAGGAGCACCAGCTAGCCACGTTGCACAGCCAGCCTGAGGGCAGCCTGAGG AACACGTC...Mouse Beta
-----------GA-AT-GC-A--A------------G---G-A------- ---G-GAG-CC-...Human Beta
------------A-A--G----T-------------T-----------------G---------...Rat Beta

TGCCACCCCTGCCACCCAC                                          ...Mouse Beta
CAT--TG--CA--------                                          ...Human Beta
-------------------                                          ...Rat Beta
```

the divergence of mice, rats and men, the following procedure was adopted. Within the constraints of the genetic code for a conserved amino acid sequence some nucleotides are more free to mutate than others. Thus, unlike the non-coding regions, where there is theoretically no constraint on mutation, each codon must be multiplied by a factor reflecting the effective number of positions at which neutral mutation can occur. This factor is equal to 1.0 for each codon of the four codon families, such as that for Ala, 0.67 for the three codon family of Ile, 0.33 for the two codon families, such as that of Asp, and 0 for the unique codons of Met and Trp. Each codon in the six codon families of Ser, Leu and Arg was assessed separately. Using these values it was calculated that of the 852 nucleotides encoding mouse β-tropomyosin, the effective number of positions at which neutral mutations can occur is 160.

When the divergence of the effective neutral nucleotides in the coding regions of mouse and rat β-tropomyosin were calculated it was found to be 14%. This compared to a 9% divergence in the 3' non-coding region and a 10.5% divergence in the 5' non-coding region. This suggests that there has been some pressure to conserve the non-coding regions of the mRNA since if there had been no pressure then one would expect them to show a divergence at least as great as that shown by the coding regions. Similarly a comparison between the mouse and human cDNA sequences showed a 35.5% divergence in the coding region compared to a 30% divergence in the 3' non-coding region and a 27.5% divergence in the 5' non-coding region.

The extent of conservation of the 3' non-coding region of β-

tropomyosin between species is not as great as that found in the α-cardiac isoform of actin (Leader et al,1986) and in the mRNAs of some other proteins (Yaffe et al, 1985). However Gunning et al (1984) have proposed that despite divergence of the primary sequence, secondary structure motifs may be conserved. Nevertheless, analysis of the 3' non-coding region of β-tropomyosin using the computer program CINTHOM (an unpublished adaptation by P. Taylor, Institute of Virology, University of Glasgow of the diagonal matrix program of Pustell and Kafatos, 1982) has failed to find any conserved potential secondary structure. However, as will be explained in more detail later, the 5' non-coding region of β-tropomyosin contains at least two regions of sequence capable of forming some form of secondary structure with sequence elements within the coding region.

4.3 The origins of the direct and inverted repeats in the β-tropomyosin cDNAs

All the β-tropomyosin cDNA clones isolated during the course of this work contained what appear to be cloning artefacts. Two of them, pCM222 and pCM212, contained regions of invertedly repeated DNA (Figures 3.14 and 3.17, respectively) and two, pCM202/pCM232 and pCM212, contained directly repeated DNA (Figure 3.17). These two different types of artefact are thought to have arisen in different ways and therefore will be dealt with separately. The case of the inverted repeats will be discussed first.

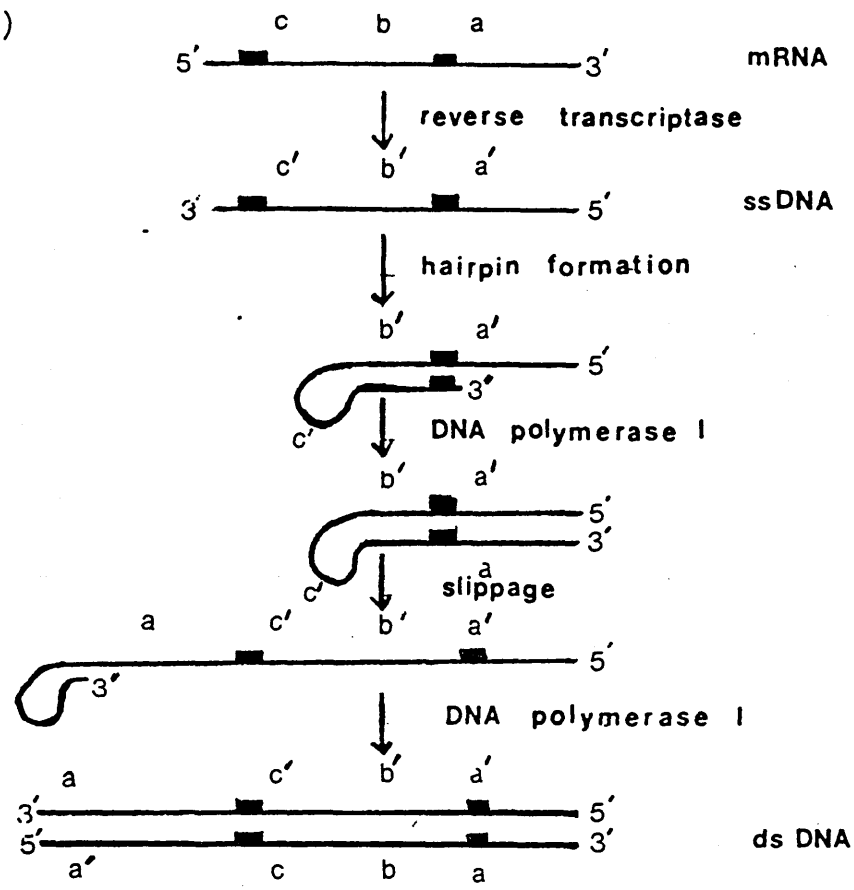
It has been proposed that regions of inverted DNA found in cDNA clones arise during the synthesis of the second strand of cDNA, and before the subsequent ligation of the cDNA to a vector (Fields and Winter, 1981; Volekaert et al, 1981). Synthesis of the second strand of cDNA using the method of Villakomeroff (1978), relies on the principle that when the RNA is removed from the reaction (usually be boiling or alkaline hydrolysis), following first strand synthesis, the 3' end of the single-stranded cDNA which is left will fold back on itself (forming a hairpin loop) to prime the synthesis of the second strand of cDNA. The second strand of cDNA is then synthesised by E.Coli DNA polymerase I. At the end of the second strand reaction the single-stranded hairpin loop is removed by S1 nuclease digestion before the double stranded cDNA is cloned into the vector. Thus the formation of almost full length cDNAs, (or those containing the 5' end of the mRNA, at least), relies on the formation of a stable hairpin loop at the extreme 3' end of the

single-stranded first strand of cDNA (presuming, of course, that a full length first strand had been synthesised). However, in many cases the extreme 3' end of the first strand of cDNA cannot form a stable hairpin loop with an adjacent region of sequence, but does so with an appropriate sequence distal to it. Inverted repeats, like those found in pCM212 and pCM222, are thought to arise when the extreme 3' end of the first strand of cDNA forms the hairpin loop structure with an appropriate sequence distal to it. However before the single-stranded loop is lost through S1 nuclease digestion one of two possible events is thought to happen (Figure 4.4). One is that the region of double-stranded cDNA denatures or "slips" so that one long single-stranded cDNA is formed once again with the extreme 3' end of this now being able to form the hairpin loop with a proximal sequence. This then primes second-strand synthesis and results in a double-stranded cDNA containing an inverted repeat immediately preceding the original sequence (Figure 4.4a). The second possibility is that the 5' — 3' exonuclease activity of DNA Polymerase I is involved in the generation of the inverted repeat. It is proposed that during second strand synthesis the original first strand of cDNA becomes nicked and denatures from the second strand, thus leaving both a region of single-stranded DNA and of double-stranded DNA. The point at which the nicking occurs then serves as a primer for complementary strand synthesis (Figure 4.4b). The reason why denaturation or "slippage" of the hairpin structure should occur is unclear, but may be as a result of the instability of the hairpin structure because of the formation of a large single-stranded loop. For example in pCM212 this would contain approximately 600 bases and in pCM222 approximately 380 bases.

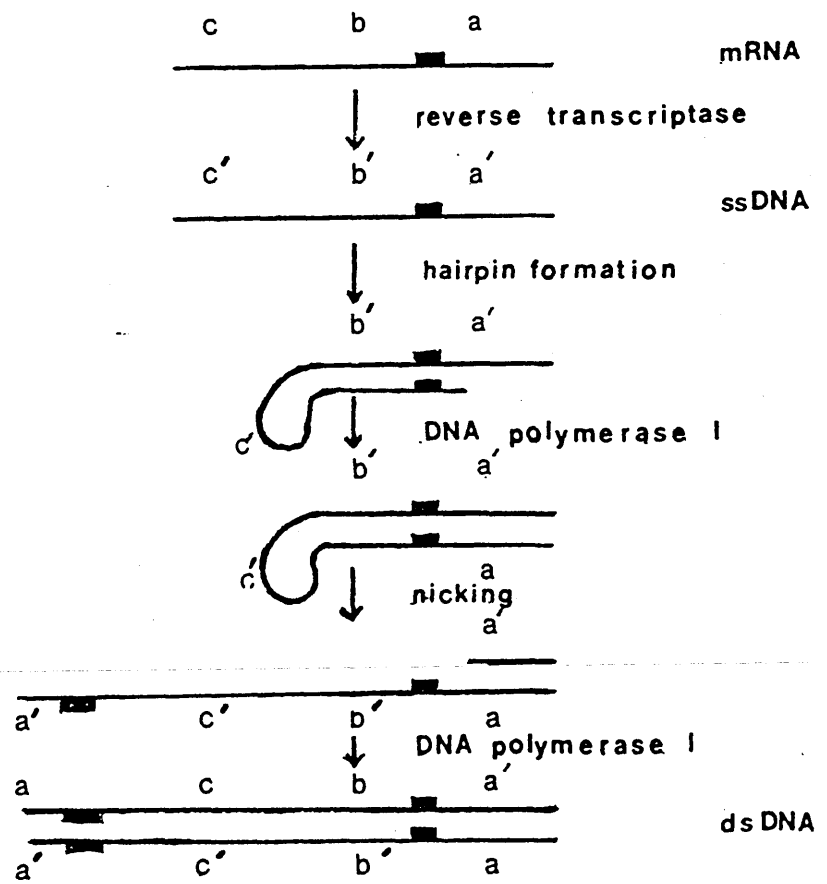
Figure 4.4 Schematic representations of the process leading to inverted repeat sequences in cDNA clones

Two theoretical schemes for the generation of inverted DNA repeats in cDNA clones are shown. (a) The inherent instability of a large hairpin loop structure may cause slippage or denaturation during second strand synthesis. (b) The 5'→3' exonuclease activity of the E.Coli DNA polymerase I has been implicated in the generation of inverted repeat sequences. Again the inherent instability of a large hairpin loop would cause unfolding of the molecule during second-strand synthesis revealing a new 3' end which is used to prime the synthesis of a new second strand.

(a)


c b a
5′ 3′
mRNA
reverse transcriptase
c′ b′ a′
3′ 5′
ssDNA
hairpin formation
b′ a′
5′
3′
c′
DNA polymerase I
b′ a′
5′
3′
c′
a
slippage
a c′ b′ a′
5′
3′
DNA polymerase I
a c′ b′ a′
3′ 5′
5′ 3′
a′ c b a
ds DNA

(b)

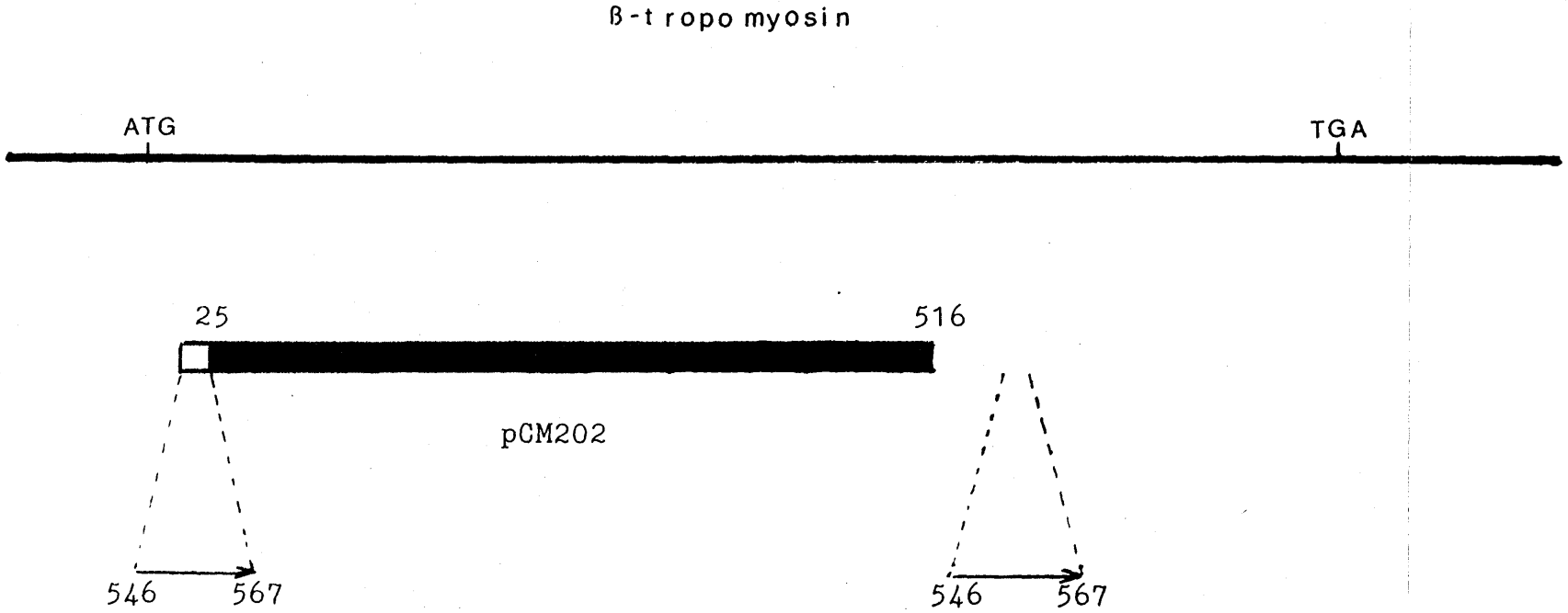

c b a
mRNA
reverse transcriptase
c′ b′ a′
ssDNA
hairpin formation
b′ a′
c′
DNA polymerase I
b′ a′
c′
a
nicking
a′
a′ c′ b′ a
DNA polymerase I
a c b a′
ds DNA
a′ c′ b′ a

Figures 3.20 and 3.16(ii) indicate the sequences which are presumed to have been the original 3' ends of the first strands of both pCM212 and pCM222 respectively and shows that they are complementary to nucleotides 463 to 470 and 289 to 297, of the β-tropomyosin coding region. Analysis of the extreme 5' non-coding sequences of both pCM212 and pCM222 in relation to the remaining sequence in the clone, using the computer program CINTHOM indicated that hairpin structures were unlikely to form between any regions other than those indicated in Figures 3.20 and 3.16.

The two direct repeats, found in clones pCM202 and pCM212 are thought to have distinct origins and therefore will be dealt with separately. It is difficult to present a theoretical model for the origin of the direct repeat in pCM202. The only precedent for such a repeat is a tandem repeat found in the 3' non-coding region of a mouse skeletal muscle actin cDNA clone (Leader et al, 1986). However there are certain differences worth noting. The direct repeat of 70 nucleotides found in the skeletal muscle actin cDNA occurs tandemly, that is the 70 nucleotides are repeated immediately without any intervening DNA sequence. Furthermore the tandem repeat is found internally within a larger sequence with over 1kb of sequence 5' to the repeat and 140 bases 3' to the repeat. The direct repeat in pCM202, however, involves the duplicative transposition of 21 nucleotides of DNA to a position 500 nucleotides upstream of where they are found in native β-tropomyosin. In addition there is only one copy of this sequence in the cDNA clone (Figure 4.5). Nevertheless, I can offer no better explanation than that of Leader et al (1986), that during first strand synthesis,

Figure 4.5 The direct repeat DNA found in clone pCM202

A diagrammatic representation of clone pCM202 is presented. The first 21 nucleotides are a direct repeat of the sequence from nucleotide 546 to 567 of β-tropomyosin. Note that pCM202 is truncated before it reaches this region.


ß-tropomyosin
ATG
TGA
25
516
pCM202
546
567
546
567

the cDNA became dislodged from the mRNA template and that reattachment occurred approximately 500 nucleotides 3' to the point at which detachment had occurred. However no potential secondary structure interactions can be found to support this model.

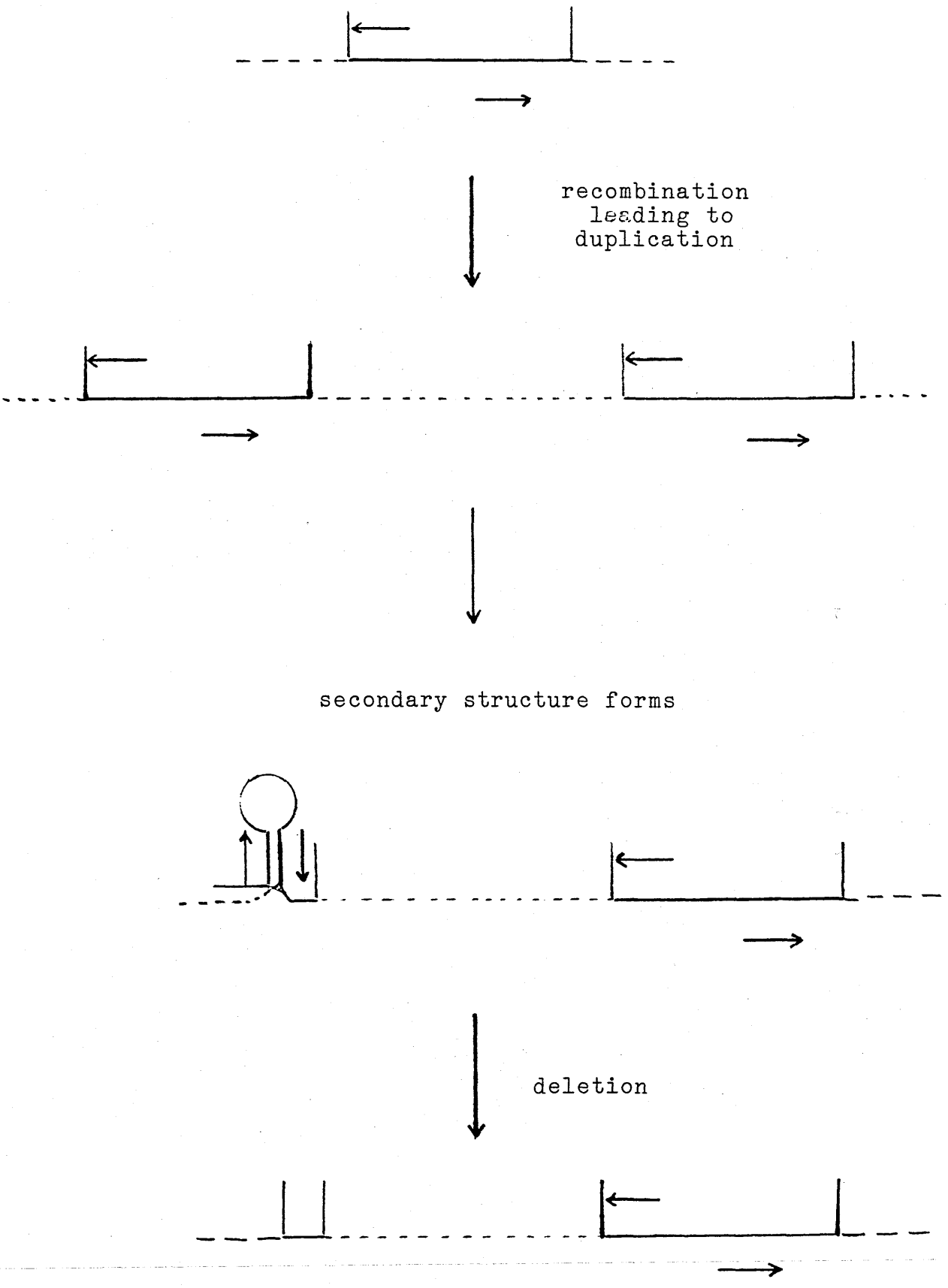
The second direct repeat, found in pCM212, is thought to have arisen after the transformation of E. coli with the original cDNA clone. The original cDNA clone is thought to have consisted of the β-tropomyosin sequence together with the associated inverted repeat and bound by two G/C cloning tails (nucleotides 2184 to 3572 of the pCM212 sequence presented in Figure 3.19). Thus although the strain of E.Coli used was a rec A⁻strain of bacteria it is proposed that the direct repeat arose as a result of an initial recombination event which duplicated the β-tropomyosin sequence together with part of the pBR322 sequence, and then the subsequent deletion of part of one of the β-tropomyosin copies. Figure 4.6 represents a theoretical model of the secondary structure formation which may have been involved in the deletion of one of the copies of the β-tropomyosin sequence.

4.4 The CpG dinucleotide distribution in β-tropomyosin cDNA

One feature of the mouse β-tropomyosin cDNA sequence that was of interest was the distribution of CpG dinucleotides. The relevance of this is to DNA methylation which has been implicated in the control of expression of mammalian genes. Mammalian DNA can be methylated most commonly resulting in a 5-methylcytosine (^{m}C) residue occurring predominantly in the dinucleotide CpG. The

Figure 4.6 Theoretical secondary structure interactions proposed for clone pCM212

Clone pCM212 is considered to have originally comprised of the β-tropomyosin sequence together with a region of invertedly repeated DNA (the inverted repeated DNA is indicated by arrows). The region of directly repeated DNA together with the two extra G/C cloning tails actually found in clone pCM212 are considered to have arisen as a result of a recombination event *in vivo* which led to the duplication of the β-tropomyosin sequence together with the inverted repeat and part of the pBR322 vector and then the subsequent deletion of part of one of the tropomyosin sequences. The figure opposite illustrates that the invertedly repeated DNA is able to form a stem and loop structure, with its corresponding sequence in β-tropomyosin. The subsequent deletion of this structure would give rise to a cDNA clone with the characteristics of that found in the clone pCM212.


recombination
leading to
duplication
secondary structure forms
deletion

incidence of CpGs in the genome is less than expected from its base composition, but this has been ascribed to the ^{m}Cs being deaminated to produce T residues. This would thus result in the replacement of CpG by TpG, with CpA appearing in the complementary strand (Salser, 1977; Bird, 1980). Consequently the occurrence of the dinucleotide CpG has been used as a gauge of the methylation state of the DNA. Although it is considered that vertebrate genomes, in general, contain a deficit of CpG dinucleotides, it has emerged that the 5' flanking regions of several mammalian house-keeping genes have "GC-rich islands". Therefore to explain the high incidence of CpG residues in these regions it has been suggested that the DNA must be unmethylated. However Peter et al (1988) have recently suggested that there may be hemimethylation of the DNA in the 5' exons of actin genes. Their proposal was based on a deficit of CpG dinucleotides occurring in the codon position [2,3], but not the in [3,1] position, the reasoning involved being explained in more detail below. This pattern was not found in several other cDNAs examined, and so it was of interest to analyse the coding region of the mouse β-tropomyosin cDNA for the occurrence of CpG dinucleotides with reference to the codon position ([2,3],[3,1]) in which they occur.

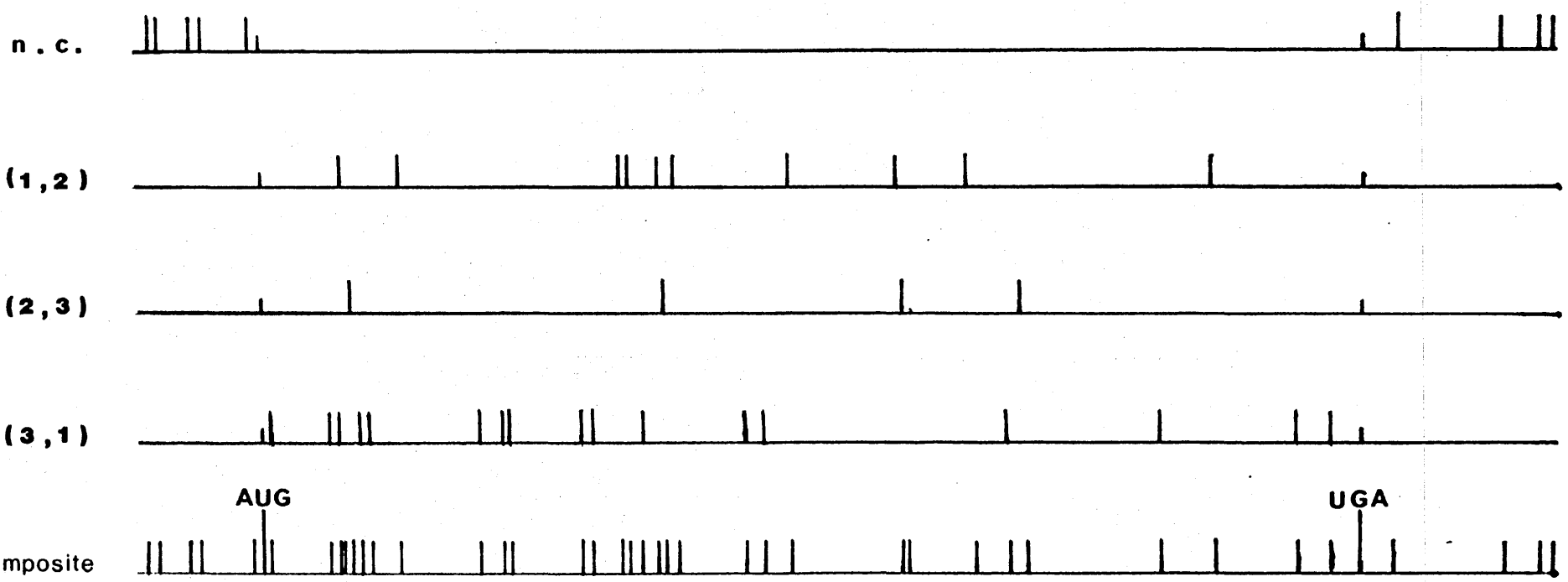
There are four occurrences of CpG in the [2,3] codon position. When this is compared to the thirteen theoretically possible occurrences (assuming the amino-acid sequence is a constraining factor) an obvious deficit is found. The number of theoretically possible occurrences was calculated by counting the number of codons containing a C in position two and then dividing this number by four

to find the number which would theoretically have a G in position three. This assumes that there is a random distribution of all four bases in position three, but in fact, since the mouse β-tropomyosin sequence contains a higher occurrence of G residues in the third position (41.8%) than any of the other bases, the observed deficit, calculated for a random 25% distribution of G residues in position three, is an underestimation.

The theoretical maximum occurrence of CpG dinucleotides in the [3,1] position was calculated using a factor which reflects the constraint of the genetic code on a conserved amino-acid sequence. Each codon which has a G residue in position one was examined together with the codon immediately preceding it. The preceding codon was then given a value which reflected the chance of it having a C residue in position three. Thus for four codon families such as Ala this value was 0.25. for the three codon family of Ile the value was 0.33. For the two codon families such as Asp the value was 0.5, but for the two codon families such as Glu the value was 0. The six codon families of Arg and Leu were given a value of 0.17 and the six codon families of Ser a value of 0.33. Finally the unique codons of Met and Trp were given a value of 0. The summation of all these values calculated for the β-tropomyosin sequence (24) was then compared to the actual occurrence of CpG dinucleotides in the [3,1] position (19) and once again a deficit was apparent. The true deficit may in fact be greater since β-tropomyosin contains 27% C residues in position three. However, when the regional distribution of CpG dinucleotides in the [3,1] position was mapped (Figure 4.7) it was found that they were clustered in the 5' half of

Figure 4.7 The regional distribution of CpG dinucleotides in the nucleotide sequence of mouse β-tropomyosin cDNA

The occurrence of CpG dinucleotides in the codon positions [1,2], [2,3] and [3,1] of the mouse β-tropomyosin cDNA sequence were mapped. These are shown together with a composite showing total occurrences. Each CpG dinucleotide is indicated by a vertical line. The initiator methionine and termination codons are indicated. The table below indicates the occurrences of CpG in each codon position in relation to the expected occurrence as discussed in the text.



codon position	expected occurrence	actual occurrence	variation
1,2	66.66%	71%	+5%
2,3	25%	7.8%	-69%
3,1	17.5%	13.8%	-21%

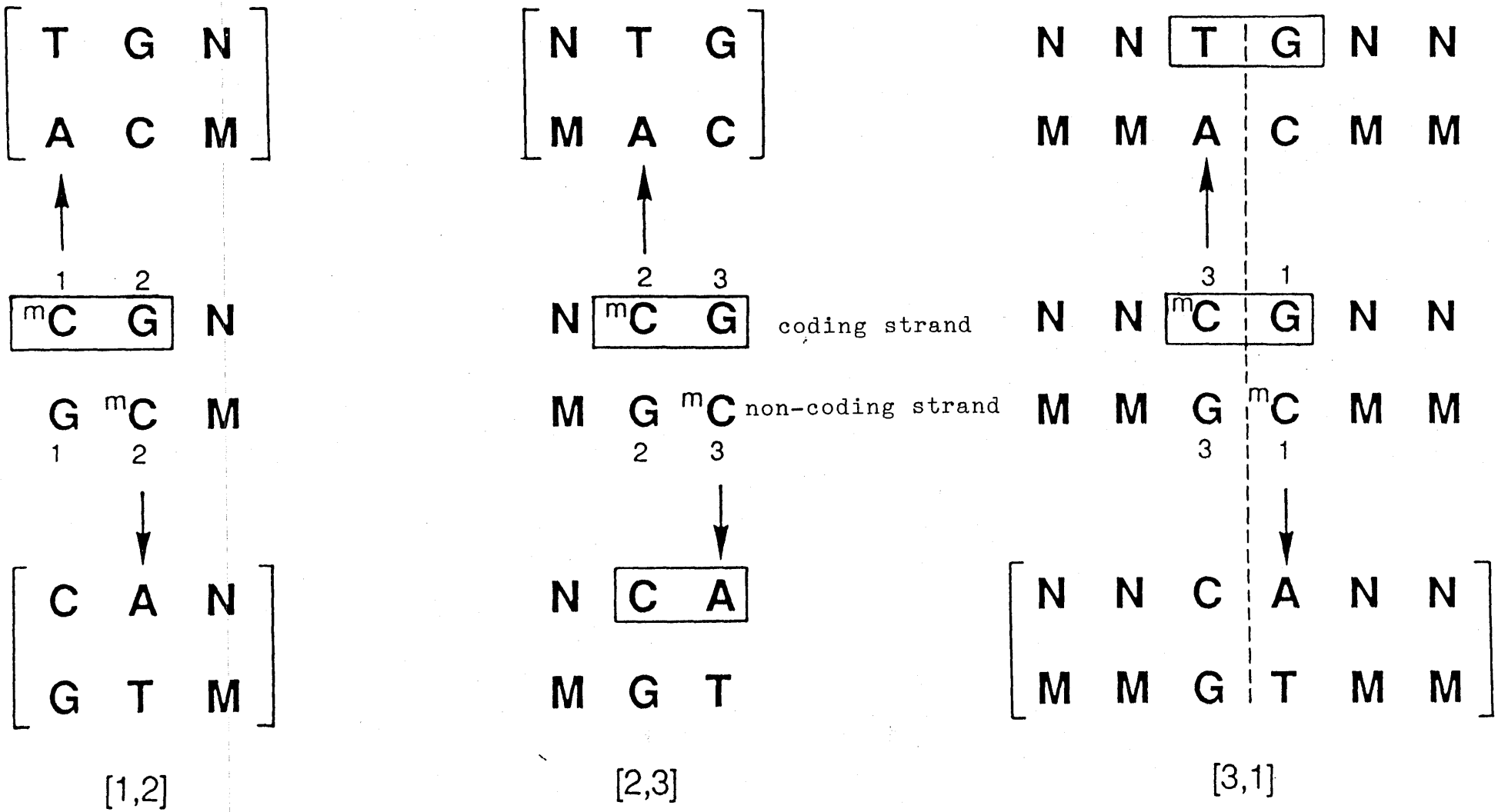
the cDNA. In these exons no deficit is apparent (for amino-acids 1-150, 14 occurrences out of an expected 12 occurrences), whereas in the 3' half of the cDNA the deficit tends to that found for the [2,3] position. This is similar to the findings of Peter et al (1988) for actin cDNAs. In the absence of any physical or biological rationale for the selection either against NCG codons, or in favour of NNC codons followed by a G, Peter et al (1988) interpreted the deficit in terms of strand-specific hemimethylation as follows.

As explained diagrammatically in Figure 4.8, in order to conserve the amino-acid sequence one must select against deaminations of ^{m}CpGs which result in an amino-acid change. Therefore if the coding strand is methylated, deaminations of ^{m}CpGs in the [1,2] and [2,3] positions will be selected against, thus no deficit of CpG dinucleotides in these codon positions will be evident. However if the non-coding strand is methylated then deaminations of ^{m}CpG in the [1,2] and [3,1] positions will result in amino-acid changes and thus would be selected against, whereas deamination of ^{m}CpG in the [2,3] will result only in a silent nucleotide mutation. Therefore one might expect to find no CpG deficits in the [1,2] and [3,1] positions but a theoretical CpG deficit in the [2,3] position would be possible.

This idea of hemimethylation is, of course, difficult to prove since only a fraction of methylated CpGs are detectable by restriction enzyme analysis which in any case could not distinguish between hemimethylated and fully methylated DNA. At present the

Figure 4.8 The theoretical effect of hemimethylation of DNA on amino-acid sequences

A diagrammatic scheme indicating the possible consequences, for an amino acid sequence, of the hemimethylation of germ line DNA. Note that if the coding strand of DNA is methylated deamination of a 5-methylcytosine in the dinucleotide CpG in codon positions [1,2] and [2,3] would lead to an amino-acid change (indicated by parentheses). Similarly if the non-coding strand of DNA is methylated, deamination of the 5-methylcytosine of CpG dinucleotides in the codon positions [3,1] and [1,2] would result in amino-acid changes. The consequence this has for the distribution of CpG dinucleotides in a sequence of DNA is discussed in the text


T G N
A C M
1 2
mC G N
G mC M
1 2
C A N
G T M
[1,2]
N T G
M A C
2 3
N mC G
coding strand
M G mC
non-coding strand
2 3
N C A
M G T
[2,3]
N N T G N N
M M A C M M
3 1
N N mC G N N
M M G mC M M
3 1
N N C A N N
M M G T M M
[3,1]

only method of test this hypothesis would appear to be the direct sequencing of genomic DNA using the chemical scission reactions of Maxam and Gilbert (1980).

It is difficult to say what implications the hemimethylation of the 5' regions of DNA would have on the expression of β-tropomyosin. Clearly if β-tropomyosin and TM_{36} are subject to alternative splicing, the hemimethylation of the DNA cannot be used to distinguish muscle genes from non-muscle genes for example, since both isoforms share common 5' exons. However since the mRNAs of both isoforms appear to have identical 5' termini it is more likely that the control mechanisms for alternative splicing of the pre-mRNA lie within the 3' half of the mRNA. Thus one imagines that hemimethylation, if it occurs, would only be involved in the (de)activation of the gene as a whole and not differential expression of β-tropomyosin the TM_{36}.

REFERENCES

Bains, W; Ponte, P; Blau, H; and Kedes, L. Mol. Cell. Biol. 4: 1449-1453 (1984)

Barton, P; Cohen, A; Robert, B; Fiszman, M.Y.; Bonhomme, F; Guenet, J-L; Leader, D.P; and Buckingham, M.E. J. Biol. Chem. 260: 8578-8584 (1985)

Bird, A.P; Nucl. Acids Res. 8: 1499-1501 (1980)

Birnboim, H.C; and Doly, J. Nucl. Acid Res. 7: 1513-1523 (1979)

Blau, H.M; Chin, C-P; and Webster, C. Cell 32: 1171-1180 (1983)

Borelli, E; Hen, R; and Chambon, P. Nature 312: 608-612 (1984)

Bourne, G.H. (Ed) "The Structure and Function of Muscle" 2nd Edition Academic Press (1973)

Breitbart, R.E; Nguyen, H.T; Medford, R.M; Destree, A.T; Mahdavi, V; and Nadal-Ginard, B. Cell 41: 67-82 (1985)

Breitbart, R.E; Andreadis, A; and Nadal-Ginard, B. Ann. Rev. Biochem. 56: 467-495 (1987)

Briggs, M.M; Klevit, R.E; and Schachat, F.H. J. Biol. Chem. 259: 10369-10375 (1984)

Buckingham, M.E; and Minty, A. In "Eukaryotic Genes: Their Structure, Activity and Regulation". Eds. McLean, N; Gregory, S.P; and Flavell, R.A. Butterworth Press Vol.21 365-395 (1983)

Caravatti, M; Minty, A; Robert, B; Montarras, D; Weydert, A; Cohen, A; Daubas, P; and Buckingham, M. J. Mol. Biol. 160: 59-76 (1982)

Carmon, Y; Czosneck, H; Nudel, U; Shani, M; and Yaffe, D. Nucl. Acids Res. 10: 3085-3098 (1982)

Chepelinsky, A.B; King, C.R; Zelenka, P.S; and Piatigorsky, J. Proc. Natl. Acad. Sci. USA 82: 2334-2338 (1985)

Chin, C-P; and Blau, H.M. Cell 40: 417-424 (1985)

Ciliberto, G; Dente, L; and Cortese, R. Cell 41: 531-540 (1985)

Close, R. Physiol. Rev. 52: 129-197 (1972)

Cohen, S.N; Chang, A.C.Y; Hsu, L. Proc. Natl. Acad. Sci. USA 69: 2110-2114 (1972)

Cummins, P; and Perry, S.V. Biochem. J. 141: 43-49 (1974)

Czosneck, H; Nudel, U; Shani, M; Barkar, P.E; Pravtcheva, D.D; Ruddle, F.H; and Yaffe, D. The EMBO Journal 1: 1299-1305 (1982)

Czosneck, H; Nudel, U; Mayer, Y; Barkar, P.E; Pravtcheva, D.D: Ruddle, F.H; and Yaffe, D. The EMBO Journal 2: 1977-1979 (1983)

Darnell, J; Lodish, H; and Baltimore, D. In "Molecular Cell Biology" Scientific American Books Inc. (1986)

Delain, D.L; Meienhofer, M.C; Proux, D; and Schapira, F. Differentiation 1: 349-354 (1973)

Dynan, W.S; and Tjian, R. Nature 316: 774-778 (1985)

Edmonds, M; Vaughn, M.Jr; and Nakazato, H. Proc. Natl. Acad. Sci. USA 68: 1336-1340 (1971)

Ellis, R.N; De Feo, D: Farth, M.E; and Scolnick, E.M. Mol. and Cell Biol. 2: 1339-1345 (1983)

Eppenberger, H.M; Eppenberger, M; Richterich, R; and Aebi, H. Dev. Biol. 10: 1-16 (1964)

Fields, S; and Winter, G. Gene 15: 207-214 (1981)

Fine, R.E; and Blitz, A.L. J. Mol. Biol. 95: 447-454 (1975)

Gelbeiter, J; Zeff, R.A; Schulze, D.H; Pease, L.R; Weiss, E.H; Hallor, A.L; Flavell, R.A; and Nathenson, S.G. Mol. Cell Biol. 6: 645-652 (1986)

Getz, M.J; Reiman, H.M; Siegal, G.P; Quinlan, T.J; Proper, J; Elder, P.K; and Moses, H.L. Cell 11: 909-922 (1976)

Gonoi, T: Hasegawa, S; Kuromin, H; and Hagihara, Y. In "Muscular Dystrophy: Biomedical Aspects" Eds Ebashi, S; and Ozawa, E. Springer-Verlag, Berlin (1983)

Gorman, C.M; Rigby, P; and Lane, D. Cell 42: 519-526 (1985)

Grunstein, M; and Hogness, D. Proc. Natl. Acad. Sci. USA 72: 3961-3965 (1975)

Gunning, P; Ponte, P; Kedes, L; Hickey, R.J; and Skoultchi, A.I. Cell 36: 709-715 (1984)

Hastie, N.D; and Bishop, J.O. Cell 9: 761-774 (1976)

Heeley, D.H; Dhoot, G.K; and Perry, S.V. Biochem. J. 226: 461-468 (1985)

Helfman, D.M; Feramisco, J.R; Ricci, W.M; and Hughes, S.H. J. Biol. Chem. 259: 14136-14143 (1984)

Helfman, D.M; Cheley, S: Knismanen, E; Finn, L.A; and Yamawaki-Kataoka, Y. Mol. and Cell Biol. 6: 3582-3593 (1986)

Hoh, J.F.Y; McGrath, P.A; and Hale, H.T. J. Mol. Cell Cardiol. 11: 1053-1076 (1978)

Holmes, D.S; and Quigley, M. Anal. Biochem. 114: 193-197 (1981)

Holt, J.C; and Lowey, S. Biochemistry 16: 4398-4402 (1977)

Horecker, B.L; Tsolas, O; and Lae, C.Y. In "The Enzymes" Ed. Boyer, P.D. Vol.7 oo 213-258. Academic Press, New York (1972)

Huynh, T.V; Young, R.A: and Davis, R.W. In "DNA cloning" Ed. Gover, D.M. IRL Press (1984)

Ibsen, K.H. Cancer Res. 37: 341-352 (1977)

Imai, H; Hirai, S.I; Hirono, M. and Hiraboyashi, H. J. Biochem. (Tokyo) 99: 923-930 (1986)

Izumo, S; Nadal-Ginard, B: and Mahdavi, V. Science 231: 597-600 (1986)

Kano, M; Suteo, Y; Satoh, T; and Kawakami, M. In "Muscular Dystrophy: Biomedical Aspects" Eds. Ebashi, S; and Ozawa, E. Springer-Verlay, Berlin (1983)

Keshet, I; Yisraeli, J; and Cedar, H. Proc. Natl. Acad. Sci. USA 82: 2560-2564 (1985)

Kirby, K.S. Biochem. J. 64: 405-410 (1956)

Kurata, N; Matsushima, T: and Sugimura, T. Biochem. Biophys. Res. Commun. 48: 473-479 (1972)

Layzer, R.B; and Epstein, C.J. Am. J. Hum. Genet. 24: 533-543 (1972)

Leader, D.P; Gall, I; Campbell, P; and Frischauf, A.M. DNA 5: 235-238 (1986)

Lebherz, H.G; and Rutter, W.J. Biochemistry 8: 109-121 (1969)

Lewin, B. In "Genes" John Wiley and Sons Inc. New York (1984)

Lompre, A-M; Nadal-Ginard, B; and Mahdavi, U. J. Biol. Chem. 259: 6437-6446 (1984)

Low, B. Proc. Natl. Acad. Sci. USA 60: 160-167 (1968)

Lowey, S; and Risby, D; Nature 234: 81-85 (1971)

MacLeod, A.R. Eur. J. Biochem. 126: 293-297 (1982)

MacLeod, A.R; Houlker, C; Reinach, F.C; Smillie, L.B; Talbot, K; Modi, G; and Walsh, F.S. Proc. Natl. Acad. Sci. USA 82: 7835-7839 (1985)

MacLeod, A.R; Houlker, C; Reinach, F.C; and Talbot, K. Nucl. Acids Res. 14: 8413-8426 (1986)

MacLeod, A.R. BioEssays 6: 208-212 (1987)

MacLeod, A.R; Talbot, K; Smillie, L.B; and Houlker, C. J. Mol. Biol. 194: 1-10 (1987)

MacLeod, A.R; and Gooding, C. Moll and Cell Biol. 8: 433-440 (1988)

Mak, A.S; Smillie, L.B; and Stewart, G.R. J. Biol. Chem. 255: 3647-3655 (1980)

Maxam, A.M; and Gilbert, W. Methods in Enzymol. 65: 499-560 (1980)

Melloul, D.D; Aloni, B; Calvo, J; Yaffe, D; and Nudel, U. EMBO J. 3: 983-990 (1984)

Meselson, M; and Yuan, R. Nature 217: 1110-1114 (1968)

Minty, A.J; Alonso, S; Caravatti, M; and Buckingham, M.E. Cell 30: 185-192 (1982)

Minty, A.J; Hardeman, E; Gunning, P; Bains, W; Blau, H; and Kedes, L.H. UCLA Symp. Mol. Cell Biol. 29: 507-519 (1986)

Minty, A.J; and Kedes, L. Mol. Cell Biol. 6: 2137-2145 (1986)

Miwa, T: Boxer, L.M. and Kedes, L. Proc. Natl. Acad. Sci. USA 84: 6702-6706 (1987)

Moosekar, M.S. Cell 35: 11-13 (1983)

Moore, K.J; and Bulfield, G. Biochem. Genet. 19: 771-781 (1981)

Novikoff, A.B; and Holtzman, E. in "Cells and Organelles". Holt, Rinehart and Winston. New York (1970)

Ott, M.O; Sperling, L; Hermobel, P; Yaniv, M; and Weiss, M.C. EMBO J. 3: 2505-2510 (1984)

Passmore, R; and Robson, J.S. Eds. "A companion to medical studies. Volume 1" Blackwell Scientific Publications (1968)

Periasamy, M; Strehler, E.E; Garfinkel, L.I; Gubits, R.M; Ruiz-Opazo, N; and Nadal-Ginard, B. J. Biol. Chem. 259: 13595-13604 (1984)

Peter, B; Man, Y.M; Begg, C.E; Gall, I; and Leader, D.P. Submitted

Peters, J; Nash, H.R; Eicher, E.M; and Bulfield, G. Biochem. Genet. 19: 757-769 (1981)

Phillips, W.D; and Bennett, M.R. Dev. Bio. 106 457-468 (1984)

Potter, J.D; and Johnson, D.A. In "Calcium and Cell Function" Ed. Cheung, W.Y. pp.145-173. Academic Press. New York (1982)

Pustel, J; and Kafatos, F.C. Nucl. Acids Res. 10: 4765-4782 (1982)

Reinach, F.C; Masaki, T; and Fischman, D.A. J. Cel Biol. 96: 297-303 (1983)

Rigby, P.W.J; Dieckmann, M; Rhodes, C; and Berg, P. J. Mol. Biol. 113: 237-251 (1977)

Robert, B; Weydert, A; Caravatti, M; Minty, A; Cohen, A; Daubas, P; Gros, F; and Buckingham, M. Proc. Natl. Acad. Sci USA 79: 2437-2441 (1982)

Robert, B; Daubas, P; Akimenko, M-A; Cohen, A; Garner, I; Guenet, J-L; and Buckingham, M.E. Cell 39: 129-140 (1984)

Robert, B; Barton, P; Minty, A; Daubas, P; Weydert, A; Bonhomme, F; Catalan, J; Chazottes, D; Guenet, J-L; and Buckingham, M. Nature 314: 181-183 (1985)

Saez, L; and Leinwand, L.A. Nucl. Acids Res. 14: 2951-2969 (1986)

Saez, L.J; Gianola, K.M; McNally, E.M; Feghali, R; Eddy, R: Shows, T.B; and Leinwand, L.A. Nucl. Acids Res. 15:

Salser, W. Cold Spring Harbor Symp. Quant. Biol. 42: 985-1002 (1977)

Sanders, C; and Smillie, L.B. J.Biol. Chem. 260: 7264-7275 (1985)

Sanger, F; Nicklen, S; Coulson, A.R. Proc. Natl. Acad. Sci. USA 74: 5463-5467 (1977)

Sarkar, S; Sreter, F; and Gergely, J. Proc. Natl. Acad. Sci. USA 68: 946-950 (1971)

Sato, K; Morris, H.P; and Weinhouse, S. Science 178: 879-881 (1972)

Schwartz, R.J; and Rothblum, K.N. Biochemistry 20: 4122-4129 (1981)

Shani, M; Admon, S; and Yaffe, D. Nucl. Acids Res. 12: 7225-7234 (1984)

Southern, E. J. Mol. Biol. 98: 503-517 (1975)

Starr, R; and Offer, G. J. Mol. Biol. 170: 657-680 (1983)

Stone, D; and Smillie, L.B. M. Biol. Chem. 253: 1137-1148 (1978)

Storti, R.V; Coen, D.M; and Rich, A. Cell 8: 521-527 (1976)

Talbot, K; and MacLeod, A.R. J. Mol. Biol. 164: 159-174 (1983)

Thomas, P.S. Proc. Natl. Acad. Sci. USA 77: 5201-5205 (1980)

Toyota, N; and Shimada, Y. Cell 33: 297-304 (1983)

Trayer, I.P; and Perry, S.V. Biochem. Z 345: 87-92 (1966)

Tsutsumi, K-I; Mukai, T; Hidaka, S; Miyahara, H; Tsutsumi, R; Tanaka, T; Hori, K; and Ishikawa K. J. Biol. Chem. 258: 6537-6542 (1983)

Turner, D.C; Maier, V; and Eppenberger, H.M. Dev. Biol. 37: 63-89 (1974)

Vandekerckhove, J; and Weber, K. Differentiation 14: 123-133 (1979)

Vandekerckhove, J; Franke, W; and Weber, K. J. Mol. Biol. 152: 413-426 (1981)

Villakomeroff, L; Efstratiodis, A; Broome, S; Lomedico, P; Tizard, R; Naker, S.P; Chick, W.L; and Gibert, W. Proc. Natl. Acad. Sci. USA 75: 3727-3731 (1978)

Volckaert, G; Tavernier, J; Derynck, R; Devos, R; and Fiers, W. Gene 15: 215-223 (1981)

Wagner, P.P; and Giniger, E. Nature 292: 560-562 (1981)

Walker, M.D; Edlund, T; Boulet, A.M; and Rutter, W.J. Nature 306: 557-561 (1983)

Weyderet, A; Daubas, P; Caravatti, M; Minty, A; Bugaisky, G; Cohen, A; Robert, B; and Buckingham, M. J. Biol. Chem. 258: 13867-13874 (1983)

Weydert, A; Daubas, P; Lazaridis, I; Barton, P; Garner, I; Leader, D.P; Bonhomme, F; Catalan, J; Simon, D; Guenet, J-L; Gros, F; and Buckingham, M.E. Proc. Natl. Acad. Sci. USA 82: 7183-7187 (1985)

Weydert, A; Barton, P; Harris, A.J; Pinset, C; and Buckingham, M.E. Cell 49: 121-129 (1987)

Weintraub, H. Cell 42: 705-711 (1985)

Whalen, R.G; Sell, S.M; Erikson, A; and Thornell, L.G. Dev. Biol. 91: 478-484 (1982)

Wickens, M.P; Buell, G.N; and Schimke, R.T. J. Biol. Chem. 253: 2483-2495 (1978)

Wilkinson, J.M; and Grand, R.J.A. Nature 271: 31-35 (1978)

Wilkinson, J.M; Moir, A.J.G; and Waterfield, M.D. Eur. J. Biochem. 143: 47-53 (1984)

Yaffe, D; and Saxel, O. Nature 270: 725-727 (1977)

Yaffe, D; Nudel, V; Mayer, Y; and Neuman, S. Nucl. Acids Res. 13: 3723-3737 (1985)

Yanisch-Perron, C; Viera, J; and Messing, J. Gene 33: 103-119 (1985)

Young, R.A; and Davis, R.W. Nature 22: 778-782 (1980)

Young, B.D; Birnie, G.D; and Paul, J. Biochemistry 15: 2823-2828 (1976)

Ziter, F.A. Exp. Neurol. 43: 539-546 (1974)