ANALYSIS OF THE STUCTURE AND EXPRESSION OF MAMMALIAN MYOGLOBIN GENES

MELANIE PRICE

Thesis submitted for the degree of

Doctor of Philosophy

in the University of Leicester

1987

UMI Number: U527120

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U527120
Published by ProQuest LLC 2015. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

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

X9502433697

PUBLICATIONS:

Some of this work has been published:

Blanchetot, A., Price, M. and Jeffreys, A.J. 1986. The mouse myoglobin gene: Characterisation and sequence comparison with other mammalian myoglobin genes. Eur. J. Biochem. 159:469-474.

Weller, P.A., Price, M., Isenberg, H., Edwards, Y.H. and Jeffreys, A.J. 1986. Myoglobin expression: early induction and subsequent modulation of myoglobin and myoglobin mRNA during myogenesis. Mol. Cell. Biol. 6:4539-4547.

ACKNOWLEDGEMENTS

I would like to say how much I have enjoyed working with Dr. Alec Jeffreys and thank him for the opportunity he gave me to be in his laboratory and for the invaluable training he has given me in the field of molecular biology.

I am very grateful to members of Dr. Jeffreys laboratory, especially Mrs. Vicky Wilson and Mrs. Zilla Wong for their kindness and technical help, Dr. Polly Weller and Dr. Stephen Harris for helpful discussions and to other members of the genetics department, especially Baljinder Mankoo and Mick Pocklington for a pleasant working environment.

Special thanks to Mrs. Tania Cresswell-Maynard and Dr. Jo Whittaker for advice and assistance with tissue culture work, to Dr. John Brookfield for advice on calculations and to Mrs. Sheila Mackley for help with the typing of this thesis.

Finally I would like to express my gratitude to Dr. Serge Casaregola for constant support and encouragement.

I acknowledge the receipt of an MRC studentship.

This thesis is dedicated to my

mother and father

ABBREVIATIONS

BCIG	5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside
BSA	bovine serum albumin
Ci	curie
cpm	count per minute
DEPC	diethylpyrocarbonate
DMSO	dimethylsulphoxide
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	diaminoethanetetra-acetic acid
IMS	industrial methylated spirit
IPTG	isopropyl-β-D-galacto-pyranoside
kb	kilo base pairs
MHC	myosin heavy chain
MLC	myosin light chain
MOPS	3-(N-morpholino)propanesulphonic acid
MY	million years
dNTP	2'-deoxynucleoside triphosphate. N= adenosine, guanosine, thymidine, cytidine
ddNTP	2'3'-dideoxynucleoside triphosphate
PBS	phosphate buffered saline
PEG	polyethyleneglycol
PVP	polyvinylpyrrolodine
RNase	ribonuclease
SDS	sodium dodecyl sulphate
SSC	saline sodium citrate
TEMED	N,N,N'N'-tetramethyl-ethylinediamine
Tn	troponin
Tris	tris(hydroxymethyl)aminomethane
HCC	hexamine cobalt chloride

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 MYOGLOBIN

Myoglobin is a member of the globin family which consists of a number of related respiratory proteins. The presence of globins in organisms as diverse as bacteria (Wakabayashi, Matsubara and Webster, 1986), plants (Landsmann et al., 1986) and higher eukaryotes (Mill, 1972, Jeffreys et al., 1983) suggests that the putative ancestral globin gene is extremely ancient.

In vertebrates, myoglobin is tissue-specific being found in striated muscles. This Chapter concentrates on myoglobin and haemoglobin genes as examples of globin genes which are incorporated into very different developmental programs: myogenic and erythroid respectively.

Structure of myoglobin

Myoglobin is a relatively small globular protein with a molecular weight of 17,800 Daltons. It consists of a single polypeptide chain of 153 amino acid residues folded about its prosthetic group haem. Haem is a substituted protoporphyrin IX ring with a central iron atom and is found in a number of other respiratory proteins, for example the cytochromes. The site of oxygen binding in myoglobin is the iron atom of the haem which is in the ferrous state (Fe^{2+}).

The three dimensional structure of sperm whale myoglobin was determined by Kendrew and his colleagues in a number of stages of increasing resolution, using the technique of X-ray diffraction (see Kendrew et al., 1958, 1960, 1961, Kendrew, 1963 and Nobbs et al., 1966).The 3-D structure has been redetermined more recently by Takano et

al. (1977). Some of the noteworthy features of myoglobin deduced from these studies are that myoglobin is an extremely compact protein with approximately 75% of the main polypeptide chain folded in a right-handed α-helical conformation with most of the polar amino-acid residues on the outer surface and the non-polar residues inside. There are eight major helical segments.

Myoglobin binds reversibly with a single molecule of oxygen and the reversibility of this reaction, which is basic to the physiological function of myoglobin, is made possible by the enclosure of haem within a pocket of hydrophobic residues. The resulting non-polar haem binding site excludes water and therefore protects the ferrous state of the haem from oxidation by water to the ferric state which is unable to bind oxygen.

The folding pattern of the polypeptide chain around the haem is a fundamental design for an oxygen carrier. Tetrameric haemoglobin is composed of four polypeptide chains (two α-like and two β-like subunits) in contrast to the single chain of myoglobin. The three-dimensional structure of haemoglobin revealed that the α and β chains of haemoglobin are very similar to myoglobin which is rather striking considering only 24 out of 114 amino acid residues in their chains are identical (Perutz, Kendrew and Watson, 1965).

The amino acid sequence of myoglobin has been determined for a large number of species (see Romero-Herrera et al., 1978). All mammalian myoglobins sequenced so far contain 153 amino acid residues. There is no evidence for developmental or tissue-specific myoglobin isoforms: human and foetal myoglobins have identical amino acid sequences (Schneiderman, 1962) and myoglobin of both cardiac and skeletal muscle is identical (Romero-Herrera and Lehmann, 1971, Tipler, Edwards and Hopkinson, 1978).

The functions of myoglobin

i) Myoglobin facilitates diffusion of oxygen in muscle

In vertebrates, haemoglobin and myoglobin carry oxygen which is required for aerobic respiration. Tetrameric haemoglobins transport oxygen in the blood while monomeric myoglobin is the principal haemoprotein of both skeletal and cardiac muscle serving to facilitate the diffusion of oxygen in muscle.

The different characteristics of haemoglobin and myoglobin which result in their suitability to their particular roles in vertebrate tissues is reflected by the shape of their respective oxygen dissociation curves (Antonini, 1965). The oxygen dissociation curve of myoglobin is hyperbolic which is expected for a simple 1:1 association of haem and oxygen. Haemoglobin has a sigmoidal oxygen dissociation curve resulting from oxygen binding to the four subunits of haemoglobin being co-operative; this enables haemoglobin to be a more efficient oxygen transporter and allows it to deliver twice as much oxygen in a low oxygen environment as it would if its oxygen binding sites were independant. However, for any given partial pressure of oxygen, the saturation of the oxygen binding sites is higher for myoglobin than haemoglobin, meaning that myoglobin has a higher affinity for oxygen.

Oxygen binding to haemoglobin is also affected by pH and levels of carbon dioxide whereas myoglobin shows no change in oxygen binding over a wide range of pH and carbon dioxide levels. Acidity and increased carbon dioxide levels enhance the release of oxygen from haemoglobin and therefore the increased pH and carbon dioxide levels in the capillaries of metabolically active tissue such as muscle promotes the release of oxygen which is then taken up by myoglobin which has a higher oxygen affinity.

Early experiments by Millikan (1939) and Wittenberg (1970 and

references within) demonstrated indirectly that myoglobin contributed in some way towards oxygen entry into muscle. For example, Millikan (1939) showed that myoglobin in muscle undergoes rapid oxygenation and deoxygenation in response to fluctuations in oxygen supply and demand of the activated muscle. Wittenberg (1970) proposed that an adequate supply of oxygen to the muscle mitochondria during contraction was dependant upon three components: the diffusion path length of oxygen from the capillaries to the muscle mitochondria, the oxygen pressure gradient existing between these two and the rate of oxygen diffusion which included myoglobin facilitated diffusion and the natural diffusivity of oxygen. With the onset of contraction, muscle oxygen demand increases immediately (Millikan, 1939) and the steady state (oxygen supply and demand are balanced) is achieved by increasing oxygen supply to the contracting muscle. Oxygen supply is increased by the widening of the blood capillaries: this alters the diffusion distance and increases blood flow thereby delivering more oxygen and changing the oxygen pressure gradient. Myoglobin is then considered to increase the inward diffusion of the oxygen into muscle fibres but it was not clear how myoglobin might bring this about. Physical model systems (see Wittenberg, 1970 for a review of these systems) established that myoglobin was able to facilitate oxygen diffusion through solution by reversible combination with oxygen and translational diffusion of the oxymyoglobin molecules. It was difficult to demonstrate that myoglobin served this function in muscle tissue and direct evidence that oxygen supply to muscle was dependant upon myoglobin was lacking at this stage. In 1975, Wittenberg, Wittenberg and Caldwell were able to demonstrate, using muscle bundles from the red muscle fibres of pigeon breast and several inhibitors of myoglobin, that myoglobin indeed acts to transport a significant fraction of oxygen consumed by the muscle mitochondria in steady state conditions.

They also established that myoglobin in muscle fibre bundles is only partially oxygenated implying a steep gradient of myoglobin oxygenation which is a requirement for any carrier mediated oxygen transport and which was previously demonstrated in working cardiac and skeletal muscle (Wittenberg, 1970).

ii) Myoglobin as a long-term oxygen store

Diving mammals and birds exhibit an exceptional tolerance to sustained periods of hypoxia compared with mammals that do not dive and frequently show elevated levels of myoglobin in their skeletal or cardiac muscle or both (Wittenberg, 1970). As pointed out by Lawrie (1950), these levels exceed by five to ten fold the levels expected from the respiratory requirements of the tissue. The correlation between high myoglobin concentrations and the diving ability of a number of animals led Irving (1939) and Scholander (1940) to propose the idea that myoglobin might serve as a long-term oxygen store during diving.

In general, the highest concentrations of myoglobin are found in those mammals that are known to perform long and deep dives. For example, seals generally contain higher levels of myoglobin than dolphins, the former diving for longer periods of time (Eichelberger et al., 1939). Phocoid seals have twice the level of myoglobin, amongst many other respiratory adaptions, than Otariid seals which correlates with longer and deeper dives carried out by Phocoid seals (Lenfant, Johansen and Torrance, 1970), and both seals have higher myoglobin concentrations and dive for longer periods than the penguin. Sperm whales are able to dive for up to 120 minutes and have ten times as much myoglobin in their skeletal muscles as found in terrestrial mammals. During a dive, blood supply to the muscles is cut off and oxygen consumption is minimised by economy of motion. Even with the increased

levels of myoglobin, sperm whale muscles go into oxygen debt during prolonged dives (Wittenberg, 1970).

There are other adaptions which allow tolerance to hypoxia in diving mammals and birds including oxygen economy factors such as cardiovascular and respiratory responses which tend to conserve available oxygen stores and differences in the oxygen storage capacity of the lung (Lenfant, Johansen and Torrance, 1970).

Increased levels of myoglobin are also found in subterranean mammals for example the mole-rat (Nevo, 1979) and the pocket gopher (Butler and Jones, 1982). Animals exposed to hypoxic environments therefore, have adapted their concentrations of myoglobin in response to particular physiological requirements.

Adaption of myoglobin levels to hypoxic environments are not universal. Diving reptiles do not show elevated myoglobin levels (see Butler and Jones, 1982) suggesting that they have evolved alternate adaptive mechanisms to hypoxia to cope with their particular requirements.

It is worth mentioning at this point some other specific adaptions of myoglobin concentrations during conditions of hypoxia. Myoglobin concentrations are increased as a result of acclimatisation to high altitudes. Vaughan and Pace (1959) were able to demonstrate that rat total body myoglobin concentration increased by 70% when rats were taken from sea-level to high altitudes or reared at high altitudes. There is an interesting seasonal variation in myoglobin levels in the body muscles of the snow-shoe hare which increase during the late winter months compared to the myoglobin levels observed in the summer (Rosenmann and Morrison, 1965).

iii) Myoglobin as a short term oxygen store

It has been proposed that myoglobin is able to serve as a short term oxygen store to buffer any fluctuations in the rate of oxygen flow from the capillaries to the mitochondria, important in the stabilisation of oxygen supply to the beating heart (see Wittenberg et al, 1975, Longo, Koos and Power, 1973).

1.2 THE GLOBIN GENE FAMILY

Globin gene organisation

Myoglobin and haemoglobin are distant relatives in the globin gene family as revealed by structural and amino-acid homologies between them. Members of the globin gene family are thought to have arisen by successive duplications of an ancestral globin gene with subsequent divergence of the duplicated genes to form new family members (see Jeffreys et al., 1983). The tetrameric haemoglobins of vertebrates arose with an initial α/β-globin gene duplication early in vertebrate evolution, about 450 MY ago (Hunt, Hurst-Calderone and Dayhoff, 1978, Czelusniak et al., 1982). Successive duplications of the α/β genes occured to form two groups of closely related genes which are dispersed to different chromosomal locations in mammals and birds (Efstratiadis et al., 1980). In man, the β-like globin cluster is on chromosome 11 and consists of five functional genes which are developmentally regulated and which encode a variety of specialised embryonic, foetal and adult haemoglobins. The genes are organised 5'-3' on the chromosome in the order of their time of expression during development (Collins and Weismann, 1984). The α-like globin cluster is on chromosome 16 and consists of three functional and developmentally regulated genes.

In contrast to mammals and birds, the α- and β-globin genes are found as a single cluster in the amphibian Xenopus tropicalis (Jeffreys

et al., 1980). Duplicated α- and β-globin genes can therefore remain tightly linked over long (450 MYs) evolutionary periods and this provides evidence that the α- and β-globin gene clusters originally arose as a tandem gene duplication with the dispersal of α- and β-globin genes in a common ancestor leading to the bird and mammalian lineages.

Amino acid sequence divergence of haemoglobin and myoglobin have suggested that the myoglobin gene arose by duplication of a primitive globin gene about 800 MY ago (Hunt, Hurst-Calderone and Dayhoff, 1978) whereas phylogenetic analyses have suggested that myoglobin has appeared more recently, (500 MY ago) early in the evolution of the vertebrates. In both estimates, the duplication precedes that which gave rise to the α- and β-globin genes (Czelusniak et al., 1982).

Recently, two mammalian myoglobin genes have been isolated and characterised. The first myoglobin gene to be cloned was from the grey seal (Wood, Blanchetot and Jeffreys, 1982, Blanchetot et al., 1983). Grey seal skeletal muscle was chosen since it contains very high levels of myoglobin and myoglobin mRNA (about eight times more than human skeletal muscle, Weller et al., 1986). The second myoglobin gene was isolated from human skeletal muscle (Weller et al., 1984) using seal myoglobin gene exons as hybridisation probes. In contrast to the large number of genes encoding the various haemoglobins, a single gene encodes myoglobin of both skeletal and cardiac muscle in the seal and man.

The linkage arrangement between the haemoglobin and myoglobin genes in man was analysed using man-rodent somatic cell hybrids (Jeffreys et al., 1984). The single functional myoglobin gene was assigned to chromosome 22 and is therefore not linked to the α- or β-like globin genes but represents a third dispersed globin locus. By analogy with the α/β gene duplication it is possible that the haemoglobin/myoglobin duplication arose as a tandem duplication followed by dispersal to

separate chromosomes. In view of the tight linkage maintained between the α- and β-globin genes in Xenopus tropicalis it might be possible that other vertebrate groups would still maintain linkage between myoglobin and haemoglobin genes. It would be interesting therefore to establish the location of the myoglobin gene relative to the single α/β-globin gene cluster in Xenopus tropicalis.

Globin gene structure

The prototypical vertebrate globin gene structure is ancient and consists of three coding regions (exons) interrupted at precise locations by two non-coding regions (introns). This common structure is found in all vertebrate α- and β-globin genes (Fig 1.1, Jeffreys et al., 1983). A variation in this gene structure is found in plant globin genes specifying leghaemoglobin; a monomeric myoglobin-like protein which serves to keep oxygen tension low in nitrogen fixing root nodules to ensure an adequate supply of oxygen to oxygen-sensitive Rhizobium (Appleby, 1974). The leghaemoglobin gene contains a central globin intron in addition to the two introns normally found in vertebrate globin genes (Fig 1.1). The resulting second and third leghaemoglobin exons encode the structural modules which form the haem-binding site in leghaemoglobin, supporting the correlation between exons and protein functional and structural units in the theory that gene functions may have arisen by an assortment of various exons encoding different functions (Blake, 1978). In remarkable contrast, the larval haemoglobin gene of the insect Chironomus thummi thummi does not contain introns and has a signal sequence upstream required for the secretion of the globin into the insect haemolymph (Antoine and Niessing, 1984).

In summary, globin gene structure is very stable in vertebrates. It is not known whether the extra intron in the leghaemoglobin gene has

Figure 1.1

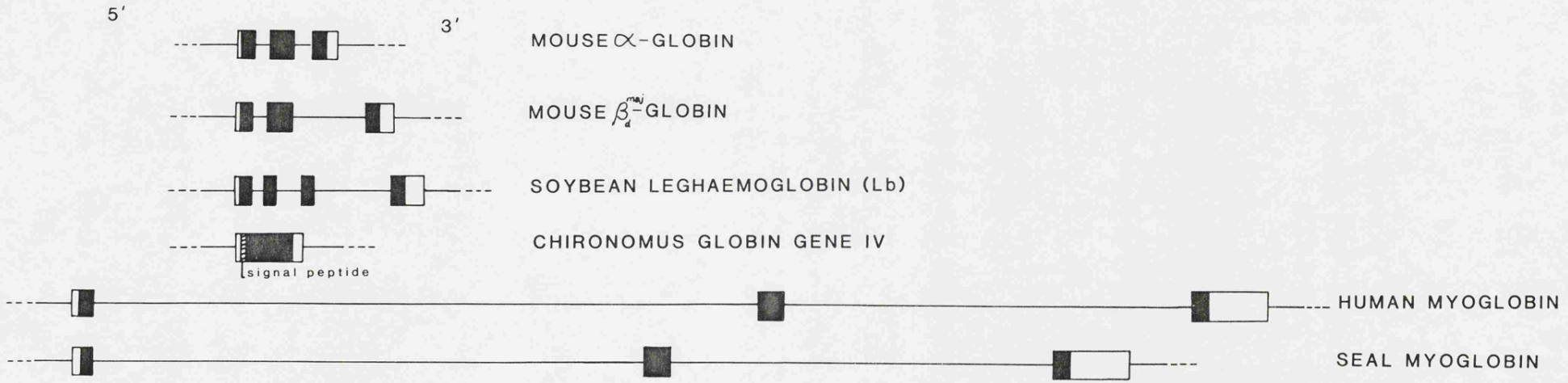
Some members of the globin gene family

The gene structure of some members of the globin gene family are represented diagramatically to scale, including, vertebrate α-and β-globin genes (Nishioka and Leder, 1979, Konkel, Maizel and Leder, 1979), a leghaemoglobin gene (Brisson and Verma, 1982), an insect globin gene (Antoine and Neissing, 1986) and two mammalian myoglobin genes (Blanchetot et al., 1983, Weller et al., 1984). Coding regions are represented as black boxes, nontranslated mRNA sequences as open boxes and flanking DNA as lines.

Note the much longer introns and 3'-nontranslated mRNA sequences of myoglobin genes.

Scale: ——

400 bp


5′
3′
MOUSE α-GLOBIN
MOUSE β_d^{min}-GLOBIN
SOYBEAN LEGHAEMOGLOBIN (Lb)
CHIRONOMUS GLOBIN GENE IV
signal peptide
HUMAN MYOGLOBIN
SEAL MYOGLOBIN

been gained recently by insertion into a globin-like gene in plants or whether an ancestral globin gene of plants and animals contained three introns and the central intron has been eliminated early in animal evolution. The finding that the leghaemoglobin gene is widespread in the plant kingdom tends to support the latter proposal (Landsmann et al., 1986). In the insect Chironomus, all globin introns have been lost and this may have occured by a similar mechanism as that which gave rise to processed pseudogenes (Vanin, 1984) with some kind of gene conversion event to maintain the gene transcription signals.

Gene lengths and intron sizes are remarkably well conserved in haemoglobin and leghaemoglobin genes (see Blanchetot et al., 1983). Mammalian β-like globin genes have intron sizes of 116-132 bp and 628-906 bp respectively while α-like globin genes have two similar sized introns, 103-104 bp in length although the human embryonic ζ gene has larger introns due to the presence of tandemly repeated sequences in the introns (Proudfoot, Gill and Maniatis, 1982). The stability in intron length of the α- and β-globin genes led to the suggestion that this might reflect a functional role for intron length rather than sequence (see Van den Berg et al., 1978).

The characterisation of the seal and human myoglobin genes led to some rather striking differences in the structure of myoglobin and haemoglobin genes (Fig. 1.1). Myoglobin genes are interrupted by introns at precisely the position found for the introns in α- and β- globin genes demonstrating that the three exon/two intron structure was established before the haemoglobin/myoglobin divergence 500-800 MY ago. However, the introns of myoglobin genes are much longer than their haemoglobin counterparts. Intron 1 of the human and seal genes is 5.8 and 4.8 kb respectively compared to 116-120 bp of the β-like globin genes. Similarly, intron 2 of the human and seal genes is much longer being 3.6

and 3.4 kb respectively compared to 628-906 bp for the β-like globin genes. The major increase in myoglobin intron length would tend to suggest that there is not a selective constraint on globin intron size unless myoglobin introns have become elongated as a result of adaption to their inclusion into a different developmental program. Another unusual feature of myoglobin genes is that they have very long 3'-nontranslated mRNA sequences.

Globin promoters and globin gene expression

i) The promoter region

A number of strategies are employed to deduce the sequences required for transcriptional promoting activity of eukaryotic genes transcribed by RNA polymerase II. First, sequence comparisons of several genes lead to the identification of conserved sequence elements in the 5'-flanking region of a large number of genes which could reflect common control sequences (for example, see Minty and Newmark, 1980) . Comparisons between the same gene in different species have identified similar conserved sequence elements which could reflect gene specific regulatory sequences (see Daubas *et al.*, 1985 for example). Secondly, mutations have been generated in the 5'-flanking region and the effect of the various mutations on transcriptional promoting activity tested in suitable *in vitro* expression systems. Thirdly, for some genes, genetic defects in gene expression have been found to be due to mutations in the promoter region, providing evidence for a role of the particular sequence in gene transcription.

There are two regions of DNA immediately upstream of the transcriptional start site which are highly conserved in both sequence and position; the "TATA" box located 25-30 bp before the mRNA CAP site (Breathnach and Chambon, 1981) and in some genes the "CAAT" box located

70-80 bp upstream of the CAP site (Efstratiadis et al., 1980). The TATA box appears to determine the precise start of transcription initiation both in vitro and in vivo (Breathnach and Chambon, 1981). In contrast, mutation in the CAAT box region has been shown to influence the efficiency of mRNA synthesis of many eukaryotic genes (Grosveld et al., 1982, Charnay, Mellon and Maniatis, 1985). A DNA binding protein has been identified that specifically binds to the CAAT motif in the Herpes simplex virus thymidine kinase promoter to increase transcription (Jones, Yamamoto and Tjian, 1985). Another less ubiquitous sequence motif identified is a GC-rich hexanucleotide found in a variety of viral and cellular promoter regions. A cellular transcription factor Sp1 stimulates transcription when it binds to the GC-rich region in virus promoters (Dynan and Tjian, 1985). The elements described above do not appear to determine the cell-type specific expression of particular genes or the inducibility of particular genes in response to extrinsic factors. The activity of many promoters appears to be modulated further by more specialised cis-acting elements. A class of control elements, "enhancers", first described in viruses are distinguished by their ability to dramatically activate transcription units independently of their orientation or their precise positioning relative to the transcription unit (Serfling, Jasin and Schaffner, 1985). A large number of cellular enhancers have now been identified exhibiting similar properties to the viral enhancers (Rogers and Saunders, 1985). Most of the genes that have associated enhancers are tissue-specific, developmentally regulated and/or inducible. For example, a tissue-specific enhancer located within 400 bp of 5'-flanking region in the insulin gene is required for the cell-specific expression of insulin in pancreatic islet cells (Walker et al., 1983).

A "typical" eukaryotic promoter therefore can be described as

being composed of several types of sequence elements: i)Enhancer elements or, for promoters which do not have enhancer elements, other specific upstream sequences, ii)more general sequence elements located 50-200 bp before the CAP site (e.g CAAT sequence) and iii)the TATA sequence and CAP site required for the accurate initiation of transcription. Each of these sequence elements is probably recognised by a DNA binding proteins (trans-acting factors) (Sassone-Corsi and Borrelli, 1986, Dynan and Tjian, 1985). Enhancer and gene specific upstream sequence binding proteins may be cell specific and each region could bind one or more of these proteins providing a wide range of specificities by different combinatorial effects (Brown, 1984). Proteins recognising the TATA and perhaps CAAT sequences may be less diverse since these elements are conserved for most promoters.

The promoters of "house-keeping" genes appear to deviate from the consensus organisation described above. For example, promoters of the genes encoding the HPRT and DHFR have a very high GC content and lack a TATA box. The GC-rich motifs of these genes, found shortly before the mRNA CAP site, are potential binding sites for transcription factor Sp1 and Sp1 has been shown to bind to the GC-rich region of the DHFR gene in vitro (Dynan, 1986).

Sequences related to the canonical TATA and CAAT boxes are highly conserved among mammalian α- and β-globin genes and are also found in leghaemoglobin genes (Efstratiadis et al., 1980, Proudfoot, 1979, Brisson and Verma, 1982). The TATA box has been shown to be required for specific transcription initiation in the β-globin gene and is also required for maximal promoter activity in vitro (Grosveld et al., 1982, Dierks et al., 1983). Dierks et al. (1983) identified an additional sequence element located ~100 bp before the CAP site in the β-globin promoter required for efficient transcription of the β-globin gene. This

region consists of a tandem repeat of a "CACCC" element and is found in the majority of other β-like globin genes. The physiological importance of these regions was further supported by the finding that a human β-globin gene containing an A→C transversion in the TATA box results in a β^+ thalassemia. Moreover, a β^o-thalassemia gene has been sequenced and shown to contain a C→G transversion in one of the CACCC regions (see Flavell and Grosveld, 1983)

Sequences further upstream of 100 bp have been shown to be conserved in β-globin genes which suggests that other sequences may be involved in β-globin gene expression than those described here (Moschonas, deBoer and Flavell, 1982).

The myoglobin promoter region has not been characterised by in vitro expression analysis but a comparison between myoglobin and haemoglobin promoters reveals that the myoglobin promoter is atypical. While a TATA box is present about 30 bp before the CAP site in both seal and human myoglobin genes, in the region containing the CAAT box and CACCC element of the β-globin promoter, human and seal myoglobin genes contain an unusually purine-rich sequence (>90% AG) termed the GAGA box (Blanchetot et al., 1983, Weller et al., 1984). These regions exhibit internal repetition, for example the seal sequence is $(GGA)_7$ $(GA)_6$.

ii) Globin gene expression

Vertebrate globin genes are expressed at different times during development. The eight functional globin genes (Section 1.2) are expressed to produce six haemoglobins ($\zeta_2\varepsilon_2$, $\zeta_2\gamma_2$, $\alpha_2\varepsilon_2$, $\alpha_2\gamma_2$, $\alpha_2\delta_2$, $\alpha_2\beta_2$). The analysis of haemoglobin promoter function was carried out in heterologous cell types and could not be expected to lead to an identification of all the sequences that function in the developmental regulation of the genes (see Introduction to Chapter 5).

Two developmental switches occur in haemoglobin production in man (see Collins and Weissmann, 1984, Karlsson and Nienhuis, 1985): the embryonic to foetal switch and the foetal to adult switch. The embryonic haemoglobins ($\zeta_2\varepsilon_2$, $\zeta_2\gamma_2$, $\alpha_2\varepsilon_2$) are produced in the placental yolk sac. There is a transition of erythropoiesis from the yolk sac to the foetal liver but the embryonic to foetal switch does not coincide with this as foetal haemoglobins ($\alpha_2\gamma_2$) are synthesised in the yolk sac also and the $\varepsilon \rightarrow \gamma$ switch (β-like globins) appears to occur slightly later than the globin gene switch ($\zeta \rightarrow \alpha$) in the α-cluster. Synthesis of adult β,δ-globins ($\alpha_2\delta_2$, $\alpha_2\beta_2$) occurs at low levels early in erythropoiesis but the major switch from foetal to adult synthesis occurs around birth. Once again the onset of erythropoiesis in the spleen or bone marrow does not coincide with any significant changes in the levels of γ- and β-globins. The adult α-globin is synthesised throughout foetal development and post-natal life.

An analysis of haemoglobin gene switches at the molecular level is difficult since no model systems are available which express all the globin genes. The introduction of globin genes into mouse erythroleukemia (MEL) cells has begun to provide some information on the sequences which might regulate developmental gene expression. MEL cells are transformed erythroid cells which appear to be arrested at an early stage of erythroid development. The addition of several chemical inducers to MEL cells leads to the induction of erythroid maturation; a variety of phenotypic changes occur including the activation of the adult β-globin gene. β-like globin genes transferred into MEL cells have revealed that there is a specific induction of adult β-globin gene expression during differentiation of these cells and human embryonic and foetal genes are not regulated (Wright et al, 1983), indicating that sequences within the transferred adult β-globin gene contain information

controlling this induction of gene expression. Gene transfer experiments into MEL cells also demonstrated that sequences both 5' and 3' to the mRNA CAP site of the β-globin gene are important in gene activation during MEL cell development (Wright et al., 1984). Sequences 3' to the initiation codon are also required for α-globin gene expression in this system (Charnay et al., 1984). The gene transfer experiments revealed that α-and β-globin genes are subject to different levels of control. The β-globin gene appears to be regulated at two distinct levels: it requires a "derepression" step which might be the formation of an active chromatin configuration from an inactive configuration, followed by an "activation" step which appears to require gene specific activators binding to specific sequences before the derepressed gene may be expressed. The α-globin gene however, appears to require only the "derepression" step.

In contrast to the MEL cell line, the human erythroid cell line K562 expresses the embryonic and foetal haemoglobin genes but only very low levels of the adult genes. This provides a complementary system to the MEL cell system which has been used to study the mechanisms of adult globin gene promoter inactivity in these cells (see Khazaie et al., 1986).

Transfection of globin genes into primary erythroid cells has been carried out recently (Hesse et al., 1986) which provides a more biologically relevant environment for gene expression analysis and globin genes may be introduced into primary cells at defined developmental stages. Transfection of chicken β-globin-CAT gene constructions into primary avian erythroid cells demonstrated that transferred genes can be expressed in these primary cultures and also led to the identification of a putative enhancer element in the 3'-flanking region of the β-globin gene. The presence of the enhancer sequence in the 3'-flanking region of

the chicken β-globin gene has been confirmed by Ok-Ryun and Engel (1986) and it appears to play a major role in the cell-specific and developmental-stage specific regulation of the adult β-globin gene.

Although useful information may be gained from these cell culture systems, it is more likely that information about developmental switches will come from a study of globin genes and modified derivatives after transfer into the genome of organisms at the very beginning of development. A hybrid 5'mouse-3'human β-globin gene which is specifically activated during MEL cell maturation contains the sequences required for tissue-specific expression and to remain inactive in primitive erythroid cells of transgenic mice and be specifically activated later in development in foetal and adult erythroid cells (Magram, Chada and Constantini, 1985). The construction of a variety of embryonic/adult, non-globin/globin gene hybrids and analysis in this system should generate information on specific sequences involved in developmental regulation. It is noteworthy that the human foetal γ-globin gene is expressed as an embryonic gene during development of transgenic mice (mice have no separate globin genes expressed specifically at the foetal developmental stage) suggesting that the human gene responds to trans-acting factors which activate the mouse embryonic globin genes (Chada, Magram and Constantini, 1986). Results using trans-species constructs must therefore be interpreted with care. It is likely, despite the demonstration that transferred globin genes containing limited amounts of flanking DNA can be expressed and regulated in the various systems, that the arrangement of the globin genes on their respective chromosomes, i.e. their "chromatin configuration" will also play a role in the developmental processes (see Collins and Weismann, 1984).

1.3 MUSCLE: FUNCTION AND STRUCTURE

Myoglobin is the principle haemoprotein of both skeletal and cardiac muscle in vertebrates. In contrast to the wealth of information concerning the expression of haemoglobin genes, relatively little is known about myoglobin gene expression. However, before describing what is known about myoglobin gene expression, the myogenic developmental system in which the myoglobin gene has been incorporated is considered.

Muscle function

Muscle is defined as "a machine that converts chemical energy to mechanical energy at constant temperature". In vertebrates, skeletal muscle is responsible for the production of rapid movements under voluntary control. Closely related forms, cardiac and smooth muscle, have become specialised for other involuntary movements such as heart pumping and gut peristalsis.

Skeletal muscle structure

i) The muscle sarcomere

Skeletal muscles consist of long, thin muscle fibres. These fibres are single, very large multinucleate cells formed by the fusion of many separate cells (see below). About two thirds of the mass of the fibre is made up of myofibrils which extend the entire length of the muscle fibre. Each myofibril is made up of a regularly repeating unit, the sarcomere, which is the basic unit of muscle function. When muscle is viewed in longitudinal section at high magnification, each sarcomere is seen as a series of dark and light stripes. The darker stripes are the A band and the lighter stripes the I bands while the dense line in the centre of the I band that separates one sarcomere from the next is the Z line (Fig.1.2). Sarcomeres of adjacent myofibrils are aligned in

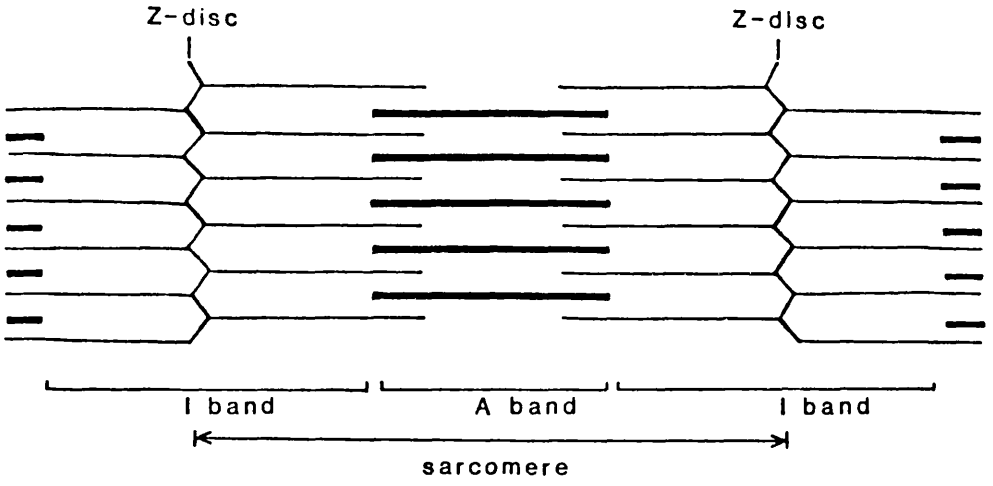
Figure 1.2

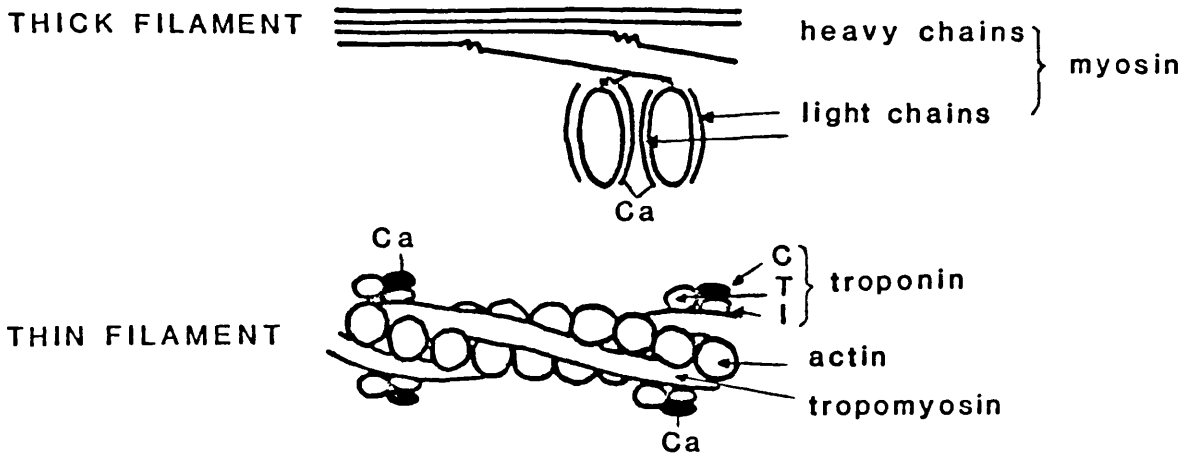
Top Schematic diagram of a sarcomere

The thick filaments of the A band are depicted by thick lines and the thin filament extending across the I bands and part of the way into the A bands by thin lines. The Z-disc which bisects the I bands is shown. Besides the major components of the sarcomere (contractile and regulatory proteins), there are other minor components in the sarcomere. Myomesin and M-protein are found in the centre of the A band (in the region called the M-line) (Perriard et al., 1986). C-protein is located in the A bands and is found as different isoforms in different muscle and fibre-types (see Obitana et al., 1986). In the Z-disc α-actinin and desmin are found and both play roles in the structure of the sarcomere. Muscle-form creatine kinase binds to the M line.

Bottom The major protein components of the muscle sarcomere

The interaction between the major contractile proteins of the sarcomere is shown (taken from Buckingham and Minty, 1983). Parts of the thick and thin filaments are shown to include one cross-bridge (domain of myosin molecule). The outer light chains are the alkali light chains, LC1 and LC3, and the inner light chains binding Ca^{2+} ions, the phosphorylatable light chain 2. Ca^{2+} ions bind to troponin C in the regulation of muscle contraction.


Z-disc
Z-disc
I band
A band
I band
sarcomere


THICK FILAMENT
heavy chains
myosin
light chains
Ca
Ca
C
T
I
troponin
THIN FILAMENT
actin
tropomyosin
Ca

register to give the whole muscle a striated appearance. The molecular basis for the cross-striations, represented diagrammatically in Fig.1.2, was revealed by electron microscopy. Each sarcomere was found to contain two sets of parallel and partially overlapping protein filaments, the thick filaments extending from one A band to the other and the thin filaments extending across the I bands.

The thin filament consists principally of sarcomeric actin, a representative of a highly conserved family of actins, members of which are found in all vertebrate and invertebrate cells (Buckingham, 1977). Actin associates as filaments consisting of two strands of globular molecules twisted into a helix and together with acessory proteins makes up the thin filament (Fig.1.2).

The thick filaments are made up principally of sarcomeric myosin. Myosin is also present in virtually every cell in the vertebrate and invertebrate body. Skeletal muscle myosin consists of two identical heavy chains (M.W. ~ 200,000 Daltons) and two pairs of light chains (M.W. ~20,000 and 16,000 Daltons). Myosin is able to hydrolyse ATP and its ATPase activity is stimulated by actin.

Contractile force is generated in the muscle sarcomere by the interaction of myosin and actin, a force which is in effect directional since it is exerted on the actin filaments which are anchored in the Z region. The thick and thin filaments slide past each other as the muscle contracts and the molecular basis for muscle contraction is detailed in a review by Huxley (1973). Force can only be generated if ATP is immediately available and the sarcomere is therefore associated with enzyme complexes that generate ATP from various substrates. These include the enzymes of the glycolytic pathway, mitochondria for oxidative ATP formation and creatine kinase which is able to generate ATP from the metabolism of the phosphocreatine stores in muscle.

Muscle contraction is initiated in the sarcomere by a rise in the concentration of intracellular Ca^{2+} ions released by the sarcoplasmic reticulum in response to a nervous stimulation of the muscle fibre. The excitation of the fibre is initiated at the motor end plate (junction between the nerve and muscle), spreads rapidly over the surface of the fibre and is transmitted into the interior by the T tubules. The excitation is coupled to the sarcoplasmic reticulum (SR) and the SR responds by release of stored Ca^{2+} ions from its interior compartment into the myoplasm (Endo, 1977). Ca^{2+} ions diffuse and bind to a set of specialised proteins associated with the actin filaments, tropomyosin and the troponin complex, to initiate muscle contraction (reviewed by Adelstein and Eisenberg, 1980).

Tropomyosin is a dimeric double stranded α-helical rod which is aligned nearly parallel to the long axis of the thin filament (Fig.1.2). Troponin (Tn) is a complex of three polypeptide chains: Tn C, Tn I and Tn T. The Tn complex is located at regular intervals along the thin filament, the period being set by tropomyosin (Fig.1.2). Tn T binds to tropomyosin while Tn I binds to actin and Tn C to Ca^{2+} ions. The interaction of actin and myosin is inhibited by troponin and tropomyosin in the absence of Ca^{2+} ions. Released Ca^{2+} ions bind to Tn C and cause conformational changes that are transmitted to tropomyosin and then to actin allowing the interaction of myosin and actin and hence contraction to occur (Hitchcock, Huxley and Szent-Gyorgi, 1973). Contraction ceases when Ca^{2+} ions are removed from the regulatory binding sites.

While this section has described the structure and contraction of skeletal muscle, the two other major types of muscle, cardiac muscle and smooth muscle, also contain actin and myosin and contract by a sliding filament mechanism. Cardiac muscle has a striated appearance which

reflects a very similar organisation of the actin and myosin filaments. Smooth muscle, however, has no striations and consists of long tapering cells that have one nucleus and contain both thick and thin filaments aligned along the axis of the cells. The thick and thin filaments in this case are not arranged in a strictly ordered pattern found in skeletal and cardiac muscles and do not appear to form myofibrils. The movement of the filaments in smooth muscle is dependant on a calcium ion-regulated phosphorylation of myosin (Adelstein and Eisenberg, 1980).

ii)The fibre-type diversity of skeletal muscle

The skeletal muscle fibres described above taken from an adult mammal do not all look alike. Close inspection at the molecular and biochemical level reveals that skeletal muscle is an extremely heterogenous tissue. The diversity results from a spectrum of fibres with different physiological and metabolic properties which are suited to the wide range of functional demands on muscle. In spite of the heterogeneity, some broad principles do begin to emerge when fibres are analysed by a number of criteria and this has led to several systems suggested to classify these different fibre-types (see Eisenberg, 1985).

Immunological studies and experiments on isolated muscle fibres have established that in adult mammals two major fibre types can be distinguished which are composed of different forms of the contractile proteins (for example, see Dhoot and Perry, 1979). Certain muscles are predominantly made up of one or other of these fibre types and such muscles have been shown to have a relatively "fast" or "slow" time course of contraction (Burke et al., 1971). The contractile proteins found in these fibre types are referred to as fast and slow forms and the fibres as fast (Type II) or slow (Type I). Barany (1967) demonstrated that in adult muscles the maximum velocity of contraction is directly

proportional to the ATPase activity of the actomyosin extracted from the muscle, thereby establishing an important functional relationship between the physiology of the muscle and the proteins present. Acid preincubation of muscle fibres has allowed histochemical separation of two fast subtypes of skeletal muscle fibres, typeIIA and IIB, as the myosin ATPases they contain use ATP differently after incubation in solutions of altered pH (Brooke and Kaiser, 1974).

Type I and Type II fibres may be further subdivided on the basis of their metabolic properties and perhaps the most straightforward classification is that proposed by Peter et al. (1972) which incorporates both the contractile and metabolic properties in the nomenclature. They describe three main fibre types refered to as slow-twitch oxidative (SO), fast-twitch oxidative-glycolytic (FOG) and fast-twitch glycolytic (FG) fibres, although recognising that there are some problems in reliably differentiating between FOG and FG fibres using histochemical analysis for mitochondrial oxidative enzymes (Nemeth et al., 1979). FOG fibres are thought to contain the Type IIB myosin isoform and FG the Type IIA isoform although this classification breaks down for a number of muscles (Jolesz and Sreter, 1981).

SO fibres have an increased reliance on oxidative phosphorylation as an energy source and have a high mitochondrial content. These fibres have a rich capillary supply and are red in colour, reflecting their high levels of myoglobin (Peter et al., 1972). In fully grown mammals, SO fibres have a smaller diameter than FG fibres reflecting their greater dependance on oxygen delivery and a need for a smaller oxygen diffusion distance (Zak, 1981).

FG fibres have a low oxidative capacity and increased levels of glycolytic enzymes and metabolise glycogen by anaerobic glycolysis to produce energy. They have a sparse blood supply and a whiter colour.

The T-tubule system is more extensive in FG fibres than in SO fibres (Eisenberg, 1985).

FOG fibres tend to be intermediate between SO and FG fibres in their metabolic enzyme and myoglobin contents.

This fibre typing system should not be taken as a rigid classification since with any classification, muscle fibres demonstrate a continuum of overlapping properties with respect to their various constituents (Pette, 1980). The balance between the different enzymes and proteins of a fibre matches the metabolic and structural requirements of the fibre to carry out work. SO fibres are recruited for sustained force generation and are fatigue resistant whereas FG fibres, with a less efficient metabolic system, are reruited occasionally for peak force production and are less fatigue resistant than both FOG and SO fibres. FOG fibres with higher levels of myoglobin and oxidative enzymes are more fatigue resistant.

Cardiac muscle which contracts continuously, has a high capacity for aerobic metabolism, obtaining energy almost entirely from oxidation of fatty acids. Consequently heart muscle has high levels of myoglobin and many mitochondria (Holloszy and Booth, 1976).

Most mammalian muscles are composed of mixed populations of the different fibre types but some muscles are predominantly composed of one fibre type, for example the soleus muscle in many mammals is predominantly SO fibres.

Muscle fibres are organised into motor units. A motor unit consists of a single motor neuron and many muscle fibres. Fibres are called into action by their motor neuron and contract in an all or nothing response. The nerve connection to the particular fibre must be matched to the appropriate type of fibre; the axon transmitting commands for quick, intermittant action for example, must activate FG fibres.

Buller, Eccles and Eccles (1960), showed convincingly that certain properties of the muscle were dependant on the source of its innervation. Denervating a fast and a slow muscle and then cross-reinnervating the two muscles resulted in the contraction time of the slow muscle becoming more like that of a fast muscle and vice versa. Later experiments with this system demonstrated that the impulse frequency and the quantity of impulses were responsible for the differences in the speed related properties of slow and fast muscles (reviewed in Salmons and Sreter, 1976). Motor units of adult mammals are physically and biochemically homogeneous (Pette, 1985). These observations have led to the hypothesis that it is the specific neural input which is of primary importance in regulating the phenotype expression (both contractile and metabolic properties) of the muscle fibre (Pette, 1985).

Contractile protein isoforms

Almost every protein of the contractile-regulatory myofibril has isoforms (see Buckingham, 1977, 1985 and Buckingham and Minty, 1983 for reviews). Differences in performance that exist between the adult muscle fibres are in part determined by the presence of these different isoforms and this is obvious in the case of myosin isoforms where there is a correlation with the speed of contraction and the V_{max} of the myosin ATPase.

The multiple isoforms of contractile proteins are encoded by members of multigene families each of which probably arose by gene duplication during evolution (see Buckingham, 1985). The individual members of the gene families are subject to fibre-specific, tissue-specific and developmental stage-specific regulation. For example, the actins comprise a group of proteins which are highly conserved in evolution and six actin isoforms have been identified in

mammals and these isoforms differ from each other by only a few amino acids (Vandekerckhove and Weber, 1978,1979). Two smooth muscle (α-and ϒ-smooth) and two striated muscle (α-skeletal and α-cardiac) isoforms are components of the sarcomere and two non-muscle isoforms (β- and ϒ-cytoplasmic) are components of the cytoskeleton. The non-muscle isoforms are found in most cells in the vertebrate body while muscle isoforms are found only in particular muscle types.

Non-muscle isoforms also exist for myosin heavy chain, myosin light chain and tropomyosin proteins (Buckingham, 1977).

At least 10 myosin heavy chain (MHC) isoforms exist in vertebrates and the genes encoding all 10 different isoforms have been isolated (Mahdavi et al., 1986). Seven of the isoforms are found in skeletal and cardiac muscles and are the predominant myosin isoforms in the different adult muscle tissues (for example, cardiac α-MHC, cardiac β-MHC = MHC_{slow}), in different adult fibre types (for example, $MHC_{fastIIA}$, $MHC_{fastIIB}$, MHC_{slow}) and specific isoforms can also be demonstrated in developing muscle tissue (for example, $MHC_{embryonic}$, $MHC_{neonatal}$ Whalen et al., 1981). It is interesting that the same gene encodes the heavy chain expressed in slow contracting muscle and in ventricular cardiac muscle in rats, particularly since the gene may be regulated independantly in the different striated muscles; the gene is expressed in cardiac tissue at the foetal stage while it is the major adult form in skeletal muscle (Lompre, Nadal-Ginard and Mahdavi, 1984).

The diversity of the contractile proteins is in some cases augmented by alternative RNA splicing of individual gene transcripts of different members of the multigene families (Nadal-Ginard et al., 1986).

Isoforms of tropomyosin are composed of two subunits α and β which are found in different molar ratios in individual muscle types, for example ratios of α:β of 3.8:1 and 1:1 in fast and slow fibres

respectively. S1 nuclease mapping studies of rat muscle RNA have revealed the existence of three distinct α-tropomyosin messenger RNAs, each encoding a different protein (Ruiz-Opazo, Weinberger and Nadal-Ginard, 1985). The isoforms are tissue-specific and developmentally regulated. The three mRNAs are produced by alternative splicing of primary transcripts which have different 3'-nontranslated sequences due to the use of different polyadenylation sites but are otherwise identical. The differences in the primary transcript structure may be sufficient to dictate the alternative splice pathways. Alternative splicing is also used as a mechanism to generate myosin light chains 1 and 3 in fast skeletal muscle (Nabeshima et al., 1984, Robert et al., 1984), three developmentally regulated myosin heavy chain gene transcripts in Drosophila (Rozek and Davidson, 1983) and to generate multiple, regulated troponin (Tn) T isoforms. Different Tn T isoforms are present in different muscle tissue types and at various developmental stages (see Dhoot and Perry, 1980). Some of this diversity is generated by differential transcription of separate genes for skeletal fast, skeletal slow and cardiac Tn T, with additional isoforms generated by differential RNA splicing. Breibert et al. (1985) demonstrated that a minimum of ten Tn T isoforms can be produced by the single rat skeletal fast gene in a tissue specific and developmentally regulated manner. They have also shown that for at least two of these isoforms, the mRNA transcript is identical, and therefore in this case differential splicing must involve control by trans-acting factors.

Even with the few examples described here, muscle cells demonstrate the capacity to generate a large number of isoforms by differential transcription of members of a multigene family and subtly different isoforms by alternative splicing of particular members of the multigene family. Both greatly increase the potential for sarcomere

diversity and therefore allow the sarcomere to perform specialised functions but despite documentation of the large numbers of muscle isoforms, data relating the structure of a particular isoform with a physiological role in muscle are lacking.

Different isoforms also exist for enzymes of muscle metabolism (Pearson, 1980). No isoforms exist for myoglobin (Section 1.2) and interestingly, enzymes involved in aerobic metabolism in muscle (for example, the β-oxidation enzymes, electron transfer enzymes) are not found in isozymic forms (Hochachka, 1985).

1.4 THE DEVELOPMENT OF MUSCLE

Introduction

In this Section, the development of vertebrate skeletal muscle is described but the general pattern of myogenesis in cardiac and skeletal muscles is similar (see Zak, 1974 for a more detailed description of cardiac muscle development).

Muscle development occurs in two general phases (see Fig.1.3). In the first phase, a mesodermal stem cell gives rise to a determined myoblast destined for myogenesis. Precursor myoblast cells fuse spontaneously with other myoblasts to form multinucleated myotubes and this morphological differentiation is accompanied by the activation of many muscle-specific proteins including the muscle isoforms of the contractile proteins. The initial formation of skeletal muscle *in vivo* in the mouse is most easily followed in the limbs. Limb bud formation occurs at 11-12 days in the mouse and the onset of muscle contractility is observed from 15-16 days (Ontell, 1982). This stage of muscle development may be studied in cell culture where the initial formation of myotubes by fusion of dividing myoblasts takes place fairly synchronously compared to the asynchronous fusion in the embryo.

Figure 1.3

The major phases of muscle myogenesis

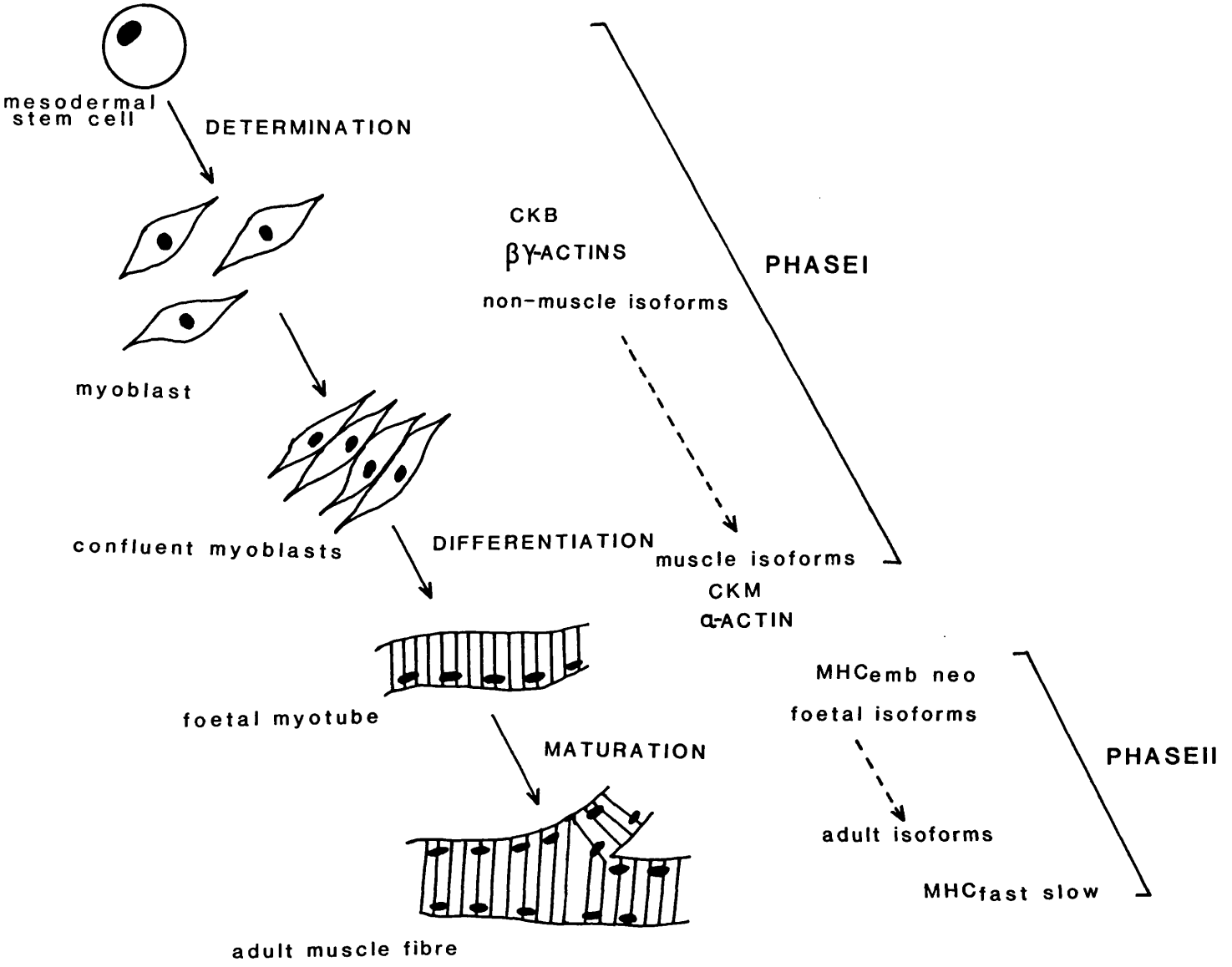
Phase 1

A mesodermal stem cell gives rise to a committed myoblast (determination). Myoblasts proliferate until they become confluent, withdraw from the cell cycle and then fuse to form myotubes with the synthesis of a battery of muscle specific proteins (differentiation).

Phase 2

The immature, multinucleated myotube (foetal myotube) undergoes a series of phenotypic and morphological transitions before the adult myotube is formed, containing the adult isoforms characteristic of the fully differentiated fibre (maturation).

Isoform transitions of muscle specific proteins occuring during muscle development include the non-muscle creatine kinase isoform (CKB) to the muscle creatine kinase isoform (CKM) and β,γ-cytoplasmic actin isoforms to the α-skeletal actin isoform during myoblast differentiation, and MHC_{neo} and $MHC_{embryonic}$ isoforms found early in development to MHC_{slow}, MHC_{fast} isoforms characteristic of the different adult muscle fibres.


mesodermal stem cell
DETERMINATION
myoblast
confluent myoblasts
DIFFERENTIATION
foetal myotube
MATURATION
adult muscle fibre
CKB
βγ-ACTINS
non-muscle isoforms
PHASEI
muscle isoforms
CKM
α-ACTIN
MHCemb neo
foetal isoforms
adult isoforms
MHCfast slow
PHASEII

The second phase of muscle development involves the maturation of the formed myotubes to acquire the properties of the different adult muscle fibres. Since these processes require hormonal and neuronal factors, this phase of development is probably best studied in animal models so that the influence of the various extrinsic factors on muscle maturation may be determined.

Phase 1. Induction of myogenic differentiation

i)Myoblast commitment

It has been suggested that the process of myoblast commitment is under fairly simple control after analysis of commitment in the cell line 10T1/2 (Konciezny and Emerson, 1985). Taylor and Jones (1979) first reported that exposure of 10T1/2 cells to 5-azacytidine to induce hypomethylation caused a conversion of these cells into determined muscle, fat and cartilage cell lineages which are able to retain their determined capacity to differentiate when analysed in clonal cell culture. A clonal cell lineage analysis of 5-azacytidine treated 10T1/2 cells (Konciezny and Emerson, 1984) established that 10T1/2 is a multipotential stem cell blocked in the expression of three independent pathways of mesodermal determination. 5-azacytidine removes this block by the demethylation of specific loci to produce genetically stable lineages of myoblasts, chondrocytes and adipocytes. Myoblasts arise most often after 5-azacytidine treatment which led to the suggestion that undermethylation of one or a few loci is sufficient to convert 10T1/2 cells to the myoblast lineage. Recently, Lasser, Paterson and Weintraub (1986) have demonstrated that genomic DNA isolated from myoblasts derived from 5-azacytidine treated 10T1/2 cells is able to convert transfected parental 10T1/2 cells to the myogenic lineage, whereas the same cells transfected with 10T1/2 genomic DNA cannot be converted to the myoblast

lineage. These findings imply that the 10T1/2 cell is only one step away from being a committed myoblast and raises the exciting possibility that a specific locus regulating myoblast commitment could be identified. As these authors point out, until this locus is cloned and characterised, a number of possibilities exist as to its exact function.

ii)The myoblast to myotube transition

Tissue culture techniques have dramatically advanced analysis of the biochemical events involved in the transition of dividing myoblasts to form post-mitotic multinucleated myotubes. A large number of permanent myoblast cell lines are now available, mainly derived from rodents (see Yaffe, 1969 and Pearson, 1980 for reviews) and although there may be differences with the *in vivo* situation (for example, see Hickey *et al.*, 1986), the fact that myoblasts consistently undergo fusion to form myotubes *in vitro* provides a useful model system for the early stages of myogenesis.

Myoblasts proliferate until they withdraw from the cell cycle and cease DNA synthesis (commitment) and then fuse in the G1 compartment of the cell-cycle to form post-mitotic myotubes (terminal differentiation). Concomitantly with the cellular events, a battery of myofibrillar genes is expressed (biochemical differentiation -see below). Permanent cell cultures apparently condition the medium before differentiation and remove a factor which inhibits differentiation (Linkhart, Clegg and Hauschka, 1981). Manipulation of medium composition is used to control the synchrony and extent of differentiation: "growth" medium is replaced with a low serum "differentiation" medium (Konigsberg, 1977).

Myoblast fusion is very cell specific, fusion of rat heart or kidney cells and rat myoblasts will not occur whereas rat and rabbit myoblasts are able to fuse (see Wakelam, 1985). The guidance of myogenic

cells during alignment and before fusion appears to be provided in part by fibronectin (Chiquet et al., 1981). Recognition and alignment is followed by adhesion and membrane union and irreversible adhesion is obligatory for fusion. The initiation of fusion is accompanied by the establishment of electrical continuity in the cells and membrane structures similar to gap junctions have been shown to be formed in vivo (Rash and Staehlin, 1974).

The factors which induce myotube formation remain obscure although the role of a number of molecules which influence the process have been described. Transferrin, for example, is required for both growth and differentiation of myoblasts (Ozawa et al., 1986) and insulin and somatomedins appear to stimulate both myoblast proliferation and fusion and may well play a role in the normal development of muscle (Ewton and Florini, 1981). Like many other fusion systems myoblast fusion exhibits a specific requirement for Ca^{2+} ions. Calcium deficient medium permits proliferation and alignment of myoblasts and myoblasts achieve fusion competence but does not allow fusion to occur. Re-introduction of Ca^{2+} ions to the medium causes rapid and extensive fusion which provides a means of acheiving synchronous cell fusion (Turner et al., 1976). Fusion involves modifications of lipids and membrane proteins (see Wakelam, 1985 for a review). As myoblasts fuse, the muscle cell surface proteins undergo many changes in location and molecular weight. If metalloendoproteinase substrates or inhibitors are added to fusion competent myoblasts at the same time as calcium addition, myoblast fusion is prevented, which led to the proposal of a role of a metalloendoprotease in myoblast fusion (Couch and Strittmatter, 1983). Strittmatter and Couch (1986) have subsequently isolated the soluble metalloendoproteinase activity which is required for myoblast fusion.

iii) Myoblast differentiation and the role of the cell cycle

Linkhart et al. (1981) have demonstrated, using a permanent clonal myoblast cell line (MM14), that myoblast proliferation and terminal differentiation are coordinately controlled by mitogens. When myoblasts deplete the medium of a specific mitogenic activity, extensive myotube formation and elaboration of muscle-specific gene products occur. In support of these findings, Compton, Merrifield and Konigsberg (1986) have shown that cell arrest in G1 is specifically caused by the depletion of growth factors in the medium and this arrest is necessary for initiation and maintenance of muscle-specific gene expression. These groups therefore propose that myoblast differentiation is not rigidly determined but can be regulated to a remarkable degree by extrinsic factors, mainly the level of mitogens in the medium. At every mitosis, myoblasts have the option to either differentiate or re-enter the cell cycle, the decision being dependent upon the level of growth factors in the medium. Other models to explain the timing mechanism propose that myoblasts are genetically programmed to undergo a limited number of divisions and a unique "quantal" mitosis is required before myoblasts are able to differentiate (Holtzer, 1978). Quinn, Holtzer and Nameroff (1985) prefer an intrinsic lineage-based program as a model for myoblast differentiation. Specifically their model proposes that a self-renewing stem cell exists in the muscle lineage which can generate committed myoblast precursors. After commitment to terminal differentiation precursor cells "count" exactly four cell divisions prior to the generation of post-mitotic differentiated muscle cells and this pattern of terminal divisions is not affected by growth in mitogen depleted medium. Although their model successfully accounts for the asynchronous time course of terminal differentiation in developing muscles and mass cultures, other results would argue against it.

It has been demonstrated by Nguyen, Medford and Nadal-Ginard (1983) using a myogenic cell line that is temperature sensitive for fusion and irreversible withdrawal from the cell cycle and by manipulation of the Ca^{2+} concentration in the culture medium of wild-type myoblasts (Emerson and Beckner, 1975) that muscle-specific genes can be induced in myoblasts in the absence of cell fusion. The mono-nucleated but biochemically differentiated cells can be stimulated to re-enter the cell cycle in the presence of high levels of growth factors and muscle-specific genes are then deinduced. The reversibility of muscle gene induction indicates that withdrawal from the cell cycle is neither necessary nor the result of muscle-specific gene induction; these two events are independent, neither alone is sufficient and both are required for the production of the terminally differentiated phenotype. The reversibility of muscle gene induction would tend to exclude an irreversible terminal division mechanism of differentiation. Also, the finding that myoblasts arrested in G1 can be subsequently stimulated to differentiate in the absence of DNA synthesis (Nadal-Ginard, 1978, Whalen et al., 1986) is contrary to a "mitotic-clock" model. The controversy concerning the role of the cell cycle and myoblast differentiation continues but as Quinn and Nameroff (1986) point out differences may be due to the use of established versus primary cell lines.

iv) Induction of muscle-specific gene expression

The visual fusion of myoblasts to form multinucleated myotubes is accompanied by a series of biochemical changes (see Pearson, 1980 for a review). Myotubes acquire a different phenotype than that of myoblasts as demonstrated by differences in mRNA and protein populations (Affara et al., 1980a,b, Garrels, 1979). The major proteins of the contractile apparatus accumulate in large amounts and are organised into the

sarcomere. Although myosin, actin and tropomyosin are found in all other eukaryotic cells (see Section 1.3), the proteins found after fusion are specifically muscle isoforms with the concomitant decrease in the non-muscle isoforms, and therefore a qualitative change in the type of contractile protein occurs.

Similar isoform shifts are observed in a variety of enzymes of intermediary metabolism. New forms of adenylate kinase, aldolase, glycogen phosphorylase and glycogen synthase appear (Pearson, 1980). Most of these enzymes undergo a qualitative and quantitative increase on fusion. For example, creatine kinase, is found as the muscle specific MM form after fusion changing from the BB (Brain) form in myoblasts and the total enzymatic activity increases about 600 fold in less than 48 hours in the mouse cell line, MM14 (Chamberlain, Jaynes and Hauschka, 1984).

As mentioned above, changes also occur in the composition of the plasma membrane during fusion. Invaginations appear which develop into the T tubule system (Pearson, 1980) and one of the most striking changes is the increase in the amount of acetyl choline receptor required for the nerve-muscle interactions (Merlie et al., 1975).

Devlin and Emerson (1978) demonstrated that primary cultures of quail myoblasts synthesise many of the major contractile proteins in a coordinate fashion around the time of myoblast fusion and in later studies (Hastings and Emerson, 1982) cDNA clone analysis revealed that a coordinate accumulation of corresponding contractile protein mRNAs occurred. The availability of a large number of cloned muscle specific gene probes has supported the findings that muscle gene mRNAs accumulate rapidly and non-muscle mRNAs cease to accumulate, the appearance of the transcripts closely following the appearance of the proteins. Coordinated accumulation has not been reported in all cases however. There have been reports of significant delays between the appearance of

some transcripts. Thus there is a delay of 24 hours between the appearance of myosin light chains and myosin heavy chain mRNAs (Caravatti et al., 1982) and a similar 24 hour delay has been reported between the appearance of myosin light chains and α- and β-tropomyosins (see Whalen, 1980). These findings suggest that perhaps more than one coordinated event is required for the induction of all muscle specific genes during differentiation.

In both established cell lines and primary cell lines in culture, the transition that occurs is from the non-muscle phenotype (myoblast) to the muscle phenotype found in embryonic/foetal skeletal muscle and the associated events are able to take place in the absence of innervation.

v) Regulatory mechanisms involved in muscle gene activation

A number of observations have suggested that the induction of muscle-specific genes appears to be determined mainly at the level of transcription (Buckingham, 1985) which appears to be the case for a large number of moderately expressed tissue-specific genes (Derman et al., 1981). Carmon et al. (1982) demonstrated that the genes coding for contractile proteins are relatively insensitive to DNase I in nuclei from dividing myoblasts and become more sensitive to DNase I in nuclei from muscle fibres. A shift to increased DNase I sensitivity of a gene appears to be associated with a progression from an inactive transcriptional state to an active state (Mathis, Oudet and Chambon, 1980). The accumulation of both muscle specific proteins and their mRNAs in myotubes and their absence in myoblasts also suggests that regulation is primarily at the level of transcription. This has been demonstrated directly for a number of genes (see summary of Chapter 4); for example, MHC gene transcription was studied in isolated nuclei which demonstrated that a major increase in transcription occurred at the time of cell

fusion (Medford, Nguyen and Nadal-Ginard, 1983).

The activation of transcription of apparently unrelated genes during the terminal differentiation provides an interesting system in which to analyse the sequences required for this stage-specific (and tissue-specific) expression as well as to search for global regulatory mechanisms which may be acting on these genes. Transfer of muscle-specific genes into proliferating myoblasts is beginning to provide information on control of gene activation at the molecular level.

One of the first transfections of a muscle gene into muscle cells in culture revealed that sequences immediately proximal to muscle genes are important for their developmental regulation. Melloul et al. (1984) introduced a hybrid 5'skeletal muscle α-actin 3'globin gene fusion stably into the genome of a rat myoblast cell line (L8). In many of the isolated clones which had integrated the gene fusion, expression of the fused gene was increased upon subsequent differentiation of the myoblast clones to form myotubes. This was shown at the level of specific mRNA transcripts. Also, the 5'skeletal actin portion of the gene fusion when fused to the gene coding for bacterial chloramphenicol acetyl transferase (CAT) could direct high levels of CAT activity in myotubes indicating that sequences contained in the 5'-flanking region were sufficient for tissue- and stage-specific expression. The increase in activity was specific to the α-actin gene fusion and only observed in differentiating myoblasts. The authors were also able to demonstrate that correct regulation of the rat muscle gene fusion occured after introduction into chick myoblasts, indicating that the regulatory sequences in the 5'-flanking region have been maintained over long evolutionary periods and indeed sequence comparisons of the 5'-flanking regions of these genes reveals highly conserved sequence elements (Nudel et al., 1985). A similar transfection of a skeletal actin gene into a mouse myoblast cell

line C2C12 however, failed to show any increase in expression from the exogenous gene during differentiation (Seiler-Tuyns, Eldridge and Paterson, 1984). A human α-cardiac actin gene fusion was introduced into L6 and C2C12 myogenic cell lines by Minty and Kedes (1986) and subsequent deletion analysis of the cardiac actin promoter and 5'-flanking region revealed that two separate sequence elements appear to modulate the expression of this gene in a region located 480 to 40 bp before the CAP site. The more 5' of the two sequences contains a sequence similar to those found in viral and cellular enhancers and is required for high level, tissue-specific expression of the α-cardiac actin gene. This region does not behave as a "normal" enhancer in that it is unable to activate heterologous promoters and is orientation dependant, only increasing expression from the cardiac actin promoter when at a defined distance 5' of the promoter. The second sequence element located downstream consists of four regions with a common sequence pattern which are highly conserved in sequence but not in number or location in other members of the actin gene family. The common sequence element is based on CC(Ar)GG (Ar="A-rich" sequence) and in other actin genes this sequence had previously been assigned as the actin CCAAT box. Minty and Kedes (1986) propose that the CCAAT box(es) in actin genes play a role in tissue specific transcription rather than in basal gene expression.

When the human cardiac actin gene fusion was introduced into C2C12 myoblasts, Minty, Blau and Kedes (1986), observed a similar result to that of Seiler-Tuyns et al. (1984). Induction of expression of the gene fusion was not observed during differentiation of this cell line in contrast to activation of expression in the L6 myoblasts, but expression was constitutive in both C2C12 myoblasts and myotubes. Minty et al. (1986) proposed that C2C12 myoblasts contain muscle specific factors which interact with upstream regions of the cardiac actin gene (probably

the regions defined above) and that these factors are present at similar levels in both C2C12 myoblasts and myotubes whereas in "normal" (L6) myoblasts they only accumulate in the differentiated state. The endogenous cardiac actin gene is not activated in C2C12 myoblasts, (but see Melloul *et al.*, (1984)) it therefore follows that the gene is probably subject to a second level of regulation that represses the endogenous gene but not the transfected gene. This could be some form of constraint at the level of chromatin structure. Therefore expression of the cardiac actin gene requires modifications necessary in a first stage to "allow" the gene to interact in a second stage with factors which modulate gene transcription. A similar model has been proposed by Charnay *et al.* (1984) for the regulation of the β-globin gene (Section 1.2) and Melloul *et al.* (1984) to explain the "less stringent" control of transferred muscle genes in myoblasts.

Bergsma *et al.* (1986) have transferred a chicken α-skeletal actin gene fusion into primary chick myoblasts and have found that only 200 bp of 5'-flanking DNA are necessary for the tissue specific and developmental stage specific expression of this gene. When the gene fusion is injected into *Xenopus* oocytes, 5'-flanking DNA can be reduced to 107 bp and this contains the sequences required for the non-regulated transcription of the gene.

In the first example of gene transfer into myoblasts in culture of a non-contractile protein gene, Jaynes *et al.* (1986) have demonstrated that a muscle creatine kinase gene fusion, containing about 3 kb of creatine kinase gene 5'-flanking region fused to the CAT coding sequence, is expressed and appropriately regulated during myoblast differentiation *in vitro*.

Konieczny and Emerson (1985) stably transfected a quail TnI gene 6.4 kb in length (530 bp of 5'-flanking DNA) into 10T1/2 cells and 10T1/2

derived myoblasts (Section 1.4) and compared the expression after determination and differentiation of the multipotential stem cell and differentiation of the determined myoblasts. They demonstrated that the transfected and endogenous genes were transcriptionally inactive in both 10T1/2 cells and myoblasts but when both differentiated to form myotubes, the endogenous and exogenous genes were activated identically which would tend to suggest that the Tn I gene does not have to go through the determined stage in order to be expressed appropriately.

The existence of diffusible *trans*-acting factors which regulate muscle gene activation has been demonstrated in experiments involving the fusion of myotubes with non-muscle cells to form heterokaryons. Fusion of myotubes with terminally differentiated chondrocytes, keratinocytes, hepatocytes (Blau *et al.*, 1985) and neural cells (Wright, 1984) demonstrated that each of these cell types can be induced to express muscle genes indicating the remarkable degree of plasticity of the nucleus of differentiated cells. The nuclei in heterokaryons remain separate and intact and therefore activation of gene expression is likely to be due to diffusible *trans*-acting factors. In the non-muscle cell the nuclei remain post-mitotic and skeletal muscle genes are therefore activated regardless of their chromatin configuration. Interestingly, fusion of myotubes and HeLa cells requires the pre-treatment of HeLa cells with 5-azacytidine before muscle genes may be induced, suggesting that at least in one case muscle genes must be altered in some way (probably demethylation of essential bases) before genes may be expressed (Chui and Blau, 1984).

The experiments described suggest that muscle-specific genes are regulated in a similar manner to other eukaryotic genes in that *cis*-acting sequences generally present in the 5'-flanking region of the gene are most likely recognised and bound by *trans*-acting factors to

control gene expression.

In contrast to the information regarding gene activation during myoblast differentiation, relatively little is known about the mechanism(s) by which certain genes are switched off during differentiation. Thymidine kinase (tk) is a replication associated enzyme whose level of activity disappears rapidly on myoblast differentiation. Transfection of a tk gene into differentiating myoblasts revealed that sequences responsible for the down-regulation of tk gene expression lie entirely within the gene itself (Merril, Hauschka and McKnight, 1984).

A similar analysis with a transfected β-cytoplasmic actin gene demonstrated that sequences in the 3'-flanking region of the gene are required for the down-regulation of β-actin gene expression observed during myoblast fusion (Paterson, Eldridge and Seiler-Tuyns, 1986).

Phase 2: maturation of muscle fibres

The production of embryonic/foetal myotubes is followed by a period of maturation to produce an adult fibre (Fig. 1.3). Once again striking changes can be observed at the morphological and biochemical levels.

i) Changes at the morphological level

The formation of initial primary myotubes is autonomous, occurring in the absence of the nerve which does not penetrate the muscle until later development (Harris, 1981). Primary myotubes separate to become independent units of contraction and when this occurs, a secondary generation of myoblasts becomes apparent which use the walls of the primary myotubes as a scaffold upon which they fuse and differentiate to form secondary myotubes (Ontell, 1982, Kelly and Rubenstein, 1986).

Primary myotubes also provide the initial target for neuromuscular synaptogenesis and through the innervation of separated primary myotubes the nerve provides the correct environment for secondary myotube formation.

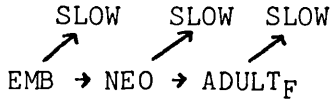
There is a rapid rate of addition of secondary myotubes in late gestation (approximately one week before birth in rodents; Ontell, 1982) and fibres continue to form up until one week after birth in some muscles of the rat (Betz, Caldwell and Ribchester, 1979). This rapid burst of fibre formation then ceases but there is a continued proliferation of myoblasts on the walls of the fibres. These myoblasts are "satellite" myoblasts and remain on the walls of muscle fibres throughout growth and adult life. In the adult, satellite cells remain quiescent unless stimulated to fuse to form myotubes in response to muscle damage (see Strohman et al., 1986). At birth the diameter of the fibres is small but with advancing development it increases together with the fibres length (see Zak, 1981). The number of muscle fibres remains the same since fibres are unable to divide (Strickland and Goldspink, 1973). Satellite cells increase the number of nuclei by fusion to the preformed fibres. Both nerve and muscle fibres in the period from gestation to early adulthood are immature and the speed of contraction of developing fibres is essentially slow although fast muscle isoforms of the contractile proteins may be present (see Gauthier, Lowey and Hobbs, 1978, Kelly and Rubenstein, 1986). The fibres are heterogenous with respect to their morphology reflecting their asynchronous formation and the isoform composition of the immature fibre does not approximate that of the adult. At the level of the neuromuscular junction, fibres are contacted by several axons per fibre in the foetal and neonatal period (Redfern, 1970) but this polyneuronal innervation regresses to a mono-synaptic configuration by approximately two weeks after birth in the rat (Bennet

and Pettigrew, 1974, Miata and Yoshioka, 1980). With advancing development muscle fibres gradually attain their various different properties and can eventually be distinguished into two broad categories: those that perfom sustained activity (slow) and those that are recruited for movement occassionally (fast). These two fibre types may be further subdivided (Section 1.3).

ii)Changes at the biochemical level

Muscles are rarely homogeneous in phenotype either during development or at adult stages.

In the soleus muscle of rat which develops into a muscle composed predominantly of slow fibres in the adult, both fast and slow type myosin light chains (MLC) are present soon after birth, although in the adult fibres MLC_{slow} isoforms predominate. Adult slow myosin heavy chain (MHC) is also found in foetal and newborn fibres, as well as the developmental MHC isoforms (MHC_{emb}, MHC_{neo}) (see Butler-Browne and Whalen, 1984). Using an immunocytochemical approach Whalen (1985) has determined the myosin heavy chain transitions which occur in single muscle fibres of the developing soleus muscle. It had previously been shown that myosin heavy chain isoforms undergo a transition in which MHC_{emb} is replaced by the MHC_{neo} which is finally replaced by the MHC_{fast} isoform in rat hind limb muscle (Whalen et al., 1981). The immunocytochemical approach revealed that an asynchronous transition from one form to the other occurred within the same fibre (see Whalen 1985). When the analysis was extended to the development of slow fibres, their data led to the proposal that these fibres will also undergo MHC transitions typical of fast fibres unless the fibres received a signal to stimulate the accumulation of slow myosin according to the scheme:



This provides a unified scheme for MHC transitions in both fast and slow fibres as well as explaining the coexistence of isoforms in the same fibre. By changing the proportion of fibres induced to accumulate slow myosin, this scheme could explain fibre MHC composition in any mixed muscle and therefore the development of slow muscle fibres might be generally applicable to other muscles.

Changes occuring during development in the isoform composition of other contractile proteins have not been reported at the single fibre level but at the level of the mRNA of the particular contractile protein gene or contractile protein in a particular muscle type.

No developmental isoforms of the actins that accumulate at specific stages have been described (Buckingham, 1985). The two striated actin proteins, α-cardiac and α-skeletal actin are accumulated in adult cardiac and skeletal muscle respectively (Section 1.3). In developing muscle however, both striated muscle mRNAs accumulate significantly (Minty et al., 1982, Mayer et al., 1984). For example, in 18 day old foetal mouse skeletal muscle, 30-40% of the muscle actin transcripts are cardiac whereas in newborn muscle cardiac actin transcripts fall to approximately 20% of the total and in adult skeletal muscle they represent less than 1% of the transcripts (Buckinham, 1985). This coaccumulation of transcripts has also been demonstrated in vitro in myogenic cell lines (Bains et al., 1984). The myosin light chain family similarly expresses a cardiac gene during skeletal muscle development; the major light chain of the adult heart atria is found in foetal skeletal muscle (Buckingham, 1985). The time course of expression of the cardiac actin and myosin light chain is not coordinated and the skeletal light chain is not expressed in developing cardiac tissue whereas the

skeletal actin is expressed in the heart of small mammals. This suggests that these genes are controlled by different regulatory mechanisms. During the maturation of mammalian cardiac and skeletal muscle the actin and myosin multigene families therefore follow different developmental strategies and for several of the contractile proteins, the organism uses the existence of multiple isoforms to provide developmentally regulated forms for some of its muscles while using the same forms as the major adult type in other muscles.

In developing muscle, there is also the added complication of the changes in alternative splicing pathways of particular muscle-specific genes which generates further isoform transitions (Nadal-Ginard et al., 1986).

Physiological signals influencing isoform transitions during muscle development

The physiological signals which might regulate the various isoform transitions are only just becoming apparent. By interfering with muscle development in various ways, a number of the mechanisms which might control isoform transitions have been inferred. Some of the MHC transitions occuring during development appear to depend on the type of innervation the muscle fibre receives while others may be influenced by hormones. The embryonic-neonatal-adult transition is probably nerve independent since dennervation does not block the appearance of fast adult myosin (Butler-Browne et al., 1982). Considerable evidence however suggests that slow MHC appearance requires innervation, at least in the post-natal period, although the embryonic to slow transition may not be nerve dependent. The role of the nerve in controlling isoform composition during the early period of fibre development is not as clearly defined as in the adult.

The levels of thyroid hormone seem to be an important factor controlling the presence of adult fast myosin. If rats are treated with anti-thyroid drugs during gestation and into the first month after birth, the hypothyroid offspring will accumulate neonatal myosin but not adult fast myosin, whereas 1 month old control rats contain adult fast MHC as the predominant isoform. The effect of thyroid hormone on MHC gene expression has been determined by Mahdavi et al. (1986) who have demonstrated that the hormone affects MHC gene expression in all muscles. The overall effect of a thyroidectomy is to induce the expression of genes encoding MHC_{slow} isoforms whereas hyperthyroidism has the opposite effect. The analysis of various MHC genes in different muscles demonstrated that the thyroid hormone response of any particular gene is not intrinsic to the gene but appears to be determined by the muscle in which the gene is expressed. For example, cardiac α-MHC gene expression does not change in the atrium in either the hypo- or hyper-thyroid state whereas in the ventrical there is a complete de-induction of this gene in response to hypothyroidism.

Regulatory mechanisms controlling isoform transitions

The mechanisms by which hormonal and neuronal effects influence isoform gene expression remain obscure. Mahdavi et al. (1986) demonstrated a close correlation between the S1 nuclease patterns of MHC transcripts and MHC isoform levels during fibre development suggesting that changes in the MHC phenotype are accounted by changes in the level of their respective mRNAs. There is no information on the mechanisms by which increases or decreases in mRNA levels are brought about. The types of experiments to link a physiological signal with a change in gene expression in the nucleus are not simple. A number of laboratories have been developing myoblast culture conditions to enable long-term fibre

maturation in an attempt to try to mimic some of the later developmental transitions in an *in vitro* system (Toyota and Shimada, 1983, Strohman *et al.*, 1986). Another possible method to study developmental transitions at the molecular level is to analyse expression from muscle-specific genes and derivatives of them in transgenic mice. Recently, a chimeric plasmid containing about two-thirds of the rat skeletal muscle α-actin gene and 730 bp of 5'-flanking sequence fused to the 3' end of an embryonic globin gene has been inserted into mouse genomes by microinjection of fertilised eggs (Shani, 1986). Sufficient *cis*-acting sequences are contained within this gene fusion for its tissue-specific and developmentally regulated expression in transgenic mice. The actin gene fusion also responds to developmental switches involved in isoform transitions: the actin gene fusion is expressed in cardiac muscle at levels higher or equivalent to those observed in skeletal muscle at the foetal stage, whereas in older mice, transcripts from the gene fusion are significantly lower in cardiac compared to skeletal muscle, reflecting actin gene transcript transitions observed for the endogenous skeletal α-actin gene (see above).

The localisation of the contractile protein genes has revealed that genes within the multi-gene family encoding actin and myosin light chains are dispersed throughout the genome which precludes a simple mechanism of *cis*-acting regulation for coordinate activation observed for several of these genes (Buckingham, 1985). In contrast, the genes encoding the myosin heavy chain multi-gene family are clustered on a single chromosome in the order in which they are expressed during development (Mahdahvi *et al.*, 1986) as is observed for the α- and β-like globin gene clusters. This organisation of MHC genes within the same chromatin domain may be significant for their sequential expression during muscle development.

1.5 THE ADAPTABILITY OF MUSCLE

Adaption due to changing functional demands

When the normal developmental changes in muscle are assumed to be complete, skeletal and cardiac muscle show the ability to adapt further to changes in functional demands imposed on the muscle (reviewed in Pette, 1980). This property of muscle is often referred to as plasticity and results in changes of phenotype as a result of changes in gene expression.

Endurance training by individuals results in an increase in the volume density of the mitochondria in all muscle fibres and increased activity of citric acid cycle, respiratory chain and intramitochondrial enzymes. Lipid stores and β-oxidation enzymes are found in higher quantities whereas the glycogen store and glycolytic enzymes are decreased. The improvements in the oxidative capacity are closely matched by an increase in capillary density, cardiac output and an increase in myoglobin concentrations (Pattengale and Holloszy, 1967). The capacity for oxygen transport is therefore adjusted to the increased rate of oxygen utilisation in endurance trained muscles. However, changes in contractile protein isoforms are not thought to occur (Holloszy and Booth, 1976). Strength training on the other hand does not improve the oxidative capacity of muscle but in these muscles, energy rich phosphates and glycogen are increased as well as the enzymes utilising these substrates.

Adaption of muscle due to imposed stimuli

Fast muscle fibres may gradually aquire the properties of slow fibres if they are subjected to a continuous low frequency electrical stimulation (Eisenberg, 1985). Among the first changes to occur are a generalised reduction in the T-tubule and SR systems and an increase in

mitochondrial content. There is a transition of fast to slow myosin isoforms and a distribution of tropomyosin subunit isoforms characteristic of that of slow muscle (Jolesz and Sreter, 1981). Muscles fatigue much less after the continuous stimulation and the early changes in fatiguability are paralleled by an increase in capillary density, blood flow, oxidative enzyme activity and myoglobin. After approximately 8 weeks the stimulated muscle is almost completely made up of type I fibres and are now more red in colour due to increased myoglobin levels. The removal of the stimulation at any stage results in the reformation of the fast muscle fibres. This type of experiment is similar to the cross-reinnervation experiments of Buller, Eccles and Eccles (1960) (Section 1.3) and they are important in demonstrating that the activity of the nerve is responsible for the phenotype expression of an adult muscle fibre.

1.6 MYOGLOBIN GENE EXPRESSION DURING MUSCLE DEVELOPMENT

The appearance of myoglobin and myoglobin mRNA during myogenesis

There is a large amount of data in the literature concerning the appearance of contractile proteins and contractile protein mRNAs during development, but little is known about the appearance of non-contractile proteins and mRNAs which must also be activated at some stage during myogenesis to complete the muscle phenotype.

Some literature exists on the appearance of myoglobin and myoglobin mRNA during muscle development. Myoglobin was shown to appear in the foetal heart early in development (Longo *et al.*, 1973, Tipler, Edwards and Hopkinson, 1978). The appearance of myoglobin in skeletal muscle was reported to occur at a later stage (Kagen and Christian, 1966, Tipler *et al.*, 1978) although the exact stage during development was uncertain (Longo *et al.*, 1973).

Levels of myoglobin were shown to increase sharply with age in the skeletal muscle of new-born puppies, with a much smaller increase in myoglobin levels in the heart (Musin, 1969). Recently, the appearance of myoglobin and myoglobin mRNA during mouse development and at various stages of human development have been analysed using more sensitive techniques in an attempt to determine more precisely the time at which myglobin is first detectable in skeletal and cardiac muscles (Weller et al., 1986). The results of this study confirmed the myoglobin levels are elevated in the foetal heart and demonstrated that myoglobin and myoglobin mRNA appear very early in development in both cardiac and skeletal muscles, earlier than previously demonstrated in skeletal muscle. Myoglobin and myoglobin mRNA are detectable in mouse skeletal muscle at about 14 days of gestation and in human skeletal muscle at 10 weeks of gestation. In the mouse, levels of myoglobin and myoglobin mRNA rise 50-100 fold during pre- and post-natal development before reaching adult levels. The increase in mRNA levels is in contrast to the level of total contractile protein mRNAs whose levels do not vary significantly during later myogenesis, although individual isoforms are subject to up and down modulation. The increase in myoglobin mRNA levels during development may be linked to the diffrentiation of slow muscle fibres which are formed over a similar time period (Ontell, 1982). Increasing levels of mRNA during development have also been described for a number of muscle-specific isoenzymes. Schweighoffer et al. (1985) demonstrated that expression of muscle-specific phosphorylase, creatine kinase and aldolase increase in a coordinate manner from the seventeenth day of foetal life in the rat. Glyceraldehyde-3-phospate dehydrogenase (G3PD) mRNA levels (and enzyme activity) increase in a similar manner to that observed for myoglobin mRNA levels in developing skeletal muscles. G3PD mRNA levels are also elevated in the foetal heart (Edwards et al.,

1985a). It is probable in both cases that the differential expression in foetal cardiac versus foetal skeletal muscle parallels the relative activity of these muscles and their metabolic requirements during development.

The differential regulation of myoglobin gene expression in foetal cardiac and skeletal muscles indicates that the myoglobin gene may be independently regulated in different muscle types and is reminiscent of the differential regulation of the β-cardiac MHC gene ($=MHC_{slow}$) in developing cardiac and skeletal muscles and the α-cardiac MHC gene in different cardiac muscle in response to thyroid hormone.

Myoglobin levels may be controlled in part by the nerve, as suggested by the finding that myoglobin levels increase shortly after dennervation in rat hind limb muscles and returns to normal during reinnervation (Askmark, Carlson and Roxin, 1984).

Levels of myoglobin in foetal and adult skeletal and cardiac muscles are largely determined by the size of the myoglobin mRNA pools (Weller et al., 1986). This is also true of the levels of myoglobin in grey seal skeletal muscle which contains about seven times as much myoglobin as human skeletal muscle as an adaption to a diving physiology (Section 1.1) and contains a correspondingly elevated level of myoglobin mRNA. It seems likely that the adaption to high myoglobin levels in the seal has arisen through enhanced transcriptional activity of the myoglobin gene and/or reduced turnover of myoglobin mRNA.

In contrast to this, mouse skeletal and cardiac muscle contain very low levels of myoglobin, but have myoglobin mRNA levels similar to those found in the human suggesting that the translational efficiency of mouse myoglobin mRNA is low or instead there might be an increased turn over of myoglobin protein mouse muscle.

1.7 OBJECT OF RESEARCH

Myoglobin genes are subject to numerous adaptions both within an animal and between species. The object of this research was to develop an *in vitro* expression system in which some of the mechanisms regulating myoglobin gene expression *in vivo* could be studied at the molecular level.

CHAPTER 2

MATERIALS AND METHODS

2.1 DNA, RNA AND CELL LINES

DBA/2 and C57/BL genomic DNAs were previously prepared and supplied by Dr. Susan Adams (ICI/Joint Laboratory, University of Leicester). Adult seal skeletal muscle total RNA was prepared and supplied by Dr. Polly Weller (Department of Genetics, University of Leicester). RNA samples were kindly prepared and provided by Ms. Clare Pidduck (Department of Biochemistry, Leicester): total RNA from pancreatic islet cells, Ms. Liz Allen (Department of Biochemistry, Leicester): total RNA from human myeloma cells, DBA/2 mouse liver and DBA/2 mouse cardiac tissue and Dr. Andrew Carter (ICI/Joint Laboratory, University of Leicester): total RNA from baby mouse kidney cells.

The mouse myoblast cell line, G8-1, was a gift from Dr. Mike Webb (Institute of Neurology, University of London) and the mouse thymidine kinase deficient L cells (LtK^- fibroblast cell line) from Ms. Jeanie Brookes (ICI/Joint Laboratory, University of Leicester).

2.2 PLASMIDS, BACTERIOPHAGE AND BACTERIAL STRAINS

2.2.1 Recombinant Plasmids

A mouse α-skeletal actin cDNA probe, pAM91 (Minty et al., 1981), and a human myosin light chain cDNA probe pMLC were gifts from Dr. Polly Weller (Department of Genetics, University of Leicester), and a human myosin heavy chain cDNA probe (Edwards et al., 1985b) was a gift from Dr Yvonne Edwards (University College, London). The chloramphenicol acetyl transferase based vectors pRSV-CAT (Gorman et al., 1982), pSVO-CAT and pSV2-CAT (Gorman, Moffat and Howard, 1982) were obtained from Dr. Gillian

Smith (ICI/Joint Laboratory, University of Leicester) and plasmid 4I (Wieringa *et al.*, 1983) was a gift from Prof. C. Weissmann (University of Zurich, Switzerland).

2.2.2 Nonrecombinant Plasmids

Nonrecombinant plasmids pAT153 (Twigg and Sherratt, 1980) and pUC13 (Messing, 1983) were used in constructing gene fusions.

2.2.3 Bacteriophage and Bacterial Strains

M13 sequencing vectors, M13mp18 and M13mp19 (Yanish-Perron *et al.*, 1985) were used in shotgun cloning of DNA's for M13 sequencing for generating ^{32}P-labelled single stranded cDNA hybridisation probes.

E. coli bacterial strains used were:-

JM83 (Messing and Vieira, 1982): λ^- *ara*, Δ(*lac-pro*), *rpsL*, *thi* Φ80d*lac* ZΔM15

JM101 (Messing *et al.*, 1981): (Δ(*lac-pro*), *supE*44, *thi*, F'*traD*36, *proAB*$^+$, *lacI*^{q}ZΔM15)

2.3 ENZYMES, ANTIBIOTICS, CHEMICALS AND REAGENTS

The source of the most important of these are listed below. Other chemicals not listed were all analytical grade.

	Source
N-2-Hydroxyethylpiperazine -N-2-ethane-sulphonic acid (HEPES)	Sigma
Acetyl Co Enzyme A (Lithium salt)	"
Dextran sulphate (Sodium salt)	"
Dithiothreitol	"
Spermidine trichloride	"
Ficol 400	"
Salmon sperm DNA	"
Bovine serum albumin	"
Diomethyldichlorosilane	"
Isopropyl-β-D-galacto-pyranoside (IPTG)	"
N,N,N',N'-tetramethyl-ethylenediamine (TEMED)	"
Polyethylene glycol (PEG)	"
Diethylpyrocarbonate (DEPC)	"
3-(N-morpholino) propane sulphonic acid (MOPS)	"
Thymine	"
Ampicillin	"

	Source
Chloramphenicol	"
Lysozyme	"
Coomassie brilliant blue G	"
Deoxyribonucleotide triphosphates	Bethesda Research Labs
M13mp18 and M13mp19 RF DNAs	"
Urea	"
S1 nuclease	"
Restriction enzymes, unless otherwise stated	"
DNA polymerase I (Klenow fragment)	Boehringer
Calf intestinal phosphatase	"
Calf liver tRNA	"
Dideoxynucleoside triphosphates	"
DNase I	Worthington
T4 DNA ligase	New England Biolabs
Avian myeloblastosis virus reverse transcriptase	Life Sciences Inc.
Acrylamide	Unisciences
Bisacrylamide	Bio-Rad Laboratories
Oligo(dT) cellulose	Collaborative Research
Agarose (standard and low melting point)	"
E. coli DNA polymerase I	Amersham
[α-^{32}P]dCTP, [^{35}S]methionine, [α-^{32}P]dATP	"
D-Threo-[Dichloroacetyl-1-C^{14}] chloramphenicol	"
Nonidet P-40	BDH
Formamide	"
5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (BCIG)	Bachem Inc.
Horse serum	Gibco/BRL
Foetal calf serum	"
Dulbecco's modified eagles medium (DMEM)	"
1 x Trypsin/EDTA	"
PBS tablets	Flow Laboratories
2,5-diphenyloxazole (PPO)	Fisons
Polyvinylpyrrolidine (PVP)	"
1,4-di-2(4-methyl-5-phenyl oxazolyl (POPOP)	"

2.4 LIQUID AND SOLID MEDIA

2.4.1 For bacteria

Luria broth (10g Difco Bacto Tryptone; 5g Difco Bacto yeast extract; 5g NaCl per litre of distilled water). Luria agar plates were prepared by solidifying liquid medium with 15g Difco Bacto-agar per litre.

BBL agar used for M13 phage work contained 10g BBL trypticase, 5g NaCl and 5g $MgSO_4$ per litre supplemented with 10mM $MgCl_2$ was solidified with either 15g of Difco Bacto agar per litre or 6g per litre for soft agar overlays.

Glucose supplemented minimal medium used to maintain strain JM101 contained 42mM Na_2HPO_4, 22mM KH_2PO_4, 18mM NH_4Cl, 8mM NaCl, 22mM glucose, 0.1mM $CaCl_2$, 1mM $MgSO_4$ and 3μM thiamine HCl.

2.4.2 For cell culture

Dulbecco's modification of Eagles minimum essential medium (DMEM, see Morton, 1970) was used to culture mouse myoblast and fibroblast cells and was obtained as a working strength solution containing sodium bicarbonate and glutamine from Gibco and stored at 4°C. DMEM was supplemented with foetal calf serum and horse serum to the appropriate percentage. Serum was similarly obtained from Gibco and stored at -20°C in aliquots.

Serum batches had all been mycoplasma screened. New batches of foetal calf serum were tested for growth promoting characteristics in a number of cell lines by members of the ICI/Joint Laboratory. New batches of horse serum were tested for their differentiation promoting ability when added as a 5% solution in DMEM + 0.5% FCS to confluent myoblasts (see Section 2.11).

Phosphate buffered saline (PBS) was prepared from tablets (Flow Laboratories) according to the manufacturers instructions.

2.5 GENERAL TECHNIQUES FOR HANDLING DNA AND RNA

2.5.1 Phenol Extraction

DNA and RNA solutions were mixed with 0.5-1 volume of phenol; chloroform; isoamyl alcohol; 8 hydroxyquinoline (100:100:4:0.4, w:v:v:w) saturated with 10 mM Tris HCl, pH 7.5. The upper aqueous phase, containing DNA or RNA was removed carefully to avoid disturbing the interphase and the phenol layer was re-extracted with an equal volume of 10mM Tris-HCl, pH 7.5. The phenol was AR grade and not redistilled.

2.5.2 Ethanol Precipitation

DNA and RNA were precipitated from solution by the addition of 0.1 volume of 2 M sodium acetate pH 5.6, and 2.5 volumes of ethanol. The solution was mixed well and chilled in a dry ice/IMS bath for approximately 5 minutes. Precipitated DNA or RNA was pelleted at either 10,000 rpm for 30 minutes at 0°C (Sorvall HB4 rotor), or in an Eppendorf centrifuge for 5 minutes at maximum speed. The supernatant was removed carefully and the pellet rinsed with 80% ethanol, centrifuged briefly and the 80% ethanol discarded. DNA and RNA pellets were dried under vacuum for approximately 5 mins before being redissolved as appropriate for further manipulation.

2.5.3 Note on the Handling of RNA

Any solutions, glassware, centrifuge tubes which were to come into contact with RNA were made up or pre-treated with, a 0.1% solution of diethylpyrocarbonate (DEPC) in distilled water. Solutions and equipment were incubated overnight at 37°C in the DEPC solution before inactivation of DEPC by autoclaving. Solutions containing Tris were not subjected to treatment with DEPC.

2.5.4 Restriction Endonuclease Digestion

DNAs at a final concentration of 0.1 mg/ml were incubated with 1 unit/µg of DNA of the appropriate restriction enzyme in the manufacturers recommended buffer at 37°C for 1 hour unless otherwise stated. Spermine trichloride was added routinely to a final concentration of 4 mM to enhance the efficiency of restriction endonuclease digestion. This is particularly important for DNA previously recovered from agarose gels (Bouche, 1981). Complete digestion of the DNA was monitored by

electrophoresing an aliquot of the digestion mixture containing ~0.2 µg of DNA in a suitable agarose gel against marker DNA of known molecular weight. If digestion was not complete, further restriction endonuclease was added and the DNA solution incubated for longer. After complete digestion, EDTA was added to a final concentraton of 20 mM. The DNA was then recovered by phenol extraction and ethanol precipitation, and redissolved as appropriate for further manipulation.

2.5.5 DNA and RNA Electrophoresis

DNA

Horizontal slab gels with loading slots from 3-7 mm long were electrophoresed in buffer (40 mM Tris-acetate, 2 mM EDTA, pH 7.7) containing 0.5 µg/ml ethidium bromide (Aaij and Borst, 1972). The size of the gel varied with the number of DNA samples to be electrophoresed from 5 x 7 cm (mini-gels) to 20 x 20 cm (blotting gels). The concentration of agarose was varied betwen 0.5 and 3% (w/v), depending upon the DNA fragment sizes to be separated. Molecular weight markers were either λ x *Hind*III or pBR322 x *Sau*3A, or pAT153 x *Hinf*I or combinations. DNA samples were mixed with 0.5 volume of agarose beads (0.2% agarose gel in 20 mM EDTA, 10% glycerol and a small amount of bromophenol blue as an electrophoretic marker; this was prepared according to Schaffner *et al.* (1976)). Gels were electrophoresed at 100v for a short time or at 15-20v overnight, as appropriate, until the bromophenol blue had migrated approximately two thirds of the length of the gel.

RNA

The quality and approximate concentration of RNA samples were tested by electrophoresis in horizontal agarose gels as for DNA samples

but with the addition of a heat shock step at 65°C for 3 minutes prior to loading the RNA on the gel. The size markers were as above for DNA gels except they were single-stranded (denatured by the addition of 0.1 volume of 1.5 M NaOH, 0.1 M EDTA, 5 minutes before loading). RNA was separated well on 1.5% agarose gels.

2.5.6 Recovery of DNA from Preparative Agarose Gels

Prepartion of DNA samples, gel loading and electrophoresis were as described by Jeffreys et al. (1980). The quantity of DNA loaded and gel slot size were chosen so that no more than 0.25 µg of DNA per mm^2 of gel slot surface was loaded. This prevented smearing of DNA due to overloading. DNA in preparative gels was visualised by the fluoresence of bound ethidium bromide under long wave ultraviolet light. The appropriate DNA fragment or size selected DNA fragments were cut out from the gel in small gel slices, DNA was then electrophoresed from the gel slice(s) on to Whatman DE81 DEAE-cellulose paper, and recovered by a modification of the method of Dretzen et al. (1981). DE81 paper was prepared by soaking in 2.5 M NaCl and rinsing several times in distilled water, and stored in 1 mM EDTA at 4°C. The agarose gel slices were wrapped in the prepared DE81 paper and replaced in the gel in slots cut away from DNA in other areas of the gel. The DNA was electrophoresed on to the paper at high voltage (200 v) for approximately 15 minutes. The gel slices were removed from the paper and transfer of the DNA to DE81 paper was visualised by ethidium bromide fluoresence under long wave ultraviolet light. The DNA contained in the paper was recovered by vortexing the paper in high salt buffer (high salt buffer= 1 M NaCl, 50 mM Tris-HCl, pH 7.5, 1 mM EDTA) until it disintegrated. The paper solution was incubated at 37°C for 15 mins followed by 3 mins at 65°C to release DNA from the paper. DNA in solution was separated from the

fragmented paper by centrifugation through a small polyallomer wool column in a bench top MSE centrifuge at maximum speed for approximately one minute. The DNA was recovered by 2 phenol extractions and 2 ethanol precipitations before further manipulation.

2.5.7 Recovery of Small DNA Fragments from Acrylamide Gels

DNA fragments to be isolated were run on 8% acrylamide gels as described in Maniatis, Fritsch and Sambrook (1982). DNA fragments were visualised by the fluoresence of bound ethidium bromide under long wave ultraviolet light and the required fragment cut out from the gel. DNA was then recovered from the gel slice recovered as described in Maniatis et al. (1982). The gel slice was chopped into very fine pieces with a scalpel blade, the pieces transferred to an eppendorf tube and 1 volume of elution buffer added (elution buffer= 0.5 M ammonium acetate, 1 mM EDTA, pH 8.0). Gel pieces were incubated at 37°C overnight with rotation and then centrifuged in an Eppendorf centrifuge to pellet the gel fragments away from the DNA in solution. The supernatant was removed with a drawn out Pasteur pipette into a fresh Eppendorf tube and an additional 0.5 volume of elution buffer was added to the gel pellet. After vortexing and centrifugation, the second supernatant was combined with the first. To remove any acrylamide fragments left in the supernatant, it was passed once through a polyallomer wool plug in a pipette tip. DNA was then recovered by ethanol precipitation twice before redissolving in a small volume of TE (TE= 10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The DNA solution was centrifuged for 5 minutes and the supernatant removed into a fresh Eppendorf tube, ready for further manipulation.

2.5.7 Preparation of Plasmid DNA

Small scale preparations of plasmid DNA were carried out using a modification of the small scale alkaline extraction method of Birnboim and Doly (1979) for analysis of plasmid recombinants. Large scale preparations of plasmid DNAs for use in eukaryotic cell transfection experiments were essentially a scaled up version of the small-scale plasmid preparation procedure with the additon of a caesium chloride gradient centrifugation purification step. The method is based upon that of Ish-Horowicz and Burke (1981).

Small-scale plasmid preparations

A 1.5 ml culture of the plasmid containing strain of E. coli was grown to stationary phase overnight in Luria broth supplemented with the appropriate selective antibiotic which was always 100 µg/ml ampicillin. The culture was pelleted by a 30 second spin in an Eppendorf bench centrifuge, the supernatant flicked off and the cell pellet suspended in 100 µl of ice cold lysis solution (lysis solution= 25 mM Tris-HCl, 10 mM EDTA, 50 mM sucrose, pH 8.0 containing freshly added lysosyme to a final concentration of 1 mg/ml). The solution was kept on ice for 10 mins, then 200 µl of alkaline-SDS (alkaline/SDS= 0.2 M NaOH, 1% SDS prepared just before use) were added and the solution mixed and left on ice for a further 5 minutes. Chromosomal DNA and most of the proteins were precipitated by the addition of 150 µl of 3 M potassium acetate, pH 5.2 and leaving on ice for 10 minutes. The solution was spun for 5 mins in an Eppendorf centrifuge and the supernatant removed into a fresh tube avoiding the precipitated material. Plasmid DNA was recovered from the supernatant by ethanol precipitation and DNA was purified by phenol extraction and ethanol precipitation.

Approximately 5-10 µg of plasmid DNA may be prepared by this

method. Very little E. coli chromosomal DNA or cellular protein contaminate the plasmid DNA but the preparation contains large amounts of RNA. The plasmid DNA can be used directly in restriction endonuclease digests, with the addition of RNase to remove the RNA in order to confirm the presence of the required DNA insert and its orientation within the plasmid DNA.

Large Scale Plasmid Preparation

The E. coli containing the appropriate plasmid was grown with shaking overnight at 37°C in 10 ml of Luria broth with 100 µg/ml ampicillin. 8 mls of the overnight culture were used to innoculate 400 mls of prewarmed Luria broth plus antibiotic in a 2 litre flask. The culture was grown to an O.D. of 1 at 600 nm and then 80 mg of chloramphenicol were added and the culture allowed to shake overnight.

The cells were pelleted by centrifugation at 5,000 rpm in a GS3 rotor for 10 minutes at room temperature, the supernatant discarded and the cells resuspended in 12 ml of lysis solution (lysis solution= 50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA) and kept at room temperature for 10 minutes. 24 mls of alkaline-SDS solution were added to the lysed cells which were mixed gently and placed on ice for 5 minutes. 12 mls of 5 M potassium acetate pH 4.8 were then added, the solution mixed gently and returned to ice for a further 15 mins. Chromosomal DNA and proteins were pelleted by centrifugation at 7,000 rpm in a GS3 rotor at 4°C for 10 minutes and the supernatant decanted from the precipitate through a polyallomer wool plug in a plastic funnel. The volume of supernatant recovered was recorded and 0.6 volumes of propan-2-ol added with mixing and the solution left at room temperature for 10 mins. Nucleic acids were precipitated by centrifugation as above at 4°C. The pellet was washed with 10 ml of 80% ethanol and resuspended in 3 ml of TE. For each

ml of solution 1 g of caesium chloride (CsCl) and 0.1 ml of ethidium bromide (10 mg/ml) were added, the solution mixed thoroughly and centrifuged at 10,000 rpm for 10 mins at room temperature in an HB4 rotor. The pellet and pellicule were discarded. The density of the solution was checked using a refractometer and adjusted by adding CsCl or TE if necessary to give a refractive index of 1.3874. The solution was loaded into "Quick seal" tubes and topped up with CsCl solution of 1.57 g/ml density. The tube was centrifuged in a 75 Ti rotor with a balance at 65,000 rpm for ~16 hours at room temperature. The plasmid band was recovered from the gradient with a syringe and 18g needle and ethidium bromide was removed by repeated extractions with propan-2-ol saturated with sodium chloride/H_2O. The plasmid DNA was dialysed extensively against TE to remove the CsCl and the concentration calculated by measurement of O.D. at 260 nm. If necessary the plasmid DNA was ethanol precipitated. The quality of the plasmid was determined by agarose gel electrophoresis.

Plasmid DNA to be used in transfection was checked to be greater than 50% in the supercoiled form and if not the preparation repeated. In some transfection experiments plasmid DNAs were subjected to two cycles of CsCl gradient centrifugation.

2.5.8 Alkaline phosphatase treatment of DNA

1 µg of plasmid DNA was digested at suitable cloning sites with the appropriate restriction endonuclease(s). 15 minutes before the end of the digestion period 1 U of calf intestinal phosphatase was added and incubation continued for 15 minutes at 37°C.

DNA was heated to 65°C for 10 minutes in the presence of 0.5% SDS, 20 mM EDTA, pH 8.0 to remove the phosphatase and then recovered by two phenol extractions and two ethanol precipitations. DNA was vacuum

dried and dissolved in water ready for further use.

To dephosphorylate blunt-ended DNA, 3 U of calf intestinal phosphatase were added to the digestion mix after first heat-shocking the digested DNA by heating to 65°C for 5 minutes and cooling on ice. DNA was recovered as described above.

2.5.9 Generation of blunt-ended DNA fragments using S1 nuclease and reverse transcriptase

Isolated DNA fragments were treated with S1 nuclease to remove overhanging DNA ends created by restriction endonuclease cleavage.

Test digests containing ~0.5 µg of pBR322 x Sau3A DNA fragments and alkaline denatured salmon sperm DNA (boiled prior to use to ensure DNA was single-stranded) were carried out with varying concentrations of S1 nuclease (ranging between 0.01 and 20 U/µl) in order to determine the concentration of S1 nuclease required to digest single-stranded DNA but not double-stranded linear DNA. Reactions were carried out in a total volume of 10 µl of S1 buffer (S1 buffer= 50 mM NaCl, 30 mM Sodium acetate, pH 4.6, 1 mM $ZnSo_4$ and 5% (v:v) glycerol) and incubated at 37°C for 30 minutes before the reaction was terminated by the addition of 1 µl of 0.5 M EDTA pH 8.0.

DNA products were electrophoresed in a 1.2% agarose gel. The most appropriate concentration of S1 nuclease which removed single-stranded DNA but did not digest double-stranded DNA was ~0.1 U/µl. This concentration was used to blunt DNA ends of DNA fragments isolated from restriction endonuclease digestions in scaled up (20 µl) but otherwise identical reaction conditions. After treatment with S1 nuclease, DNA was recovered by phenol extraction and ethanol precipitation.

Reverse transcriptase

For some DNA fragments, protruding DNA ends generated by restriction endonuclease digestion were filled-in with dNTPs using reverse transcriptase. DNA was incubated in reverse transcriptase buffer (reverse transcriptase buffer= 1 M Tris-HCl, pH 8.3, 1 M KCl, 250 mM $MgCl_2$, 20 mM dNTPs and 700 mM 2-mercaptoethanol) with 2 U/µlDNA of reverse trancriptase for 1 hour at 37°C. DNA was then recovered after phenol extraction by one ethanol precipitation.

2.5.10 Ligation of Fragments into Plasmid DNA

Fragments isolated from restriction endonuclease digestions of recombinant plasmid DNAs and myoglobin gene fragments isolated and treated with S1 nuclease as described above were ligated with the appropriate linearised phosphatased vector in a molar ratio of 2:1 (fragment:vector).

Ligation reactions containing 1 µg of DNA or less were set up in a 20 µl of ligase buffer (ligase buffer= 50 mM Tris-HCl, 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, pH 7.8; stored as a 10 x concentrate at -80°C) with 4 mM spermidine, 1 mM HCC and 50 U of T4 DNA ligase. HCC was added to stimulate intermolecular ligation of DNA ends (Rusche and Howard-Flanders, 1985).

2.5.11 Transformation

Transformation of E.coli JM83 with recombinant plasmid DNAs was carried out according to the method of Kushner (1978) which is described in Section 2.9.

After the heat-shock treatment, transformed JM83 cells were added to 1 ml of Luria broth and incubated at 37°C for 1 hour. Cells were then centrifuged in an Eppendorf centrifuge for 1 minute, the cell pellet

resuspended in 0.2 ml of Luria broth and cells plated out onto selective media plates (Luria agar with 100 μg/ml ampicillin).

When bacterial cells were transformed with recombinants obtained after blunt-end ligation of S1 nuclease treated fragments into the appropriate vector, all of the transformed cells were plated onto selective media. For cohesive-end DNA ligations, a dilution series of the transformed cells was carried out prior to plating out. After drying, plates were incubated overnight at 37°C.

2.5.12 Screening of Transformants

The method used to screen transformants was a modification of the filter hybridisation method of Grunstein and Hogness (1975), which has been described in detail by Barrie, (Ph.D thesis, 1982). Probes used for the detection of the desired recombinant clones were the appropriate fragments isolated to be cloned.

2.5.13 Photography and Autoradiography

DNA and RNA in agarose gels was visualised by the fluoresence of bound ethidium bromide using a short wave ultraviolet transilluminator (Ultra-violet Products Inc., California, U.S.A.). Gels were photographed using a polaroid MP-3 land camera and Polaroid 4 x 5 type-52 or type-57 film. Autoradiographs were photographed on a white light transilluminator, using a Nikon F camera and Kodak AHU 35 mm film which was processed according to the manufacturers instructions using Kodak D19 developer and May and Bakers' Amfix fixer.

All autoradiography used Kodak X-ray film (35 x 40 or 13 x 18 cm X-Omat RP or 13 x 18 cm X-Omat AR). Exposure times to film were judged by cpm (counts per minute) detected using a hand-held mini-monitor (Mini-Instruments Ltd., g-m monitor, type 5'10). Exposure to X-ray film

in the presence of an intensifying screen (Ilford Tungstate) was at -80°C, otherwise at room temperature.

X-ray film was developed by immersion for 3 mins in developer (Kodak DX-80), followed by a rinse in diluted acetic acid and fixation for 3 minutes in Kodak FX-40 fixer plus HX-40 hardener.

2.6 TRANSFER OF DNA AND RNA TO NITROCELLULOSE OR HYBOND NYLON FILTERS

Southern blotting

For Southern blotting, DNA was electrophoresed double stranded in 0.8% agarose gels with 7 mm slots. Before DNA was transferred from the gel to nitrocellulose or Hybond Nylon (Hybond-N) filters (Amersham International), the DNA in the gel was denatured by acid/alkali treatment. After being photographed, gels were soaked twice in 0.25 M HCl for 7 minutes, to reduce the size of the DNA fragments *in situ* by partial depurination. The DNA was denatured by soaking twice in 0.5 M NaOH, 1 M NaCl for 15 minutes. This was followed by two 15 minute washes in 0.5 M Tris-HCl, 3 M NaCl, pH 7.5 to neutralise. Transfer to both hybond-N and nitrocellulose was by a modification of the method of Southern (1975), as described by Barrie (Ph.D thesis, 1982) except that after transfer to Hybond-N, the filter was wrapped in saran wrap and DNA cross linked to the filter by exposure of the wrapped filter, DNA side down to short wave ultra-violet light on a transilluminator for 5 minutes.

Northern blotting

RNA was transferred to nitrocellulose filters by an adaption of the method of Thomas (1980). Horizontal slab gels were used, as for DNA, with the same sized slots run at 1%. Agarose was dissolved in 0.5 gel volume of water, to which was added 0.5 volume of 2 x gel buffer (2 x gel

buffer= 4.4 M formaldehyde, 50 mM sodium phosphate buffer, pH 7.6). The RNA samples to be loaded were first vacuum dried then redissolved in 10µl 50% formamide, 2.2 M formaldehyde, 20 mM sodium phosphate buffer, pH 7.6 and denatured at 60°C for 10 minutes. After cooling on ice, 1 volume of agarose beads was added before loading (RNA agarose beads were a 0.2% suspension of agarose in 2.2 M formaldehyde, 20 mM sodium phosphate buffer, pH 7.6, 10% glycerol and a very small amount of bromophenol blue). DNA size markers were run single-stranded, as described in Section 2.5. Gels were electrophoresed in 2.2 M formaldehyde, 10 mM sodium phosphate buffer pH 7.6 at 100v, until the bromophenol blue had migrated about 8 cm from the slots.

After electrophoresis, some of the formaldehyde was driven out of the gel by shaking gently in water at 60°C for 10 minutes, then in water at room temperature for 10 minutes. Gels were soaked in 20 x SSC (saline sodium citrate; 1 x SSC= 0.15 M NaCl, 15 mM trisodium citrate, pH 7.0). RNA was transferred to Nitrocellulose as for DNA (Barrie, Ph.D. thesis 1982), except the nitrocellulose and first sheet of Whatman 3 MM paper were soaked in 20 x SSC, rather than 3 x SSC and, after transfer, the nitrocellulose filter was baked at 80°C for only 2 hours.

Transfer of RNA to Hybond-N was carried out according to the manufacturers instructions. RNA samples to be analysed were incubated at 65°C for 5 minutes in the following solution: RNA (final volume 12 µl), 25 µl formamide (specially purified for biochemistry, BDH), 5 µl 10X MOPS buffer (10 x MOPS buffer = 0.2 M MOPS (3-(N-Morpholino) propane-sulphonic acid) 0.05 M Na acetate, pH 7.0, 0.01 M EDTA) and 8 µl of formaldehyde. The RNA solution was then chilled on ice and 5 µl of 5% (v/v) glycerol containing 0.1 mg/ml bromophenol blue added. Samples were electrophoresed in a 1% agarose gel. Agarose was dissolved in water containing 10 ml 10 x MOPS buffer per 75 ml of water and cooled to 50°C

before adding 17 ml formaldehyde (40% v/v solution). Gels were electrophoresed in 1 x MOPS buffer, until the bromophenol blue had migrated ~ 8 cm from the slots. After electrophoresis RNA was transferred directly to Hybond-N as for DNA (Barrie, Ph.D thesis 1982) except the filter and first sheet of Whatmann 3MM paper were soaked in 1 x MOPS buffer, rather than 3 x SSC. After transfer, Hybond-N filters were not rinsed but RNA was UV cross-linked to the filter as described above for cross-linking of DNA to Hybond-N.

For RNA dot blots, RNA was spotted directly on to nitrocellulose filters. An appropriate volume of RNA was vacuum dried and redissolved in 5 µl of TE. RNA was heat-shocked at 65°C for 3 minutes and then the whole 5 µl spotted on to a nitrocellulose filter. Carrier tRNA was not added since RNA samples spotted on to the filters contained at least 1µg poly(A)$^+$ RNA. Filters were allowed to dry at room temperature, and then baked at 80°C for 2 hours before hybridisation.

2.7 LABELLING OF DNA PROBES WITH ^{32}P

2.7.1 Nick-translation

The method used to label certain DNA probes by nick translation was that of Weller *et al.* (1984). 50-100 ng of DNA in 5 µl of water were denatured by heating in a boiling water bath for 3 minutes before being plunged into ice. The following reagents were then added to the DNA:

2.5 µl 10 x nick mix (0.5 M Tris-HCl, pH 7.5, 50 mM $MgCl_2$, 0.1 M 2-mercaptoethanol)

2 µl each of 50 µM dATP, dGTP, dTTP

1 µl 0.1M spermidine

1 µl 8 ng/ml DNase I (freshly diluted from a 1 mg/ml stock in 10 mM Tris-HCl pH7.5)

1.5 µl [α-^{32}P]dCTP (0.37 MBq/µl, 111 TBq/mMol)

5 units *E. coli* DNA polymerase I

7 µl H_2O

The reaction mix was mixed well and incubated at 15°C for 90 minutes. The reaction was stopped by the addition of 25 µl of 0.5% SDS, 12.5 mM EDTA, 10 mM Tris-HCl, pH 7.5 and then the DNA extracted with phenol. 100 µg of high molecular weight salmon sperm DNA (prepared as described by Harris, Ph.D thesis, 1985) were added as a carrier and the DNA was ethanol precipitated after a short chill on ice without centrifugation. The supernatant was removed and the DNA redissolved in 0.2 ml 10 mM Tris-HCl, pH 7.5 and reprecipitated. The DNA was rinsed twice with 80% ethanol. Specific activities of 10^7 to 10^8 dpm/µg were routinely acheived by this method.

2.7.2 Preparation of ^{32}P-labelled single-stranded probes from M13 recombinants

The method to label DNA fragments contained within M13 DNA was that described by Jeffreys *et al.* (1985). Approximately 0.4 µg of single-stranded M13 recombinant DNA was annealed with 4 ng of 17-mer sequencing primer (Duckworth *et al.*, 1981) in 10 µl of 10 mM $MgCl_2$, 10 mM Tris-HCl, pH 8.0 at 60°C for 30 minutes. Primer extension was carried out by the addition of 20µl of [80 µM dATP, 80 µM dTTP, 80 µM dGTP, 10 mM Tris-HCl, 0.1 mM EDTA (pH8.0)] plus 3 µl (1.11 MBq) [α-^{32}P]dCTP (111 TBq/mmol) and 3.5 U of Klenow polymerase. After mixing well, primer extension was allowed to proceed for 15 minutes at 37°C. 2.5 µl of 0.5 mM dCTP was added after this time and chasing was allowed to proceed for 15 minutes at 37°C. The M13 DNA was then cut at a suitable restriction endonuclease site either in the insert, or in the M13 polylinker 5'- to

the insert. The addition of 1.5 M NaOH, 0.1 M EDTA denatured M13 DNA to release the ^{32}P- labelled insert fragment. This labelled single-stranded DNA fragment extending from the primer was recovered by electrophoresis in a 1.2% low melting point agarose gel with 7 mm slots. If the insert band was not visible, the gel was covered with aluminium foil, overlaid with X-ray film and autoradiographed for 5 minutes to detect the labelled insert band. The fragment contained in agarose was excised from the gel and was melted at 100°C for 2 minutes with 1 mg of alkali-denatured salmon sperm carrier DNA (prepared as described by Harris, Ph.D thesis, 1985) and 500 µl of water. The melted probe was then added directly to a pre-warmed hybridisation chamber containing hybridisation solution for either Northern or Southern hybridisations (see below). With this labelling method, specific activities acheived were usually in excess of 10^9 cpm/µg DNA.

2.8 FILTER HYBRIDISATION

2.8.1 DNA filters

Nitrocellulose filter hybridisations were carried out as described in detail by Barrie (Ph.D. thesis, 1982). Round filters containing lysed bacterial colony DNA and filter strips containing genomic DNAs were prehybridised at 65°C in a gently rocking bath in the following solutions in this order (all solutions were degassed under vacuum for 10 minutes prior to use except the first two):

10 minutes	3 x SSC	
45 "	1 x Denhardts	0.2% Ficol 400, 0.2% polyvinlypyrrolidine, 0.2% BSA in 3 x SSC
30 "	1 x CFHM	CFHM = complete filter hybridisation mix (1 x Denhardts plus 0.1% SDS and 50 µg/ml denatured salmon sperm DNA in 1 x SSC)
30 "	1 x CFHM+PEG	CFHM + 6% Polyethylene glycol

Hybond N filters were pre-hybridised and hybridised according to a modification of the method of Church and Gilbert (1984).

Genomic Southern blot filters were cut into strips and prehybridised for 5-30 mins at 65°C in degassed hybridisation buffer. (Hybridisation buffer= 1 mM EDTA, 7% SDS, 0.5 M $NaHPO_4$, pH 7.2). Filters were then transferred to fresh hybridisation buffer containg the probe in a hybridisation chamber, carried out as described below. The .^{32}P-labelled probe DNA (≦10 ng/ml) was added to the hybridisation solutions (1 x CFHM + 6% PEG (nitrocellulose filters) or 1 mM EDTA, 7% SDS, 0.5 M $NaHPO_4$, pH 7.2), (Hybond N) after being denatured by heating to 100°C for 5 minutes (nick translated probes) or 2 minutes (single-stranded M13 probes). Genomic DNA filters were hybridised at 65°C overnight, while colony hybridisation filters were generally hybridised at 65°C for 3-5 hours.

Unbound labelled DNA was washed off nitrocellulose filters in several changes of 1 x CFHM at 65°C, changing the wash solution several times for five minutes, then twice after 15 minutes and once after one hour and until no more radioactivity could be detected in the washings.

To remove unbound labelled DNA from Hybond N filters, filters were washed either in a high stringency wash (for probes which were 100% homologous to DNA on the filter) or a low stringency wash (for less homology). For the high stringency wash filters were first rinsed in 40 mM $NaHPO_4$, pH 7.2, 1% SDS at 65°C 3 times for 5 minutes and then 3 washes of 15 minutes at 65°C. Filters were then transferred to 0.1 x SSC, 0.01% SDS and washed for one hour at 65°C with 2-3 changes of the solution.

For low stringency wash, the conditions were as above with the first wash in 0.5 M $NaHPO_4$, 1% SDS and the second in 1 x SSC, 0.1% SDS, 10 µg/ml competitor DNA (prepared as described by Harris, Ph.D thesis,

1985).

After washing both types of filters, filters were rinsed in 3 x SSC, blotted dry on 3 mM paper, allowed to dry at room temperature and then reconstructed onto a glass plate or film cassette for autoradiography.

Hybond N filters to be rehybridised with other probes were washed in 0.4 M NaOH for 5 min at 45°C. Filters were then transferred into 0.1 x SSC, 0.1% SDS, 0.2 M Tris-HCl, pH 7.5 and incubated for a further 30 minutes at 45°C. Filters were blotted dry and hybridised with a new probe as described above.

2.8.2 RNA filters

RNA filter hybridisations (Northern blot analysis) were carried out using the method of Thomas (1980) for both nitrocellulose and Hybond N filters. Filters were cut into strips and prehybridised at 42°C with gentle agitation for 8-20 hours in prehybridisation buffer (50% formamide (v:v), 5 x SSC 250 µg/ml denatured salmon sperm DNA, 1 x Denhardts, 50 mM sodium phosphate buffer, pH 6.5). Filters were transferred to hybridisation buffer (hybridisation buffer= 4 parts prehybridisation buffer plus 1 part 50% dextran sulphate) to which the probe had been added, exactly as for DNA filter hybridisations. RNA filters were always hybridised at 42°C overnight.

To remove unbound labelled DNA, filters were washed at room temperature in soap boxes in 4 changes of 2 x SSC, 0.1% SDS, followed by 4 changes of 1 x SSC, 0.1% SDS and finally in 2-3 changes of either 0.1 x SSC, 0.1% SDS, if the probe sequence was 100% homologous to the particular RNA on the filters or a lower stringency 0.25 x SSC, 0.1% SDS for less homology. The final washes at 0.1 x SSC or 0.25 x SSC were carried out at 50°C in a shaking waterbath. Filters were blotted on

Whatman 3 MM paper and allowed to dry at room temperature before reconstruction and autoradiography. Nitrocellulose filters to be rehybridised were washed extensively in water at 42°C to remove hybridisaed ^{32}P-labelled DNA. Hybond N filters were washed for 1-2 hours at 65°C in 5 mM Tris-HCl, pH 8.0, 2 mM EDTA, 0.1 x Denhardts solution. Filter strips containing single-stranded DNA mobility markers from Northern blots were hybridised using RNA filter hybridisation conditions.

2.9 M13 SEQUENCING

2.9.1 Preparation of M13 recombinants

Sonication of Plasmid DNA and Isolated endonuclease fragments

Large restriction endonuclease fragments to be sequenced were isolated as described in Section 2.5 and self-ligated overnight before sonication. Sonication was performed in a sonicating waterbath (Kerry Ultrasonics Ltd) containing 1-2cm of water. A 1.5ml Eppendorf tube containing ≤15µg of recombinant plasmid DNA or 5µg of isolated restriction endoclease fragment (previously self-ligated) in a total volume of 30µl (made up with water) was placed on the bottom of the waterbath for 4 x 30 second bursts of sonication. Between each burst of sonication the solution was placed on ice and given a quick spin to bring the solution back to the bottom of the tube. The appearance of a "mist" on the sides of the tube was an indication of successful sonication. A 1µl aliquot was electrophoresed against pBR332 x Sau3A markers on an agarose gel to check that after sonication the majority of the DNA was a smear between 1.2 and 0.6 kb in size. The procedure was repeated until this size range was achieved. DNA was recovered after phenol extraction by two ethanol precipitations, vacuum dried then redissolved into 10µl of H_2O ready for end-repair.

End-repair of sonicated DNA and size-selection

The sonicated DNA was end-repaired by incubation overnight at 15°C in the following mixture;

DNA	(in H_2O)	20µl
10x ligase buffer	(500 mM Tris-HCl,100 mM $MgCl_2$ 100 mM DTT,pH 7.5)	3µl
TM buffer	(100mM Tris-HCl,100mM $MgCl_2$,pH 7.5)	3µl
Spermidine	0.1 M	1.2µl
sequence chase mix	(0.25 mM solution of each dNTPin TM buffer)	2µl
DNA polymerase I	(10 units of large fragment Klenow enzyme)	2µl

After end-repair sonicated plasmid of fragment DNA was electroporesed in a 1.5% preparative agarose gel against pBR322 x Sau3A markers. DNA between 400-900 bp was collected on DE81 paper, recovered (Section 2.5) and redissolved in 20 µl of H_2O

Preparation of stock M13 vector DNA

2µg of M13mp18 or M13mp19 RF DNA were cleaved at the desired cloning site by a restriction endonuclease(s) (for blunt ended subtrates the enzyme used was SmaI) then recovered after phenol extraction by ethanol precipitation, vacuum dried and redissolved into 10 mM Tris-HCl, pH 7.5. The DNA was then treated with calf intestinal alkaline phosphatase as described in Section 2.5 and the DNA diluted to 20µg/ml with 10mM Tris-HCl, pH 7.5.

Ligation of sonicated DNA into M13 vector DNA

The following ligation reaction mixes were prepared;

Size selected DNA partials	1µl	2µl	4µl
Phosphatased M13 vector DNA	1µl	1µl	1µl
10 mM ATP	1µl	1µl	1µl
20 mM Spermidine	2µl	2µl	2µl
10x Ligase buffer	1µl	1µl	1µl
H_2O	4µl	3µl	1µl

400 units of T4-DNA ligase were added to each ligation mixture which were incubated overnight at 15°C. A further 0.5µl of 10 mM ATP and 100 units of T4 DNA ligase were then added and ligations incubated at 4°C for a further day (after which ligations could be stored at -80°C indefinitely).

Transfection of recominant M13 into the E.coli strain JM101

Competent cells were prepared by a modification of the method of Kushner (1978).

E.coli JM101 was grown overnight with shaking, at 37°C in Luria broth containing thiamine (2 µg/ml). 0.5ml of bacterial culture were diluted 1/100 in the same medium and grown to an OD_{600}=0.3. Excess culture was kept at room temperature (for later use) while 1.4 ml aliquots of cells (one aliquot for each ligation reaction) were pelleted by a 30 second spin in an Eppendorf centrifuge. The supernatant was flicked off and the cells resuspended gently in 0.5 ml of MR (MR= 10 mM MOPS, 10 mM Rubidium chloride (RbCl) pH 7.0). The cells were pelleted again and the supernatant removed as before. Cells were resuspended in 0.5 ml of MRC (MRC= 100 mM MOPS, 10 mM RbCl, 50 mM $CaCl_2$, pH 6.5) and left on ice for 45 minutes. After another 30 second spin pelleted cells were resuspended in 0.15 ml of MRC and kept on ice. 3 µl of

dimethylsulphoxide (DMSO) and 5 µl of the ligation mix were added to each tube of now "competent" cells. The mixture was left on ice for 1 hour, heat-shocked at 55°C for 35 seconds, cooled on ice for 1 minute, then held at room temperature. Before plating out the cells were transferred to a glass tube containing 200 µl of JM101 log phase cells (from the initial culture kept at room temperature), 25 µl of 25 mg/ml BCIG (in dimethylformamide) and 25 µl of IPTG in(H_2O). 3 ml of LUB soft agar were added and cells plated out on LUB agar plates and incubated overnight at 37°C.

White and blue "plaques" develop overnight in the bacterial lawn corresponding to recombinant and nonrecombinant M13 transformants respectively. "Plaque" in this sense refers to an area of reduced bacterial lawn growth due to *in vivo* M13 replication rather than the cell death associated with the virulant replication cycle of lambda. White plaques were screened for recombinant sequences and single-stranded DNA sequencing templates prepared from positives as described by Weller *et al*. (1984).

2.9.2 Sequencing of M13 recombinants

Sequencing of M13 recombinant clones was based on the methods of Sanger *et al*. (1980) and Biggin, Gibson and Hong, (1983) for M13-dideoxyribonucleotide chain-termination using [α-^{32}P]-dAPT and [A-^{35}S]-dATP respectively. Quantities quoted are for 15 sequencing templates. Reactions were performed in 1.5 ml Eppendorf tubes and all centrifugations were done in an Eppendorf bench centrifuge.

Cloned template DNAs were incubated at 60°C for 10 minutes prior to annealing to ensure they were completely dissolved. Each clone was annealed to the 17-mer universal primer by adding 5 µl of clone DNA to 5 µl of primer mix (7.2 µl of 2 µg/ml 17-mer primer, 64 µl H_2O, 8 µl TM

buffer (100 mM Tris-HCl, pH 8.0, 100 mM $MgCl_2$) and incubating at 60°C for 60 minutes with a quick centrifugation after 30 minutes. The annealed clones were held at room temperature until the sequencing reactions were started. For each clone, 4 reaction tubes were prepared containing 1µl of annealed clone plus 1 µl of either a "T","C","G", or "A" NTP mix added to the bottom of the tube.

NTP mixes for sequencing were as follows

	"T"	"C"	"G"	"A"
0.5mM dTTP	12.5	250	250	250
0.5mM dGTP	250	250	12.5	250
0.5mM dCTP	250	12.5	250	250
10mM ddTTP	12.5(6.2)			
10mM ddCTP		2(4)		
10mM ddGTP			4(8)	
10mM ddATP				0.75(1.2)
TE buffer	500	500	500	250

(M13 TE buffer= 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA). 0.5 mM dNTP and 10 mM ddNTPs were prepared in TE buffer. The figures in brackets are the amount of ddNTP added to the mix when ^{35}S-adATP was the radiolabelled substrate.

Tubes were placed in a 37°C water bath and 1µl of freshly prepared "Klenow" mix (^{32}P- Klenow mix= 117 µl TE buffer, 16 U Klenow polymerase, 10 µl α-^{32}P-dATP, 0.7 µl 50 µM dATP; ^{35}S-Klenow mix= 114 µl H_2O, 10 U Klenow polymerase, 10 µl α ^{35}S-dATP) was added. The solution was mixed gently and incubated further at 37°C for 20 minutes. After 20 minutes (which is approximately the time taken to add the Klenow mix to 60 tubes) 1µl of sequence chase mix (sequence chase mix= 0.25mM each of dATP, dGTP,

dTTP, dCTP made up in M13 TE buffer) was added to each tube (in the same order as before), mixed and incubated a further 20 minutes at 37°C. At this point the reaction tubes can be prepared for loading onto a sequencing gel by addition of 4µl of formamide dye (stock solution containing 10 ml deionised formamide, 10 mg xylene cyanol FF, 10 mg bromophenol blue, 0.2 ml 0.5 M EDTA, pH 8.0). ^{32}P-labelled substrates were run as soon as possible while ^{35}S-labelled substrates could be stored for several days at -20°C or -80°C if the formamide dye had not been added.

M13 sequencing gels

Preparation and running of 40cm 6%(w/v) polyacrylamide buffer gradient gels were essentially as described by Biggin, Hong and Coulson,(1983). A "sharks" tooth comb was used which enabled 15-18 clones to be run on a single gel. The standard mixes used to prepare two buffer gradient gels are given below.

	0.5x	2.5x
acrylamide	17.1g	2.28g
bisacrylamide	0.9g	0.12g
urea	150g	20g
sucrose	-	2g
10 x TBE	15ml	10ml
H_2O up to final vol	300ml	40ml

The solutions were filtered through a Whatman no. 1 filter using a Buchner funnel and then the following solutions were added::

	0.5x (150ml)	2.5x (20ml)
10% APS	1.05ml	0.14ml
TEMED (prior to on pouring)	72µl	9.6µl

Sequencing reaction mixes were boiled for 3 minutes and 1.5 µl of

the (^{32}P) or 2.5 µl(^{35}S) samples loaded onto the gel. After electrophoresis for 3-8 hours at 1.4-1.7 kv the gels were fixed in a 10% (v:v) methanol, 10% (v:v) acetic acid solution for 15 minutes, then dried on to Whatmann 3MM paper using a Bio-rad Gel drier. Gels were autoradiographed at room temperature for 16 hours-4 days.

2.10 COMPUTING

Various computing facilities were used in the course of this work. The dot-matrices presented were prepared using a DNA manipulation package written in Fortran 77 by Dr. Z. Nugent and run on Leicester University facilities comprising a Digital Equipment Corporation Vax 8600 computer and a Calcomp 936 graph plotter. Alignment of DNA sequences was also performed on the Vax 8600 using Basic program written by Dr. C. Boyd.

M13 clone sequences were aligned against reference sequences and each other using a version of the Staden (1980) programme modified to run on a Vax11/785. The Vax11/785 was also used in conjunction with the word-processing package Word 11 to format the DNA sequences as presented. A BBC/ACORN microcomputer was used to run a package of DNA sequence manipulation programmes written in BBC Basic by Dr. A. J. Jeffreys. The dot-matrix program in this package is based on an algorithm of White *et al*. (1984).

2.11 CELL CULTURE TECHNIQUES

2.11.1 Growth conditions

The cell culture procedures described below are for the growth of the mouse myoblast cell line G8-1. Specific details of growth conditions for this cell line were obtained from Dr. M. Webb (Institute of Neurology, London) and other general cell culture procedures were as described by Mrs. Tania Cresswell-Maynard (ICI/Joint laboratory,

University of Leicester) or as described in Freshney (1983).

The G8-1 mouse myoblast cell line was grown as monolayers of cells at 37°C and in 5% CO_2 in Dulbeccos modification of Eagles medium (DMEM) supplemented with 10% foetal calf serum. When myoblasts are confluent they may be stimulated to fuse and differentiate by the replacement of the normal growth medium with low serum containing medium (DMEM supplemented with 5% horse serum and 0.5% foetal calf serum).

The thymidine kinase deficient mouse L cells (Ltk^- fibroblast cell line) used in this work were similarly grown in DMEM supplemented with 10% foetal calf serum at 37°C and in 5% CO_2. All cell culture manipulations were carried out in a Class II Microbiological safety cabinet (Medical Air Technology Ltd.). All glassware to be used in making up solutions for use with cells was treated in a solution of 10% chromic acid overnight, rinsed extensively in distilled water before drying and sterilisation at 200°C for 5 hours. All solutions to be used on cells were made up in double distilled autoclaved water and filter sterilised using 0.2µm Acrodisc filter units (Gelman Sciences Inc.).

2.11.3 Subculture of Proliferating Myoblasts and Fibroblasts

Subconfluent cells in dishes or flasks were rinsed with PBS (~2ml per 10 cm dish) and the rinse discarded. Approximately 2 ml per 10 cm dish of 1 x Trypsin/EDTA (Gibco) were added carefully to the cell monolayer and the dish or flask tilted from side to sided to completely cover the cells. The trypsin solution was removed after ~ 15 seconds and the dish or flask containing the cells placed back in the incubator for up to 3 minutes until cells had 'rounded-up' as judged under a phase contrast microscope. Medium was added to the rounded cells, usually 3 mls were added per 10 cm dish, and cells dispersed by repeated pipetting over the cell monolayer. It was always ensured that a single cell

suspension was achieved for uniform growth after reseeding and also that all cells were removed from the dish by this procedure to prevent selection for trypsin sensitive cells. The number of cells per dish were counted using a haemocytometer and were diluted to the appropriate seeding concentration by adding the appropriate volume of cells to a set volume of medium in a culture flask or dish (or when more precise seeding densities were required, cells were diluted to the total volume required and distributed to dishes or flasks). For transfection (Section 2.14), cells were seeded at 1 x 10^6, for routine subculture cells were diluted 1:10 (G8-1 myoblasts) and 1:20 (clone 1 myoblasts and fibroblasts) without counting. Flasks and dishes were returned as soon as possible to the incubator after sub-culture.

2.11.3 Media Changing of Cells

Medium was removed from the cell monolayer and replaced with the same amount of fresh medium which was previously warmed to room temperature from 4^oC storage. This was carried out every three days for low density, proliferating cells and for fusing myoblast cultures.

2.11.4 Growth curve

Myoblasts were trypsinised as for regular sub-culture and cells diluted as appropriate to seed fourteen 5 cm dishes at 1 x 10^4 cells per dish. Cells were mixed well and returned to the incubator. After 24 hours, two dishes of cells were counted using a haemocytometer. At low cell concentrations, trypsinised cells were centrifuged at 800 rpm in a Digifuge centrifuge for 5 minutes and the cell pellet resuspended, without frothing, in a very small volume of medium (about 20-50µl). At higher cell concentrations, cells were counted directly without the centrifugation step. Cells were sampled every 24 hours and fresh medium

was added every 2 days. Myoblasts clumped together after about 7 days growth due to the initiation of cell fusion. A growth curve of total cells versus growth in days was plotted to give an indication of the doubling time.

2.11.5 Isolation of Clones from Diluted Cells

To carry out cell cloning, cells were first diluted at 1:1000, 1:100, 1:10 into 5 ml of medium in small (5 cm) culture dishes and mixed well to disperse cells. Diluted cells were incubated for several days until colonies of cells became visible which contained a few hundred cells and appeared as a small round area under the microscope. Colonies were picked using sterile cloning cylinders. Medium was removed from the dish containing colonies and cells washed with PBS which was discarded. Cloning cylinders, greased lightly with vaseline on their bottom surface, were placed over the colonies which are easily visible by eye after removal of medium from the dish. A drop of trypsin was added inside the cylinder and removed immediately. Cells inside the cylinder were observed under the microscope until they had rounded up and were then removed in a few drops of media added inside of the cylinder. Cells from the single colonies were transferred into 2 cm culture dishes containing 1 ml of medium. The cells were allowed to grow to confluency after which time cells were subcultured into a larger sized dish and finally the number of dishes of cells built up to approximately thirty 10 cm dishes of cells. When cells were confluent (~3-5 x 10^6 cells per dish) they were frozen down at 3 x 10^6 cells per ampoule into liquid nitrogen (see below). The characteristics of the particular cell line (for example, the fusion of myoblasts to form myotubes) were checked before freezing. Possible mycoplasma contamination of cells was checked before freezing by Tania Cresswell-Maynard (ICI/Joint Laboratory, University of Leicester).

2.11.6 Freezing and Thawing of Cells

G8 myoblasts were grown up and frozen as soon as possible after arrival of the cells. Myoblasts and Ltk^- fibroblasts were frozen in a solution containing 10% DMSO and 10% foetal calf serum in liquid nitrogen. Cells were grown to a density of 3-5 x 10^6, trypsinised from the culture dish and pelleted by centrifugation at 800 rpm for 5 mins in a Digifuge centrifuge before resuspension in PBS and a second centrifugation. The PBS supernatant was removed and cells were gently resuspended in 1 ml of freezing medium (freezing medium= 8 ml normal growth medium, 1 ml FCS, 1 ml cryoprotective agent-10% DMSO). Cells were counted using a haemocytometer and volumes adjusted to give ~3 x 10^6 cells per ml of freezing medium in one ampoule. Before transfer into liquid nitrogen, ampoules were frozen at -70°C overnight.

Cells were thawed rapidly from liquid nitrogen storage by placing ampoules directly into a 37°C water bath. When thawed, cells were transferred to a 10 cm dish containing 10 ml of growth medium which had previously been warmed to 37°C. Cells were mixed gently and after ~6 hours cells were checked under a phase contrast microscope to have attached to the bottom surface of the dish. If cells had attached, the medium above them was replaced with fresh medium and cells incubated overnight before subculturing the next morning.

2.11.7 Photography of myoblasts and myotubes

Myoblasts and myotubes were visualised by bright-field microscopy using a phase contrast microscope and photographed using an Olympus OM1 N camera attached to the microscope.

2.12 PREPARATION OF RNA FROM MAMMALIAN CELLS IN CULTURE

2.12.1 Total RNA

Preparation of total RNA from cell monolayers was based on the method of Chirgwin *et al.* (1974) with modifications described below.

Up to 5 10 cm dishes of cells were used for large preparations but generally 1-2 10 cm dishes of sub-confluent myoblasts or differentiated myotubes were used. Cell monolayers were washed once with ice cold PBS and drained. 1 ml of lysis buffer (lysis buffer= 7 M urea, 2% SDS, 0.35 M sodium chloride, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) was added per plate and lysed cells were pooled in a sterile flask. An equal volume of phenol:chloroform was added and mixed into the cell lysate until an emulsion formed. The mixture was centrifuged at 10,000 rpm in an HB4 rotor for 10 mins at 4°C and the upper aqueous layer removed carefully from the phenol layer. The aqueous layer was extracted with chloroform:isoamyl alcohol (50:1 v:v) until it was clear. 0.4g of caesium chloride (CsCl) were added per ml of aqueous phase and when the CsCl was dissolved, this mixture was layered gently over 2 mls of a CsCl cushion in an SW50 centrifuge tube. (CsCl cushion= 5.7 M CsCl, 0.1 M EDTA, filtered using a 0.2μm millipore filter). The CsCl gradient was centrifuged in an SW50 rotor at 32,000 rpm for 15 hours at 20°C to pellet RNA. After the centrifugation the supernatant was removed by careful pipetting. DNA is contained within the CsCl cushion and is apparent as a viscous region in the supernatant. DNA may be obtained from the viscous supernatant by dialysing extensively against TE to remove the CsCl. After removal of the supernatant from the tubes, the RNA pellets were drained by inverting the tubes. 2 mls of ice cold 10 mM Tris-HCl, pH 7.5 were added to resuspend the RNA pellets. Resuspension at this stage took a long time and was acheived by pipetting the RNA solution for several minutes with a Gilson pipetteman while kept on ice.

When dissolved, RNA was ethanol precipitated overnight at -20°C. The precipitated RNA was pelleted at 10,000 rpm in an HB4 rotor for 20 minutes at 0°C. RNA pellets were washed with 80% ethanol, 0.1 M sodium acetate pH 7.0 and dried briefly under vacuum. Dried pellets were resuspended in the appropriate volume of DEPC treated water and stored frozen in aliquots at -70°C. RNA concentrations were determined by measurement of the optical density at 260 nm using a Cecil Instruments CE202 ultraviolet spectrophotometer with CE135 micro-sipette control attachment. An A_{260nm} of 24 is equivalent to a concentration of 1 mg/ml of RNA. Quality of RNA preparations were tested by agarose gel electrophoresis in 1.5% agarose gels.

2.12.2 Cytoplasmic RNA

Cytoplasmic RNA was prepared from cell monolayers by the method of Gorman et al. (1982). Cells were trypsinised from the culture dish, pelleted and washed with ice cold PBS. Cell pellets (usually frozen pellets stored at -80°C) were lysed in approximately 6 volumes of lysis buffer. (Lysis buffer= 10 mM Tris-HCl pH 7.4, 10 mM sodium chloride, 3 mM magnesium chloride, 0.5% Nonidet P-40) for 5-10 minutes at 0°C. Nuclei were removed by centrifugation at 1900 rpm in an HB4 rotor for 5 minutes at 4°C. Cytoplasmic RNA was isolated from the supernatant by addition of an equal volume of RNA extraction buffer (RNA extraction buffer= 7 M urea, 0.35 M sodium chloride, 10 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS and ~30 µg/ml of tRNA as a carrier). This mixture was phenol extracted twice and then extracted twice with chloroform:isoamyl alcohol (50:1 v:v). After ethanol precipitation of the RNA at -20°C overnight, RNA was collected by centrifugation at 10,000 rpm in an HB4 rotor for 20 mins at 0°C. RNA pellets were washed in 80% ethanol, dried briefly under vacuum and resuspended in the appropriate volume of

DEPC treated water and stored as above. RNA concentrations and quality were determined as above.

2.12.3 Polyadenylated RNA

Polyadenylated RNA (poly(A)$^+$ RNA) was purified by oligo(dT) cellulose chromatography as described by Aviv and Leder (1972). Columns of 50 mg of oligo d(T) cellulose (bed volume 0.2-0.3 ml) were prepared in siliconised Pasteur pipettes plugged with polyallomer wool and equilibrated with several bed volumes of 1 x binding buffer (1 x binding buffer= 0.5 M Lithium chloride, 0.5% SDS, 1 mM EDTA, 10 mM Tris-HCl, pH 7.5). Total and cytoplasmic RNAs were first heat-shocked for 3 minutes at 65^{o}C and cooled on ice before the addition of an equal volume of 2 x binding buffer. RNA was loaded onto the column and the eluate collected in a sterile siliconised tube, heat-shocked and reloaded on to the column. This procedure was repeated twice. Unbound RNA (the poly(A)$^-$ fraction) was washed through the column with ~20 bed volumes of 1 x binding buffer and then the poly(A)$^+$ RNA eluted with ~10 bed volumes of elution buffer (elution buffer= 0.05% SDS, 1 mM EDTA, 10 mM Tris-HCl, pH 7.5).

Poly(A)$^+$ RNA was recovered by ethanol precipitation with a -20^{o}C chill overnight and centrifugation at 10,000 rpm in an HB4 rotor for 30 minutes at 0^{o}C. RNA pellets were washed with 80% ethanol, dried briefly under vacuum and redissolved in a small volume of DEPC treated water. Concentrations and quality were determined as above.

2.13 PREPARATION OF GENOMIC DNA FROM MAMMALIAN CELLS IN CULTURE

Genomic DNA was prepared from cell monolayers using a procedure based on the method of Kirby (1957).

Generally, ten 10 cm dishes of cells were used for each

preparation. Media was removed from the cell monolayers which were then rinsed briefly on ice with 2 mls of 1 x SSC per dish. Cells were lysed by the addition of 2 mls of SE buffer containing 1% SDS (SE buffer= 150 mM NaCl, 100 mM EDTA, pH 8.0). Lysed cells were pooled in a conical flask, phenol extracted with gentle shaking to avoid shearing of high molecular weight DNA and centrifuged at 13,000g for 5 minutes at 4°C. The upper aqueous phase was decanted from the phenol layer and phenol extraction was repeated. Nucleic acids, remaining proteins and carbohydrates were precipitated from the aqueous phase by ethanol without chilling. The flocculent precipitate formed was removed from the ethanol solution using a 'hooked' siliconised Pasteur pipette and the precipitate redissolved in 0.1 x TNE (TNE = 50 mM Tris-HCl, 5 mM EDTA, 100 mM NaCl, pH 7.5) at 4°C. This usually took overnight. The DNA solution was then incubated for 15 minutes at 37°C with 100 µg/ml of pancreatic RNase A (from a 20 mg/ml stock in 0.15M sodium chloride which had previously been treated at 80°C for 15 minutes to inactivate DNase).

0.1 vol of 10% SDS, 0.05 volume of 20 x TNE pH8.0 were then added with 100 µg/ml of proteinase K and the solution incubated for a further 15 minutes at 37°C. The DNA solution was phenol extracted, ethanol precipitated, washed as described above and redissolved in a small volume of 10 mM Tris-HCl pH 7.5. Ethanol precipitation, washing and redissolving were repeated at least three times and the DNA solution dialysed overnight at 4°C against two litres of 2 mM Tris-HCl, 0.1 mM EDTA, pH 7.5 with several changes of this solution.

DNA concentration was determined as for RNA concentration. An absorbance of 20 at 260 nm is equal to 1 mg/ml of double-stranded DNA.

The quality of the DNA preparation was determined by agarose gel electrophoresis of 0.5 µg of native DNA. Samples were run against λ DNA x HindIII molecular weight markers. The DNA obtained by this method had

a double-stranded size greater than 23 kb.

2.14 INTRODUCTION OF EXOGENOUS DNA INTO MOUSE MYOBLASTS

The procedure used to transfect myoblasts was the calcium phosphate precipitation technique of Graham and Van der Eb (1973) with modifications described by Wigler et al. (1978) and Gorman, Moffat and Howard (1982).

Myoblasts were thawed from liquid nitrogen storage as described above and subcultured only once if possible but always less than six times before seeding the required number of 10 cm dishes the day before the transfection at a density of 1×10^6 cells per dish.

If cells were spread evenly over the surface of the dish they were used in the transfection experiment. Cells were refed with 8 mls of fresh growth medium three hours before the addition of the calcium phosphate ($CaPO_4$)-DNA precipitates which were prepared as follows.

A total of 20 µg of DNA, (10 µg of the test plasmid and 10 µg of sonicated salmon sperm carrier DNA) were used to transfect each 10 cm dish of myoblasts. The quality of both test plasmid and carrier DNA is important for successful transfections. Plasmid DNAs were prepared as described in Section 2.5 and always tested before use by agarose gel electrophoresis to be more than 50% in the supercoiled form and if not were not used in transfections as plasmid DNAs which have become substantially nicked have been shown to give unreproducible transfection efficiencies (Gorman, Moffat and Howard, 1982).

The salmon sperm carrier DNA was prepared as described by Harris (Ph.D thesis, 1985) with additional phenol extraction and ethanol precipitation steps. The molecular weight of the salmon sperm DNA was reduced by several passages through a 19g needle and DNA purified by a further phenol extraction and ethanol precipitation. Care was taken not

to extensively break up the high molecular weight DNA since if the molecular weight is less than 10 Kb a reduction in uptake and expression of exogeneous DNA has been reported (Spandidos and Wilkie, 1983). Carrier DNA prepared and used in all transfections gave a smear of fragments beginning from the slot of the gel when DNA was electrophoresed on a 1% agarose gel.

DNAs to be used in transfections were dissolved in sterile 0.1 x TE (stored at 4°C) to give a final volume of 450 µl. 50 µl of 2.5M calcium chloride (freshly prepared, filter sterilized and cooled to 4°C just before use) were added slowly to the DNA solution and the resulting solution mixed well. At this stage manipulations were carried out in sterile Eppendorf tubes. After the addition of the calcium chloride solution, the whole solution was transferred to a sterile plastic 10 ml centrifuge tube (Falcon, cone bottom). 0.5 ml of 2 x HBS (2 x HBS= 280 mM NaCl, 50 mM HEPES, 1.5 mM Na_2PHO_4, pH 7.1; solution stored at 4°C for no longer than one month, pH is critical) was added dropwise into the DNA-calcium chloride solution from a 1 ml pipette using a pipette aid and the centrifuge tube shaken during the addition. After all the 2 x HBS had been added, the solution was sucked up and down the 1 ml pipette a number of times and then air blown gently into the solution through the pipette for several seconds. After ~15 minutes at room temperature a $CaPO_4$-DNA precipitate appeared as a "bluish haze" in the solution. Any coarse, floccular precipitates were discarded. Such coarse precipitates rarely formed unless solution volumes were scaled up above a 2-3 ml final volume.

After 15-20 minutes at room temperature, the whole precipitate made above was added to a 10 cm dish of cells, mixed and cells returned to the incubator as soon as possible. Monolayers observed one hour after the addition of the precipitate showed a grainy precipitate over the

surface of the cells when viewed under a phase contrast microscope. The cells were exposed to the $CaPO_4$-DNA precipitate either overnight (~ 15 hours) or for 6 hours. After the overnight incubation, the precipitate was removed from above the cells and the cells refed with fresh growth medium. After the 6 hour incubation, cells were subjected to a glycerol shock. They were first washed twice with PBS and the wash discarded. 2 mls of 20% glycerol were then added carefully down the sides of each 10 cm dish and the dish tilted from side to side to spread the glycerol over the monolayer surface. After 2.5 mins, the glycerol was carefully removed and cells washed twice with PBS. Fresh medium was then added to the cells.

The next day cells had their medium changed and ~ 48 hours after the addition of the precipitate cells were harvested by trypsinisation and washed in PBS. Cell pellets were frozen and stored at -80°C before they were used to prepare RNA (Section 2.12), or to make cell extracts for enzyme assay (Section 2.15).

When myoblasts were shifted to low serum medium after transfection, the medium was changed the morning after transfection to low serum medium. Cells were then harvested by trypsinisation over the time course of myoblast differentiation and cell pellets stored as above.

The same transfection procedure was used to transfect Ltk^- fibroblasts but fibroblasts were seeded at a density of 2×10^6 the day before transfection and the glycerol shock after 6 hours transfection was for 1 minute only.

2.15 ANALYSIS OF TRANSIENT EXPRESSION OF EXOGENOUS DNA IN MOUSE MYOBLASTS

2.15.1 Analysis of Specific RNA Transcripts

RNA was prepared from transfected cells as described in Section 2.12 and analysed on Northern blots with the appropriate DNA hybridisation probe as described in Sections 2.6-2.8.

2.15.2 Preparation of Cell Extracts

Cell extracts were prepared as described by Ott et al., 1984 and modifications are described. The cell pellet (sometimes frozen) from each 10 cm dish transfected (containing around 5 x 10^6 cells) was resuspended by gentle pipetting in 50 µl of freeze-thaw buffer (freeze-thaw buffer= 250 mM Tris-HCl, 5 mM DTT and 5% (v:v) glycerol, pH 7.8). Suspended cells were transferred into an Eppendorf tube on ice and then subjected to freeze-thaw lysis. Pellets were frozen for ~15 seconds in liquid nitrogen and then thawed for ~2 minutes at room temperature and then 1 minute at 37°C. The procedure was repeated twice and cells placed back on ice immediately. After the three cycles of freeze-thaw, cells were viewed under a phase-contrast microscope to check for lysis. Lysed cells appeared spherical but with very 'ragged edges' and had lost phase brightness. Lysed cells were then centrifuged at full speed in an Eppendorf centrifuge for 15 minutes at 4°C. The supernatant was removed into a fresh Eppendorf tube on ice. Extracts were used immediately.

Protein concentrations in cell extracts were determined as described by Spector (1978).

2.15.3 CAT Assay

The CAT assay was carried out as described by Gorman, Moffat and

Howard (1982) and any modifications are described below.

The required volume of cell extract to be assayed was added to a fresh screw-top Eppendorf tube on ice. Equivalent volumes of extract containing about the same concentration of protein were assayed. The cell extract to be assayed was heated at 60°C for 5 minutes to remove contaminating deacetylases (R. Aft, D. Yaffe, personal communication). The components of the assay mixture (total volume 180 µl) were 100 µl 0.25 M Tris-HCl, pH 7.5, 0.2 µCi ^{14}C-chloramphenicol (specific activity: 1.85 GBq/mmol, unless otherwise otherwise stated) and cell extract (25-50 µl). The solution was incubated at 37°C for 5 minutes and the reaction started by the addition of acetyl coenzyme A to 4.4 mM. The reaction was allowed to proceed for 4 hours (or overnight in some cases) and then assays were terminated by plunging tubes onto ice and adding 1 ml of ice cold ethyl acetate which extracts chloramphenicol and its derivatives. Solutions were mixed well by vortexing and the two layers separated by centrifugation for 1 minute in an Eppendorf centrifuge. The upper organic layer was removed into a glass tube and a further 0.5 ml of ice cold ethyl acetate added to the aqueous layer and the extraction repeated. The combined organic layers were dried down under vacuum and then taken up in 20 µl of ice cold ethyl acetate. The total 20 µl (in most cases) containing the radioactive chloramphenicol and derivatives were spotted onto a silica gel plate (sil G, CAMLABS) 2 cm from the bottom edge using a drawn out Pasteur pipette to deliver small volumes at a time. The spots were allowed to dry and the silica gel plate placed in a chromatography tank and run with chloroform:methanol (95:5 v:v) ascending.

After the solvent front had run approximately three quarters the distance up the silica gel plate (~1-1.5 hours), the plate was removed and allowed to dry before autoradiography. After autoradiography spots

corresponding to radioactive signals were cut out of the silica gel plate and counted. Counting was carried out by adding equal sized spots to 2 mls of scintillation fluid (Scintillation fluid= 4g of PPO and 60.5mg of POPOP in 1 litre of toluene) in a scintillation vial. Vials were counted in a Packard liquid scintillation counter for one minute each and counting was repeated at least once.

Data were expressed as the percentage of chloramphenicol acetylated per volume of extract used or as the total amount of acetylated chloramphenicol products formed per volume of extract used.

Purified chloramphenicol acetyl transferase (CAT) protein (Type III) used as a control in CAT assays, was obtained from Dr. Colin Kleanthos (Department of Biochemistry, University of Leicester).

2.16 CONTAINMENT

All experiments carried out in this thesis were conducted with reference to the Genetic Manipulation Advisory Groups guidelines on saftey and containment conditions.

CHAPTER 3

CHARACTERISATION OF THE MOUSE MYOGLOBIN GENE AND SEQUENCE COMPARISON WITH OTHER MAMMALIAN MYOGLOBIN GENES

3.1 Introduction

The recently characterised seal and human myoglobin genes exhibit similarities in their organisation which are markedly different to those of other globin genes (see Section 1.2). To further define whether this is a stable organisation of mammalian myoglobin genes, the mouse myoglobin gene was isolated and characterised.

The mouse myoglobin gene was isolated from the mouse myoblast cell line G8-1 (Christian *et al.*, 1977) to provide a basis for analysing myoglobin gene expression. Myoblast cell lines can mimic myogenesis *in vitro* (see Section 1.4) and there was evidence that the myoglobin gene was expressed during differentiation of this particular myoblast cell line (see Chapter 4). The cloned G8-1 mouse myoglobin gene would allow the exactly homologous gene to be re-introduced into its parental cell type in order to begin myoglobin gene expression analysis.

Cloned seal myoglobin gene exons were used as hybridisation probes to isolate the mouse myoglobin gene from a G8-1 myoblast genomic DNA library.

3.2 A single myoglobin gene in the mouse

Southern blots of digested mouse genomic DNA revealed that there is a single myoglobin gene in the mouse (Fig.3.1). Probes containing exon 1 or exon 2 of the human myoglobin gene (see legend of Fig 3.1 for the details of the probes used) were hybridised to a Southern blot of mouse genomic DNA's from C57BL/10 liver, DBA/2 liver, G8 myoblast and G8

Figure 3.1

Top Southern blot hybridisations of mouse genomic DNA with exon probes isolated from the human myoglobin gene

5µg samples of mouse genomic DNA from myoblasts (MB), myotubes (MT), DBA/12 (D), C57/BL10 (C) mouse livers were digested with BglII and HindIII and electrophoresed in a 0.8% horizontal agarose gel. After acid/alkali denaturation of the gel, DNA fragments were transferred to Hybond N by Southern blotting. The filter was hybridised with a human myoglobin exon 1 probe in 0.5M sodium phosphate, 7% SDS at 65°C overnight followed by washing in 1 x SSC at 65°C and autoradiography. The filter was rehybridised with a human myoglobin exon 2 probe in the same conditions, except the final wash stringency was 0.1 x SSC.

The human myoglobin exon 1 probe was the 216 bp PvuII fragment (Table 3.1) The human myoglobin 2 probe was the 261 bp single-stranded antisense DNA fragment generated from M13HEx2 (Table 3.1).

Bottom A genomic restriction map of the mouse myoglobin gene

A genomic restriction map of the mouse myoglobin gene was determined from the two purified λ clones λ3G8 and λ7G8. Mouse myoglobin exons were located by Southern blot hybridisation of λ DNA with seal myoglobin exon probes. EcoRI and BamHI restriction fragments containing exons 1 and 2 and 3 respectively were subcloned into pUC13 to give pλ3G8E4 and pλ7G8B3 (Blanchetot et al., 1986).

Coding regions are indicated by filled boxes and nontranslated mRNA sequences by open boxes. Diagrams of the genes are represented to scale. The position of a B1 element in the mouse myoglobin 5'-flanking region is represented by a hatched box. The repeat region refers to a simple repeat structure found in both seal and human myoglobin genes (see Section 3.5).

Probe	Length in bp	Probe comprised of
Human exon 1	216	63 bp of 5'-nontranslated region, exon 1 and 58 bp of intron 1
Human exon 2		
M13 HEx2	316	212 bp of exon 2, 42 bp of intron 2 remainder primer and poly-linker
	261	204 bp exon 2, remainder primer and poly-linker
Mouse exon 1		
M13 MEx1	190	119 bp of exon 1, 13 bp of 5'-flanking DNA, remainder primer and poly-linker

myotube digested with BglII and HindIII. Digests were run in 0.8% agarose gels and transferred to Hybond N filters by Southern blotting (Section 2.6). Both probes revealed a single major hybridising fragment (Fig. 3.1). This result was confirmed when cloned mouse myoglobin gene exon probes (see Section 3.3) were hybridised with genomic DNA from myoblasts, myotubes, C57BL/6 mouse liver and BALB/C mouse liver and analysis of double digests of mouse myoblast genomic DNA provided a genomic restriction map which is shown in Fig. 3.1. It can be concluded from these results that, as with human and seal myoglobins, mouse myoglobin is encoded by a single gene.

If lower stringency washes were used with the human myoglobin gene exon 2 probe, a complicated set of hybridising fragments were detected in mouse genomic DNA, similar to the exon 2 related family described by Weller et al. (1984) for the human myoglobin gene (not shown). The exon 2 related family has yet to be characterised since it is not certain whether the phenomenon is due to a real family of myoglobin exon 2 derivatives or whether it is due to the presence of a "repeat" sequence element in the probe that is detecting similar repeats dispersed around the genome.

3.3 Cloning and sequencing the mouse myoglobin gene

The mouse myoglobin gene was cloned from genomic DNA prepared from G8 mouse myoblasts as described in Section 2.13. The mouse gene cloning was carried out by Dr. A Blanchetot using seal myoglobin gene exons 1 and 3 as hybridisation probes and this is described in detail in Blanchetot, Price and Jeffreys, 1986. The mouse myoglobin gene was isolated from the G8 myoblast genomic DNA library in two overlapping lambda (λ) recombinants, λ3G8 and λ7G8, shown in Fig. 3.1. This Figure also illustrates the specific restriction endonuclease fragments which were

subcloned into pUC13 (Messing, 1983) to be sequenced after location of the exons by hybridisation with seal myoglobin exon probes. The recombinant pUC13 plasmids containing the mouse myoglobin gene exons were either sonicated and then end-repaired or partially digested with Sau3A or AluI and resulting fragments ~400-900 bp in size were ligated into the SmaI site (end-repaired fragments) or the BamHI site (partial digest fragments) of M13mp19 (Yanish-Perron, Vieira and Messing, 1985). Nucleotide sequencing of the M13 clones containing the particular regions of interest was carried out in collaboration with Dr. A. Blanchetot using the dideoxy chain termination method of Sanger et al. (1980) and Biggin, Gibson and Hong (1983) as described in Section 2.9.

A total of 4.59 kb of mouse myoglobin DNA were sequenced including the coding regions, 1.2 kb of 5'-flanking sequence, 1.55 kb of intron sequence and 0.25 kb of 3'-flanking sequence. All M13 clones were sequenced at least twice and sequence was confirmed on both DNA strands. The entire sequence is presented in Fig.3.2.

3.4 Organisation of the mouse myoglobin gene

The derived amino acid sequence of mouse myoglobin from the DNA sequence agrees precisely with the recently established mouse myoglobin sequence (D. Harris, personal communication) verifying that the functional mouse myoglobin gene which specifies both skeletal and cardiac muscle myoglobin has been isolated.

The mouse myoglobin gene is interrupted by two introns at codon 31 and between codons 105 and 106, corresponding exactly with the intron positions in human and seal myoglobin genes and all characterized vertebrate α- and β-globin genes. The splice junction sequences of the mouse gene conform to the GT/AG rule of Breathnach and Chambon (1981) and homology to seal and human myoglobin gene sequences extends (~7-15 bp)

Figure 3.2

DNA sequence of the mouse myoglobin gene

Sequences present in the mature myoglobin RNA are shown in capital letters. Positions of the CAP site and poly(A) addition site are labelled and were deduced from the homologous sequence locations in the seal and human myoglobin genes (Blanchetot et al., 1983, Weller et al., 1984). The 5' TATA box and 3' AATAA polyadenylation signal are underlined. The polypyrimidine tract present in the mouse myoglobin gene 5'-flanking region is shown in italic lower case letters and the B1 element found further upstream in the 5'-flanking region is shown in italic uppercase letters bordered by arrows underlining its direct repeat sequences. Within the B1 element, the RNA polymerase III promoter regions are underlined and the putative TATA box is overscored by dashes.

exon1

```
         10         20         30         40         50         60......   70         80         90        100
tttaccctaaggacttctcagggaatttcgagaacatggcGCCGGGCGTGGTGGCGAGCCTTTAATCCCAGCACTTGGGAGGCAGAGGCAGAGGCAGGTG

GATTTCTGAGTTCGAGGCCAGCCTGGTCTACAAAGTGAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCTGTCTCGAAAAAACAAAACAAAACAAAC   200

AAAACAAAACAGAAACAAAAACAAAACAACAAAAAAAGGAAAAAAAAAAAAAAgaacatggcaaagagaagagggtctgtgcctggcagagtccgggag   300

tattggctgccatcaccctgctggaagtgccacatgctgcaccaagggaagtggtggggtaggagatcacaaggaatgggaggagctgagggtttcatgg   400
agcatgggagcacaaccttgcagagaaagccaggagcaagctcagagctgggcattcttcaagtcagagacttgtcttggttggttcagccacaagcaca   500
gacctgttgtggctcctttgtaccactttcagctgggcttagggcatcagcaaagtaagaagagggtggcaagatggcttagtgggtatgaagacttgtc   600
accacggctgaggatctgagtaagattcctgggtctttttgaaaggaaagaatcaatgcctgaaagttgtcttctgacctttacaaacatgctgtagcgc   700

actctctttctctctctctctctctctctctctctctctctctctctctcgtacatgtacatgcacacacacacgtattaaataaacaaaagcagt   800

cagaaagtgagaggaggagacttctaggggtcagcaagatgcctgtgcccaagcttagaaacatgacaagtccctgtggggcagcagacagggggagaat   900
tgggtttcagctgctggcagtggcttcggtgccctttctgtgggcttgtgcaagtccagacagtgacctggctgcttcccccctccccacccccctactgcc   1000
ccaagctgtcctagctcccacaatggcacttgccccaaaatagctgcccatgtgagggccagagaaaggcagagattagacccctggaggttgagcacgg   1100

                                                                                                CAP
tagcaggaagggcatgtggcacccagtgatcctggccagactagcatctgggaaggtataaaagcccttcaggaccaggtggcctcaaATCTCAGCTGAG   1200

                                              IniGlyLeuSerAspGlyGluTrpGlnLeuValLeuAsnValTrpGlyLysValGlu
AGCCAGCCCACACTCTCTCCTTTTGTCCTCTTTAGAAGCCACCATGGGGCTCAGTGATGGGGAGTGGCAGCTGGTGCTGAATGTCTGGGGGAAGGTGGAG   1300

AlaAspLeuAlaGlyHisGlyGlnGluValLeuIleGl(y)
GCCGACCTTGCTGGCCATGGACAGGAAGTCCTCATCGGgtaagaggaagagacccttatctcaatacctgtttgccaaggttcaccatatggcgacaagg   1400

tggaagctggcagggtagaccagttggctgcaatgatgaggtcagatttcacttgcttctaagtgctggagactggagtagagaaataaaaaggggggatc   1500
tatccatctccgctgacgtcagcaaccaaagcaagcttagcatcaaacagtccaccagatgaatcctattacggcagagagtcagaataatgagggcagt   1600
gatagatctcgtctactagatgtggactgaaagcttgggtaatgtggctctcggctgcaaaggatcctagacatgcccagaaaaattgcagattctgagc   1700
ctgggcaagggggtagggttggagggttgcgaggctcatacacagggcctcactgctcacccagaggc                                   1800
```

exon2/exon3

```
         10         20         30         40         50         60         70         80         90        100
ggtcatttacattgccagtggaaagtggttggggttagatctgaagtttatgctcatgtagaggggtctcaacaacagccacagatagtcaggaggtcta
aaacttcaggtttcagcatcttcaggcaggaagcaatagaggcatccctgacacagatgtcttgaggaatagagggtacttctgagtcatatagatccaa   200

                                                                       (Gl)yLeuPheLysThrHisProGluThrLeuA
tggctgcctgccttctctgcacactagccctgagatccacgggtctcactccatctcttcttgtcccacagTCTGTTTAAGACTCACCCTGAGACCCTGG   300

spLysPheAspLysPheLysAsnLeuLysSerGluGluAspMetLysGlySerGluAspLeuLysLysHisGlyCysThrValLeuThrAlaLeuGlyTh
ATAAGTTTGACAAGTTCAAGAACTTGAAGTCAGAGGAAGATATGAAGGGCTCAGAGGACCTGAAGAAGCATGGTTGCACCGTGCTCACAGCCCTGGGTAC   400

rIleLeuLysLysLysGlyGlnHisAlaAlaGluIleGlnProLeuAlaGlnSerHisAlaThrLysHisLysIleProValLysTyrLeuGlu
CATCCTGAAGAAGAAGGGACAACATGCTGCCGAGATCCAGCCTCTAGCCCAATCACACGCCACAAAGCACAAGATCCCGGTCAAGTACCTGGAGgtaggc   500

ggccagagcaagtctccagggcagagatataaatcccagcttagccactcaatacgagtggcctgctttctccccactaagctttctccccagttctcct   600
ctataacctaccccagccctgttcccctgagggtgctgaggagccacacagcatccaatttaatggcttgcccaaatgctgggccagcaccttagtgtga   700
aatatgactaatacacaacttgccttctcctcagaagcgggagctcttgttggacccactggactggacagatgtggaaactgaggcacaggcagcttgg   800
gcaagtcactgaggcaggattcacactcagggcgaatcagttctgtcagagagcggtagacgctttgtttctctgatgctctgctgccccagtccatgat   900
ggttcagacaccagttggaaagtggcctaaagagaggccagcagccctcagtggggtcctgtctgtcttcatttgcaactggagtcacagggcagagca   1000
gggcagagcagggcagagcagggcagggccagacacatacccagcctccctgaagctggggctggttccgcagacccaggctccacagacgtttctgtat   1100
ctgtgaattgtgaggcagatttcccagaggtgtatttgcaaagatgtcctgtgcctctgggtgcacttgtgtccctcggtgtccccttctgctgtgtccc   1200
actctcatctctgctgtgtcacttgcacataagggccccagcccaaaacagcagctgtacccccactcctgagctagcaggggtgtgtccacagtccctg   1300
ggtctaatctcagcttactatctacctgctacatgacctcagcctggtcacttggtagtctcttcagtacagcagccaaagctgtacgggaccaaactct   1400
aagtcaatgacgataagtaagtgttatgatagtttcatcgtgagcctttggattctgggtggtcccgatccccatctaggcaatgtcggtacctgatctc   1500
ccctggcccctctccccttgttagtacctgttcctcagtcccaggtttagggataaaatcacatgatgggaaaagcctcctcccaagtgtcctgattgac   1600
agtggtgggctagccgatgctgagcaaggtatcatgggccggctgctgggcctgcatcctgaagtagtgtgggcacacgagaggtgagaaaggggcccca   1700
gagcgtccctgagtttgagtgaccctccctgactgcatgacttcatgcacacagctatacctgtctttgcctcagtttccccatttacagagcagggatg   1800
gtggtggtgcctgcttccaggtagtcaggtttaaaggagttggttcatggtaattgcatggaatggcacttggtacatcgtaagtgtgcaaggaacagtc   1900

                                                                                             PheIleSe
ctatagtaaggagaaggtcagtgagtacacacccccttagctcatggcttgctccgtcccctgacctacaacctcttgtccctttcttgcagTTTATCTC   2000

rGluIleIleIleGluValLeuLysLysArgHisSerGlyAspPheGlyAlaAspAlaGlnGlyAlaMetSerLysAlaLeuGluLeuPheArgAsnAsp
AGAAATTATCATTGAAGTCCTGAAGAAGAGACATTCCGGGGACTTTGGAGCAGATGCTCAGGGCGCCATGAGCAAGGCCCTGGAGCTCTTCCGGAATGAC   2100

IleAlaAlaLysTyrLysGluLeuGlyPheGlnGlyTer
ATTGCCGCCAAGTACAAGGAGCTAGGCTTCCAGGGCTGAGCCATGGGCTCCCACTGTCCAGCCCACCAAGCTGGGACCCAGTGTTGTGTAGCAAGTAGCG   2200

TGTGCAGTGTTCTACCTTAGCAGAGAACAGAAGAGGGGAGCATAGTGTGGCATCCACCCAGTCCCCAGGGGACAGGGCTCTGGGCTGTGTTACCCTCCAG   2300

CCCAGAGGTGCAAAGTGGCCTTCGTCAGCTCTGCCGGGTCATGCTCAGGTCTCCTAAGTCCCAGTCCATTTTCTTCTGGTTTTGGGAAAATCTCTTTTCC   2400

ACTGTCACATTTGACCCCAAATCCAAGTCACTGACTAGCAGACCCTGACCTTTGGGCGAGATGGAGGGTTGCTTAGAGGGAGTGGAGGGTGAAAACGGGG   2500

                                                                               poly(A)
CGGTGAGCATCAAGTCTCCCACTGCTCAGCTTCCCGTTGACCCACCTTGTCTCAATAAAATATCCTGCGAGTCCTCaattcttgtctccgtccggtatat   2600

ttttatccttccggtaagggaaacatccacaaacgtcgctggaagaaatgggaaggtgcattctgggatggcgggcagacagaagatgtgagacaggccc   2700
ctttctacaatcacctctcaatgacaagcacacactgggctaggttcagggacagggaagaaagggcccccacctgggctctgatcaagatggcgggcc   2800
tgcagacagggtcagctgacc                                                                                  2900
```

over the donor/acceptor sites.

The introns of the mouse myoglobin gene, like those of the human and seal, are substantially longer than introns found in α- and β-globin genes. Intron 1 of the mouse gene is 4.5 kb in length compared to 5.8 kb and 4.8 kb for human and seal genes respectively. Intron 2, however is noticeably shorter than its human and seal counterparts being 1.5 kb in length, compared to 3.6 kb and 3.3 kb for the human and seal genes respectively (Weller et al., 1984, Blanchetot et al., 1983).

The CAP and polyadenylation sites of the mouse myoglobin gene were identified by homology with seal and human myoglobin sequences and are indicated in Fig.3.2. These two sequences have been mapped directly for the seal myoglobin gene (Blanchetot et al., 1983) and the CAP site mapped for the human myoglobin gene (Weller et al., 1986) The sites predict an mRNA of 957 nucleotides (nt) excluding the poly(A) tail (465 bp coding sequence plus 55 bp of 5'- and 437 bp of 3'-nontranslated mRNA sequences). This is approximately 100 nt shorter than human (1066 nt) and seal (1083 nt) mRNAs due to a shorter 3'-nontranslated mRNA sequence (mouse 437 bp, human 531 bp and seal 535 bp, Weller et al., 1984, Blanchetot et al., 1983). This size difference in myoglobin transcipts has been confirmed by Northern blot analysis of mouse and seal skeletal muscle RNA (see Chapter 4).

Upstream of the CAP site in the mouse myoglobin gene 5'-flanking region the only recognisable promoter element (see Section 1.2) is a conserved TATA box, 27 bp before the CAP site, which is underlined in Fig.3.2. The CCAAT promoter element found ~80-100 bp before the CAP site in β-globin genes (Dierks et al., 1983) and frequently at this position in other eukaryotic genes is absent from the mouse promoter region and from human and seal promoter regions. Also, the purine-rich region located ~67-114 bp before the CAP sites in the seal and human genes

discussed in Section 1.2 is not present in the mouse gene. A polypyrimidine tract (almost perfectly $(CT)_{26}$) is found 434 bp upstream of the CAP site. Alternating co-polymers are highly conserved in a number of eukaryotic genes (Hamada, Petrino and Kagunaga, 1982). Pyrimidine-rich stretches have been observed in the 5'-flanking region of several genes. For example, the α-2(1) collagen promoter region contains a pyrimidine rich stretch which is sensitive to S1 nuclease digestion and this S1 sensitivity has also been observed in a pyrimidine repeat sequence in a Drosophila heat-shock gene (Mckeon, Schimdt and de Crombrugghe, 1984, Mace, Pelham and Travers, 1983). The model proposed for S1 nuclease sensitivity at $(CT)_n$ sequences is that out-of-register DNA slippage occurs at these sequences with the production of single-stranded DNA loops in the slipped segment (Hentschel, 1982, Mace et al., 1983, Mckeon et al., 1984). It has not yet been tested whether the $(CT)_{26}$ sequence in the mouse myoglobin gene is sensitive to S1 nuclease.

A member of the B1 family of mouse dispersed repeated DNAs (Krayev et al., 1980) is found 925 bp before the CAP site in the 5'-flanking region. The two sequences are illustrated in Fig.3.2.

3.5 Sequence comparison between myoglobin genes

The purpose of comparing myoglobin gene sequences was to identify conserved sequence elements between these genes which might signify a functional role for these elements and to examine sequence divergence in the coding regions compared to that in non-coding DNA.

In the first instance, possible regions of homology between human and mouse non-coding DNA was investigated by dot-matrix comparisons of the two gene sequences. Computer generated dot matrices provide a good, unbiased means of detecting regions of homology shared between two genes

over their coding and non-coding sequnces (Konkel, Maizel and Leder, 1979). A matching criterion of 26 bp out of a large window size of 60 bp was used in the human x mouse comparison, as suggested by White et al. (1984), to reduce background noise but still retain homology between the two sequences. Dot matrix comparisons are shown separately over the three exons (Fig. 3.3).

Even when lower stringency match criteria were used (not shown) the only detectable homology between the two genes is over regions coding for myoglobin. The bulk of the non-coding DNA sequences, including introns (except at the intron junctions), the 3'-nontranslated and 3'-flanking sequences (except around the poly(A) addition site), and most of the 5'-flanking sequence show no clear alignment. This dot-matrix comparison is in agreement with similar comparisons of orthologous β-related globin genes of the human and mouse (see White et al., 1984) which revealed that the level of divergence between these species is so great that alignment of orthologous non-coding human and mouse sequences is extremely difficult. The only exception to this lack of similarity in non-coding human and mouse myoglobin sequences is found towards the beginning of the 5'-flanking region (Fig.3.3A). A conserved domain is obvious as a diagonal line of homology extending from the CAP site to about 400 bp upstream.

Since it is not possible to align human and mouse non-coding sequences the discrete nature of this conserved domain could not be determined since it is entirely possible that lack of homology before position -400 bp is due to a large deletion or insertion in either the human or mouse sequence. In order to determine the endpoint of this conserved domain and to analyse in more detail the distribution of substitutions across the length of the 5'-flanking region, the seal myoglobin 5'-flanking region was next sequenced as seal and human

Figure 3.3

Dot-matrix comparisons of the human myoglobin gene sequence with the mouse myoglobin gene sequence

The entire sequence available of the human myoglobin gene (6980 bp -Weller et al., 1984) was compared with the entire sequence of the mouse myoglobin gene shown in Fig. 3.2. The dot-matrix comparison was carried out separately over the three exons and associated flanking regions.

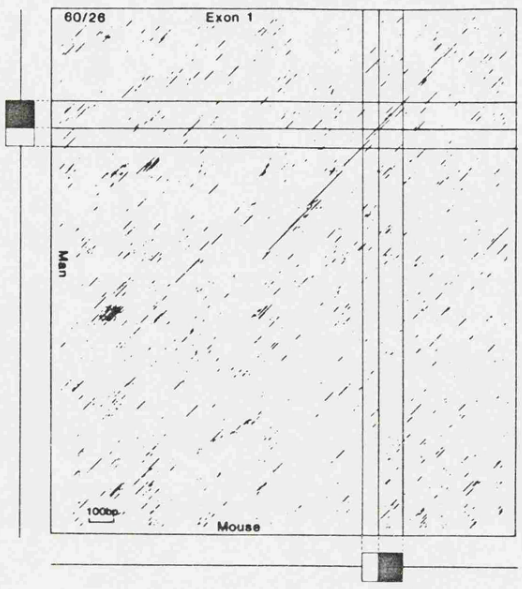
The window size in this comparison (the given number of consecutive nucleotides under comparison) was 60 bp and the minimum hit size (the number of correct matches required before a dot is plotted) was 26 bp (see White et al., 1984).

Homlogy between the two sequences appears as a diagonal line at 45° across the grid.

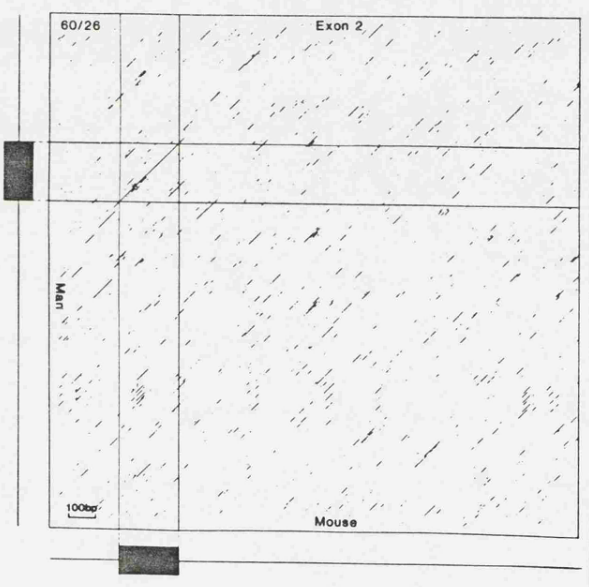
The positions of the coding sequences (filled boxes) and the 5'- and 3'-nontranslated mRNA sequences (open boxes) and intervening and flanking sequences (lines) are shown alongside the grids.

A significant feature to notice in this comparison is the homology detected in the 5'-flanking region, extending upstream (about 400 bp) of the CAP site of the genes. This contrasts with the absence of homology over the bulk of the remaining noncoding DNA sequences.

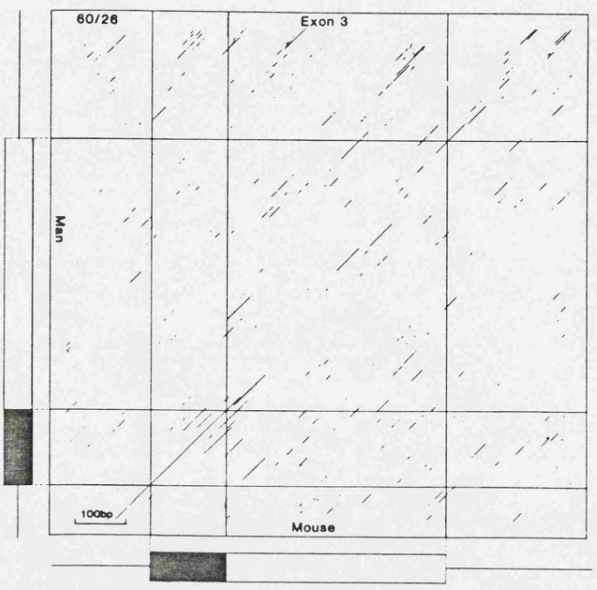
A


60/26
Exon 1
Man
100bp
Mouse

B


60/26
Exon 2
Man
100bp
Mouse

C


60/26
Exon 3
Man
100bp
Mouse

non-coding DNA may be aligned fairly easily (Weller et al., 1984) and only ~230 bp of seal 5'-flanking DNA had been previously determined preventing this comparison (Blanchetot et al., 1983).

The following sections of this Chapter present the sequence of 1.6 kb of seal 5'-flanking DNA and a sequence comparison to search for further constraint on myoglobin gene sequences in the more alignable human and seal myoglobin genes.

3.6 The seal myoglobin gene 5'-flanking region

A 1.6 kb HindIII fragment, from position -42 bp to -1620 bp before the seal myoglobin CAP site, was isolated from pSM19.5 (Blanchetot et al., 1983) by restriction with HindIII and purification of the appropriate fragment from an agarose gel using DE81 paper (Section 2.5). The fragment was self-ligated, sonicated, end-repaired and resulting blunt-ended DNA fragments around 400-900 bp in size were cloned into the SmaI site of M13mp19. The M13 clones which hybridized with the 1.6 kb HindIII fragment were sequenced using the dideoxy chain termination method of Sanger et al. (1980) and Biggin et al. (1983).

Fig.3.4 shows the complete sequence of the HindIII fragment of seal 5'-flanking DNA. This sequence was confirmed on both DNA strands.

3.7 A simple repeat region is found in the seal myoglobin gene 5'-flanking region

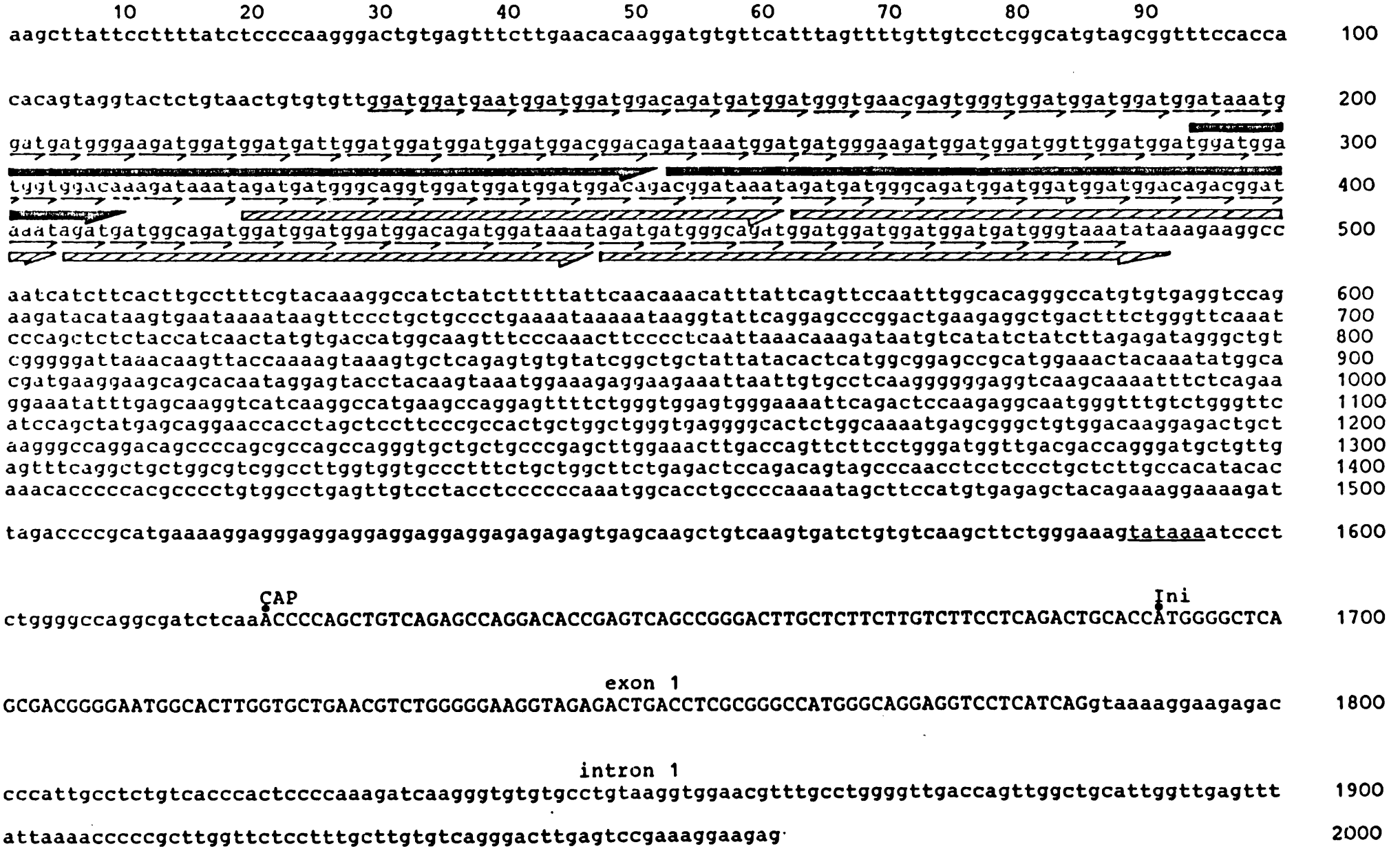
The seal myoglobin gene 5'-flanking region contains a 359 bp tandem repetitive sequence based on $(GGAT)_n$ located at 1135 bp to 1494 bp before the CAP site. This is illustrated as a series of arrows beneath the sequence in Fig. 3.4 and it can be seen that the number of small repeating units is 92. This simple repeat sequence may have arisen by chance duplications of an original dimeric GGAT sequence by an unequal

Figure 3.4

DNA sequence of 1.6 kb of the seal myoglobin 5'-flanking region

The sequenced HindIII DNA fragment of seal myoglobin gene 5'-flanking region lies from position 0 to 1573 bp in this Figure (lower case letters). For completeness, sequences up to and including 59 bp of myoglobin gene intron 1 have been added. Sequences present in the mature RNA are shown in uppercase letters. The CAP site and initiation codon (Ini) are marked. Tandem repetitive sequences found in the seal myoglobin 5'-flanking region are underlined by arrows.

seal Mb 5′ flanking sequence.



crossing-over mechanism as proposed by Smith (1976) for the evolution of DNA satellite sequences. The basic repeat unit of this sequence has undergone considerable alteration being substantially diverged from a hypothetical $(GGAT)_{92}$ structure (Fig. 3.5). The calculated divergence level is 20.3% (uncorrected for multiple hits).

Dot matrix analysis of this region (not shown) revealed that longer tandem repetitive homology blocks exist (Fig. 3.5). Two 59 bp blocks are found in tandem 67 bp from the 5'-end of the repeat unit. 11 bp downstream of this tandem repeat unit are another four tandemly repeated blocks, 43 bp in length. These higher order tandemly repeated blocks of sequence within the simple repeat unit have probably arisen by unequal exchange between different members of the GGAT array. Weller et al. (1984) found a similar simple sequence element based on $(GGAT)_{165}$ in the same location near the human myoglobin gene indicating that such relatively unstable elements can persist for long periods during evolution. Interestingly, the human repeat element also contains larger homology blocks but these are not repeated in tandem (Fig. 3.5) and have probably arisen by gene micro-conversion events.

A repeat region based on GGAT has been found upstream of a processed calmodulin pseudogene (Stein, 1986). The pseudogene is expressed in skeletal muscle although its promoter region has been lost and Stein proposes that the GGAT repeat region has been incorporated as an "enhancer' element by this pseudogene resulting in its expression.

If the mouse myoglobin gene contains a similar repeat, it does not lie within the region of the 5'-flanking region sequenced.

3.8 Comparison of human and seal 5'-flanking sequences

A dot matrix comparison between the seal myoglobin exon 1 sequence shown in Fig. 3.4 and the 3768 bp human myoglobin exon 1 sequence (Weller

Figure 3.5

A simple repeat region in the 5'-flanking of the seal myoglobin gene

Top The different constituents of the GGAT array, demonstrating the substantial divergence from a hypothetical $(GGAT)_{92}$ array. Arrows underneath the repeated sequences represent the larger tandemly repeated homology blocks within the sequence: two 59 bp blocks (single underlining) and four 43 bp blocks (double underlining) are found.

Bottom Diagrammatical representation of the larger blocks of homology found in both seal and human myoglobin gene simple repeat regions. Identical arrows indicate homologous blocks of sequence. The numbers beneath the arrows show the length in bp of the particular homologous sequence. The locations of the repeat regions before the myoglobin gene CAP sites are indicated.

$(GGAT)_2$ GAAT $(GGAT)_2$ GGAC AGAT GAT GGAT GGGT GAAC GAGT GGGT $(GGAT)_4$

AAAT GGAT GAT GGGA AGAT $(GGAT)_2$ GATT $(GGAT)_4$ $(GGAC)_2$ AGAT

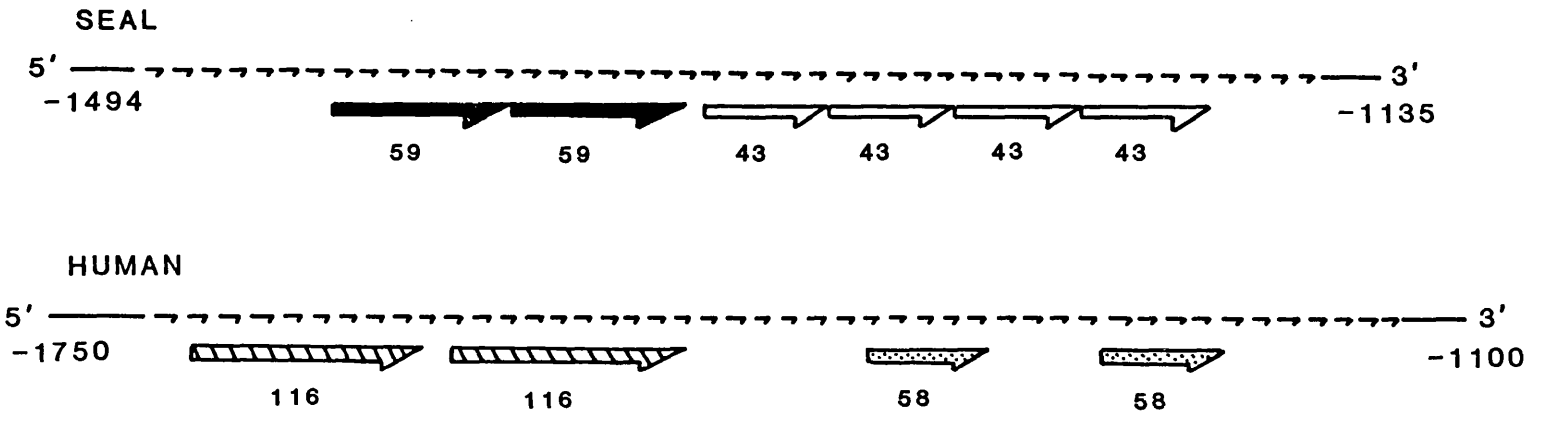
AAAT GGAT GATG GGAA GAT $(GGAT)_2$ GGTT $(GGAT)_4$ GGT GGAC AAAGAT AAAT

AGAT GAT GGGC AGGT $(GGAT)_3$ GGAC AGAC GGAT AAAT

AGAT GAT GGGC AGAT $(GGAT)_3$ GGAC AGAC GGAT AAAT

AGAT GAT GGC AGAT $(GGAT)_3$ GGAC AGAT GGAT AAAT

AGAT GAT GGGC AGAT GGAG $(GGAT)_3$ GAT GGGT AAAT



et al., 1984) is shown in Fig. 3.6. A diagonal line representing homology between the two sequences is quite clearly seen over the coding region of the genes and up to the TATA boxes. The homology breaks down in the GAGA box region but resumes again upstream over the length of the conserved region which is now embedded in more alignable human and seal non-coding DNA. In contrast to the human and mouse 5'-flanking region comparison, homology extends far upstream, beyond the conserved domain in the 5'-flanking region in human and seal comparisons, as demonstrated by the continuous presence of a diagonal line of homology in the dot-matrix grid.

The GGAT repeat region appears as a complicated series of repeated lines in this analysis.

Using the dot matrix analysis, the 5'-flanking sequences of the human and seal myoglobin genes were aligned base pair by base pair excluding the GGAT repeat region (see Fig.3.6 and legend).

The conserved region is apparent in this alignment at positions approximately 120 bp to approximately 330 bp before the CAP site where the number of identities between the two genes, illustrated in the Figure by gaps in one sequence relative to the other, appears to be increased over this region. To quantitate the apparent sequence conservation in this region, DNA sequence divergence (% mismatches) between human and seal sequences was calculated at 50 bp intervals over various regions of the aligned 5'-flanking regions and exon 1 (Table 3.2). Divergence levels between the two sequences in the conserved region drops to 11% (uncorrected for multiple hits) from 29% (uncorrected for multiple hits) in 5'-flanking DNA upstream of this region. The divergence between the two sequences is 7% (uncorrected for multiple hits) over the coding region and 15% (uncorrected for multiple hits) around the TATA box suggesting that these undiverged regions have been actively conserved by

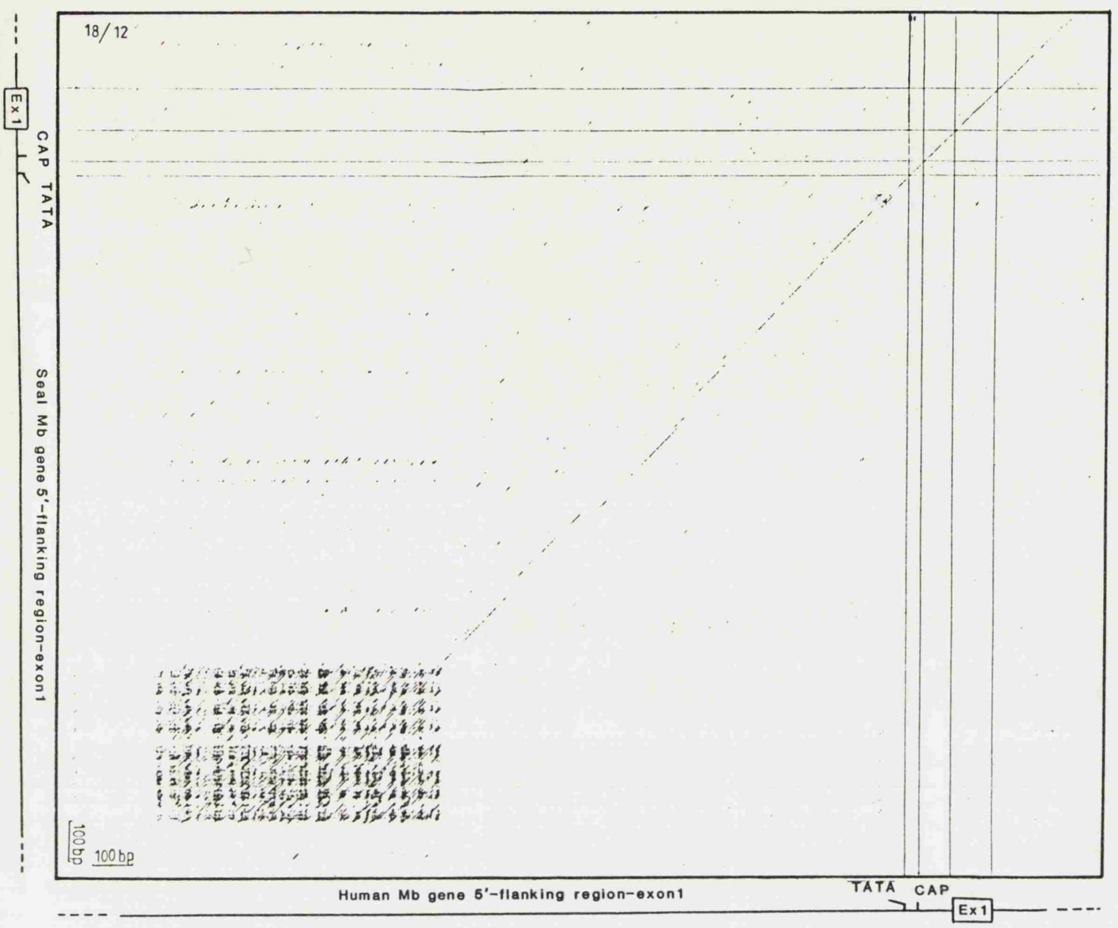
Figure 3.6

Dot-matrix comparison of human and seal myoglobin gene 5'-flanking region

The entire exon 1 and flanking sequences determined for the human myoglobin gene (3768 bp) was compared by dot-matrix analysis with the seal myoglobin gene 5'-flanking region and exon 1 sequences shown in Fig. 3.4. The window size was 18 bp and the minimum hit size 12 bp in this comparison.

DNA sequence comparison of the human myoglobin gene 5'-flanking region with the seal myoglobin gene 5'-flanking region

The human and seal exon 1 and 5'-flanking sequences used in the dot-matrix comparison but excluding the GGAT repeat sequence and sequences upstream of this were aligned using the dot-matrix analysis to position deletions and insertions between the two sequences. Gaps in the human sequence indicates an identical nucleotide to the seal sequence. A dash indicates the absence of a nucleotide in one sequence relative to the other and nucleotide differences are shown. Sequences present in the mature mRNA are shown in uppercase letters. The TATA box and purine rich region (GAGA box) are boxed. The CAP and initiation codon (ini) are indicated. The region where the identities (gaps) between the two sequences apperars to increase is double underlined with discontinuous end-points, indicating that the exact borders of the conserved region are not known.




```
       -1150                                        -1100                                          -1050
seal catcttcacttgcctttcgtacaaaggccatctatctttttattcaacaaacatttattcagttccaatttggcacagggccatgtgtgaggtccagaagatacataagtgaataaaataagttcc
man      c cac ct    t tgc aa    c c ca   c   t               a  c       aa acca      c  ag     g gg t      c gagct

                          -1000                                         -950
seal ctgctgccctgaaataaaaataaggtattcaggagcccggactgaaggc-tgactttctgggttcaaatcccagctctctaccatcaactatgtgaccat--ggcaagtttcccaaacttccc
man     cct       c-----------------tg aa g-a     g   g g   c  a           a   --g    g g    gc ta  ca tg t      t-    t t

      -900                                          -850                                              -800
seal ctcaattaaacaaagataatgtca-tatctatcttagagatagggctgtcgggggattaaacaagttaccaaaagtaaagtgct---------------------------cagagtgtgtatc
man           t gg          g  c    ac tt gg    t  tg -         tg         t c     t tgggacagggcctggcacccagcaaagtctctt  a  gct

                         -750                                         -700
seal ggctgctattatacactcatggcggagccgcatggaaacta-caaatatggcacgatgaaggaagcagcacaataggagtacctacaagtaaatggaaagaggaagaaattaattgtgcctcaagg
man            c taatgga aa  -tg      a   c gg     g atg cc ttg        atg    c   acttac   ag   g a  g       c    g g  t     a

     -650                                          -600                                              -550
seal ggggaggtcaagcaaaatttctcagaaggaaatatttgagcaaggtcatcaaggccatgaagccaggagttttctgggtggagtgggaaaattcagactccaagaggcaatgggtttgtctgggtt
man   a t t     a   c  t    g     cc       g c   g aa---------       c  -----------------------    attg   ac   c     ac

                        -500
seal catccagctatgagcaggaa-ccacctagctccttcccgccactgctggctgggtgaggggcactctggcaaaatgagcgggctgtggacaaggagactgctaagggccaggacagccccagcgcc
man    c tg     a   ca  a  t cg t    t t t    t     -        t-          g- g a aag  cgtg -  gt    g  t   -    t t tg--

                                                                               -320                                -280
seal agccagggtg-ctgctgcccgagcttggaaacttgaccagttcttcctgggatggttgacgaccagggatgctgttgagtttcaggctgctggcgtcggccttggtggtgccctttctgctggctt
man     g  cac   g g  a    a    a    - g  c t     g c    cg -     gc--  g                    t ct               tc    a

                                    -240                                 -200                                       -160
seal ctgagactccagacagtagcccaacctcctccctgctcttgccacatacacaaacacccccacgcccctgtggcctgagttgtcctacctcccccaaatggcacctgccccaaaatagcttcc-a
man  -     g        -             c-t    t     cg    c     - c   -         c     g    g a -            t           c

                                  -120                                                        -60      -50      -40
seal tgtgagagctacagaaaggaaaagattagaccccgcatgaaaaggagggaggaggaggaggagg-----agagagagtgagcaagctgtcaagtgatctgtgtcaagcttctgggaaagtataaaa
man        g                      tc c g tga     a   a a t a     gcaggg  g g ac   g  ca tg   c    t       a  cca    g

       -20     -10       CAP                                                                Ini        Exon 1
seal tc-cctctggggccaggcgatctcaaACCCCAGCTGTCAGAGCCAGGACACCGAGTCAGCCGGGACTTGCTCTTCTTGTCTTCCTCAGACTGCACCATGGGGCTCAGCGACGGGAATGGCACTTG
man  a g  ct    a     agc              TG G          C  G   CAT         T       T        G                        G

                                                                         Intron 1
seal GTGCTGAACGTCTGGGGGAAGGTAGAGACTGACCTCGCGGGCCATGGGCAGGAGGTCCTCATCAGgtaaaaggaagagaccccattgcctctgtcacccactccccaaagatcaagggtgtgcctg
man                      G   G    A  CA            A                      tt       c   c     a   t  gatca g  tgt*ca**t
```

Table 3.2

Sequence comparison of the seal and human myoglobin genes

Comparisons and sequence coordinates are based on the alignment of the seal and human myoglobin gene exon 1 and 5'-flanking region shown in Fig. 3.6. The number of gaps introduced into non-coding DNA regions in order to maximise sequence homology are shown in parentheses and these regions were not scored in divergence calculations. Corrected sequence divergence over the 5'-flanking region was calculated as described by Kimura (1981).

Regional divergence between seal and human myoglobin gene exon 1 and 5'-flanking sequence

Region of gene	co-ordinates in sequence	number of bases compared	% sequence divergence (uncorrected)
5'-flanking	-393 - -1160	668 (21)	29.35% ± 1.7
conserved region	-119 - -331	212 (7)	10.80% ± 4.5
GAGA box	-67 - -109	42 (1)	35.70% ± 7.4
TATA-- CAP	0 - -33	33 (1)	21.20% ± 7.1
CAP-- Ini	0 - 70	70	15.70% ± 4.3
Ini-- Intron 1	70 - 165	95	7.30% ± 2.6

Overall divergence in the 5'-flanking region (uncorrected) 29.35% ±1.7 Transition frequency 60.40% ± 4.8

(corrected) 36.20% 62.30%

purifying selection. To define the end-points of the sequence conservation and to assess the significance of this sequence conservation, the distribution of substitutions across human and seal aligned exons 1, 2 and 3 and associated non-coding DNA sequences was subjected to a statistical analysis using a computer program designed to give a measure of the significance of sequence conservation or divergence in a defined window size. The details of the program (written by Dr.A.J.Jeffreys) are given in the legend to Fig. 3.7. Briefly, the probability of obtaining the particular sequence divergence in a defined window size is calculated using the Fisher analysis when the computer is given an expected level of divergence between the two sequences. The log of the calculated probability is plotted as a point on a curve and a typical example of the shape of the curve over the aligned sequences is shown in Fig. 3.7. All log probability plots obtained demonstrate that the 5'-distal sequences, intervening sequences and 3'-nontranslated and flanking sequences all rise to a plateau level on the curve representing neither significantly diverged or conserved sequence. This suggests that these sequences are evolving at a constant rate, presumably at or close to the neutral rate. The uniform level of divergence calculated for the bulk of the 5'-flanking region (Table 3.2), of 29% (uncorrected) and 36% (corrected for multiple hits -see Table 3.2) suggests that this non-coding DNA is diverging rapidly by neutral drift without any localised selective constraints. This level of divergence shows no significant difference from the value of 35% for the mean level of silent site divergence in human and seal coding sequences (Weller et al., 1984) which is expected if silent substitutions accumulate freely through neutral drift.

The uniform level of divergence in the non-coding regions may be used to predict significant departures from this overall level of

Figure 3.7

How significant is the sequence conservation found in the human and seal myoglobin gene 5'-flanking conserved domain?

Aligned seal and human myoglobin gene sequences were subjected to a statistical analysis using a computer program designed to run on a BBC/ACORN micro computer. In this analysis the window size (sequence search string) was 60 bp and gaps inserted in the sequences were scored as regions of 0% homology. The expected divergence level between the two sequences was given as 29% (the average divergence across the bulk of the 5'-flanking region). The computer totals the number of identities in 60 bp strings and calculates the probability of obtaining this number of identities based on the expected number (71% of 60 bp) using the Fisher analysis:

$$P_i = (0.29)^{60-N}(0.71)^N \times \text{number of positions the identity can occur}$$

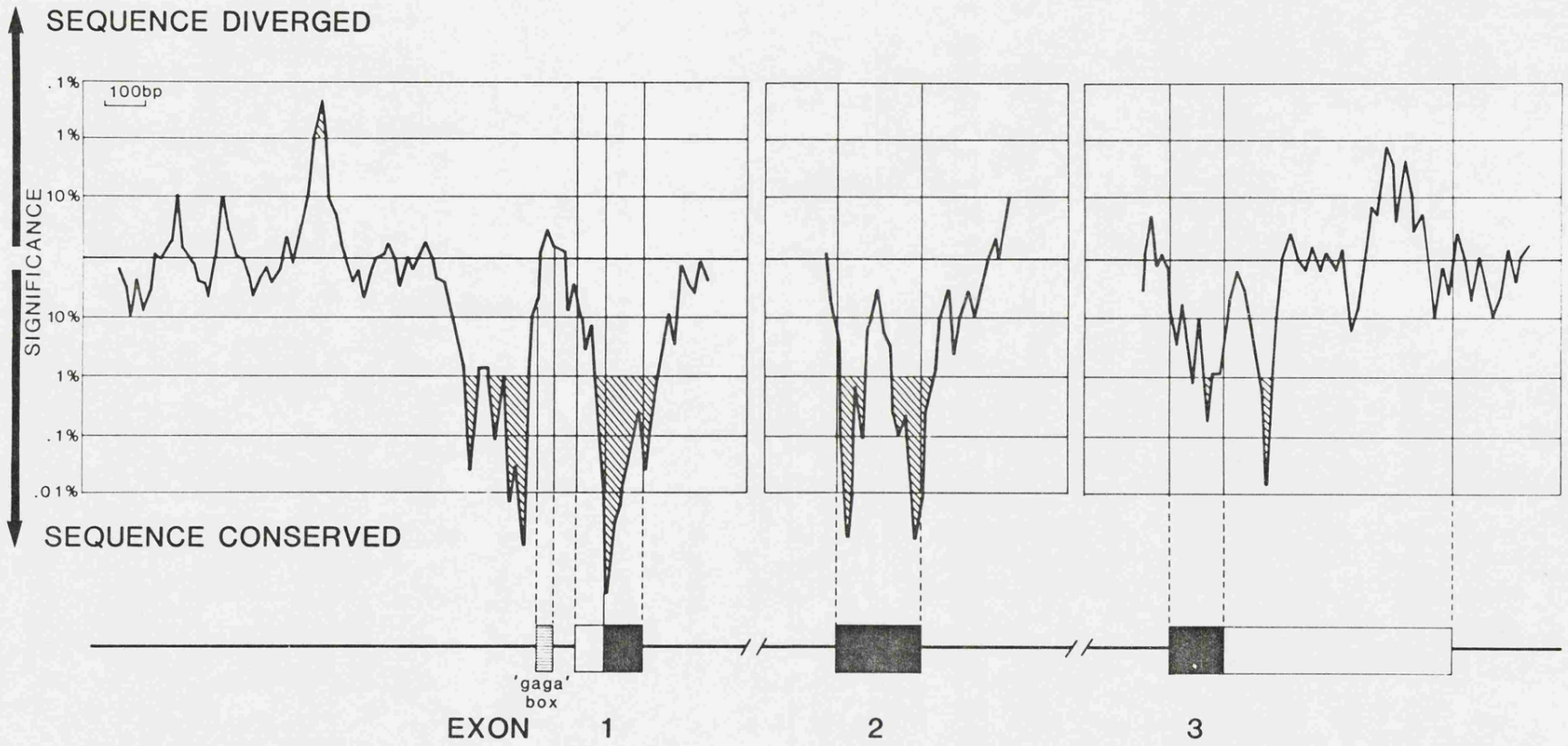
where N= number of identies

If the number of identities is 20 for example, the probabilities of obtaining 20 identities or less are totalled and plotted as a point above the zero significance (i.e. 71% of 60) line. If the number of identities is 55 for example, the probabilities of obtaining 55 or more are totalled and plotted as a point below the zero significance line. Sequences which are more diverged appear as peaks and those which are more conserved than expected appear as troughs on the graph. The significance of deviation from an expected level of divergence is shown as horizontal lines across the graph and shaded areas represent significant conservation or divergence.

The coding sequences of exons 1, 2 and 3 (filled boxes), non-translated mRNA sequences (open boxes) and flanking DNA (lines) of the myoglobin gene are shown alongside the graph to scale. The GAGA box is represented as a shaded box.

The bulk of the myoglobin gene flanking DNA is evolving at a constant rate rising to a plateau at the zero significance level. In contrast, a localised domain (125-325 bp before the CAP site) of conserved sequence is observed in the 5'-flanking region, significant at the 99.99% level.

The GAGA box is neither significantly diverged or conserved in this analysis. A diverged area of sequence (significant at the 99%) level is found in the 5'-flanking region.


SEQUENCE DIVERGED
100bp
.1%
1%
10%
SIGNIFICANCE
10%
1%
.1%
.01%
SEQUENCE CONSERVED
'gaga' box
EXON 1
2
3

divergence. Given the expected level of divergence of 29% over the aligned human and seal myoglobin gene sequences and a window size of 60 bp, the computer generated the curve depicted in Fig. 3.7. In contrast to the overall level of divergence, the conserved region represents a localised domain of sequence conservation in the 5'-flanking region. Furthermore, the conservation in this region is highly significant: the deviation from the zero probability plateau over this 200 bp window is significant at the 99.99% level. This level of significance is similar to that observed for the coding regions. Using this program, the limits of the conserved domain may be more accurately determined and the significance of the localised departure from neutral divergence established.

The corresponding non-coding DNA divergence calculations between human and mouse sequences cannot be so reliably estimated in view of the inability to align the sequences, demonstrated in the dot-matrix comparisons (Fig. 3.3). However, we may calculate the silent site substitutions between human and mouse coding sequences to give an estimate of the non-coding DNA divergence. Silent site substitution levels over aligned myoglobin exons were calculated according to Perler et al. (1980). The mean level of silent site divergence between human and mouse coding sequences suggest a neutral divergence level of 46.2% (uncorrected for multiple hits) and 63.5% (corrected for multiple hits; Perler et al., 1980). In contrast to this high level of neutral divergence, the level of substitutions in the 5'-flanking conserved domain is very low (16% uncorrected for multiple hits) again suggesting that most substitutions are being eliminated in this region by purifying selection.

The sequence comparison between the three mammalian myoglobin genes also provides strong evidence that the rate of neutral drift is not clockwise. Humans, seals and mice are presumed to have last shared a

common ancestor about 80 MY ago at the time of the mammalian radiation (Romero-Herrera et al., 1973) and yet there is a large difference in their level of neutral divergence: human/seal comparisons give a divergence level of 36% (corrected for multiple hits) and human/mouse comparisons 63.5% (corrected for multiple hits). This discrepancy in divergence rates, also documented by others (see Weller et al., 1984), suggests that the rate of neutral substitution is 2.6 fold higher in rodents than in primates and pinnipeds and is possibly due to the rate of neutral drift being geared to generation time of the species rather than absolute time (Weller et al., 1984). Another possibility is that rodents diverged earlier, approximately 160 MY ago to maintain a uniform rate of neutral substitution.

The alignment of the 5'-flanking region of human, seal and mouse myoglobin genes is shown in Fig. 3.8. The mouse conserved sequence could easily be aligned with the human and seal conserved sequences shown in Fig. 3.6. Outside of the conserved region (upstream of -390 bp in Fig. 3.3) it very difficult and eventually becomes impossible to align the mouse sequence to the human and seal sequences. From the detailed 3-way alignment, the conserved domain may be positioned at -110 bp before the CAP site (including the gaps in the sequence) extending to position --390 bp in the mouse myoglobin gene. For the human and seal genes sequence conservation is found over two regions from -120 bp to 220 bp and 245 bp to 390 bp before the CAP site, the presence of a 21 bp deletion in the mouse sequence between -223 bp to -244 bp dividing the conserved region in human and seal sequences (see Fig. 3.9). The three-way alignment reveals the prominant patterns of homology which is not so evident in the human and seal comparison.

It is unlikely that the conserved region codes for a polypeptide since there are a number of deletions and insertions between the three

Figure 3.8

DNA sequence comparison of human, seal and mouse myoglobin gene 5'-flanking regions

495 nt of 5'-flanking region, exon 1 and 40 nt of intron 1 of seal, human and mouse myoglobin genes were aligned. The mouse myoglobin sequence was fitted into the alignment of the human and seal sequences shown in Fig. 3.6. using the dot matrix analysis (Fig. 3.3) to aid the positioning of deletions and insertions in the mouse sequence relative to the human sequence. 5'-flanking region sequences could be aligned easily up to 390 nt before the myoglobin gene CAP site, upstream of this (-390 to -490), alignment was difficult and eventually becomes impossible. A dash in any of the sequences represents the absence of a nucleotide in one sequence relative to the other two. An identical nucleotide in all three sequences is represented by an asterisk beneath the approriate nucleotide in the human sequence. Nucleotides which differ in seal and mouse sequences from the human sequence are shown. The TATA box and initiation codon (ATG) sequences are underlined and the CAP site is shown. Sequences in the mature RNA are in uppercase letters and numbers above the sequence represent nucleotide position before the CAP site.

A feature to notice in this alignment is the significant increase in the long stretches of uninterrupted asterisks from position -390 to -120 nt before the CAP site compared to the patch homology observed 5' of -390 nt and 3' of -120 nt up to the TATA box.

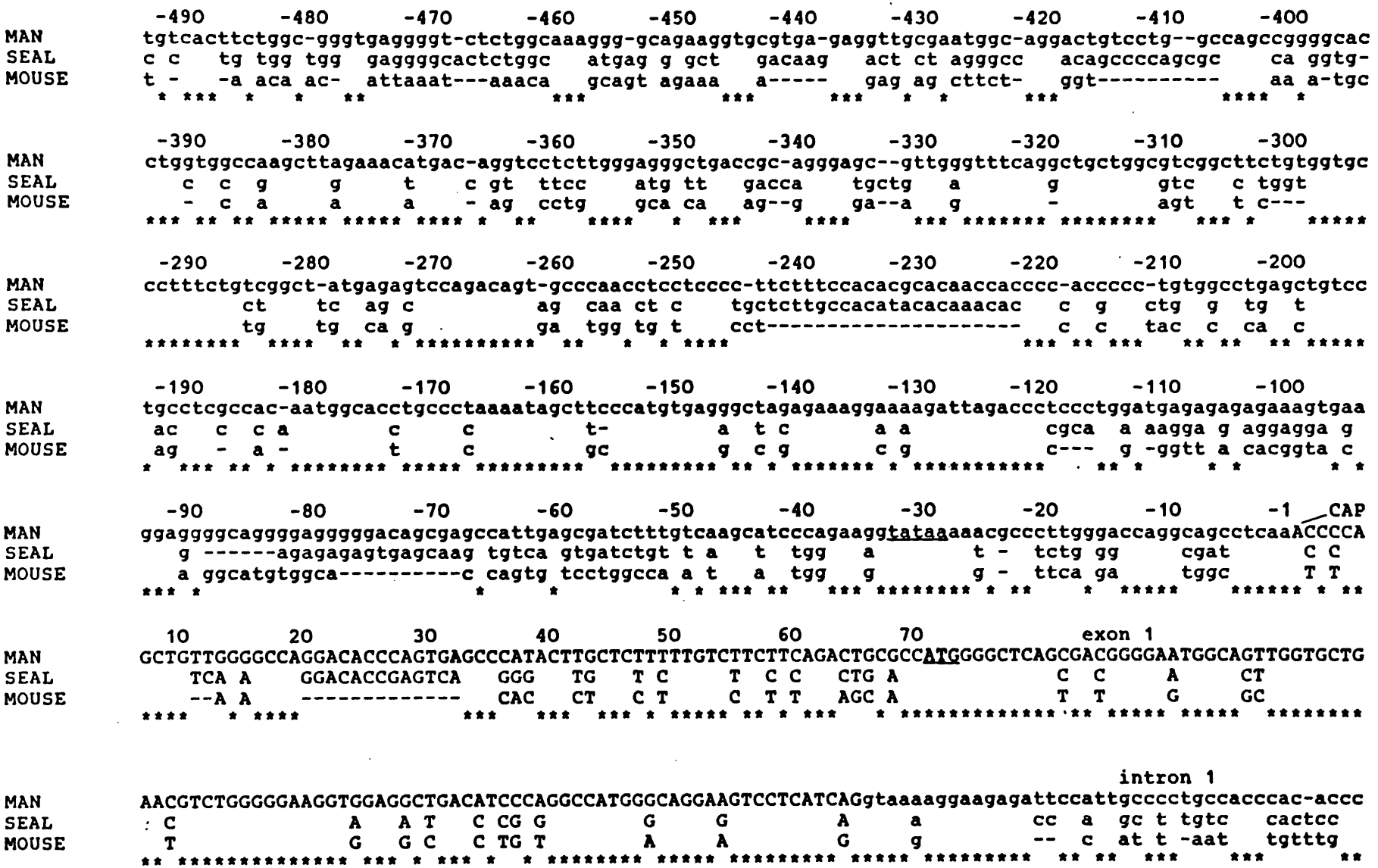

-490 -480 -470 -460 -450 -440 -430 -420 -410 -400
MAN tgtcacttctggc-gggtgaggggt-ctctggcaaaggg-gcagaaggtgcgtga-gaggttgcgaatggc-aggactgtcctg--gccagccggggcac
SEAL c c tg tgg tgg gaggggcactctggc atgag g gct gacaag act ct agggcc acagccccagcgc ca ggtg-
MOUSE t - -a aca ac- attaaat---aaaca gcagt agaaa a----- gag ag cttct- ggt---------- aa a-tgc
-390 -380 -370 -360 -350 -340 -330 -320 -310 -300
MAN ctggtggccaagcttagaaacatgac-aggtcctcttgggagggctgaccgc-agggagc--gttgggtttcaggctgctggcgtcggcttctgtggtgc
SEAL c c g g t c gt ttcc atg tt gacca tgctg a g gtc c tggt
MOUSE - c a a a - ag cctg gca ca ag--g ga--a g - agt t c---
-290 -280 -270 -260 -250 -240 -230 -220 -210 -200
MAN cctttctgtcggct-atgagagtccagacagt-gcccaacctcctcccc-ttctttccacacgcacaaccacccc-accccc-tgtggcctgagctgtcc
SEAL ct tc ag c ag caa ct c tgctcttgccacatacacaaacac c g ctg g tg t
MOUSE tg tg ca g ga tgg tg t cct---------- c c tac c ca c
-190 -180 -170 -160 -150 -140 -130 -120 -110 -100
MAN tgcctcgccac-aatggcacctgccctaaaatagcttcccatgtgagggctagagaaaggaaaagattagaccctccctggatgagagagagaaagtgaa
SEAL ac c c a c c t- a t c a a cgca a aagga g aggagga g
MOUSE ag - a - t c gc g c g c g c--- g -ggtt a cacggta c
-90 -80 -70 -60 -50 -40 -30 -20 -10 -1 CAP
MAN ggaggggcaggggaggggacagcgagccattgagcgatctttgtcaagcatcccagaaggtataaaaacgcccttgggaccaggcagcctcaaACCCCA
SEAL g ------agagagagtgagcaag tgtca gtgatctgt t a t tgg a t - tctg gg cgat C C
MOUSE a ggcatgtggca----------c cagtg tcctggcca a t a tgg g g - ttca ga tggc T T
10 20 30 40 50 60 70 exon 1
MAN GCTGTTGGGGCCAGGACACCCAGTGAGCCCATACTTGCTCTTTTTGTCTTCTTCAGACTGCGCCATGGGGCTCAGCGACGGGGAATGGCAGTTGGTGCTG
SEAL TCA A GGACACCGAGTCA GGG TG T C T C C CTG A C C A CT
MOUSE --A A ---------- CAC CT C T C T T AGC A T T G GC
intron 1
MAN AACGTCTGGGGGAAGGTGGAGGCTGACATCCCAGGCCATGGGCAGGAAGTCCTCATCAGgtaaaaggaagagattccattgcccctgccacccac-accc
SEAL C A A T C CG G G G A a cc a gc t tgtc cactcc
MOUSE T G G C C TG T A A G g -- c at t -aat tgtttg

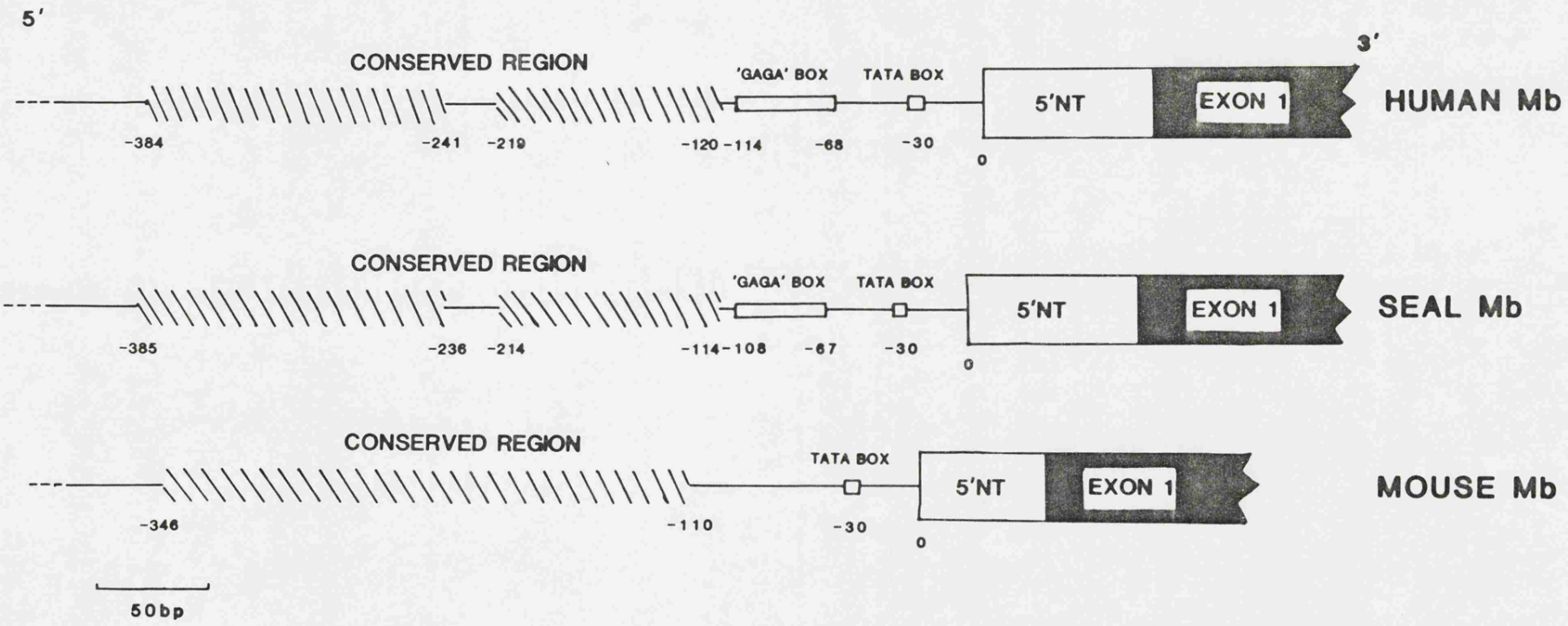
Figure 3.9

The location of the conserved domain in the 5'-flanking region of human, seal and mouse myoglobin genes

The 5' conserved domain (shaded areas) is depicted diagramatically in the human, seal and mouse myoglobin (Mb) gene 5'-flanking region. Numbers underneath each diagram represent the position in bp before the CAP (position 0 bp) and are taken from the three-way alignment shown in Fig. 3.8, excluding the gaps in the sequences. The location represents the points at which the human and seal myoglobin gene sequences can be aligned easily with the mouse sequence. The log probability plot shown in Fig. 3.7, however, suggests that sequence conservation is significant between the human and seal genes only to 325 bp before the CAP site.

The poly-purine region (GAGA box) found in the human and seal genes and the TATA boxes are shown. In the case of the TATA boxes, position -30 bp represents the approximate location of this sequence. The 5'-nontranslated mRNA sequence (5'-NT) and exon 1 (filled box) are shown. The Figure is to scale.

Note that the conserved region is in two sections for the human and seal myoglobin genes due to a small (~20 bp) deletion in the mouse sequence or an insertion in the human/seal sequences.


5′
3′
CONSERVED REGION
'GAGA' BOX
TATA BOX
5′NT
EXON 1
HUMAN Mb
-384
-241
-219
-120
-114
-68
-30
0
CONSERVED REGION
'GAGA' BOX
TATA BOX
5′NT
EXON 1
SEAL Mb
-385
-236
-214
-114
-108
-67
-30
0
CONSERVED REGION
TATA BOX
5′NT
EXON 1
MOUSE Mb
-346
-110
-30
0
50bp

sequences in this region which would introduce frame-shifts into any possible sequences encoding mRNAs.

Dot-matrix analysis of the conserved region (not shown) failed to detect any repeat sequences within this region and no inverted repeat sequences could be found which might form secondary structure.

3.9 Search for homology in the myoglobin gene conserved region with conserved regions of other muscle-specific genes

A comparison of the myoglobin gene conserved region with evolutionary conserved sequences or sequences shown to be essential for expression *in vitro* of other genes expressed specifically in muscle might identify common sequences which could signify common regulatory motifs for muscle genes.

The myoglobin gene conserved sequence was compared to such sequences using dot-matrix analysis and by eye. No significant sequence homology could be detected in any of these comparisons. A full description of sequences compared and potential homologies between other muscle-specific genes is given in Chapter 8 (Discussion).

3.10 Summary

The isolation and characterisation of the single, functional mouse myoglobin gene has established that, as with its human and seal counterparts, the gene has the same three-exon/two-intron organisation found in all characterised vertebrate α- and β-globin genes.

Sequence comparison between the three mammalian myoglobin genes led to the identification of a highly conserved upstream domain extending approximately 120 to 320 bp before the CAP site. The sequence conservation in this region is highly significant statistically and is similar to that seen in the myoglobin exons themselves. There is no

detectable homology of the myoglobin gene conserved region to conserved 5'-flanking sequence described for other muscle-specific genes.

The length of the conserved domain is longer than most evolutionary conserved sequences found in the 5'-flanking region of eukaryotic genes which are typically 15-50 bp in length, although enhancer regions can be larger (Rogers and Saunders, 1985) and Hu, Sharp and Davidson (1986) and Daubas et al. (1985) have described long highly conserved regions in the 5'-flanking region of chicken and rat α-skeletal actin genes and a mouse myosin light chain (MLC 1) gene respectively. It might be the case with the myoglobin conserved region that shorter functional sequences are embedded within a region which has been maintained as a generally homologous area. An indication that this might be correct is that all of the sequences in the conserved myoglobin 5'-flanking region are not essential as illustrated by a deletion in this region in the mouse sequence, or alternatively, the exact position of the sequences are not essential if an insertion has occurred in the human/seal sequences.

The position of the conserved domain before the transcriptional start site of the myoglobin gene and the significant sequence conservation suggest that it may play a regulatory role in myoglobin gene expression.

An analysis of myoglobin genes has provided an essential structural background from which functional studies could be undertaken.

CHAPTER 4

INDUCTION OF MYOGLOBIN GENE EXPRESSION DURING DIFFERENTIATION OF EMBRYONIC MYOBLASTS IN VITRO

4.1 Introduction

Myoglobin mRNA is detectable very early during embryogenesis in the mouse at around 14 days of gestation (see Section 1.6). This represents the time at which primary myotubes are formed, just before the extensive accumulation of secondary muscle fibres with accompanying muscle contractility (Rugh, 1968, Ontell, 1982). The early expression of myoglobin during embryogenesis suggested that the myoglobin gene might be expressed in embryonic myoblast cell lines capable of modelling in tissue culture the fusion and differentiation of skeletal muscle myoblasts to form multinucleated myotubes observed during myogenesis *in vivo* (see Section 1.4.). This model system has provided a great deal of information concerning the molecular basis for the activation of specific genes required for the establishment of a muscle phenotype and it was therefore of interest to determine whether this system could be used to study the expression of the myoglobin gene.

In this Chapter the presence of myoglobin transcripts in RNA from myoblast cell lines is investigated using the Northern blot technique and cloned myoglobin gene exons as hybridisation probes.

The particular myoblast cell line used in these experiments was a mouse line G8, originally isolated by Christian *et al* (1977). A derivative of this cell line, G8-1, was obtained from Dr. M. Webb, (Institute of Neurology, University of London) and set up in culture.

4.2 Growth characteristics of the mouse myoblast cell line G8-1

G8-1 mouse myoblasts grow continuously as monolayers of mononucleated cells with a doubling time of 24-28 hours when kept at a low cell density and grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS). When the cells are confluent, changing the medium to low serum medium (DMEM supplemented with 0.5% FCS and 5% horse serum (HS)) results in a relatively rapid (60-72hr) and synchronous fusion of the mononucleated myoblasts to form multinucleated myotubes (see Fig. 4.1), and this transition represents one of the most striking examples of terminal differentiation that can take place *in vitro*. Approximately 7-8 days after the shift to low serum medium, myotubes begin to contract spontaneously in culture. The extent to which the myotubes will contract depends upon the quality of the myotubes after fusion. Although this myoblast cell line consistently undergoes differentiation when confluent myoblasts are shifted to low serum medium or when myoblasts are left to grow to confluency and remain in their conditioned medium, the time of first appearance of visible myotubes and the level of spontaneous contraction is variable. Possible reasons for this are the growth characteristics of the cells before fusion and different batches of serum affecting the concentration of fusion-promoting factors in the cell culture medium (Buckingham 1977, Walsh and Phillips 1981). G8-1 myotubes can be maintained in culture for up to ~16 days during which time they mature from fused myoblasts into the striated contractile fibres. After this time they show signs of deterioration although it is not known why this occurs (Pearson, 1980).

In Fig. 4.1 and 4.2, two sets of myoblasts are shown undergoing fusion to form multinucleated myotubes, the original G8-1 myoblasts (Fig. 4.1) and clone 1 (Fig. 4.2), a clone derived from the original G8-1 myoblasts. This clone was isolated during routine cell cloning of G8-1

Figure 4.1

G8-1 mouse myoblasts and myotubes

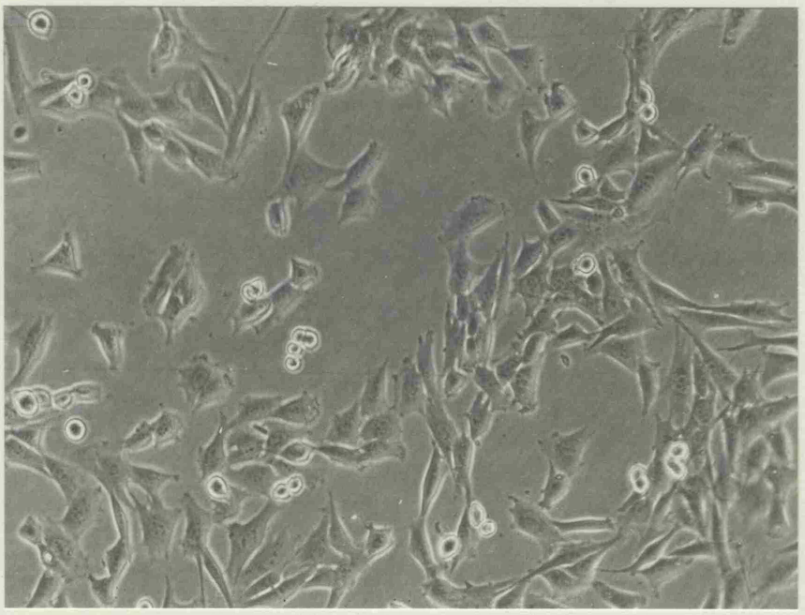
Cells were photographed under a phase contrast microscope using an Olympus camera.

A G8-1 myoblasts, 100x magnification

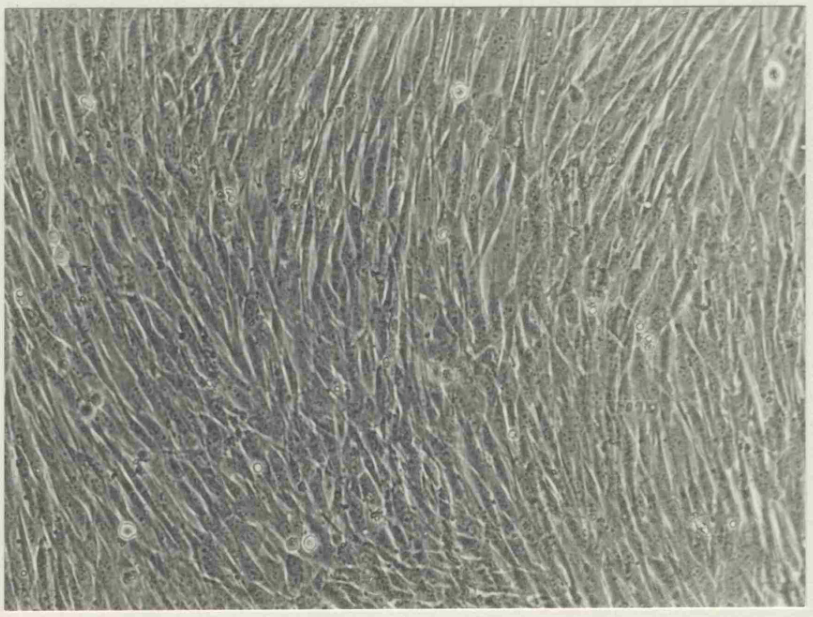
B Confluent G8-1 myoblasts, 40x magnification

C G8-1 myotubes after 5 days in low serum medium, 100x magnification.

A



B



C

Figure 4.2

Clone 1 myoblasts and myotubes

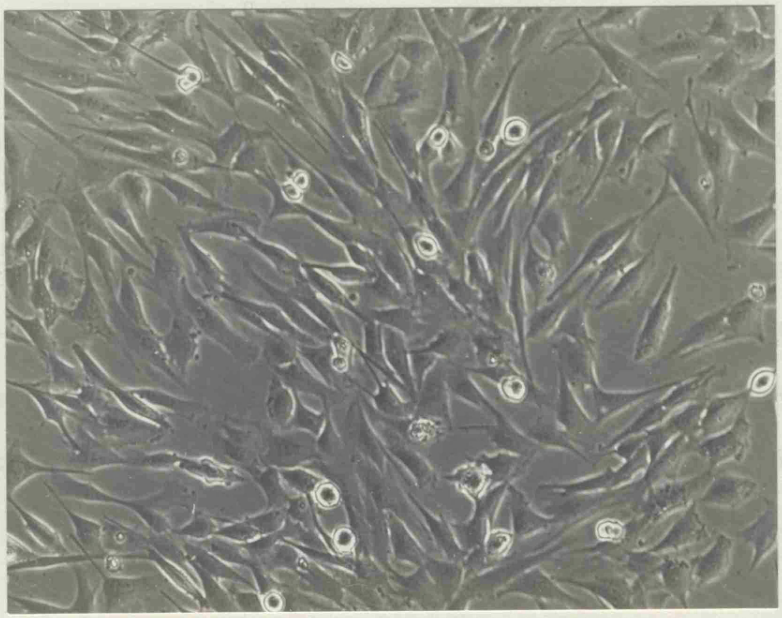
Cells were photographed as in Fig. 4.1.

A Clone 1 myoblasts, 100x magnification

B Clone 1 myotubes after 2 days in low serum medium, 200x magnification.

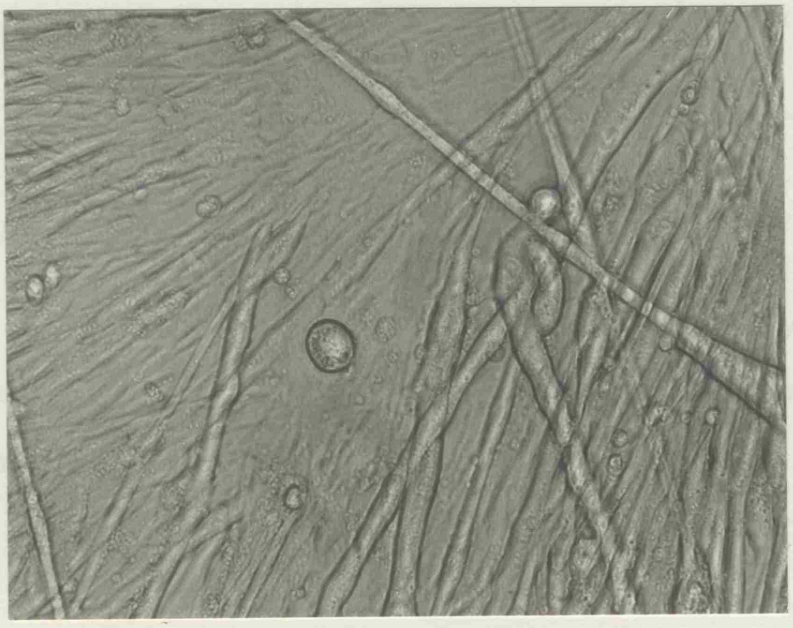
C Clone 1 myotubes after 5 days in low serum medium, 200x magnification

A



B

C



myoblasts (Section 2.11) to derive a batch of cells originating from a single cell and which should therefore have identical characteristics. Clone 1 myoblasts appeared to grow and differentiate more rapidly than the parental G8-1 myoblasts. They have a doubling time of 14-16 hours and fusion of myoblasts to form visible myotubes occurs ~24-40 hours after the addition of low serum medium to confluent myoblasts.

A typical growth time course for the myoblasts is outlined below:

G8-1 Myoblasts

Day 0	Myoblasts unfrozen into DMEM + 10% FCS. Media changed 5 hours later
Day 1	Myoblasts diluted 1:10
Day 2	Harvest sub-confluent myoblasts
Day 3	Media changed on confluent myoblasts to DMEM + 0.5% FCS/5% HS
Day 5	Myotubes beginning to form
Day 6	Low serum media refreshed, small myotubes present
Day 8	Myotubes harvested
Day 11-12	Spontaneous contraction begins to occur
Day 16-18	Degeneration of the fibres

Clone 1 Myoblasts

Day 0	Myoblasts unfrozen into DMEM + 10% FCS. Media changed 5 hours later
Day 1	Myoblasts diluted 1:20
Day 2	Harvest sub-confluent myoblasts
Day 3	Media changed on confluent myoblasts to DMEM + 0.5% FCS/5% HS
Day 4	Myotubes beginning to form
Day 5	Low serum media refreshed, mostly large myotubes present
Day 6	Myotubes harvested
Day 11	Myotubes begin to degenerate

Clone 1 myoblasts grow and differentiate more rapidly. Clone 1 myotubes rarely contract spontaneously and they degenerate earlier than G8-1 myotubes

4.3 Induction of myoglobin gene expression during fusion and differentiation of embryonic myoblasts in vitro

In order to determine whether the myoglobin gene is expressed at any point during the differentiation of original G8-1 myoblasts, RNA was prepared from pre-fusion myoblast and post-fusion myotube cultures and analysed on Northern blots using myoglobin gene exons as hybridisation probes.

Total RNA was isolated from ~2-5 x 10^7 cells of proliferating myoblast cultures which were ~75% confluent and from myotube cultures starting with the same number of original myoblasts which had been fusing in low serum medium for 5 days (Day 8 above) as described Section 2.12. Total RNA was then subjected to one cycle over an oligo dT column (Section 2.12) to enrich for poly(A)$^+$ RNA. Fig.4.3 shows an example of the quality of the RNA prepared by these methods. Typical yields from ~2 x 10^7 myoblasts was 0.4-0.5 mg total RNA and ~10µg of poly(A)$^+$ RNA.

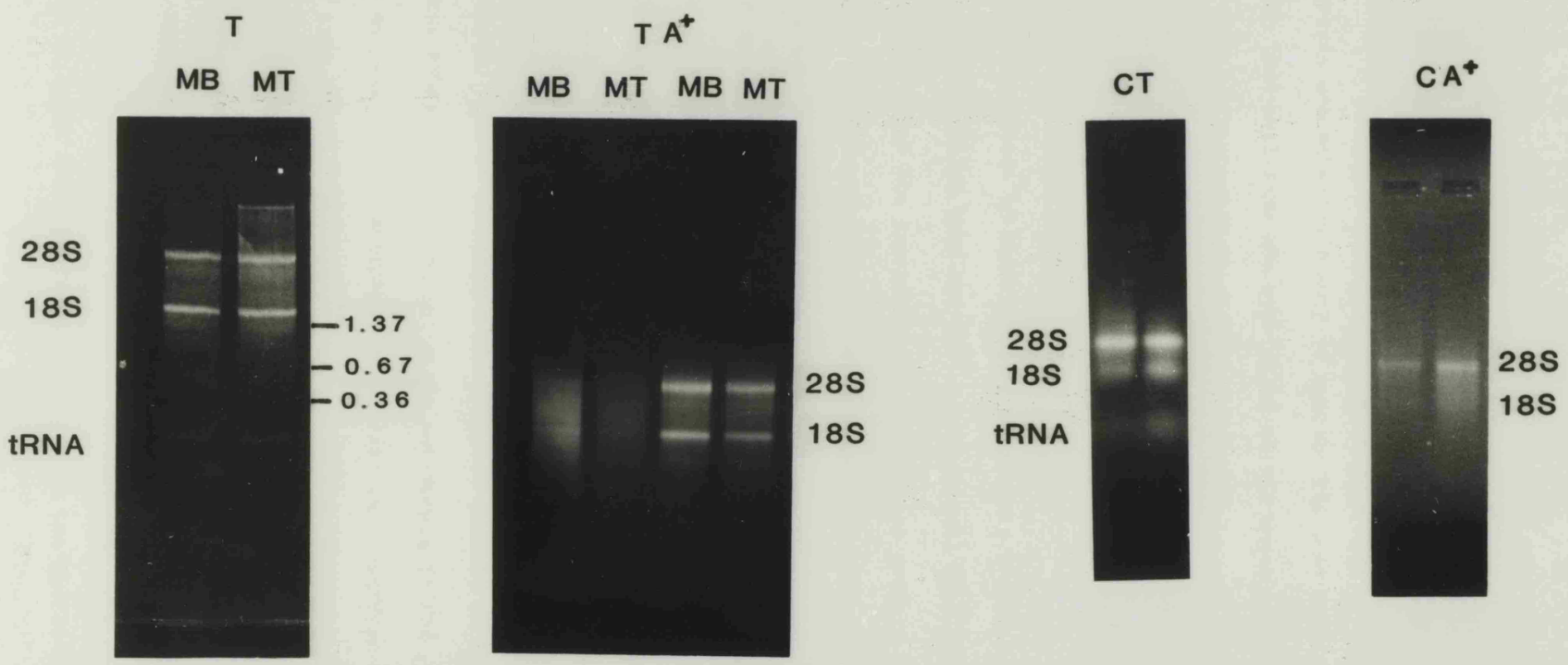
1µg of myoblast and myotube poly(A)$^+$ RNA was analysed on Northern blots using a human myoglobin gene exon 2 probe (details of the probe is given in the Figure legend of Fig.3.1). The probe detected a hybridising component ~1200 nt in size in myotube poly(A)$^+$ RNA only (Fig. 4.4, and also in total myotube RNA -data not shown). To demonstrate that this hybridising component is likely to be myoglobin gene transcripts, a filter containing the same RNA was hybridised with a mouse myoglobin gene exon 1 probe (probe details in Fig. 4.4 legend). This probe also detected the 1200 nt component in myotube poly(A)$^+$ RNA (Fig. 4.4), indicating that this is myoglobin mRNA. Longer exposures of both autoradiographs failed to reveal any signals corresponding to myoglobin mRNA in myoblast poly(A)$^+$ RNA.

Figure 4.4 also illustrates that the mouse myoglobin transcript in G8 myotubes is shorter than the seal myoglobin transcript in adult seal

Figure 4.3

Total and poly(A)$^+$ RNA extracted from G8-1 myoblast and myotube cell cultures

Total RNA (T) and cytoplasmic RNA (~0.5µg per lane) made from myoblast (MB) and myotube (MT) cultures is shown after electrophoresis in 1.5% agarose gels. Poly(A)$^+$ RNA was prepared from either total RNA (TA$^+$) or cytoplasmic RNA (CA$^+$) by passage once over an oligo d(T) cellulose column. 28S and 18S refer to the migration positions of the ribosomal RNA and tRNA to the transfer RNA. The migration positions of pBR322 x Sau3A markers (single-stranded) are shown next to the sample of total RNA. After oligo d(T) chromatography the relative intensity of the ribosomal RNA bands is reduced and no tRNA is present.


T
MB MT
28S
18S
tRNA
1.37
0.67
0.36
T A+
MB MT MB MT
28S
18S
CT
28S
18S
tRNA
C A+
28S
18S

skeletal muscle. This is due to the shorter 3' non-translated mRNA sequence in the mouse myoglobin gene (section 3.4).

Myoglobin gene expression is also induced during fusion of another rodent cell line -the rat myoblast cell line L6 (Yaffe, 1968). Myoglobin transcripts are detectable only in myotube RNA when myoblast and myotube RNA from this cell line are hybridised to the mouse myoglobin exon 1 probe (Fig.4.4 and see Weller et al., 1986).

4.4 Induction of expression of contractile protein genes during G8-1 myoblast differentiation

Hybridisation to contractile protein gene probes was carried out in order to establish that the G8-1 myoblast cell line was behaving "normally" in that the expression of contractile protein genes is induced on fusion and differentiation of myoblasts as is observed for the other myogenic cell lines (Buckingham, 1977). The appearance of contractile protein gene transcripts in RNA isolated from G8-1 myoblasts and myotubes was investigated using a human myosin light chain 1 cDNA probe and a mouse α- skeletal actin cDNA probe (probe details are given in the legend to Fig. 4.4). The myosin light chain cDNA probe hybridises to the transcripts from genes encoding myosin light chain 1 and embryonic myosin light chain. The α- skeletal actin probe is capable of hybridising with all actin mRNAs in cells including the non-muscle β- and γ-cytoplasmic actins as well as the sarcomeric α-skeletal and α-cardiac actins. 1 μg of the same poly(A)$^+$ RNA sample was analysed by Northern blot hybridisation with these probes. Fig.4.4 shows that α-skeletal actin transcripts are increased significantly in myotube RNA whereas the β- and γ-cytoplasmic actins are predominantly present in myoblast RNA, their levels being reduced significantly in myotube RNA. The cytoplasmic isoforms of actin are therefore replaced with the muscle actin isoform

Figure 4.4

Induction of expression of the myoglobin gene following fusion and differentiation of G8-1 myoblasts detected by Northern blotting

1µg samples of identical poly(A)$^+$ RNAs from G8-1 myoblasts (MB), G8-1 myotubes (MT), and 0.2µg of total adult seal skeletal muscle RNA were electrophoresed and transferred to nitrocellulose by Northern blotting as described in Section 2.6. Filters were hybridised with the probes indicated:

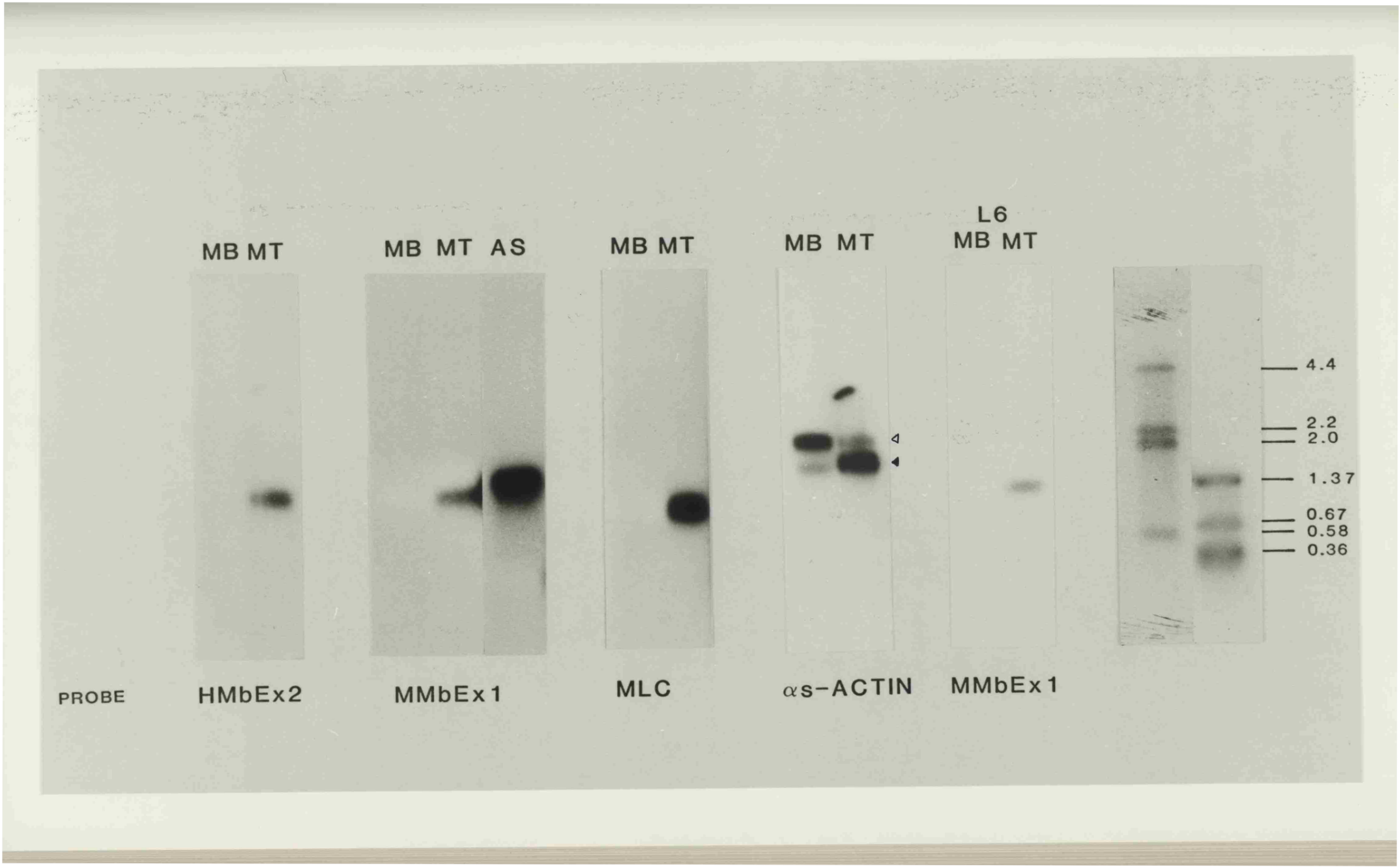
Human myoglobin gene exon 2 (HMbEX2)
316 bp single-stranded antisense DNA probe generated from M13 HEx2. Autoradiography was for 7 days with an intensifier screen

Mouse myoglobin exon 1 (MMbEx1)
190 bp single stranded antisense DNA probe generated from M13 MEx1. Autoradiography was for 5 days. Hybridisation to 0.5µg of adult seal skeletal muscle (AS) total RNA is shown (autoradiography overnight).

Human myosin light chain (MLC) cDNA probe
The MLC probe was a 600 bp PstI fragment of a MLC cDNA clone in pBR322. Autoradiography was for 5 days.

Mouse α-skeletal actin gene (α-S actin)
The mouse α-skeletal actin cDNA probe was a PstI fragment from pAM91 (Minty et al., 1981), 1350 bp in length with 90% actin coding sequences and 300 bp of nontranslated sequence. Autoradiography was was overnight. β- and γ-cytoplasmic actin (~2000 nt) and muscle α-actin (~1600 nt) transcripts are indicated by open and filled triangles respectively.

1µg samples of poly(A)$^+$ RNA from rat L6 (Yaffe, 1969) myoblasts (MB) and myotubes (MT) were hybridised with the mouse myoglobin gene exon 1 single stranded antisense DNA probe. Autoradiography was for 8 days.


MB MT
MB MT AS
MB MT
MB MT
L6
MB MT
4.4
2.2
2.0
1.37
0.67
0.58
0.36
PROBE
HMbEx2
MMbEx1
MLC
αs-ACTIN
MMbEx1

during myoblast fusion. Similarly, muscle myosin light chain transcripts accumulate to high levels in myotube RNA, and no transcripts are detectable in myoblast RNA.

4.5 Appearance of myoglobin transcripts during myoblast differentiation

The apperarance of contractile protein gene transcripts during myoblast differentiation in established cell lines is well documented (see Pearson, 1980, Buckingham, 1977). In some cases, the appearance of the muscle-isoform gene transcripts has been documented as being coordinate (Devlin and Emerson, 1979) and in others a significant delay has been reported between the appearance of some of the contractile protein gene transcripts (see Section 1.4). It would be interesting to determine the time of appearance of myoglobin gene transcripts compared to contractile protein gene transcripts in order to determine whether there are any significant differences in the kinetics of appearance of the respective transcripts.

RNA was extracted from cultures of myoblasts and myotubes at various times during myoblast differentiation and analysed on a Northern blot using the mouse myoglobin exon 1 probe, the α-skeletal actin cDNA probe and a human embryonic myosin heavy chain cDNA probe.

In this particular Northern blotting experiment, RNA was transferred to Hybond N. Subsequent rehybridisation of these filters with the MHC gene and α-skeletal actin gene probes was of poor quality due to a very high back-ground level of radioactivity on the filters after the first hybridisation which could not be successfully removed.

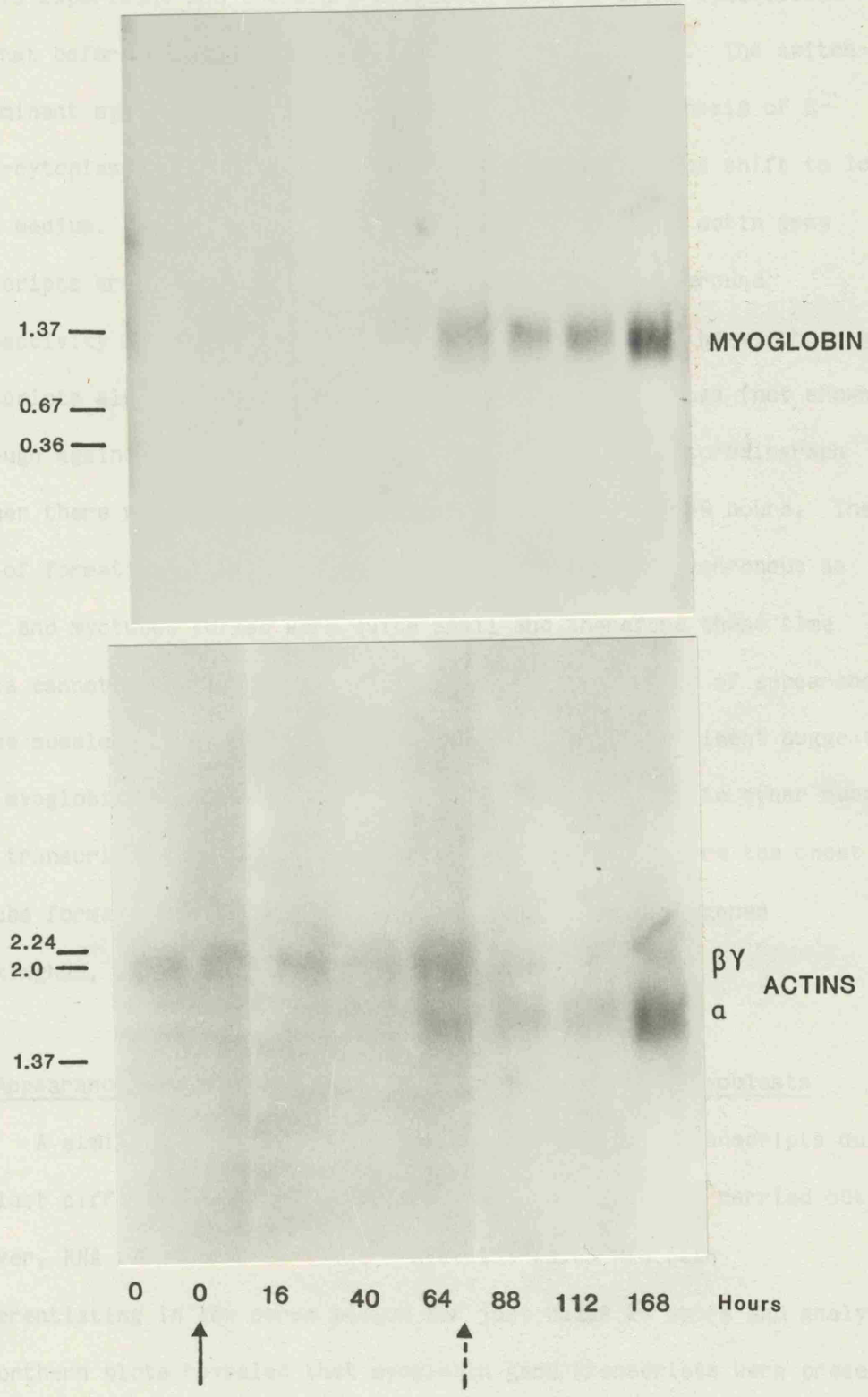
Myoglobin gene transcripts appear at around 40 hrs after changing the medium of confluent myoblasts to low serum medium (Fig. 4.5). The low level of transcripts is probably due to a lower recovery of RNA at this particular time point. Myoglobin mRNA levels are fairly similar

Figure 4.5

Appearance of myoglobin gene transcripts during myoblast differentiation

Poly(A)$^+$ RNA was extracted from myoblasts at different stages of myoblast differentiation and was analysed by Northern blotting using a mouse myoglobin gene exon 1 probe (Fig. 4.4) (top panel). The same filter was then rehybridised with a mouse α-skeletal actin cDNA probe (Fig. 4.4) (lower panel). The first two lanes on the left are non-confluent and confluent myoblasts respectively. The vertical arrow indicates the time of addition of low serum medium to confluent myoblasts and time in hours indicates the total time in low serum medium. The broken vertical arrow indicates the time of appearance of myotubes in the culture (after about 80 hours in low serum medium). The low level of myoglobin gene transcripts at 40 hours is shown by an open triangle. The migration positions of the cytoplasmic actins (β,γ) and skeletal actin (α) are shown.

Autoradiography was for five days (top panel) and 2 days (bottom panel).


1.37
0.67
0.36
MYOGLOBIN
2.24
2.0
1.37
βγ
α
ACTINS
0
0
16
40
64
88
112
168
Hours

in 64-88 hours in fusion medium samples, but are increased in the more mature myotubes at 160 hours. Visible fusion of myoblasts to form small myotubes occured at about 80 hours after the switch to low serum medium in this experiment and therefore myoglobin mRNA is being synthesised somewhat before the fusion of myoblasts to form myotubes. The switch to predominant synthesis of α-skeletal actin mRNA from synthesis of β- and γ-cytoplasmic actins occured at around 64hrs after the shift to low serum medium. It is difficult to see whether α-skeletal actin gene transcripts are present at 40 hours due to the high background radioactivity levels on the filter. Apperance of myosin heavy chain gene transcripts also occurs at 64 hours after the medium change (not shown) although again it was difficult to determine from the autoradiograph whether there was a low level of transcription at around 40 hours. The time of formation of myotubes in this case was not as synchronous as usual and myotubes formed were quite small and therefore these time points cannot be taken as an exact indication of the time of appearance of the muscle gene transcripts. Nevertheless, this experiment suggests that myoglobin gene transcripts appear at a similar time to other muscle gene transcripts during myoblast differentiation and before the onset of myotube formation as is observed for contractile protein genes (Buckingham, 1985).

4.6 Appearance of myoglobin gene transcripts in clone 1 myoblasts

A similar analysis of accumulation of myoglobin transcripts during myoblast differentiation of the clone 1 myoblasts was not carried out. However, RNA isolated from clone 1 myotubes which had been differentiating in low serum medium for just under 24 hours and analysed on Northern blots revealed that myoglobin gene transcripts were present at this early time point (see Fig 4.6). A very low level of myoglobin

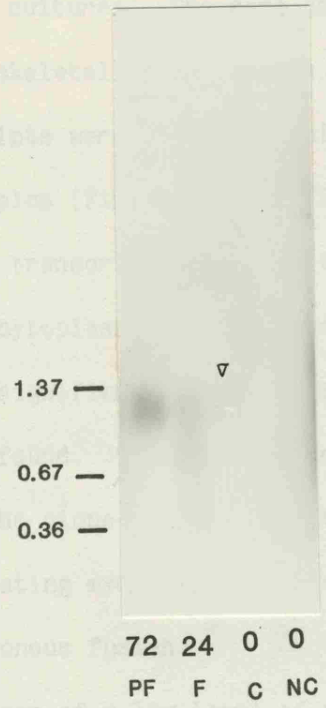
Figure 4.6

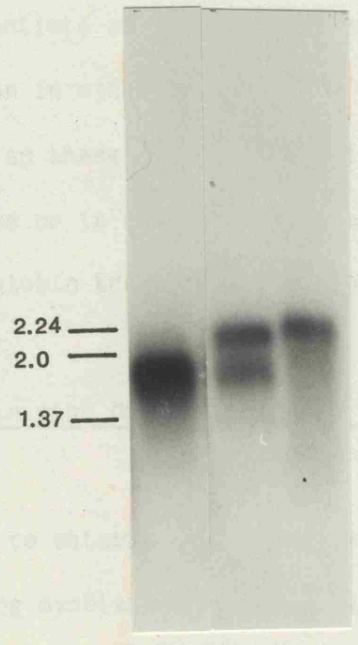
Appearance of myoglobin gene transcripts in clone 1 myoblasts

Poly(A)$^+$ RNA was prepared from clone 1 myoblasts at different times during myoblast differentiation representing: non-confluent myoblasts (NC), confluent myoblasts (C), fusing myoblasts (F) and post-fusion myotubes (PF). Time in hours of myoblasts in low serum medium are shown. RNAs were analysed on two Northern blots.

Top A 190 bp single-stranded antisense DNA, generated from M13 MEx1 (see Table 3.1) was used as hybridisation probe. Myoglobin gene transcripts present at low levels in confluent myoblasts are indicated by an open triangle. Autoradiography was for 5 days.

Bottom A mouse α-skeletal actin cDNA probe (Fig. 4.4) was used as hybridisation probe. The 24 hour poly(A)$^+$ RNA sample has been ommitted. The migration positions of the β and γ cytoplasmic actins (β,γ) and α-skeletal actin (α) are shown. Notice the high level of α-skeletal or cardiac actin gene transcripts in the confluent myoblast (C) RNA sample. Autoradiography was for 2 days.


1.37
0.67
0.36
MYOGLOBIN
72 24 0 0
Hours
PF F C NC


2.24
2.0
1.37
βγ
α
ACTINS
72 0 0
Hours
PF C NC

gene transcripts were also detectable in RNA isolated from near confluent clone 1 myoblast cultures. The same RNA analysed on a second Northern blot with the α-skeletal actin cDNA as hybridisation probe, revealed that .α-actin transcripts were present at higher levels than usual in clone 1 myoblast RNA samples (Fig. 4.6). By 72 hours, a large increase in the level of α-actin transcripts had occured with the complete disappearance of the β- and γ-cytoplasmic actin transcripts. At this time point in G8-1 myoblasts, significant levels of β- and γ-cytoplasmic gene transcripts are found. The disappearance of the cytoplasmic actins indicates that the clone 1 myotube cultures are fully differentiated, free of contaminating myoblasts. Clone 1 myoblasts therefore undergo a rapid and synchronous fusion.

The presence of a low level of myoglobin transcripts and the increased level of α-actin transcripts in near-confluent clone 1 myoblasts are probably due to the increased growth rate of these myoblasts resulting in a rapid depletion of the growth promoting factors in the medium and therefore populations of myoblasts will be triggered to begin to differentiate at an earlier stage. In order to prevent precocious fusion in myoblast cultures, the growth medium to be refreshed more frequently on these myoblasts. RNA extracted from less confluent myoblast cultures or in myoblast cultures where the medium is refreshed frequently, myoglobin transcripts are not detectable (Fig. 4.6).

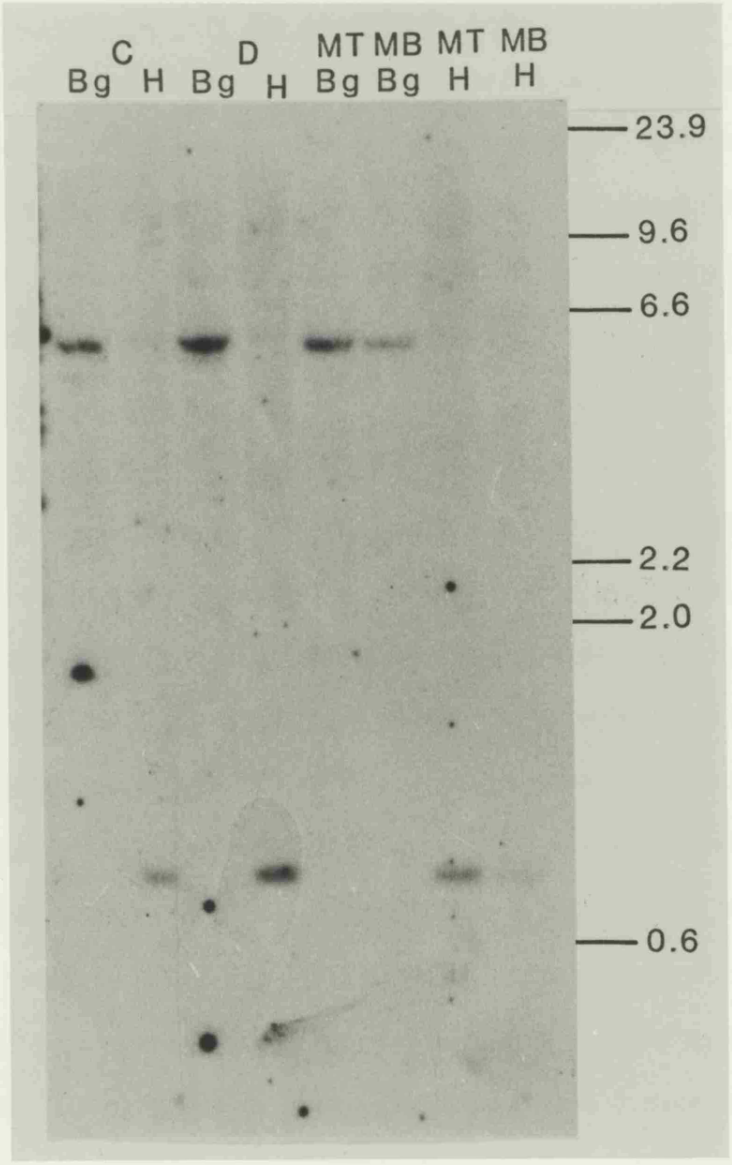
4.7 The myoglobin gene does not undergo major rearrangement during myogenesis

In order to establish whether the increase in myoglobin gene expression during myoblast differentiation is associated with any major rearrangements of the myoglobin gene, the myoglobin gene was analysed in myoblast and myotube DNA by Southern blotting. Fig.4.7 shows the

Figure 4.7

Southern blot analysis of the myoglobin gene in G8 mouse myoblast and myotube DNA

5µg samples of genomic DNA prepared from G8-1 myoblasts (MB), myotubes (MT) and from DBA/2 (D) and C57/BL10 (C) mouse liver were digested with BglII (B) and HindIII (H). After electrophoresis in a 0.8% agarose gel, DNAs were denatured *in situ* and transferred to nitrocellulose by Southern blotting. The filter was hybridised with the mouse myoglobin gene exon 1 probe (Table 1.1). Autoradiography was for 5 days.


C
Bg H
D
Bg H
MT MB
Bg Bg
MT MB
H H
23.9
9.6
6.6
2.2
2.0
0.6

Southern blot of genomic DNA from myoblast and myotube cultures and DBA/2 liver, hybridised to the mouse myoglobin gene exon 1 probe. A single hybridising fragment of approximately equal intensity relative to the DNA on the stained agarose gel is seen in each digested DNA track indicating that induction of myoglobin gene expression in G8-1 myotubes is not accompanied by gene amplification or other major rearrangement of the gene. This was also demonstrated previously (Chapter 3) when Southern blots of genomic DNA from myoblast, myotube and mouse livers were hybridised with human myoglobin gene exon probes.

4.8 Myoglobin gene expression is tissue-specific

To establish whether the expression of the myoglobin gene in differentiating myoblasts is specific to this cell type, a number of non-muscle cell line RNAs and a non-muscle tissue RNA were analysed on Northern blots using myoglobin gene exons as hybridisation probes. Total RNA was obtained from the following non-muscle cell lines: a rat pancreatic cell line (Gazdar et al.,1980), a fibroblastoid line isolated from embryonic mouse kidneys (gift from Dr. A Carter) and a human myeloma cell line (gift from Liz Allen). 10µg of each were analysed on Northern blots with the mouse myoglobin exon 1 probe and the human myoglobin exon 2 probe. Total RNA from mouse liver and heart tissue was also analysed. Neither probe detected myoglobin transcripts in cell line RNA or tissue RNA that was non-muscle (shown in Fig.4.8, using the mouse myoglobin exon 1 probe).

4.9 Summary

The early appearance of myoglobin in embryogenesis can apparently be modelled in cultured mouse and rat embryonic myoblasts which have been shown to synthesise myoglobin mRNA during differentiation in vitro to

Figure 4.8

Is myoglobin gene expression muscle-cell specific?

10μg samples of total RNAs were electrophoresed and transferred to Hybond-N by Northern blotting as described in Section 2.6. Filters were hybridised with the 190 bp single-stranded antisense DNA probe from M13MEx1 (Table 3.1) followed by washing in 0.25 x SSC. Lanes contain RNA from:

1:Pancreatic islet cells

2:Fibroblastoid cells, isolated from baby mouse kidneys

3:Human myeloma cells

4:DBA/12 mouse liver

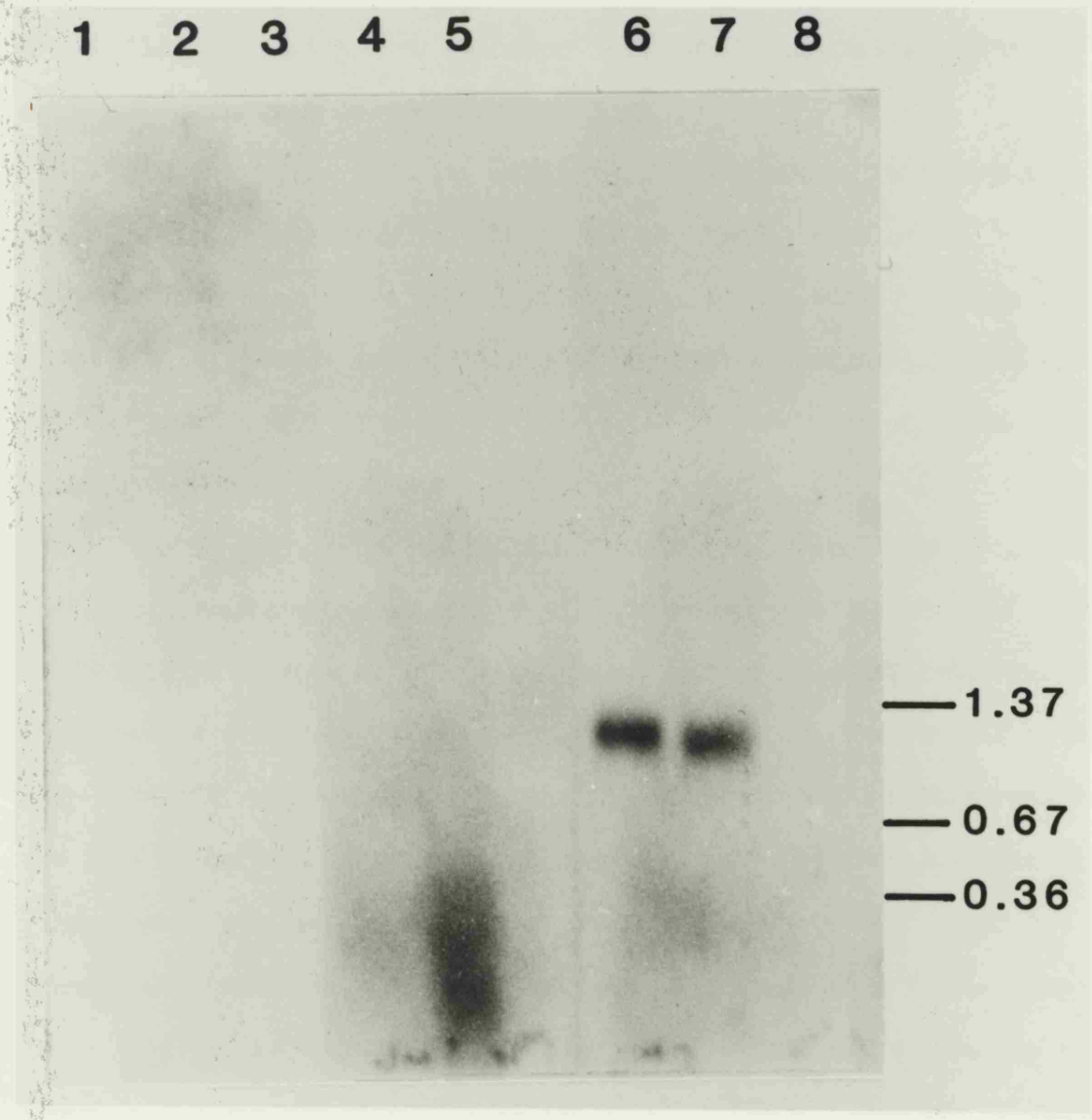
5:DBA/12 mouse heart

6:Mouse myotubes

7: " "

8:Mouse myoblasts

Autoradiography was for 8 days with an intensifier screen. Myoglobin gene transcripts are detected in myotube RNA samples and heart muscle RNA samples. The heart muscle RNA sample obtained was found to be degraded, resulting in the smear of hybridisation observed in lane 5. All other RNA samples were intact. The same result was obtained with a human myoglobin gene exon 2 probe.


1 2 3 4 5 6 7 8
1.37
0.67
0.36

form myotubes. These two myoblast cell lines have been isolated by different methods. The L6 myoblast line was isolated by selecting for cells from carcinogen induced tumours in neonatal rat thigh muscle (Yaffe, 1969) and the G8 myoblast line was isolated as a spontaneously arising transformant during prolonged culture of dissociated embryonic mouse hind limb myoblasts (Christian et al., 1977).

The G8-1 mouse myoblast cell line has been shown to behave as a normal myoblast cell line in that it induces the expression of contractile protein genes during in vitro differentiation. Myosin light chain mRNA and myosin heavy chain mRNA are synthesised during myoblast differentiation and there is a clear switch in the types of actin synthesised in G8-1 myotubes.

The accumulation of myoglobin transcripts during myogenesis in the G8-1 cell line is reproducible; myoglobin transcripts are always detectable in myotube RNA. Levels of transcript accumulation, however, vary between different batches of fused cells but it is important to note that dishes of myoblasts seeded from the same original myoblasts which are then stimulated to differentiate with the same low serum medium all give rise to similar transcript levels. The final level of myoglobin transcripts attained in myoblast cultures are consistantly higher than those obtained in foetal skeletal muscle in vivo and are similar to those found in 1-8 day old mice (see Weller et al., 1986). This may be due to the particular rodent cell lines under investigation containing defects in myoglobin gene regulation and therefore do not reflect the true in vivo situation. Cell lines with defects in contractile protein gene regulation have been described (see Hickey et al., 1986 for example). Higher transcript levels could also result if myoglobin gene transcription is being modulated by some external stimulus which later regulates myoglobin gene expression during the post-natal period.

The kinetics of accumulation of myoglobin transcripts is similar to the accumulation of contractile protein gene transcripts and both take place before visible morphological differentiation. This suggests that myoglobin and contractile protein gene expression may share some common regulatory mechanisms. The accumulation of muscle gene transcripts during myoblast differentiation is well established (Pearson, 1980, Buckingham, 1977), but the time of first appaearance of the transcripts and of visible myotubes depends upon the cell line in use. For example, in the C2C12 myoblast cell line (Bains et al., 1984) the synthesis of muscle gene transcripts and the switch to predominant synthesis of α-skeletal actin transcripts occurs just 6 hours after the switch to low serum medium and visible myotubes appear at around 24 hours. In the L6 myoblast line the appearance of muscle gene transcripts and the predominant switch to α-skeletal actin transcripts occurs at around 36 hours after change to low serum medium, myotube formation occuring at 48 hours (Endo and Nadal-Ginard, 1986). G8-1 myoblasts therefore differentiate relatively slowly in the particular conditions described here. It would be a good idea to analyse the accumulation of myoglobin transcripts in other well characterised myoblast lines to determine whether the kinetics are similar to those in G8-1.

A clonal derivative of the G8-1 cell line has been isolated which appears to differentiate more rapidly. This clone has "leaky" expression of myoglobin mRNA and higher levels of α-skeletal actin transcripts in near confluent myoblasts if myoblasts are not maintained carefully. It also appears that clone 1 myoblasts differentiate rapidly and synchronously in culture since there is an almost complete disappearance of the β- and γ-cytoplasmic actin transcripts.

Myoglobin gene expression appears to be tissue specific in that myoglobin mRNA is only detected in differentiating myoblasts and muscle

tissue. Myoglobin transcripts could not be detected in non-muscle RNA prepared from various non-muscle cell lines or a non-muscle tissue.

Gene activation during myogenesis does not involve a major rearrangement of the gene, for example gene amplification, since gene structure in myoblast and myotube genomic DNA and non-muscle genomic DNA is apparently identical.

It is not possible by Northern blot analysis to determine whether the accumulation of myoglobin mRNA during myoblast differentiation is primarily due to an increase in the rate of transcription of the myoglobin gene during differentiation. Transcriptional regulation of skeletal muscle specific isoforms has only been demonstrated directly for MHC (Medford et al., 1983), α-skeletal actin (Melloul et al., 1984, Nudel et al., 1985, Grichnk, Bergsma and Schwartz, 1986), troponin I (Konieczny and Emerson, 1985) and muscle creatine kinase (Jaynes et al., 1986).

Irrespective of whether the observed induction of myoglobin gene expression is due to trancriptional or post-transcriptional events, it should be possible to identify DNA sequences in or around the gene which are necessary for the induction of gene expression using this cell culture system.

CHAPTER 5

INTRODUCTION OF EXOGENOUS DNA INTO MOUSE MYOBLASTS

5.1 Introduction

One way towards understanding the processes which regulate the transcription of genes encoding mRNA in eukaryotic cells is to achieve natural transcription events in cell free systems reconstituted from purified components. *In vitro* systems which are capable of synthesising specific capped transcripts from genes transcribed by RNA polymerase II have been described (Weil *et al.*, 1979, Manley *et al.*, 1980) but in general these systems do not faithfully reflect regulated gene expression observed *in vivo* (see for example, Wickens and Laskey, 1981, Wilderman *et al.*, 1984). While the *in vitro* reconstitution experiments will remain the ultimate aim, basic features of such processes must first be defined by reintroducing genes into an appropriate cellular environment. For this purpose introduction of genes into tissue culture cells is extremely valuable.

Transfer of DNA into eukaryotic cells (transfection) is exactly analogous to the transformation of bacteria by the introduction of genes or other segments of DNA. In bacteria, introduction of plasmid and phage DNA has allowed the identification of *cis*- and *trans*-acting regulatory elements which regulate gene expression. Since the description of a technique that allowed a more efficient uptake of DNA by mammalian cells in culture (Graham and van der Eb, 1973), substantial advances have been made in the study of control of eukaryotic gene expression.

Early experiments to study gene expression in transfected mammalian cells used a variety of viral vectors on which to introduce the gene of interest (Hamer, 1977, Sarver *et al.*, 1981). In these vectors

parts of the viral genome were replaced with the gene under investigation. For example, in Simian Virus 40 (SV40) vectors the late or early region genes of the virus were replaced with foreign DNA. In vectors that retained all the necessary signals for transcription, including an intact mRNA leader sequence, splice signals and a poly(A) signal, transcription of the foreign gene in the virus vector was observed (Mulligan, Howard and Berg, 1979, Gruss and Khoury, 1981, Lui, Yansura and Levinson, 1982). Later viral vectors combined SV40 sequences with plasmid sequences and the gene of interest which not only overcame the size restriction on the gene to be inserted into an SV40 genome, but also allowed the vector to be propagated in bacteria before transfer into mammalian cells. Transfections utilising virus based vectors, whether in transient expression analyses or long term transformation, enabled a number of genes to be expressed in a cell type in which they are never expressed *in vivo*. For example, the rabbit β-globin gene could be expressed in HeLa cells (Grosveld *et al.*, 1982), or mouse L cells (Mantei, Boll and Weissman, 1981, Dierks *et al.*, 1981) and mouse α-globin could be expressed in monkey kidney cells (Mellon *et al.*, 1981), whereas these two genes are only expressed in erythroid tissue *in vivo*. The expression of genes in heterologous cell types was explained by the identification of an 'enhancer' sequence in the virus promoter region in similar transient expression experiments. Rabbit β-globin gene expression in HeLa cells showed a requirement for the SV40 enhancer sequence (Banerji *et al.*, 1981), although the α-globin gene does not show this requirement. In many other cases it has been shown that the enhancer sequence may over-ride the regulatory control sequences, for example those specifying cell-type expression, of a gene downstream. Renkawitz *et al.* (1982) found that the chicken lysosyme gene may be faithfully transcribed in HeLa and MCF-7 cells where lysozyme is not

normally expressed only when the SV40 enhancer region is placed upstream. In the absence of the SV40 enhancer the lysozyme gene may only be expressed when introduced into chicken oviduct cells. Furthermore, lysosyme expression in the heterologous cell types is not regulated by steroid hormones but is regulated appropriately by these hormones in chick oviduct cells. It is generally true that in the absence of a viral enhancer sequence, the gene in question is expressed at significant levels only in cell types where the gene is normally expressed *in vivo*. Although the transfection experiments with these original virus based vectors proved successful in identifying 'core' promoter elements in the 5'-flanking region of eukaryotic genes, for example the TATA and CCAAT boxes (see Section 1.2) it became clear that signals further upstream in the promoter region were probably going to be overlooked due to the inclusion of virus regulatory sequences in the vector. Viral sequences were therefore omitted from expression vectors in order to carry out studies on regulated gene expression. Recently, the introduction of genes into differentiated cell types in culture that express the cognate gene, demonstrates that appropriate regulation of the cloned gene is reproduced and it appears that sequences within the 5'-flanking region of a gene are frequently sufficient to confer regulated and cell type-specific expression on the gene in both transient or stable expression experiments (Walker *et al.*, 1983, Ott *et al.*, 1984, Melloul *et al.*, 1984, Muglia and Rothman-Denes, 1986).

In order to study regulated myoglobin gene expression, the myoglobin gene must be re-introduced into the cell type in which the endogenous myoglobin gene is expressed, namely muscle cells in culture (Chapter 4). The G8-1 mouse myoblast cell line described in Chapter 4 has not previously been used in transfection experiments and therefore it was necessary to determine whether these cells were capable of taking up

and expressing exogenously added DNA. This Chapter describes transfection of various expression vectors into mouse myoblasts and analysis of expression from these vectors over a transient expression period.

The most widely used method for introducing genes into mammalian cells is the calcium phosphate technique first described by Graham and van der Eb (1973) and this procedure was used to transfect G8 myoblasts. This technique can be used to introduce any DNA into mammalian cells for transient expression assays or long-term transformation. It involves mixing exogenous DNA with calcium chloride to which is then added a solution containing phosphate ions. The calcium phosphate-DNA co-precipitate which forms in this mixture is readily taken up by mammalian cells in culture and exogenous DNA is expressed at high levels. Expression from DNA transfected into mouse myoblasts was investigated after a transient expression period using either Northern blot analysis to detect specific transcripts in RNA isolated from the transfected myoblasts or enzyme assays to detect gene products in transfected cell extracts.

5.2 Outline of the procedure used to transfect mouse myoblasts

The calcium phosphate technique used to transfer DNA into G8 mouse myoblasts was essentially that of Graham and van der Eb (1973), with modifications described by Wigler *et al.* (1978) or Gorman, Moffat and Howard (1982) and is described in Section 2.14.

Basically, exponentially growing myoblasts were trypsinised and replated at a density of $1x10^6$ cells per 10 cm dish the day before transfection to ensure that myoblasts were in log phase the next morning. Myoblasts were always about 70% confluent at the time of transfection and transfections were always carried out at a low cell passage number-

usually on the second passage and always before the sixth passage. The normal growth medium above the cells was changed 3 hours before addition of the DNA precipitate to the cells. Plasmid DNA to be transfected was electrophoresed on agarose gels before each transfection to ensure that it was predominantly supercoiled. Care was taken to obtain very fine calcium phosphate-DNA precipitates since if this is not accomplished, a low efficiency gene transfer results (Spandidos and Wilkie, 1983). To prepare fine precipitates, the solution containing the phosphate ions was added to the well mixed DNA and calcium chloride solution dropwise from a small pipette using a pipette aid and immediately after the addition air was bubbled through the solution for several seconds using the pipette aid. A very fine precipitate usually formed after 15 mins at room temperature. Precipitates were never made in volumes larger than 2 mls since scaling-up the solutions resulted in the formation of coarse, uneven precipitates which led to low transfection efficiencies. Usually, 1 ml of precipitate containing 20 µg of DNA (plasmid plus carrier DNA) was added directly to a 10 cm dish containing 10 ml of medium. DNA precipitates were left on the cells overnight (15-20 hours, but see later) and the medium containing the DNA precipitate was then replaced with fresh medium containing 10% FCS. The cells were left to incubate for a further 24-36 hours. After this time, cells were washed, trypsinised from the culture dish and pelleted. Total RNA or cell extracts were prepared from cell pellets which had been frozen at -70°C.

5.3 Transfection of expression vector p4I into mouse myoblasts and analysis of gene expression using Northern blots and dot-blots

G8 myoblasts were transfected with an expression vector (p4I), obtained from C. Weissmann. Expression vector p4I, (Fig.5.1, Wieringa *et al.*, 1983) consists of pBR327, plus a 3,309 bp segment of SV40 DNA

containing the enhancer sequence as well as the origin of replication and the complete early transcription unit together with the mouse and rabbit β-globin genes. This particular vector was chosen for the preliminary transfection experiments since it produces a high level of transcripts from the SV40 early region genes and the rabbit and mouse β-globin genes, easily detectable on Northern blots.

Ten 10 cm dishes of sub-confluent myoblasts were transfected with 10μg of p4I + 10μg of salmon sperm carrier DNA. The same number of myoblasts were transfected with 20μg of salmon sperm carrier DNA only as a mock-transfected control. Myoblasts were transfected as described above and left to express the exogenous plasmid DNA for ~50 hours. RNA and DNA were isolated from the cell pellets of transfected and mock-transfected myoblasts as described in Section 2.12 and 2.13 respectively. DNA recovered from transfected and mock-transfected cell pellets was analysed on a Southern blot using nick-translated p4I expression vector as a hybridisation probe (Fig.5.1). Plasmid DNA was readily detected in digested and non-digested DNA from myoblasts which had been transfected but hybridisation signals were not observed in mock-transfected myoblast DNA tracks. However, the presence of the mouse β-globin gene on p4I enables ^{32}P-labelled p4I vector to detect endogenous mouse β-globin genes in the EcoRI digested mock-transfected myoblast DNA sample. The probe hybridised to two fragments, each ~10 kb in size which correspond to the fragment sizes observed for C57/L (Hbb^s) mouse β-globin genes (Sue Adams, Ph.D thesis, 1983). In the EcoRI digested p4I transfected myoblast DNA, three distinct hybridising components are observed which correspond to the expected fragment sizes of 4I plasmid DNA digested by EcoRI, although most of the DNA in this track runs as a smear of degraded plasmid DNA around 0.5 kb in size. A faint smear of hybridisation is also observed above the three expected

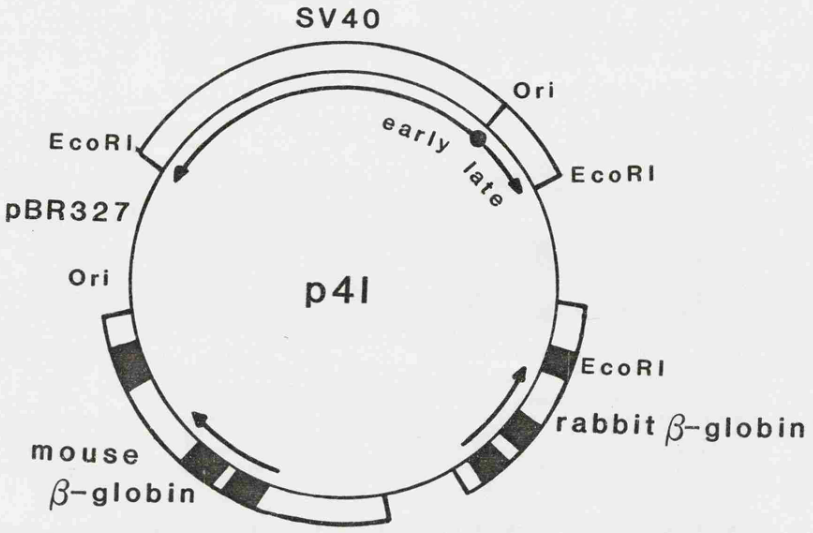
Figure 5.1

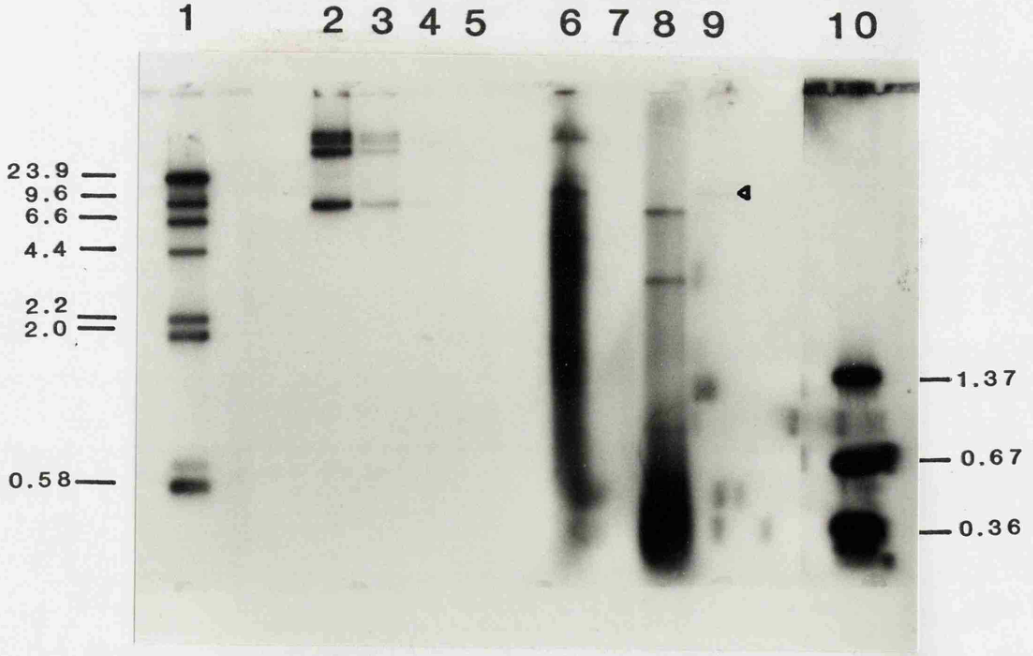
Uptake of exogenous DNA by mouse myoblasts

Top A diagrammatic representation of the expression vector, p4I used to transfect mouse myoblasts. This vector (~11 kb) comprises pBR327 plus a 3.05 kb segment of SV40 DNA (open box) containing the enhancer sequence as well as the entire early transcription unit and origin of replication (ori), together with the mouse and rabbit β-globin genes (open boxes representing non-coding DNA, black boxes coding DNA). The direction of transcription of the two globin genes and the early and late region genes of SV40 is indicated by arrows.

Bottom Genomic DNA from p4I and mock-transfected myoblasts (8 μg per gel lane) were transferred to nitrocellulose by Southern blotting (Section 2.6). The filter was hybridised with ^{32}-P labelled p4I DNA probe, followed by washing in 1 x SSC at 65°C.
Lanes 1 and 10 contain λ x HindIII and pBR322 x Sau3A molecular weight markers.
Lanes 2, 3 ,4 and 5 contain p4I vector DNA
Lane 6 undigested p4I transfected myoblast DNA
Lane 7 undigested mock-transfected myoblast DNA
Lane 8 EcoRI digested p4I transfected myoblast DNA
Lane 9 EcoRI digested mock-transfected myoblast DNA

The ~10 kb doublet DNA fragment which hybridises with p4I probe DNA in lane 9 and is likely to be the endogenous mouse β-globin gene is shown by an open triangle. Autoradiography was overnight.


SV40
Ori
EcoRI
early
late
EcoRI
pBR327
Ori
p4I
EcoRI
rabbit β-globin
mouse
β-globin


1
2 3 4 5
6 7 8 9
10
23.9
9.6
6.6
4.4
2.2
2.0
0.58
1.37
0.67
0.36

EcoRI fragment sizes of the vector indicating that the vector is hybridising to EcoRI fragments of different sizes. This probably indicates that a small amount of vector DNA has become integrated since if all the DNA had been in extrachromosomal multimers, digestion by EcoRI would have produced the three fragments only. It appeared therefore that p4I plasmid DNA was being taken up by myoblasts although it is difficult to demonstrate with Southern blot analysis that the hybridisation signal is not due to plasmid DNA which is associated with the outside of myoblast cell wall debris which might be present as a contaminant in the genomic DNA.

To demonstrate that DNA had indeed been taken up by myoblasts and was expressed, RNA was analysed on Northern blots using p4I vector as a hybridisation probe.

Total RNA was prepared from transfected and mock-transfected myoblasts and passed once over an oligo-dT column to isolate poly(A)$^+$ RNA and remove most of the contaminating plasmid DNA. Poly(A)$^+$ RNA, approximately 3µg per track, was analysed on a Northern blot, using [^{32}P]labelled p4I vector as a hybridisation probe. Fig. 5.2 shows that in the p4I transfected myoblasts RNA a strong hybridising signal is observed but no hybridisation can be detected to RNA of mock-transfected myoblasts. Furthermore the signal in the p4I transfected myoblast poly(A)$^+$ RNA track disappears after RNase treatment, indicating that the signal is not due to plasmid DNA contaminating the RNA. The transcripts however are heterogeneous, probably due to several different transcripts and partial degradation of the RNA. To ensure that the transfection of myoblasts with this vector was reproducible, the transfection experiment was repeated with a different batch of 4I plasmid DNA and transcripts analysed on RNA dot blots (Fig.5.2). An RNase sensitive signal was detected once again only in RNA from myoblasts

Figure 5.2 Expression of p4I DNA in mouse myoblasts

A. Northern blot analysis of mRNA isolated from p4I transfected and mock-transfected myoblasts

Poly(A)$^+$ RNA samples, (~3µg per gel lane), were electrophoresed and transferred to nitrocellulose by Northern blotting as described in Section 2.6. Filters were hybridised to ^{32}P-labelled p4I DNA followed by washing in 0.1 x SSC at 50°C and autoradiography.

Lanes 1 and 9 contain molecular weight markers λ x HindIII and pBR322 x Sau3A respectively. Sizes in kb are shown next to the appropriate hybridising fragment.

Lanes 2, 3 and 4 contain 100, 10 and 1 pg of uncut p4I DNA respectively run in the same conditions as the RNA.

Lanes 5-8 contain 3µg of poly(A)$^+$ RNA from:

5: Mock-transfected myoblasts. Treated with RNase

6: p4I transfected myoblasts. Treated with RNase

7: Mock-transfected myoblasts

8: p4I transfected myoblasts

The open triangle shows the presence in lane 6 of a small amount of contaminating, degraded vector DNA. Autoradiography was for three days.

B. Dot-blot analysis of p4I transfected and mock-transfected myoblasts

Samples of poly(A)$^+$ RNA from p4I and mock-transfected myoblasts were spotted onto a nitrocellulose filter as described in Section 2.6 and hybridised as in A. Spots contain ~1µg of poly(A)$^+$ RNA from:

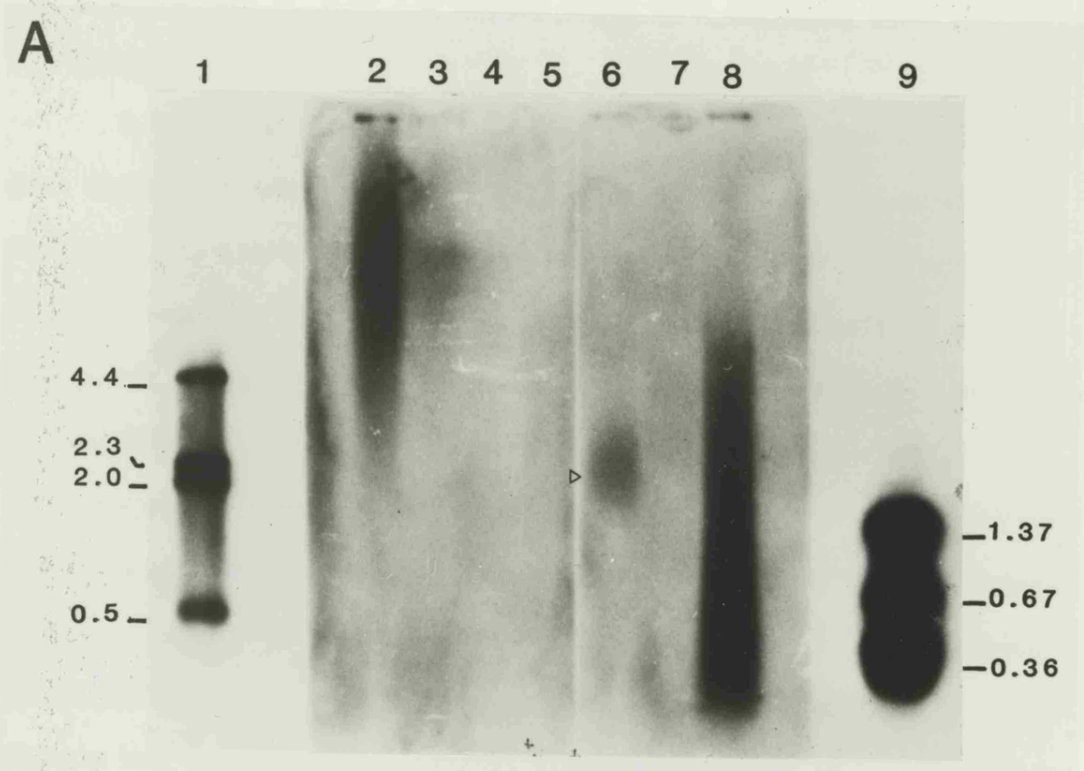
1: p4I transfected myoblasts, treated with RNase.

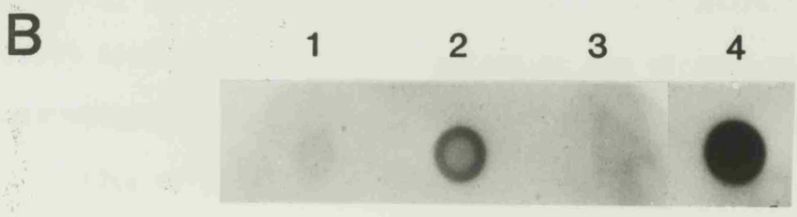
2: p4I transfected myoblasts

3: mock-transfected myoblasts

and 4: 200 pg of p4I vector DNA

Autoradiography was for 4 days.


A
1
2
3
4
5
6
7
8
9
4.4
2.3
2.0
0.5
1.37
0.67
0.36


B
1
2
3
4

transfected with p4I DNA and not in the mock-transfected control.

5.4 Transfection of mouse myoblasts with expression vector pRSV-CAT and analysis of gene expression using both Northern blots and an enzyme assay

A number of expression vectors have been developed which contain a gene coding for enzyme which is not normally expressed in eukaryotic cells and may therefore be used as a "reporter" of promoter activity when placed downstream of the promoter in question. An example of an expression vector containing a reporter function is pRSV-CAT, developed by Gorman et al. (1982). This vector contains the bacterial chloramphenicol acetyl transferase gene (CAT) from transposable element Tn9 placed under the transcriptional control of a long terminal repeat promoter (LTR) from Rous Sarcoma virus (RSV). The CAT gene has been modified for use in eukaryotic cells by the addition of SV40 splice signals and polyadenylation signal (Chapter 6). CAT is a particularly useful reporter function to monitor the promoter activity of a gene (see Chapter 6), and has been used in a wide variety of eukaryotic cells to assay promoter activity and to quantitate transfection efficiencies (Gorman et al., 1984). pRSV-CAT was used to transfect mouse myoblasts as a first step to establish the potential use of this attractive system to study myoglobin gene expression.

10µg of pRSV-CAT were used to transfect five 10 cm dishes of sub-confluent myoblasts as described above. Myoblasts were incubated for 52 hrs after addition of the DNA precipitate to allow expression of the CAT gene to occur. Cytoplasmic RNA was extracted from myoblast cell pellets of transfected and mock-transfected samples and poly(A)$^+$ RNA prepared from the cytoplasmic RNA (Section 2.12). Expression of CAT was investigated using Northern blots to detect CAT transcripts in the

cytoplasmic RNA and poly(A)$^+$ RNA. The hybridisation probe was the CAT containing fragment isolated from pRSV-CAT and contained the CAT coding segment and SV40 RNA processing signals which was ^{32}P labelled by nick-translation. A faint hybridising component can be detected in tracks corresponding to total RNA from transfected myoblasts which is absent both from transfected myoblast RNA samples which have been treated with RNase and from mock-transfected total RNA samples (Fig 5.3). It is difficult to see the band because of a high background of hybridisation to contaminating plasmid DNA but it appears to be migrating at about 1900nt. The transcript size is in agreement with a predicted CAT mRNA size of 1620 nt (not including the polyA tail). Poly(A)$^+$ RNA's from pRSV-CAT and mock-transfected myoblasts analysed on a second Northern blot contain a hybridising signal only in RNA from pRSV-CAT transfected myoblasts. The transcript is quite degraded but appears to be migrating around 1300 nt. The hybridising signal in the pRSV-CAT transfected poly(A)$^+$ RNA sample is again sensitive to RNase indicating that it is likely to be due to a specific RNA transcript in pRSV-CAT transfected myoblasts rather than contamination of the RNA samples with plasmid DNA.

Further transfections of mouse myoblasts were carried out with pRSV-CAT to determine whether CAT gene mRNA is translated. CAT activity was monitored by assaying cell extracts made from transfected myoblasts for enzyme activity. The CAT assay is a simple radioactive assay in which CAT activity is detected by the conversion of radioactive chloramphenicol to two mono-acetylated products which can be separated from chloramphenicol by thin layer chromatography and visualised by autoradiography of the silica gel plate. Assays of CAT enzyme activity in cell extracts are a more sensitive method to monitor expression from CAT gene fusions (Gorman et al., 1984) and provide a rapid method of monitoring promoter activity compared to the tedious process of RNA

Figure 5.3

Expression of pRSV-CAT DNA in mouse myoblasts

A Cytoplasmic RNA samples were electrophoresed and transferred to nitrocellulose by Northern blotting (Section 2.6). The filter was hybridised to the 1.6 kb BamHI/HindIII CAT coding fragment from pRSV-CAT. The first lane of each pair contains 20μg and the second lane, 40μg of RNA from:

Lanes 1 and 2: Mock-transfected myoblasts. Treated with RNase.

Lanes 3 and 4: pRSV-CAT transfected myoblasts. Treated with RNase.

Lanes 5 and 6: Mock-transfected myoblasts

Lanes 7 and 8: pRSV-CAT transfected myoblasts

Lane 11 contains molecular weight markers λ x HindIII and lanes 9 and 10 contain 10 pg and 100 pg of pRSV-CAT digested with HindIII

The hybridising component observed at around 1900 nt is shown by an open triangle. Exposure was for 2 days.

B Poly(A)$^+$ RNA prepared from the cytoplasmic RNA samples in A, were treated and analysed as in A. Lanes contain ~3μg of poly(A)$^+$ RNA from:

Lane 2: pRSV-CAT transfected myoblasts treated with RNase

Lane 3: pRSV-CAT transfected myoblasts

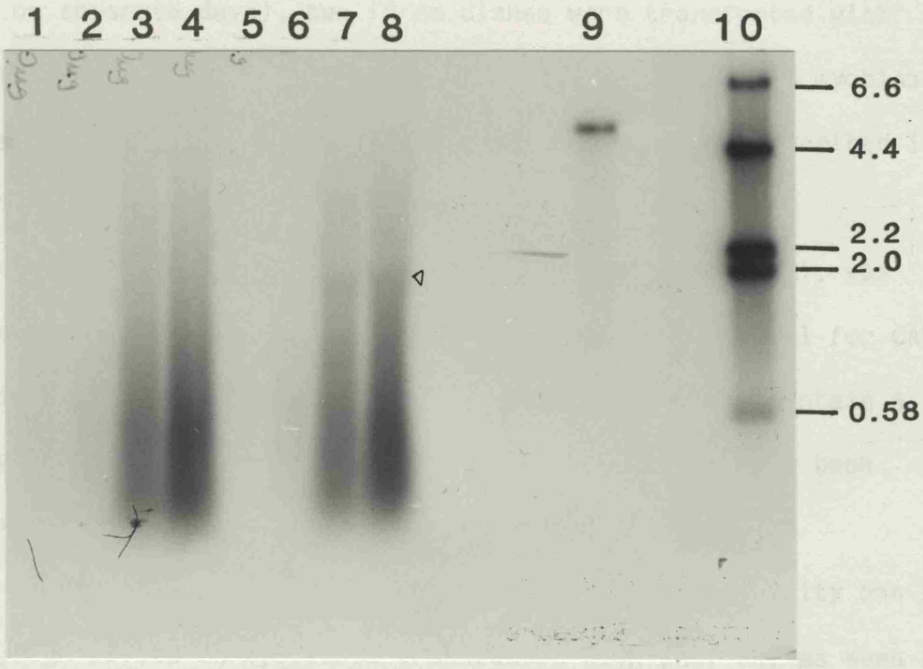
Lane 4: Mock-transfected myoblasts treated with RNase

Lane 5: Mock-transfected myoblasts

Molecular weight markers λ x HindIII and pBR322 x Sau3A are in Lanes 1 and 9 and Lanes 6, 7 and 8 contain 1, 10 and 100 pg pRSV-CAT DNA x HindIII

Exposure was for 3 days with an intensifier screen.

A


1 2 3 4 5 6 7 8 9 10
6.6
4.4
2.2
2.0
0.58

B


1 2 3 4 5 6 7 8 9
1.37
0.67
0.36
6.6
4.4
2.2
2.0
0.58

isolation and analysis.

For each transient expression experiment (two transfections were carried out on separate days), two 10 cm dishes were transfected with 10µg of pRSV-CAT. Cell extracts were prepared from transfected myoblasts after an expression period of about 48 hours and assayed as described in Section 2.15.

A second CAT based vector, pSVO-CAT (Gorman et al., 1982), was also used to transfect myoblasts. This vector acts as a control for CAT expression in CAT based expression vectors since it does not contain a eukaryotic gene promoter, the SV40 early region promoter having been deleted from this vector.

Fig. 5.4 shows the results of the CAT assays. CAT activity can be detected at high levels in myoblasts transfected with pRSV-CAT as seen by the conversion of chloramphenicol to the two mono-acetylated forms. The radioctive spots containing unmodified chloramphenicol and chloramphenicol 1- and 3-acetate were cut out of the TLC plate and counted in a scintillation counter. The percentage of conversion of chloramphenicol to its acetylated products in the two transfections are 79% and 72% using 25µl of cell extract. There is no detectable CAT activity in mock-transfected myoblast cell extracts and only backround levels (≤1% conversion of chloramphenicol) in pSVO-CAT transfected myoblast cell extracts.

The conditions of the transfection assay were varied in order to optimize transfection efficiency and to determine whether the various procedures of the transfection experiment affected the subsequent differentiation of mouse myoblasts. Several chemical facilitators have been described which increase the expression of transferred donor DNA in recipient cells (Wilkie and Spandidos, 1983). Two such facilitators, 18% DMSO and 20% glycerol, were tested for toxicity to G8 myoblasts when

Figure 5.4

Expression of pRSV-CAT DNA in mouse myoblasts

CAT assays were carried out on two sets of transfected myoblast cell extracts as described in Section 2.15. All assays used 25µl of cell extract containing the same amount of protein (around 250µg) as determined in Section 2.15 and 0.5 µCi of ^{14}C-chloramphenicol. Extracts were assayed for four hours.

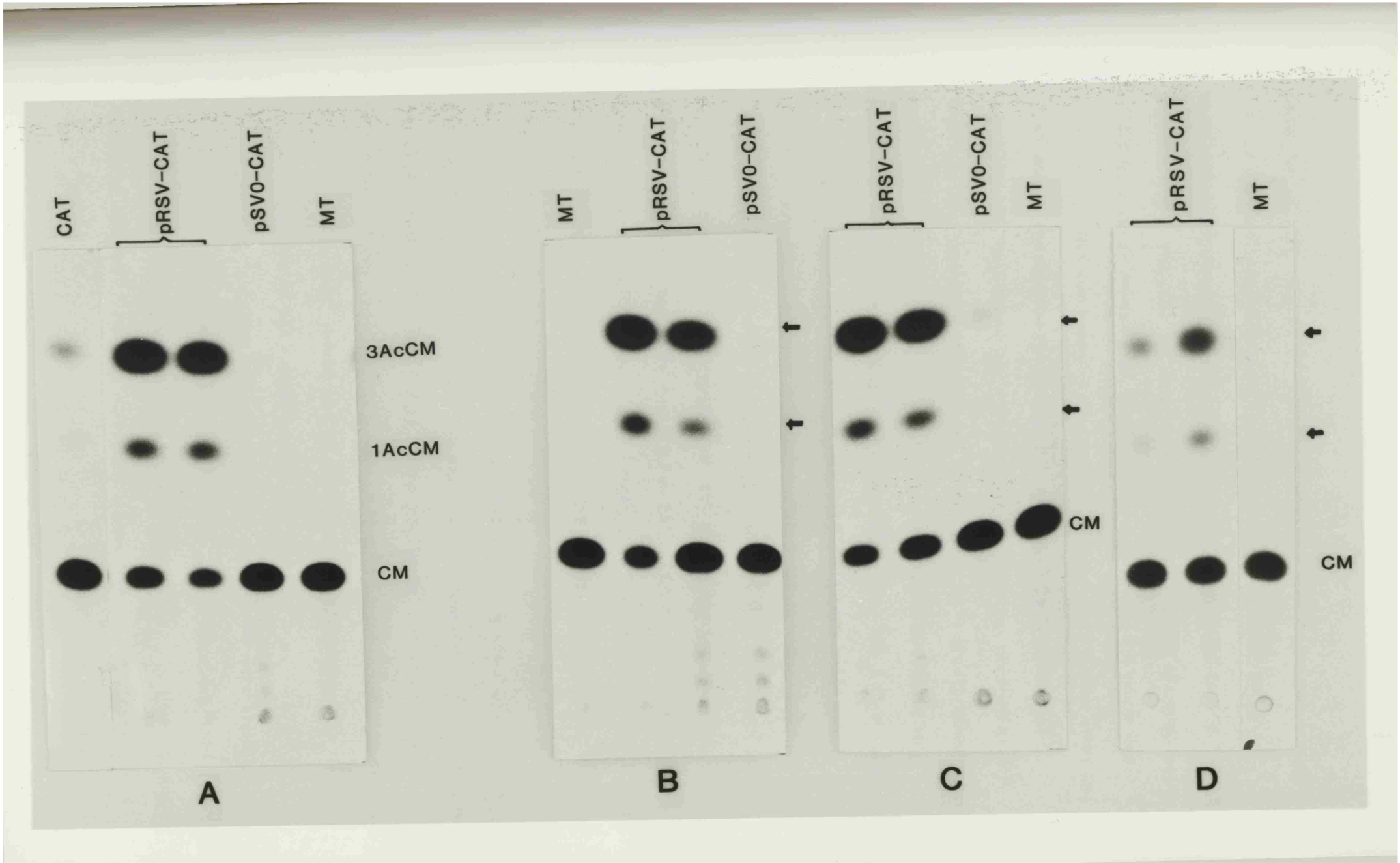
A CAT assays from myoblasts transfected with pRSV-CAT and pSV0-CAT. Myoblasts were exposed to the calcium phosphate-DNA precipitate overnight.

B CAT assays from myoblasts transfected with pRSV-CAT and pSV0-CAT. Myoblasts were exposed to the calcium phosphate-DNA precipitate overnight.

C Assays from myoblasts transfected with pRSV-CAT and pSV0-CAT. Calcium phosphate-DNA precipitates were left on myoblasts for 6 hours after which time myoblasts were subjected to a 20% glycerol shock for 2.5 minutes.

D Assays from myoblasts transfected with pRSV-CAT. DNA precipitates were left on myoblasts for 6 hours.

CM shows the position of the unmodified chloramphenicol and 1AcCM and 3AcCM the two mono-acetylated products, depicted by arrows in panels B-D. M.T. shows CAT assays carried out on mock-transfected myoblasts. Autoradiography was overnight.


CAT
pRSV-CAT
pSV0-CAT
MT
3AcCM
1AcCM
CM
A
MT
pRSV-CAT
pSV0-CAT
B
pRSV-CAT
pSV0-CAT
MT
CM
C
pRSV-CAT
MT
CM
D

added for varying lengths of time. Exposure for short periods to DMSO (~1-1.5 minutes) appeared quite damaging to the myoblasts, clone 1 especially, whereas even quite long exposures of myoblasts to 20% glycerol (~2-5 minutes) had no effect (not shown). A 20% glycerol shock of 2.5 mins after exposure of the myoblasts to DNA for six hours was chosen since this was reported to increase expression of transferred DNA in L6 and L8 rat myoblasts (Rebecca Aft and David Yaffe, personal communication, Melloul et al., 1984). Two transfection procedures were investigated: Myoblasts were exposed to DNA (pRSV-CAT) precipitates for 6 hours only. The DNA precipitate was then removed and myoblasts were refed with fresh growth medium. Alternatively, after the 6 hour exposure of the myoblasts to the DNA precipitate, the cells were subjected to the glycerol shock described and then refed with fresh medium. These conditions were compared to the overnight incubation of myoblasts with the DNA precipitate. Transfected myoblast cell extracts (from two seperate transfections) were assayed for CAT activity (Fig. 5.4). The levels of CAT activity appear to be essentially the same in extracts from myoblasts which had been exposed to the DNA precipitate overnight (64% and 36% of chloramphenicol converted) and in extracts from myoblasts which had been exposed to the DNA precipitate for 6 hours followed by a glycerol shock (76% and 45% of chloramphenicol converted). This was true for other batches of plasmid DNAs. However, in cell extracts from myoblasts that had been exposed to DNA precipitates for 6 hours without a glycerol shock, lower levels of CAT activity were observed (about 10% and 15% of conversion of chloramphenicol).

Parallel transient mock-transfection experiments were carried out to assess the effect of the calcium phosphate-DNA precipitate and glycerol shock on subsequent myoblast differentiation. It appeared that myoblasts exposed to calcium phosphate-salmon sperm DNA precipitates

overnight were much less effective at differentiating to form myotubes than myoblasts that were exposed to the precipitates for shorter periods of time even when these myoblasts were subjected to the glycerol shock step (not shown). Even the shorter exposure time delayed myoblast differentiation for about 24 hours. This was true for both parental G8-1 and derived clone 1 myoblasts. Another line of myoblasts, rat L8 myoblasts are sensitive to the calcium phosphate-DNA precipitate which cannot be left on the myoblasts for longer than a six hour period. Differentiation of transfected L8 myoblasts is also delayed for about 24 hours (R. Aft-Konisberg and D. Yaffe personal communication).

5.5 Transfection of mouse myoblasts with a second CAT vector, pSV2-CAT

In order to determine whether the CAT gene may be expressed from promoters other than the RSV LTR promoter, a second similar CAT expression vector, pSV2-CAT, was used to transfect myoblasts. In pSV2-CAT, CAT expression is driven by the SV40 early region promoter and the rest of the vector is essentially the same as pRSV-CAT (Gorman, Moffat and Howard, 1982). The levels of CAT activity in myoblast cell extracts transfected with this CAT vector are shown in Fig 5.5. As previously, no detectable CAT activity is observed in mock-transfected myoblasts and only a small amount of background activity is observed in cell extracts from pSV0-CAT transfected myoblasts. pSV2-CAT transfected myoblast cell extracts have high levels of CAT activity (in duplicate experiments, (39% and 75% conversion of chloramphenicol per 35µl of cell extract). In transfected myoblasts, the RSV LTR promoter appears to drive approximately two times more CAT expression than the SV40 early region promoter, compared to the situation with transfected CEF cells (Gorman et al., 1982), where the LTR promoter drives at least 5 times more CAT expression than the SV40 early region promoter. Level of CAT

Figure 5.5

Analysis of CAT enzymatic activity in myoblasts transfected with pSV2-CAT

Myoblasts were transfected with pSV2-CAT as described in Section 2.14 with a glycerol shock step. CAT assays were carried out for 4 hours on 35µl of transfected cell extracts, containing about 350-400µg of protein and 0.5 µCi of chloramphenicol. Autoradiography was overnight.

The position of chloramphenicol (CM) and the two monoacetylated derivatives (1AcCM, 3AcCM) are shown. Assay of CAT activity in mock-transfected myoblast(M.T.) and pSV0-CAT transfected myoblast cell extracts are shown as controls.

pSV0-CAT
pSV2-CAT
MT
3AcCM
1AcCM
CM

expression from both promoters varies from transfection to transfection but as determined by Gorman et al. (1984), variable levels can be minimised by subjecting plasmid DNAs to be transfected to two cycles of caesium chloride density centrifugation.

The transfections and CAT assays illustrated in this Chapter were carried out using clone 1 myoblasts as these are the cells of choice for later transfections in which cloned myoglobin genes will be introduced into myoblasts since clone 1 myoblasts differentiate rapidly.

Transfection of myoblasts was not always successful as demonstrated by very low levels of CAT activity in some transfected cell extracts. Similar transient expression experiments with the original G8-1 myoblasts revealed lower and even less consistent levels of CAT activity compared with clone 1 myoblasts (not shown).

5.6 Summary

G8 mouse myoblasts are able to take up exogenously added DNA. This has been demonstrated successfully with several eukaryotic expression vectors. Furthermore, G8 mouse myoblasts can express the exogenous DNA. Gene expression from these vectors was demonstrated by detection of both specific RNA transcripts from the gene(s) on Northern blots and activity of a specific gene product in transfected cell extracts.

The transfection results presented in this Chapter are essentially qualitative and give no information on transfection efficiencies although it has been established that the optimum transfection procedure that will allow subsequent differentiation is a short calcium phosphate-DNA precipitate exposure time including a glycerol shock. Since mouse myoblasts can be transfected, it is now possible to carry out similar transient transfection experiments where cloned myoglobin genes are

introduced into mouse myoblast cells in culture in an attempt to begin to study the DNA sequences that regulate myoglobin gene activation during myoblast differentiation.

CHAPTER 6

CONSTRUCTION OF RECOMBINANTS CONTAINING 5'-UPSTREAM FRAGMENTS OF THE MYOGLOBIN GENE LINKED TO THE CAT CODING SEQUENCE

6.1 Introduction

It is preferable to study the regulated expression of a gene by reintroduction into its homologous cell type since it is more likely to respond to its correct regulatory signals (see introduction to Chapter 3). However, in this case the exogenous gene must be modified in some way to distinguish it from the endogenous gene. A way to do this, and a method which has been sucessful to study gene regulation in prokaryotic systems, is to link the promoter region containing the putative control signals to a second gene segment that codes for a reporter function for example, a readily assayable enzyme normally absent from the cells in which expression analysis is to be carried out. Expression of the fused gene may then be subject to the control signals present on the regulatory region of the gene of interest and in this way, expression from the exogenous gene is clearly distinguished.

A particularly useful reporter function is the bacterial gene CAT described in Chapter 3. It was previously demonstated that CAT expression driven by virus promoters could be assayed after introduction of CAT vectors into G8 mouse myoblasts. The CAT system has been used to define the regulatory borders of a large number of eukaryotic genes (see Walker et al., 1983, Ott et al., 1984, Gorman et al., 1984, Grichnk et al., 1986) but this particular strategy to study the molecular basis for gene regulation does have some limitations. The use of the CAT system as with any other gene fusion system allows only transcriptional promoting activity in the 5'-flanking region of the gene to be investigated.

Sequences in other regions of the gene which might modulate transcriptional activity or be involved in post-transcriptional regulation would not be identified by this method. Another limitation with this system is that CAT activities may vary between transfection experiments, although the relative levels remain constant.

CAT activity has been shown to be proportional to the transcriptional strength of the promoter in appropriately constructed gene fusions (Gorman et al., 1982). Such gene fusions require that the mRNA leader sequence of the gene in question remains intact; the importance of all the necessary signals in the leader sequence for production of a stable mRNA molecule was demonstrated by Mulligan et al. (1979). Moreover, no other AUG sequences should be included in the leader sequence before the CAT start codon (Gorman et al., 1984) since it has been shown in a number of cases that inclusion of an extra AUG leads to a decrease in CAT synthesis from the gene fusion, presumably due to use of abnormal translational starts. With these points in mind, and since the regulated expression of an exogenous myoglobin gene was to be analysed after introduction into myoblasts, myoglobin-CAT fusions were constructed without the inclusion of virus enhancer sequences.

6.2 Insertion of the CAT coding sequence and a poly-linker sequence into pAT153

A two stage strategy was employed to generate a versatile cloning vector for the insertion of myoglobin promoter fragments upstream of CAT. In the first stage, the CAT gene was inserted into pAT153. The purpose of moving the CAT fragment into pAT153 was to avoid the complicated removal of the LTR promoter region from pRSV-CAT (at this time pSVO-CAT was not available). Also, pAT153 does not contain the poison sequence, present in pBR322, which inhibits gene expression in eukaryotic cells

(Lusky and Botchan, 1981).

A 1.6 kb HindIII-BamHI DNA fragment containing the CAT gene was excised from pRSV-CAT (Gorman et al., 1982). This fragment contains the entire CAT coding sequence and downstream of this the SV40 small t antigen intron and splice signals and the SV40 polyadenylation signal (Fig. 6.1). The fragment was recovered from an agarose gel using DE81 paper as described in Section 2.5. The recovered CAT fragment was introduced into the corresponding HindIII/BamHI sites of digested pAT153 by ligation. Ampicillin resistant recombinants were recovered after transformation of E.coli JM83 and plasmid DNA isolated from these resistant colonies was checked for the insertion of the CAT fragment by restriction enzyme analysis. The resulting vector was termed pAT-CAT and is depicted in Fig. 6.1.

The second stage involved the insertion of a poly-linker sequence, containing a variety of useful restriction endonuclease sites, upstream of the HindIII site in pAT-CAT. This procedure is shown diagramatically in Fig. 6.2. The particular poly-linker chosen was from pUC13 since this poly-linker contains a SmaI site which would provide a cloning site for blunt-ended DNA fragments. The poly-linker sequence would also provide several extra restriction endonuclease sites 3' and 5' of an inserted myoglobin fragment which would be useful later for creating promoter deletions at either side of the fragment. Furthermore, unlike many poly-linker sequences, the pUC13 sequence does not contain an ATG sequence.

To isolate the poly-linker, pUC13 was first restricted with EcoRI to release the 5' end of the poly-linker and the 5'-protruding termini produced by cleavage filled in with dNTPs using reverse transcriptase (Section 2.5). The blunt-ended linear plasmid DNA was next restricted with HindIII at the 3' end of the poly-linker to release the 45 bp

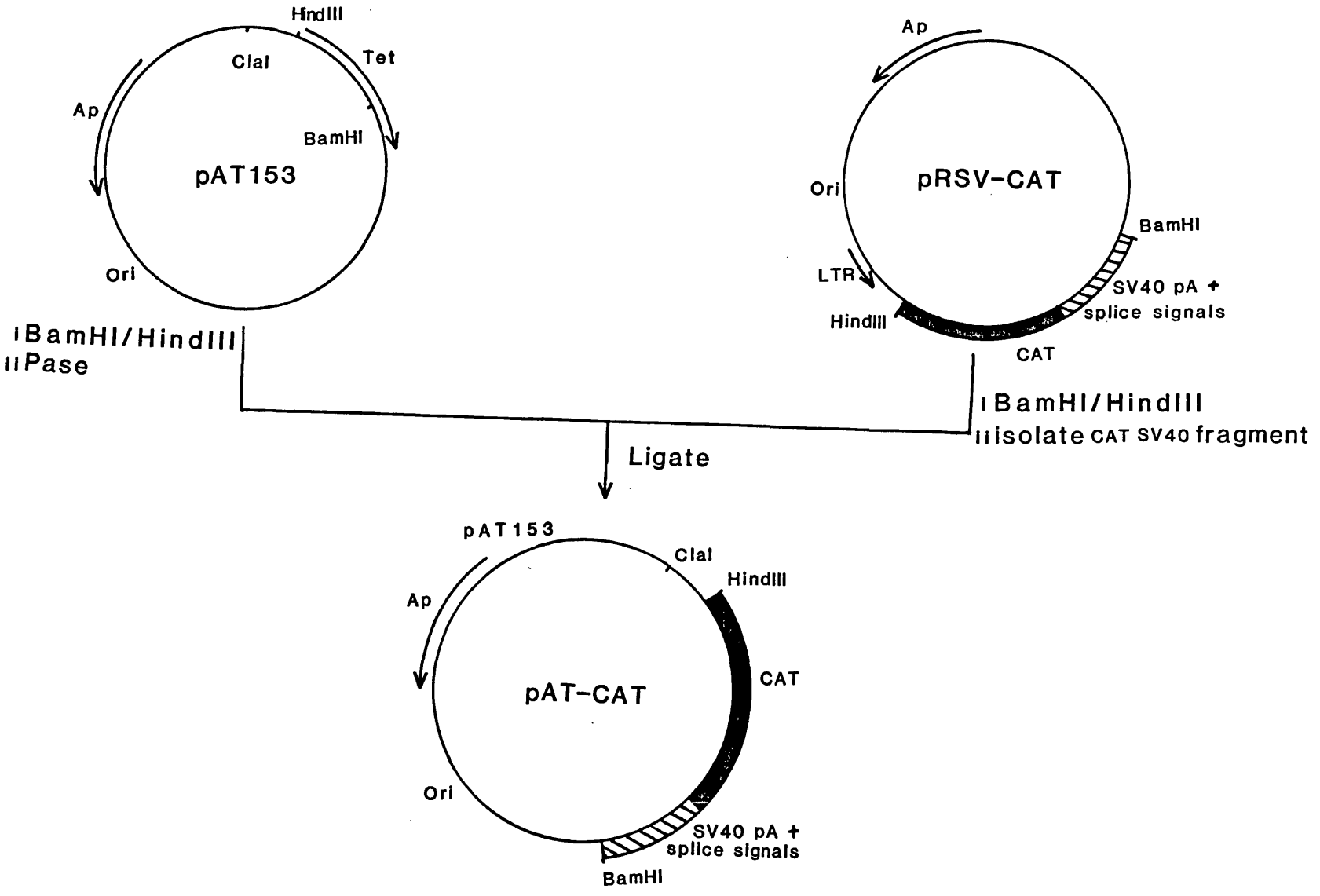
Figure 6.1

Construction of vector pAT-CAT

The 1.6 kb BamHI/HindIII CAT-SV40 DNA fragment from pRSV-CAT containing the CAT coding sequence (filled box) and SV40 poly adenylation signals (pA) and splice sites (both shaded box) was isolated as described in Section 2.5 and inserted into the corresponding sites of pAT153. Pase refers to the treatment of linear vector DNA with alkaline phosphatase to dephosphorylate DNA ends.

On the various plasmids, gene coding for ampicillin resistance (Ap) is depicted by an arrow showing the direction of transcription and the origin of bacterial replication (Ori) is shown. On pRSV-CAT, the RSV long terminal repeat promoter is shown with its direction of transcription leading into the CAT gene.

Diagrams of the plasmids in this Figure and Figures following (Figs. 6.2-6.5) are not to scale.


HindIII
ClaI
Tet
Ap
BamHI
pAT153
Ori
I BamHI/HindIII
II Pase
Ap
Ori
pRSV-CAT
BamHI
LTR
SV40 pA +
splice signals
HindIII
CAT
I BamHI/HindIII
II isolate CAT SV40 fragment
Ligate
pAT153
ClaI
HindIII
Ap
CAT
pAT-CAT
Ori
SV40 pA +
splice signals
BamHI

poly-linker fragment. This small DNA fragment was separated from the bulk of the plasmid DNA on an 8% acrylamide gel and isolated from the gel slice as described in Section 2.5. To insert the poly-linker fragment upstream of the HindIII site in pAT-CAT, pAT-CAT DNA was restricted first with ClaI and again the 5'-protruding termini filled in with dNTPs by using reverse transcriptase. Plasmid DNA was recovered by ethanol precipitation and then subjected to a second cleavage by restriction at the HindIII site. The resulting digestion products were treated with alkaline phosphatase to prevent religation of cohesive vector DNA ends in the next step. The isolated and blunt-end/HindIII poly-linker fragment was ligated to the compatible blunt-end/HindIII sites of the vector. Recombinants were recovered as ampicillin resistant colonies after transformation of E.coli JM83. Insertion of the poly-linker fragment into pAT-CAT was tested for by restriction of recombinant plasmids using restriction endonucleases for which sites were only present in poly-linker DNA sequence. The resulting vector was termed pAT-CATp.1. (Fig. 6.2).

6.3 Construction of a mouse myoglobin-CAT gene fusion

The outline of the construction of the mouse myoglobin-CAT gene fusion is illustrated diagramatically in Fig. 6.3. Restriction endonuclease sites were chosen to excise 5'-myoglobin DNA fragments of ~1 kb in size, the assumption being made that all the signals necessary for myoglobin gene expression would lie within this 1 kb region. In the three myoglobin-CAT gene fusions to be constructed, an NcoI restriction endonuclease site present within the initiation codon sequence of all three myoglobin genes could be utilized as the 3'-restriction endonuclease site. Therefore all three constructions would contain a complete 5'-mRNA leader sequence.

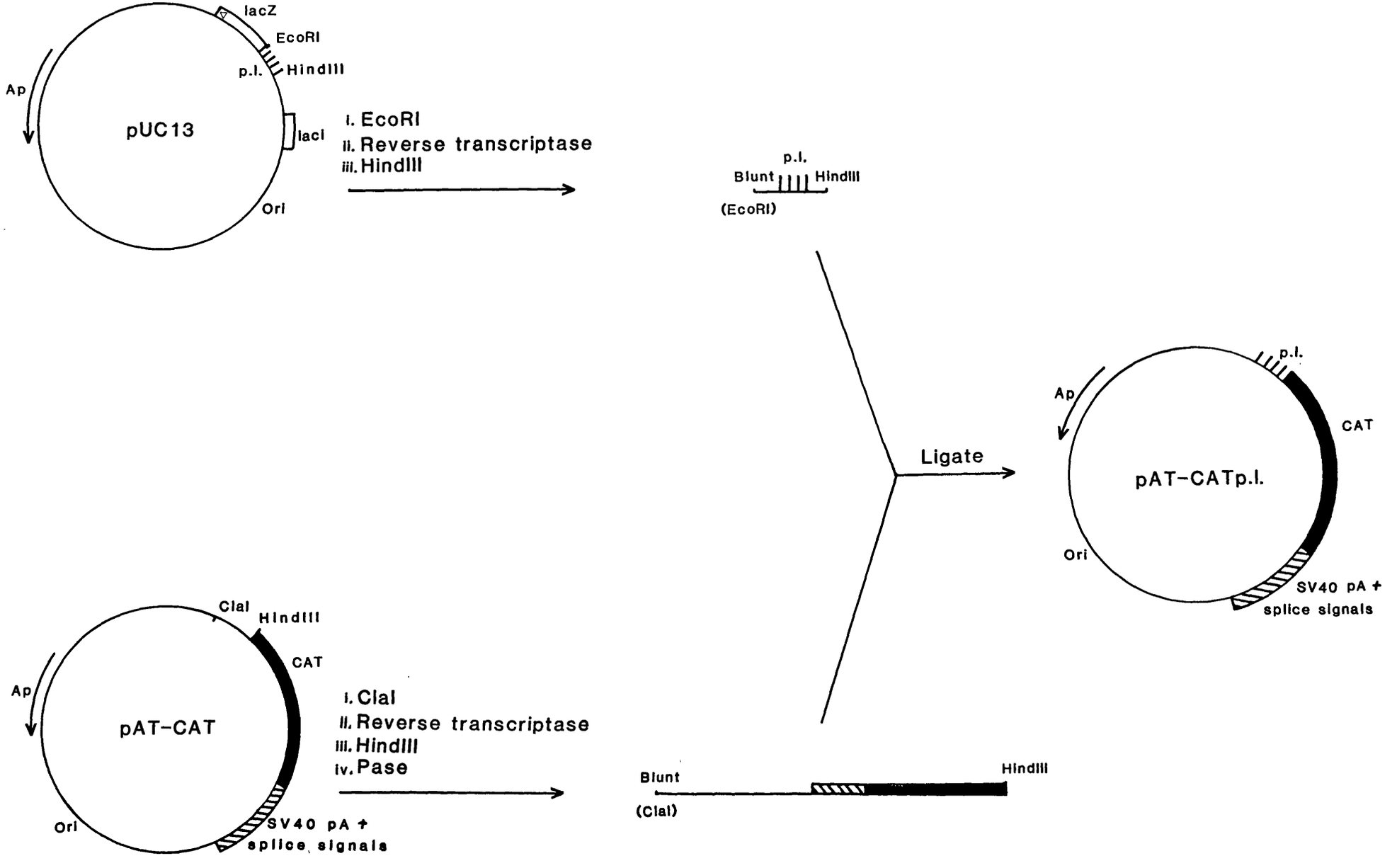
Figure 6.2

Insertion of the pUC13 poly-linker into pAT-CAT

The plasmid, pUC13 was cleaved with EcoRI, treated with reverse transcriptase to create blunt DNA ends and then cleaved with HindIII to release the 45 bp poly-linker (p.l.). The poly-linker fragment was purified as described in Section 2.5. The plasmid pAT-CAT was cleaved with ClaI, treated with reverse transcriptase to create blunt DNA ends and then cut with HindIII. The resulting vector DNA ends were dephosphorylated (Pase) and the 45 bp poly-linker was then ligated into the prepared pAT-CAT vector to produce pAT-CATp.l.

The CAT-SV40 fragment is represented as in Fig. 6.1, as are the various markers on the plasmid.

The lacZ and lacI genes of pUC13 are shown by open boxes and a small open triangle indicates the direction of transcription of the lacZ gene. Restriction endonuclease sites shown in brackets demonstrate sites which have been destroyed due to the creation of blunt DNA ends.


lacZ
EcoRI
p.l.
HindIII
Ap
pUC13
lacI
Ori
i. EcoRI
ii. Reverse transcriptase
iii. HindIII
p.l.
Blunt
HindIII
(EcoRI)
Ligate
p.l.
Ap
CAT
pAT-CATp.l.
Ori
SV40 pA +
splice signals
ClaI
HindIII
CAT
Ap
pAT-CAT
i. ClaI
ii. Reverse transcriptase
iii. HindIII
iv. Pase
Ori
SV40 pA +
splice signals
Blunt
HindIII
(ClaI)

The mouse myoglobin 5'-flanking region fragment was obtained by digestion of the EcoRI insert fragment of pλ3G8E4 (Fig. 3.1), which contains mouse myoglobin exon 1 and associated flanking sequences, with restriction endonucleases TaqI and NcoI. The mouse fragment, 1064 bp in size, contains the conserved region and the $(CT)_{26}$ sequence found in the mouse myoglobin promoter region and the intact mRNA leader sequence up to the iniation codon. The TaqI restriction enzyme site is located 1068 bp before the CAP site and therefore includes 122 bp of the 3' end of the B1 element but does not include the B1 element promoter (see Fig. 3.2). The DNA fragment was blunt-ended by brief digestion with S1 nuclease (Section 2.5). S1 nuclease was employed to create blunt DNA ends for all three myoglobin fragments due to the necessity to remove the myoglobin initiation codon sequence contained within the NcoI restriction endonuclease sequence.

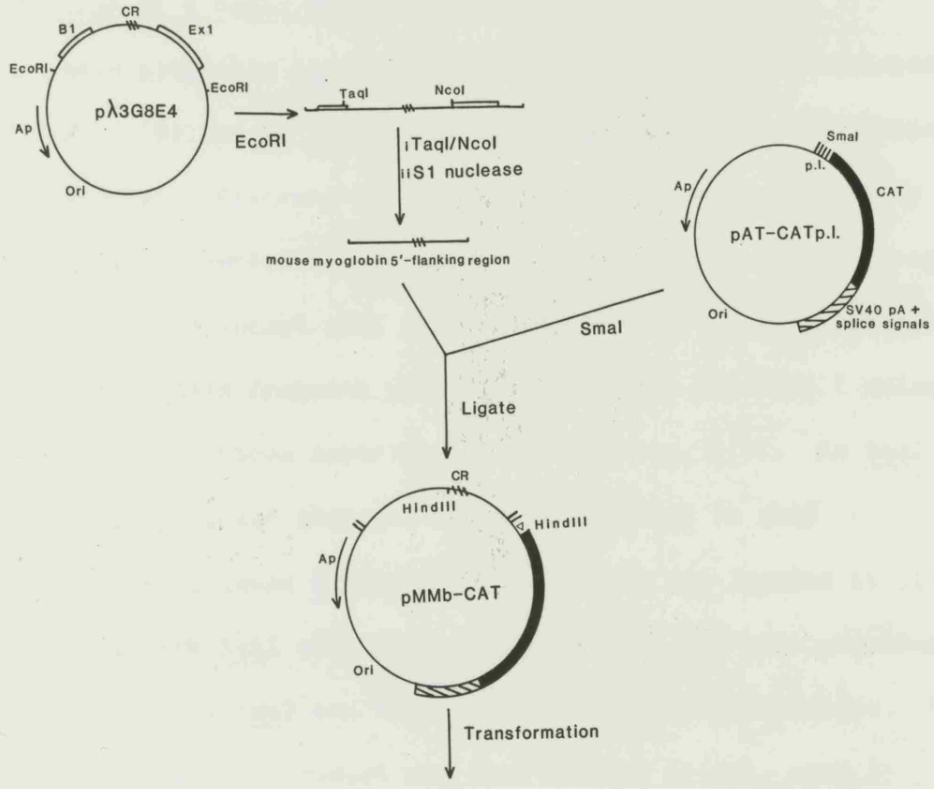
The resulting blunt-ended mouse DNA fragment was ligated into the SmaI site of digested pAT-CATp.1. Ligated recombinants were used to transform E.coli JM83. Ampicillin-resistant colonies of E.coli JM83 were purified and recombinant plasmids were checked for the insertion of the mouse 5'-flanking fragment by colony hybridisation using [^{32}-P]labelled TaqI/NcoI fragment DNA as a hybridisation probe. Plasmid DNA was isolated from colonies whose DNA hybridised to the probe. To establish the correct orientation of the insert fragment (the direction of transcription from the mouse promoter leading into the CAT coding sequence), plasmid DNAs were digested with various restriction endonucleases which would give informative digestion patterns. Recombinants containing the 5'-flanking region in the correct orientation were also restricted with NcoI to ensure that the initiation codon of the myoglobin gene had been destroyed and with various other restriction enzymes to determine the size of the insert and to discern whether

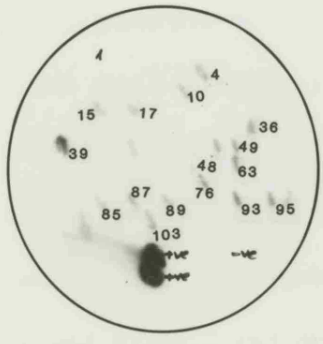
Figure 6.3

Construction of the mouse myoglobin-CAT gene fusion

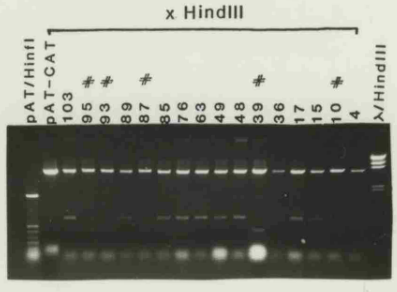
The EcoRI insert fragment of pλ3G8E4 containing myoglobin gene exon 1 and associated flanking DNA was purified and subsequently digested with TaqI and NcoI to release the promoter region up to the initiation codon and around 800 bp of 5'-flanking DNA. The TaqI/NcoI DNA fragment was treated with S1 nuclease to create blunt-ends and the resulting fragment ligated into the SmaI site of restricted pAT-CATp.1. Recombinants were transformed into E.coli JM83 and and colonies containing cloned myoglobin gene insert were identified by colony hybridisation and are shown numbered. These recombinants were digested with HindIII which cleaves the insert fragment assymetrically relative to the HindIII site of the CAT coding fragment in order to establish the orientation of the insert. Recombinants containing the myoglobin gene insert in the correct orientation are shown by hashes (#).

The conserved region (CR) is shown and the open triangle in pMMb-CAT depicts the direction of transcription of the gene fusion.





colonies containing myoglobin gene insert



insert correct orientation?

restriction enzyme sites around the fusion points were still intact (not shown). Recombinants fulfilling these criteria are depicted by hashes (##) in Fig. 6.3.

6.4 Construction of a human myoglobin-CAT gene fusion

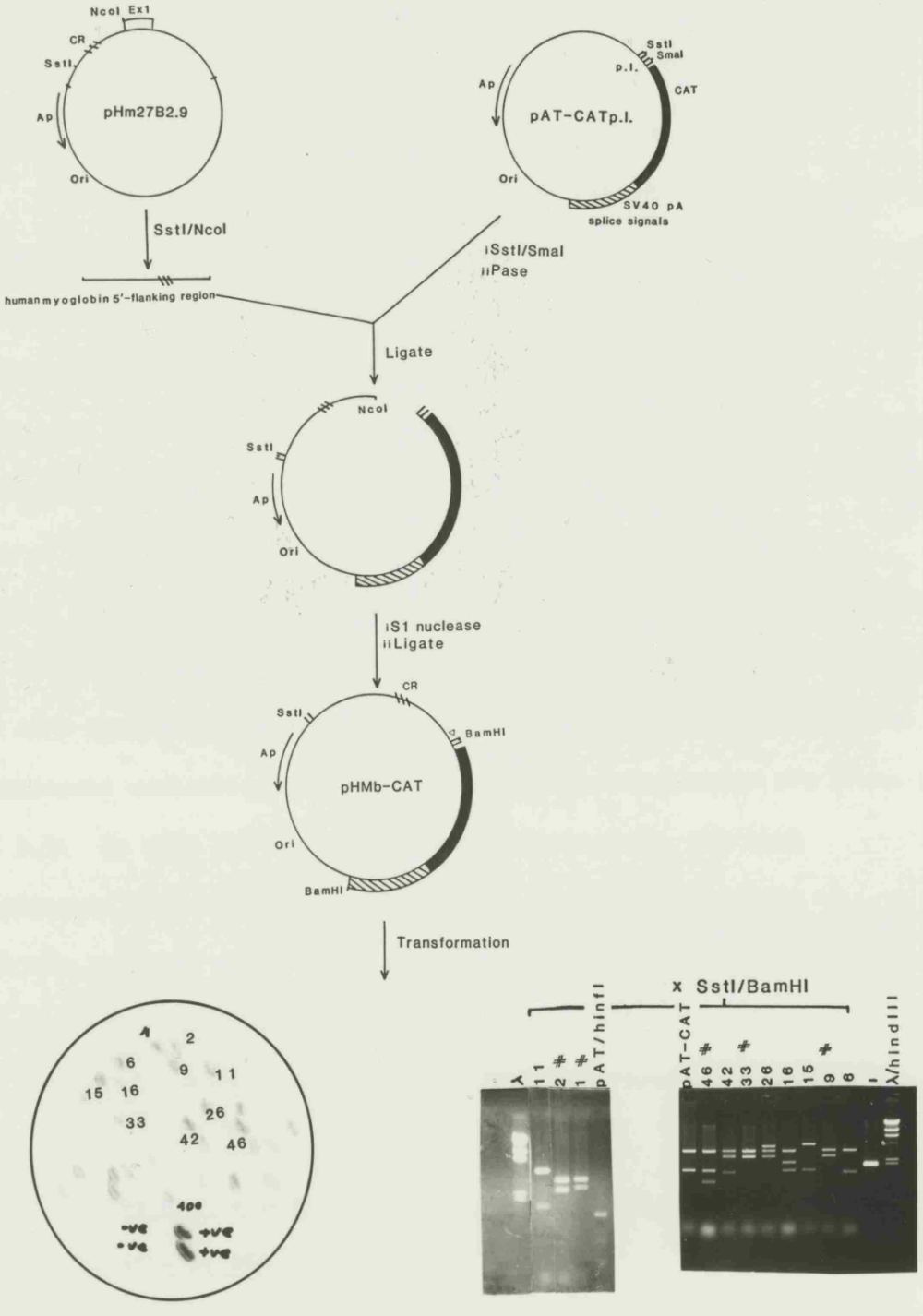
The human myoglobin gene 5'-flanking DNA fragment was isolated from pHM.27.B2.9 (Weller et al., 1984) which contains human myoglobin exon 1 and associated flanking DNA as an SstI/NcoI fragment, 1026 bp in size. The fragment carries the conserved region and the GAGA box region (see Section 1.2) and intact mRNA leader sequence to the initiation codon at the NcoI site. This fragment was introduced into pAT-CATp.1 using similar procedures to those described above (see Fig. 6.4). An SstI site present in the poly-linker sequence could be utilised in this construction. The isolated SstI/NcoI DNA fragment was ligated at its cohesive end into the SstI site of pAT-CATp.1 which had been previously digested with SstI and SmaI and treated with alkaline phosphatase. The resulting linear ligation product was then treated briefly with S1 nuclease to generate blunt-ends at the NcoI digested DNA ends. The plasmid DNA was recircularised by ligation of the blunt DNA ends. Recombinants obtained from this procedure were identified and isolated as described in Section 6.3. The human myoglobin-CAT gene fusion should have been the most straightforward construction to carry out since both fragment and plasmid vector contained an SstI site by which to orientate the insertion of the fragment into the vector. However, many of the positively hybridising colonies gave unexpected DNA fragment sizes when digested with a various of restriction endonucleases. Only five out of the twenty positively hybridising colonies which were checked appeared to contain the insert in the correct orientation (Fig. 6.4, labelled by hashes) and further restriction endonuclease digestions revealed that

Figure 6.4

Construction of the human myoglobin-CAT gene fusion

The plasmid pHM27B2.9 containing the human myoglobin gene exon 1 and associated flanking DNA was digested with NcoI and SstI to release an NcoI/SstI fragment which was purified. The SstI end of the DNA fragment was ligated to the SstI end of pAT-CATp.1., previously cut with SstI and SmaI and treated with alkaline phosphatase. The resulting linear plasmid was treated with S1 nuclease to create blunt DNA ends at the NcoI cleavage site to allow ligation to the SmaI DNA end of the vector. Recombinant plasmids were transformed into E.coli JM83 and colonies containing cloned human myoglobin gene insert were identified by colony hybridisation and are shown numbered. These recombinants were digested with SstI and BamHI to determine the size of the insert and since most recombinants had lost the BamHI site in the poly-linker, the orientation of the insert could be determined. Recombinants containing the correct size insert in the correct orientation are shown by hashes (#).

The conserved region (CR) is shown and the open triangle in pHMb-CAT depicts the direction of transcription of the gene fusion.



colonies containing myoglobin gene insert

insert correct orientation?

although the insert was about the right size in these recombinants many of the sites in the poly-linker were lost leading to some unexpected restriction digest patterns. The myoglobin gene NcoI site had been destroyed in all of the recombinants labelled with hashes.

6.5 Construction of a seal myoglobin-CAT gene fusion

The seal myoglobin 5' flanking region was isolated as an NcoI/NcoI fragment from pSM 19.5 (Blanchetot et al., 1983) which contains exon 1 and exon 2 of the seal myoglobin gene and associated flanking sequences. The fragment, 962 bp in size, contains the conserved region, GAGA box and intact mRNA leader sequence. The isolated seal myoglobin DNA fragment was treated briefly with S1 nuclease to create blunt DNA ends at the NcoI sites of the fragment and was then inserted into the SmaI site of digested pAT-CATp.1 DNA by ligation. Recombinants generated by ligation were identified and isolated as in Section 6.3 and recombinant plasmid DNAs were analysed by appropriate restriction endonuclease digestions. Recombinants containing inserts in the correct orientation are shown in Fig. 6.5. As with the mouse and human recombinants, the seal recombinants were restricted with various restriction enzymes to establish that the size of the insert was correct, that restriction sites around the fusion points remained intact and that the seal myoglobin gene NcoI site at the initiation codon had been destroyed.

6.6 Definition of the gene fusions in mouse, human and seal myoglobin-CAT recombinants by DNA sequencing

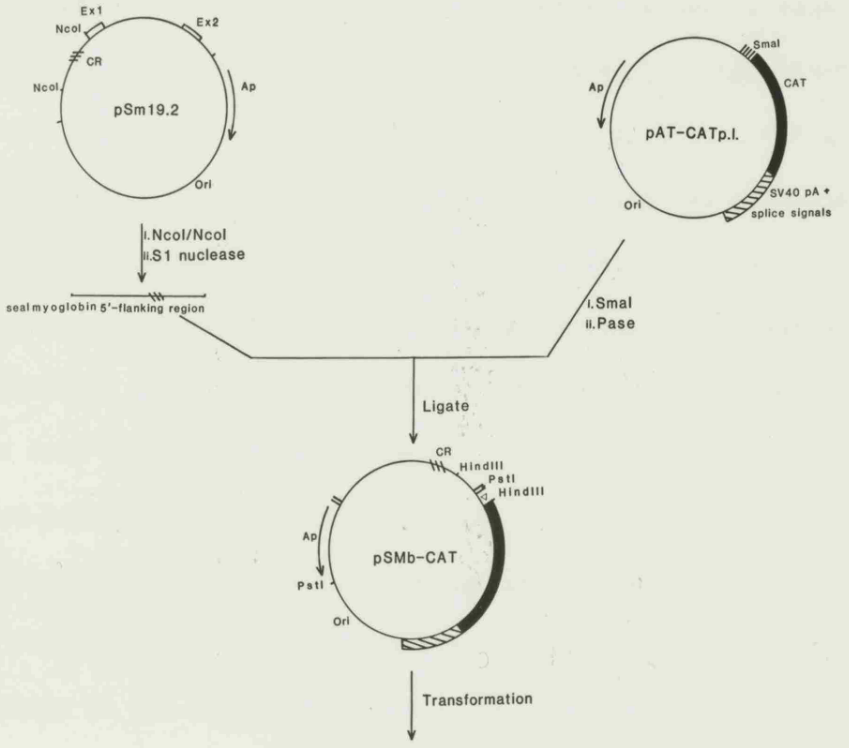
It is possible that during the various sub-cloning procedures small deletions or minor rearrangements which might have disrupted important control signals had occurred or that an ATG sequence had been inadvertently created. Such sequence changes would probably not be

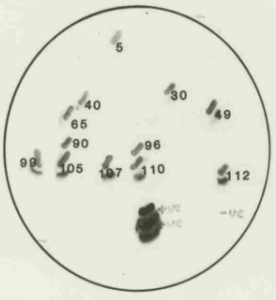
Figure 6.5

Construction of the seal myoglobin-CAT gene fusion

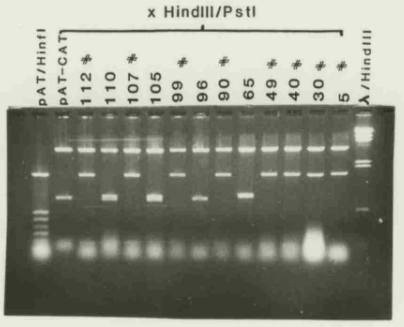
The plasmid pSM19.5 (Blanchetot et al., 1983), containing seal exon 1 and 2 and associated flanking DNA was digested with NcoI to release an NcoI/NcoI fragment which was purified and treated with S1 nuclease to create blunt DNA ends. The prepared fragment was ligated to the vector pAT-CATp.1. previously cut with SmaI and treated with alkaline phosphatase. Recombinant plasmids were transformed into E.coli JM83 and colonies containing the cloned seal myoglobin gene insert were identified by colony hybridisation and are shown numbered. These recombinants were digested with HindIII and PstI. HindIII cleaves the fragment assymetrically relative to the PstI sites of the vector. Recombinants containing the myoglobin gene insert in the correct orientation are depicted by hashes (#).

The conserved region (CR) is shown and the open triangle in pSMb-CAT depicts the direction of transcription of the gene fusion.





colonies containing myoglobin gene insert



insert correct orientation?

detected by the restriction enzyme analysis and therefore to determine precisely the sequence around the gene fusion point, a few myoglobin-CAT gene fusions were sequenced over the region where the myoglobin 5'-flanking region was joined to the CAT coding region fragment.

In Fig. 6.6 the relevant region of DNA sequenced in each recombinant is illustrated. For the human and mouse myoglobin-CAT recombinants, a PvuII fragment beginning at the myoglobin CAP site and extending about half-way into the CAT coding region was sequenced. For the seal recombinants, an RsaI fragment from 693 bp before the myoglobin CAP site to a third of the way into the CAT coding region was sequenced. Fragments to be sequenced were isolated from digests of plasmid DNA and inserted into the SmaI site of M13mp19. Resulting M13 recombinants were sequenced using the dideoxy chain termination method of Sanger (1980) (Fig. 6.5). Due to the random insertion of blunt-ended DNA fragments, most of the recombinants were sequenced on both strands. In all cases, the sequence at the fusion joint could be determined and the integrity of the TATA, CAP, 5'-untranslated region of the myoglobin gene and initiation codon sequence of the CAT gene could be deduced. Sequencing revealed that the insertion of human, mouse and seal 5'-flanking DNA fragments into the poly-linker sequence upstream of the CAT coding sequence in pAT-CATp.1. had occurred as predicted and in all cases without any changes or with changes in the DNA sequences being joined which did not interrupt known control sequences.

6.7 Summary

Expression vectors have been constructed in which the reporter gene, bacterial CAT, has been placed under the transcriptional control of the myoglobin promoter region of seal, human or mouse myoglobin genes. The hybrid CAT constructs were first examined by restriction endonuclease

Figure 6.6

DNA sequence over the fusion joints of pMMb-CAT, pSMb-CAT and pHMb-CAT recombinants

pMMb-CAT

The expected nucleotide sequence (E) over the fusion joint is shown in uppercase letters. The observed sequences of two pMMb-CAT recombinants (10 and 39) are shown in lower case letters.

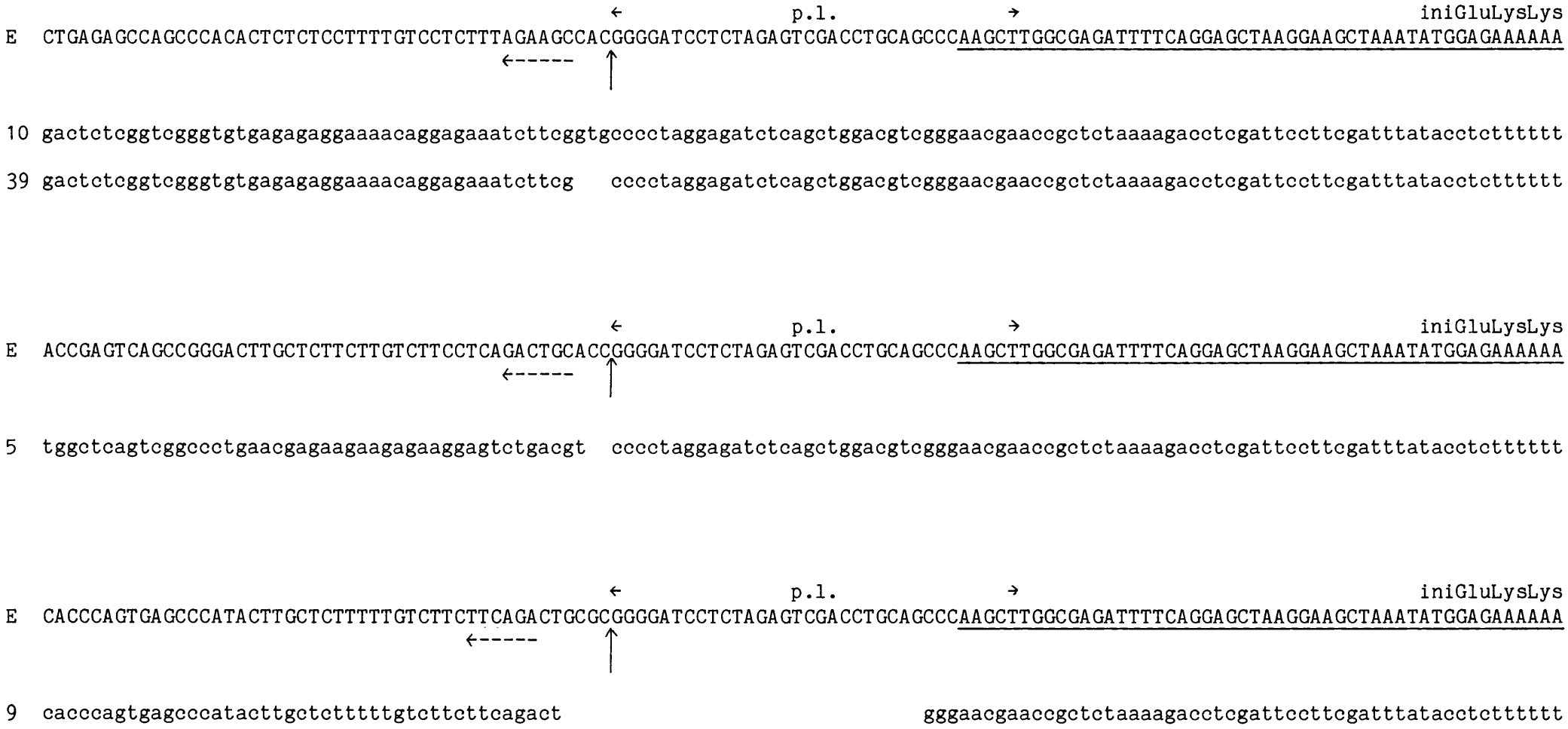
pSMb-CAT

The expected sequence (E, uppercase letters) and the observed sequence of a pSMb-CAT recombinant (5) are shown. The fragment sequenced in this case extended about 500 bp into the seal 5'-flanking DNA and all upstream sequences were demonstrated to be intact (not shown).

pHMb-CAT

The expected sequence (E, uppercase letters) and the observed sequence of a pHMb-CAT recombinant (9) are shown.

Sequences of recombinants are shown for the antisense DNA strand. All sequences were determined on both strands (except 39) and were sequenced at least twice. Gaps in the recombinant sequences indicate deleted nucleotides. On the expected sequences the CAT fragment is underlined, with the initiation codon (ini) and the first three amino acids shown and the poly-linker (p.l.) sequence is bordered by two small arrows. The broken horizontal arrow indicates sequences in the 5'-nontranslated region of the myoglobin genes and the vertical arrow indicates fusion at the NcoI and SmaI sites.


← p.l. → iniGluLysLys
E CTGAGAGCCAGCCCACACTCTCTCCTTTTGTCCTCTTTAGAAGCCACGGGGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAATATGGAGAAAAAA
←-----
10 gactctcggtcgggtgtgagagaggaaaacaggagaaatcttcggtgcccctaggagatctcagctggacgtcgggaacgaaccgctctaaaagacctcgattccttcgatttatacctcttttt
39 gactctcggtcgggtgtgagagaggaaaacaggagaaatcttcg cccctaggagatctcagctggacgtcgggaacgaaccgctctaaaagacctcgattccttcgatttatacctcttttt
← p.l. → iniGluLysLys
E ACCGAGTCAGCCGGGACTTGCTCTTCTTGTCTTCCTCAGACTGCACCGGGGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAATATGGAGAAAAAA
←-----
5 tggctcagtcggccctgaacgagaagaagagaaggagtctgacgt cccctaggagatctcagctggacgtcgggaacgaaccgctctaaaagacctcgattccttcgatttatacctcttttt
← p.l. → iniGluLysLys
E CACCCAGTGAGCCCATACTTGCTCTTTTTGTCTTCTTCAGACTGCGCGGGGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAATATGGAGAAAAAA
←-----
9 cacccagtgagcccatacttgctcttttgtcttcttcagact gggaacgaaccgctctaaaagacctcgattccttcgatttatacctcttttt

mapping and then by nucleic acid sequencing to define the gene fusion junctions. The construction strategy was successful in that no major rearrangements, disturbing control sequences, had occurred during the procedure.

The myoglobin-CAT gene fusions could now be tested to determine whether the CAT gene is expressed after introduction of the gene fusion into G8 mouse myoblasts, as the next step towards the identification of control sequences in the 5'-flanking region of myoglobin genes.

CHAPTER 7

THE EXPRESSION OF MYOGLOBIN-CAT GENE FUSIONS IN DIFFERENTIATING MYOBLASTS IN VITRO

7.1 Introduction

Having established that mouse myoblasts can take up and express exogenous DNA in transient expression experiments (Chapter 5), the myoglobin promoter-CAT gene fusions described in Chapter 6 were tested in similar transient expression experiments. There were three major questions to address in these experiments to discern whether the myoglobin gene expression could be induced under the special conditions of a transient expression assay; i) does the myoglobin promoter drive transcription when introduced into myoblasts? ii)is the transcription of the CAT gene fusion enhanced during myoblast differentiation? and iii)is the expression of the gene fusion cell-specific?

When the work on transfections of myoglobin-CAT gene fusions into myoblasts started, there was only one report of successful introduction of a muscle specific gene into myoblasts in culture and in this particular case expression from the exogenous gene was monitored in stable transformants which had integrated the muscle gene fusion into genomic DNA (Melloul *et al.*, 1984). The results of this work concluded that the information contained within the promoter region of the muscle-specific gene was sufficient to confer tissue and stage-specific expression to the gene.

I preferred to carry out a similar study on the *in vitro* expression of mammalian myoglobin genes using transient transfection experiments since these transfections may be carried out more rapidly when several promoters are to be tested at once. Another advantage is

that in transient transfections introduced DNA remains extra-chromosomal in which case the only cis-acting effects monitored are those of the construct and are not due to the site of integration of the DNA into the host chromosome. If expression does occur from any gene fusion copies which might have become integrated, since a heterogenous pool of cells are analysed in transient expression experiments, position effects would be averaged out.

Since the reported work of Melloul et al. (1984), a number of other laboratories have introduced various muscle specific genes into myoblasts in culture (Nudel et al., 1985, Konieczny and Emerson, 1985, Grichnk et al., 1986, Minty and Kedes, 1986, Jaynes et al., 1986, Bergsma et al., 1986) and most have adopted the transient transfection procedure with some success. The majority of transfections have been carried out with contractile protein genes. This chapter describes transfection of mouse myoblasts with a non-contractile muscle-specific gene fusion in an attempt to mimic the induction of expression of the endogenous gene in mouse myoblasts during myogenesis in vitro.

7.2 Can myoblast differentiation be induced over a transient expression period?

It became apparent in the early transfection experiments that the addition of the calcium phosphate-DNA precipitate to myoblasts resulted in a delay of differentiation to form myotubes of at least 24 hours (Chapter 5). A similar delay in differentiation of ~24 hours has also been noted after transfection of another myogenic cell line, L8 (Rebecca Aft and David Yaffe, personal communication), but this is not true of all myogenic cell lines (see Grichnk et al., 1986 for example).

This delay in differentiation posed the immediate problem of whether the myoglobin-CAT gene fusion could be expressed within a

transient expression period since the level of gene expression reaches a maximum at around 24 to 48 hours in a transient expression period and then falls off sharply (see Shen et al., 1982). The delay in differentiation should not prevent the expression of the myoglobin gene within a transient expression period with clone 1 myoblasts since myoglobin transcripts should still appear at around ~ 40-50 hours in fusion medium. However, it might be difficult to maintain parallel cultures of pre-fusion myoblasts for this period of time in order to analyse gene expression in pure populations of myoblasts and myotubes.

Transfections were carried out using the conditions determined in Chapter 6; myoblasts were exposed to calcium phosphate-DNA precipitates for 6 hours after which time they were subjected to a glycerol shock. After the glycerol shock, myoblasts were re-fed with proliferative medium and left overnight after which time they were re-fed with low serum fusion promoting medium and left for a further 40-50 hours.

The following experiments describe various transfections into mouse myoblasts in an attempt to acheive reproducible expression from transferred myoglobin-CAT gene fusions during differentiation within a transient expression period.

7.3 Introduction of myoglobin-CAT constructs into mouse myoblasts and assay for CAT activity during differentiation over a transient expression period

Transfection was carried out in the first instance with the mouse myoglobin-CAT gene fusion since it was expected that a transferred mouse myoglobin promoter would be more likely to function normally in its homologous cell-type.

Each dish of $1x10^6$ cells were transfected with 10μg of the mouse myoglobin-CAT fusion plasmid, pMMb-CAT. Control transfections were

carried out alongside with pSV0-CAT and carrier DNA only. After the glycerol shock, myoblasts were maintained in DMEM + 10% FCS overnight (17 hours) before one of the two dishes of myoblasts transfected with pMMb-CAT were refed with low serum, fusion-stimulating medium. The cultures were left in fusion medium for approximately 30 hours. During this time, the proliferative growth medium was changed on the remaining myoblasts once. After the 30 hour fusion period, the total period of time the cells had been expressing the exogenously added DNA was ~48 hours and myoblasts were very confluent -therefore cells were assayed at this time. Cells were harvested and freeze-thaw extracts made from cell pellets. Extracts were assayed for CAT activity and results of these assays are shown in Fig.7.1. There is a very low level of CAT activity in extracts from myoblasts transfected with pMMb-CAT, slightly above the background observed in mock-transfected cell extracts. This is true for both pre-fusion and fusing cultures and levels are approximately the same in both extracts. Myotubes were not apparent in the fusing cultures in this transfection experiment but myoblasts had begun to line-up.

The transfection experiment was repeated and the myoblasts transfected with pMMb-CAT were allowed to differentiate in low serum medium for a further 24 hours (total fusion time ~50 hours). A small amount of activity was observed in fusing cell extracts (Fig. 7.1). Myoblasts which were not stimulated to differentiate were harvested at the same time point as in the previous experiment (48 hours) and therefore there was a 24 hour difference between assay of the samples. The radioactive spots containing unmodified chloramphenicol and the acetylated derivatives were cut out from the TLC plate and counted using a scintillation counter, to give some indication of the relative increase in CAT activities. The additional fusion time resulted in a small amount of CAT expression in fusing cultures, 4.7 fold higher than in myoblast

Figure 7.1

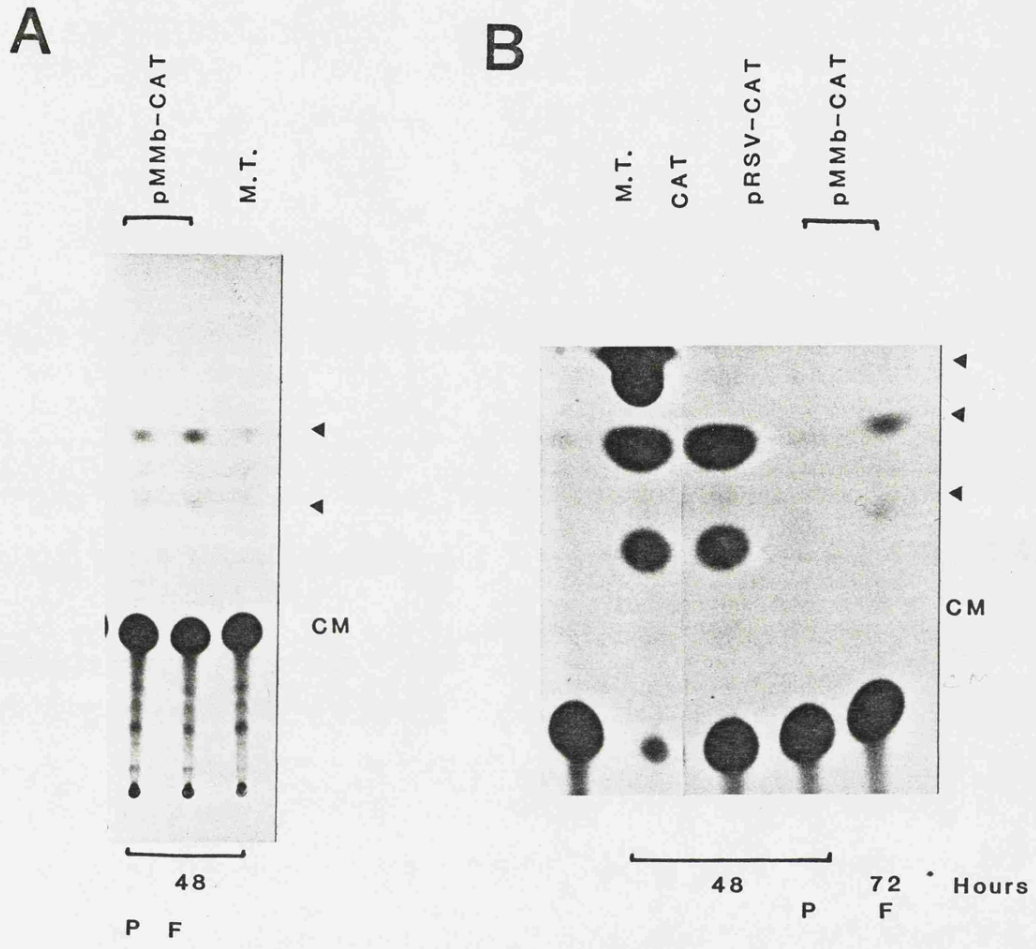
Analysis of CAT enzymatic activity in clone 1 myoblasts transfected with pMMb-CAT

Myoblasts (1x 10^6 cells), were transfected with pMMb-CAT as described in Section 2.14. About 15 hours after addition of the $CaPO_4$-DNA precipitate, one dish of myoblasts were switched to low serum medium to induce myoblast fusion. Cell extracts (50µl) were assayed for CAT activity overnight as described in Section 2.15. and using 1 µCi of ^{14}C-chloramphenicol.

A CAT assays from pre-fusion (P) and fusing (F) myoblasts transfected with pMMb-CAT. Time in hours after addition of the DNA precipitate is shown. Assay of CAT activity in mock-transfected myoblast cell extracts (M.T.) is shown as a control. Autoradiography was for 3 days.

B CAT assays from pre-fusion (P) and fusing (F) myoblasts transfected with pMMb-CAT. CAT assays carried out on pRSV-CAT transfected myoblasts (25µl of cell extract) and carrier DNA (M.T.) are shown as controls. Time in hours after addition of the DNA precipitate is shown. Assay of 0.6µg of pure CAT protein (CAT) was used as a control for assay conditions in all CAT assays performed and is shown here as an example. Autoradiography was for 3 days.

Radioactive spots containing chloramphenicol derivatives were cut out from the TLC plate and counted using a scintillation counter. The increase in counts in acetylated chloramphenicol products per 50µl of cell extract from the 48 hour sample to the 72 hour sample is 4.7 fold.


A
pMMb-CAT
M.T.
CM
48
P F
B
M.T.
CAT
pRSV-CAT
pMMb-CAT
CM
48
72
Hours
P
F

cultures. pMMb-CAT is therefore expressed in myotube cultures but at low levels compared to the expression from RSV-CAT in myoblasts.

7.4 Transfection of myoglobin-CAT constructs and assay of CAT activity over a time course

In order to determine whether the level of CAT expression could be improved, a number of changes were made in the transient expression experiment. Firstly, the seal myoglobin-CAT gene fusion was introduced into myoblasts. Seal skeletal muscle contains eight times more myoglobin mRNA compared to mouse (see Section 1.6) which raises the possibility that the seal myoglobin promoter might be more efficient than the mouse homologue. Secondly, myoblasts were left to differentiate over a longer time period. Transfected clone 1 myoblasts were assayed at a pre-fusion time point of 24 hours after addition of the DNA precipitate. Further samples were assayed at time points representing fusing (48 hours after addition of the DNA precipitate, 28 hours in fusion medium) and post-fusion (72 and 96 hours after addition of the precipitate, 52 and 76 hours in fusion medium) cultures. Similar transfections were carried out with G8-1 myoblasts to determine whether clone 1 and G8-1 myoblasts similarly expressed myoglobin-gene fusions. G8-1 myoblasts were assayed over an extended fusion time period (pre-fusion, 48 hours, post-fusion, 120, 140 hours) to be at an equivalent stage of differentiation as clone 1 myoblasts.

Myoblasts transfected with pRSV-CAT were used as a control in the transfection experiments to check for artifactual increases in transcription due to cell proliferation or transcriptional competence of a non-muscle promoter during myoblast differentiation. Fig. 7.2 shows the result of the CAT assays carried out on the transfected G8-1 ond clone 1 myoblast cell extracts. In differentiating clone 1 myoblasts

Figure 7.2

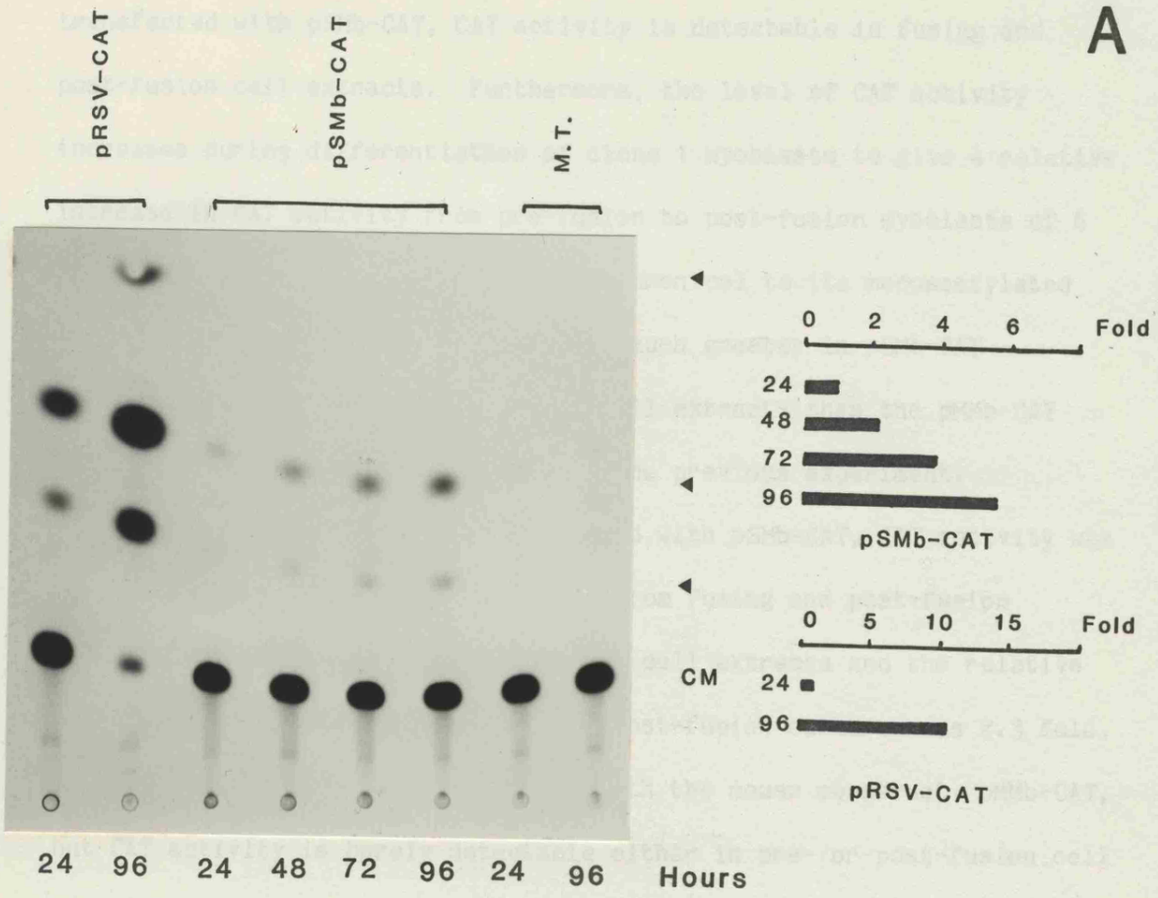
Analysis of CAT enzymatic activity in clone 1 and G8-1 myoblasts transfected with pSMb-CAT and pMMb-CAT over an extended time period

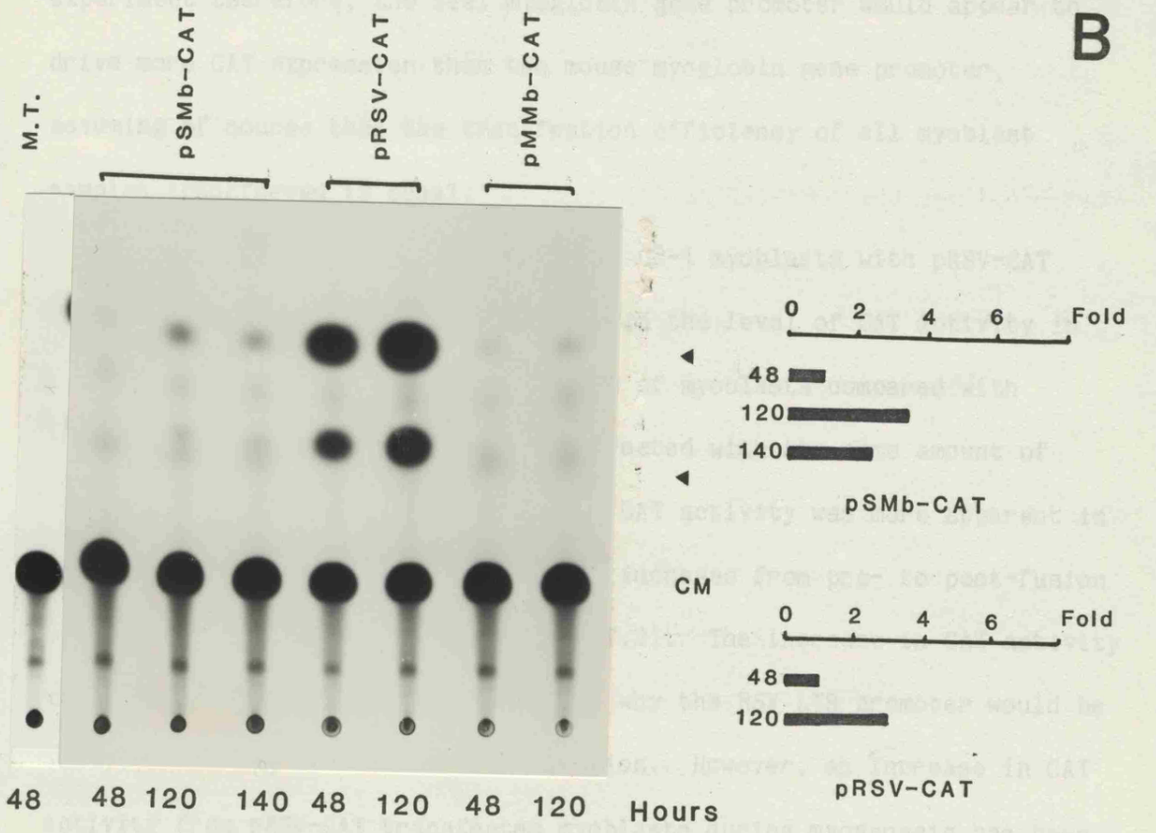
A Clone 1 myoblasts were transfected with pSMb-CAT. CAT assays of the transfected cell extracts is shown. Assays were carried out for 4 hours using 50µl of cell extract (containing 500-550µg of protein) from myoblasts transfected with pSMb-CAT or mock-transfected (M.T.) and 25µl of extract (containing ~300µg of protein) from myoblasts transfected with pRSV-CAT. The time in hours after addition of the DNA precipitate is shown. Low serum medium was added to the 48-96 hour samples ~15 hours after the addition of the DNA precipitate. Autoradiography was for three days. The graph alongside the TLC plate shows CAT activities: radioactive spots were cut out and counted in a scintillation counter to give total counts converted to acetylated products from the 48, 72 and 96 hour samples divided by the activity expressed at 24 hours (pre-fusion level); this gives the fold increase in activity.

B G8-1 myoblasts were transfected with pSMb-CAT, pMMb-CAT and pRSV-CAT and carrier DNA (M.T.) as controls. CAT assays were carried out using 50µl of extract (containing ~400-450µg of protein) from cells transfected with pSMb-CAT, pMMb-CAT and mock-transfected myoblasts (M.T.) and 25µl (~200µg of protein) from cells transfected with pRSV-CAT. Autoradiography was for three days.

The graph alongside the TLC plate shows CAT activity (as in A) from the 120 and 140 hour samples divided by the activity expressed at 48 hours, again representing the fold increase in activity.

The migration positions of chloramphenicol (CM) and the monoacetylated and diacetylated derivatives (arrow heads) are shown. 1µCi of ^{14}C-chloramphenicol was used.


A
pRSV-CAT
pSMb-CAT
M.T.
24 96 24 48 72 96 24 96
Hours
0 2 4 6 Fold
24
48
72
96
pSMb-CAT
0 5 10 15 Fold
CM
24
96
pRSV-CAT


B
M.T.
pSMb-CAT
pRSV-CAT
pMMb-CAT
48 48 120 140 48 120 48 120
Hours
0 2 4 6 Fold
48
120
140
pSMb-CAT
CM
0 2 4 6 Fold
48
120
pRSV-CAT

transfected with pSMb-CAT, CAT activity is detectable in fusing and post-fusion cell extracts. Furthermore, the level of CAT activity increases during differentiation of clone 1 myoblasts to give a relative increase in CAT activity from pre-fusion to post-fusion myoblasts of 6 fold (Fig. 7.2). Conversion of chloramphenicol to its monoacetylated derivatives does not appear to be very much greater in pSMb-CAT transfected differentiating myoblast cell extracts than the pMMb-CAT transfected myoblast cell extracts of the previous experiment.

In G8 mouse myoblasts transfected with pSMb-CAT, CAT activity was similarly detectable in cell extracts from fusing and post-fusion cultures but not in pre-fusion myoblast cell extracts and the relative increase in CAT activity from pre- to post-fusion cultures was 2.3 fold. G8-1 myoblasts were also transfected with the mouse construct, pMMb-CAT, but CAT activity is barely detectable either in pre- or post-fusion cell extracts of these transfected myoblasts (Fig. 7.2). In this particular experiment therefore, the seal myoglobin gene promoter would appear to drive more CAT expression than the mouse myoglobin gene promoter, assuming of course that the transfection efficiency of all myoblast samples transfected is equal.

Transfection of both clone 1 and G8-1 myoblasts with pRSV-CAT reveals that there is also an increase in the level of CAT activity in cell extracts from post-fusion cultures of myoblasts compared with pre-fusion cultures of myoblasts transfected with the same amount of pRSV-CAT plasmid DNA. The increase in CAT activity was more apparent in Clone 1 cell extracts where a relative increase from pre- to post-fusion cultures of 11 fold is observed (Fig. 7.2). The increase in CAT activity was surprising since it is not obvious why the RSV LTR promoter would be inducible during myoblast differentiation. However, an increase in CAT activity from pRSV-CAT transfected myoblasts during myogenesis has been

noted by others using different myogenic cell lines (W.E. Wright, personal communication). Also, Melloul et al. (1984) found a much wider range of increase in CAT activities following myoblast differentiation in their pSV2-CAT containing clones than in their β-actin-CAT containing clones. Felsani et al. (1985) have demonstrated that the adenovirus E1a enhancer is active in myotubes but not in myoblasts. Since a number of virus promoters are observed to become activated during differentiation of teratocarcinoma cells (see Gorman, Rigby and Lane, 1985 for example) it might also be true that at "later" stages of cell differentiation virus promoters are more competent to be expressed and perhaps this effect is being observed during myoblast differentiation.

7.5 Introduction of a myoglobin-CAT construct into mouse Ltk$^-$ fibroblasts

To establish whether the CAT activity observed in differentiating myoblast cell extracts is specific to this cell type, pSMb-CAT was used to transfect Ltk$^-$ mouse fibroblasts in parallel with the above transient expression time course carried out with myoblasts. The transfection of mouse fibroblasts failed to reveal any CAT activity, above background, in cell extracts prepared from fibroblasts which had been allowed to express the exogenously added DNA for 24 to 72 hours (Fig. 7.3). As expected, a high level of CAT activity was observed in fibroblast cell extracts from cells transfected with pRSV-CAT similar to the levels observed in myoblast cell extracts. Since no increase in CAT activity is observed over the transient expression time period in fibroblasts transfected with pRSV-CAT, this would tend to suggest that increase in CAT activity in differentiating pRSV-CAT transfected myoblasts is not due to cell proliferation since fibroblasts were also proliferating during the transient expression period.

It therefore appears that the CAT activity and the increase in CAT

Figure 7.3

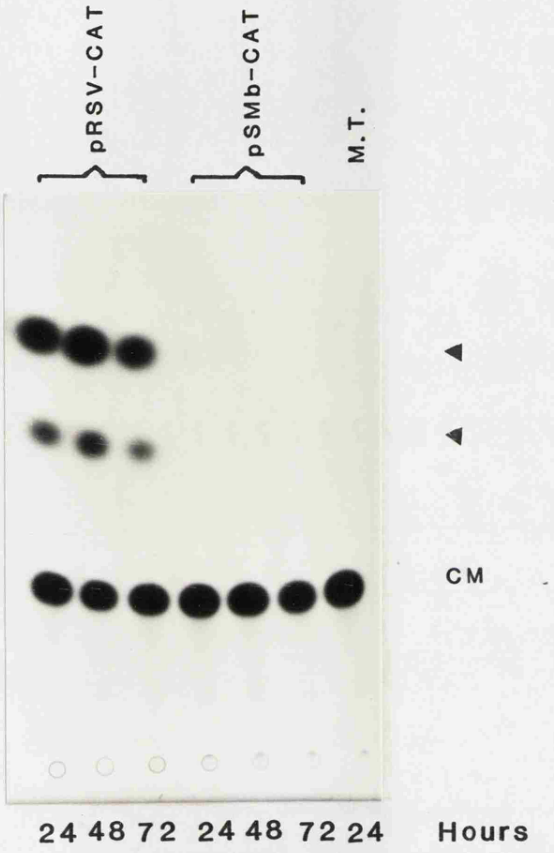
Analysis of CAT enzymatic activity in Ltk$^-$ fibroblasts transfected with pSMb-CAT

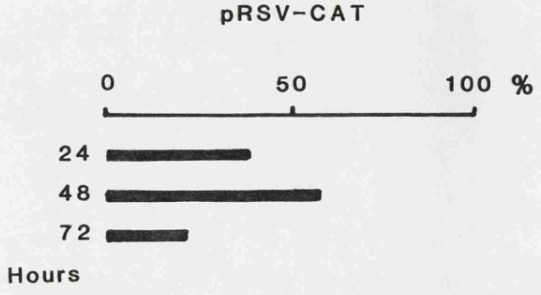
A Ltk$^-$ fibroblasts were transfected with pSMb-CAT in parallel with the transfections shown in Fig. 7.2.

Fibroblasts were also transfected with pRSV-CAT and mock-transfected (M.T.) as controls. CAT assays were carried out for four hours with 15µl (containing 200-250µg protein pRSV-CAT transfected and M.T. myoblast cell extracts) and 50µl (containing 550-600µg of protein, pSMb-CAT transfected myoblast cell extracts). Time in hours after addition of the DNA precipitate is shown. Autoradiography was for 3 days.

B The graph represents the percentage of acetylation of chloramphenicol to monoacetylated derivatives for each of the pRSV-CAT transfected myoblast cell extracts shown in A. Radioactive spots were cut out from the TLC plate and counted.

Extracts from fibroblasts transfected with pSMb-CAT have background levels of CAT activity.


pRSV-CAT
pSMb-CAT
M.T.
CM
24 48 72 24 48 72 24
Hours


pRSV-CAT
0
50
100 %
24
48
72
Hours

activity over a time period observed in fusing cultures of mouse myoblasts which have been transfected with myoglobin-CAT constructs, although low, is specific to cultures of differentiating myoblasts.

7.6 Transfection of mouse myoblasts with a promoterless construct

In order to determine whether the expression from the myoglobin-CAT fusions was dependant upon the presence of a myoglobin promoter in the construct, transfections of myoblasts were carried out using the original CAT containing vector, pAT-CATp.1., into which myoglobin promoters were subsequently inserted (Chapter 6). An equal concentration (10μg per 10 cm dish) of pAT-CATp.1. plasmid DNA was used to transfect clone 1 myoblasts. The same number of clone 1 myoblasts were transfected with pSMb-CAT, pRSV-CAT and carrier DNA only. After transfection myoblasts were stimulated to differentiate over a transient expression period of 48 to 96 hours after the addition of the DNA precipitate, the pre-fusion myoblast cell extract being assayed 24 hours after the addition of DNA. Fig. 7.4 illustrates the results of the various CAT assays. CAT activity is detectable in fusing (48 hours after addition of the DNA precipitate) and post-fusion (72 and 96 hours after the addition of the DNA precipitate) myoblast cell extracts from myoblasts transfected with pSMb-CAT but not in the pre-fusion cell extracts of the same transfected samples. CAT activity from pAT-CATp.1. transfected myoblasts is barely detectable in either pre-, fusing or post-fusion cell extracts, at 6% of that observed in post-fusion myoblast cell extracts from pSMb-CAT transfected myoblasts. Expression from the hybrid-CAT gene therefore requires the presence of the myoglobin promoter.

The activity of the endogenous myoglobin gene was assessed in this transfection experiment over the differentiation time period.

Figure 7.4

Analysis of CAT enzymatic activity in myoblasts transfected with pAT-CATp.1. and appearance of endogenous myoglobin gene transcripts

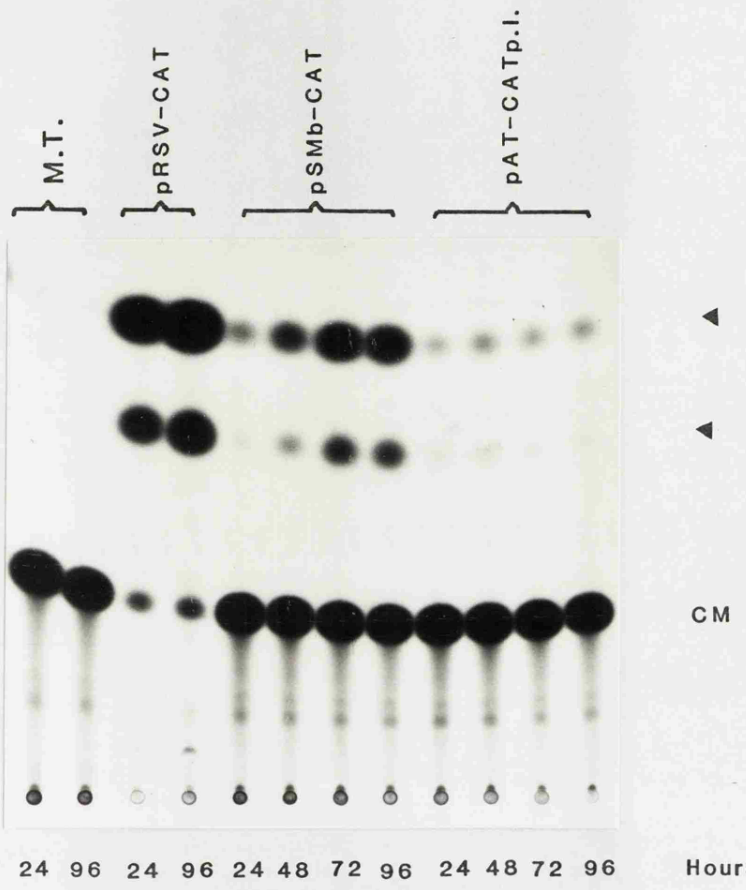
A Clone 1 myoblasts were transfected with pAT-CATp.1. which does not contain a eukaryotic promoter as described in Section 2.14. pSMb-CAT, pRSV-CAT and carrier DNA (M.T.) were used as controls. In all transfections, plasmid DNA had been subjected to two cycles of caesium chloride density centrifugation. Time in hours after addition of the DNA precipitate is shown and low serum medium was added to the 48-96 hour myoblast samples ~15 hours after the addition of the DNA precipitate. CAT assays were carried out for four hours using 40µl (containing ~400 -500 µg of protein) of cell extract for each sample. The migration positions of chloramphenicol (CM) and the monoacetylated products (arrow heads) on the TLC plate are shown.

B Graph showing the percentage of acetylation of chloramphenicol in cell extracts from pSMb-CAT and pRSV-CAT transfected myoblasts, determined by scintillation counting of the radioactive spots.

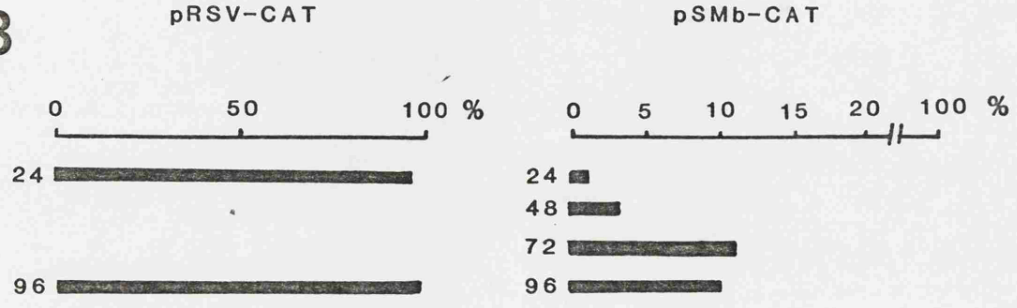
pAT-CATp.1. cell extracts converted only background levels of chloramphenicol (≤ 0.3%)

C Appearance of myoglobin gene transcripts in RNA extracted from ~ one fifth of the pSMb-CAT transfected myoblast cell extracts was analysed by Northern blotting using a human myoglobin gene exon 2 hybridisation probe. Time in hours after the addition of the DNA precipitate is shown. (The lanes are taken from the same blot as that shown in Fig. 4.5 but hybridisation was with the human exon 2 probe and the extra myoblast RNA sample is shown).

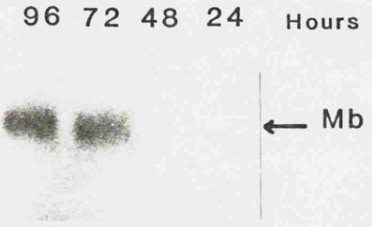
A


M.T.
pRSV-CAT
pSMb-CAT
pAT-CATp.i.
CM
24 96 24 96 24 48 72 96 24 48 72 96 Hours

B


pRSV-CAT
pSMb-CAT
0 50 100 %
0 5 10 15 20 100 %
24
96
24
48
72
96

C


96 72 48 24 Hours
Mb

Cytoplasmic RNA was analysed on Northern blots using the human myoglobin exon 2 probe (which should not cross-hybridise to the myoglobin 5'-flanking region and 5'-nontranslated mRNA sequences present in the myoglobin-CAT gene fusion). The RNA was prepared from the cell extracts of transfected myoblast samples by phenol extraction and ethanol precipitation, and analysed on agarose gels prior to blotting to ensure uniform concentrations. The Northern blot shown in Fig. 7.4 illustrates that myoglobin transcripts are apparent in the post-fusion cell extracts (fusion time: 52 and 76 hours) but not in pre-fusion or fusing cell extracts (fusion time:0 and 28 hours). The appearance of endogenous myoglobin transcripts seems to coincide therefore with the amount of CAT activity.

In this particular transfection experiment levels of CAT activity observed in cell extracts from myoblasts transfected with pRSV-CAT show high levels of CAT activity and since practically all of the chloramphenicol is converted to its mono-acetylated products or the di-acetylated product, (~98% conversion) in cell extracts assayed from pre-, fusing and post-fusion cultures it is not possible to determine whether there would be higher levels of CAT in differentiated myoblast cell extracts. This experiment makes the important point that even if levels of CAT do increase as a result of cell differentiation in myoblasts transfected with pRSV-CAT, the RSV LTR promoter is not behaving exactly as the myoglobin promoter since the LTR may drive very high levels of CAT in myoblasts whereas the myoglobin promoter may not.

7.7 Transfection of clone 1 myoblasts with pHMb-CAT

A transfection was carried out with pSMb-CAT, pAT-CATp.1., pHMb-CAT, pMMb-CAT and pSV2-CAT into clone 1 myoblasts in order to assess in one transfection experiment (rather than separate experiments) the

levels of CAT activity from the various promoters. In order to ensure more uniform transfection efficiencies, all plasmids transfected were subjected to two caesium-chloride centrifugations (see Gorman, Moffat and Howard, 1982).

The results of the various transfections demonstrate that levels of CAT activity are increased during differentiation in myoblast cell extracts from myoblasts transfected with pSMb-CAT, pHMb-CAT and pMMb-CAT (Fig. 7.5). pHMb-CAT and pMMb-CAT extracts were assayed at 30 hours after the addition of the DNA precipitate (pre-fusion) and 72 hours after the addition of the DNA precipitate (post-fusion, time in fusion medium = 46 hours). pHMb-CAT showed a relative increase of CAT activity from 24 to 72 hours of 11.2 fold and pMMb-CAT of 7.4 fold. At the 72 hour time point in myoblasts transfected with pSMb-CAT, a relative increase in CAT activity of 5.3 fold is observed and by the 96 hours CAT activity has increased 10.2 fold. In this particular experiment it does not appear that the 5'-flanking region of the seal myoglobin gene present in the pSMb-CAT gene fusion drives more efficient transcription than the 5'-flanking regions present in the mouse and human gene fusions, assuming that transfection efficiencies are equal with all three constructs.

As was observed for the levels of CAT activity from the RSV LTR in Fig. 7.4, pSV2-CAT transfected myoblast cell extracts have high levels of CAT activity (Fig 7.5). Since there is almost 100% conversion of chloramphenicol (upper limits of assay) it is not possible therefore to determine whether there would be a large difference in CAT activity in myoblasts and myotubes.

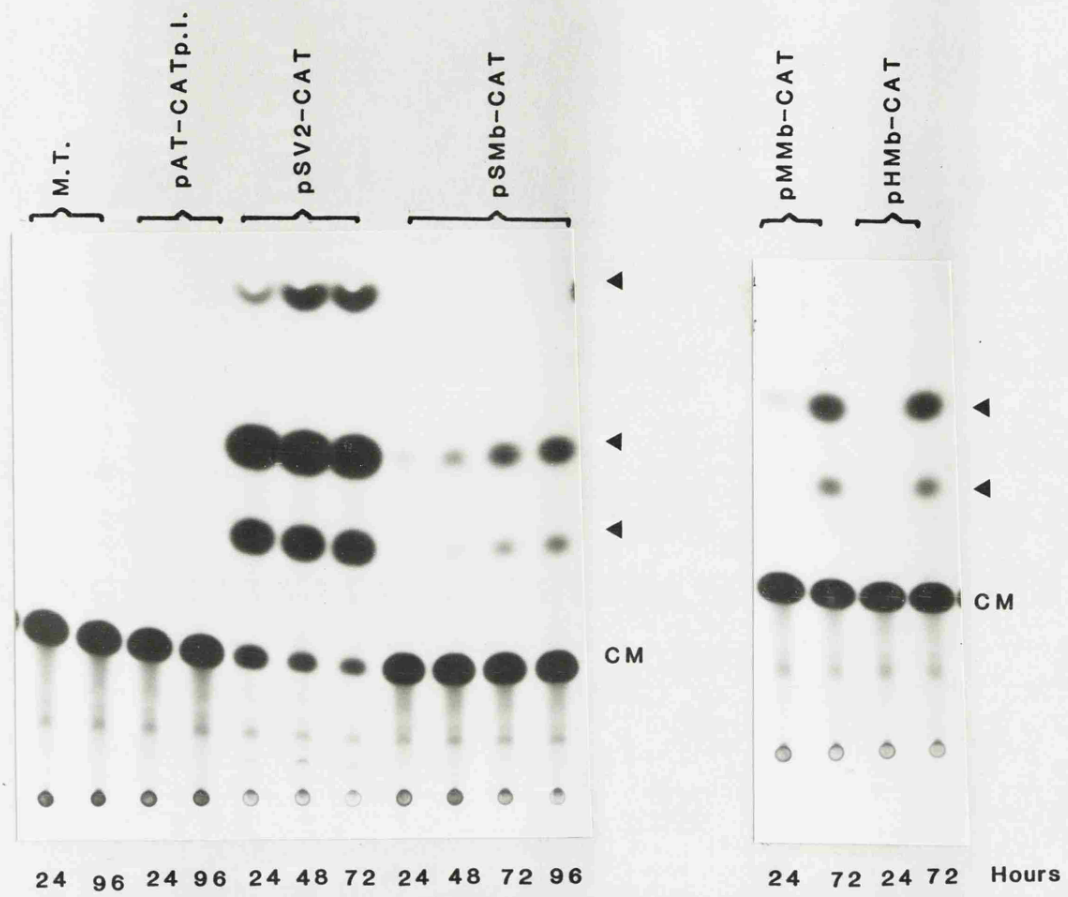
Northern blot analysis of RNA from transfected myoblast cell extracts revealed the presence of myoglobin transcripts in post-fusion (72 hour expression) myotube RNA but not in pre-fusion (30 hour expression) myoblast RNA samples (not shown).

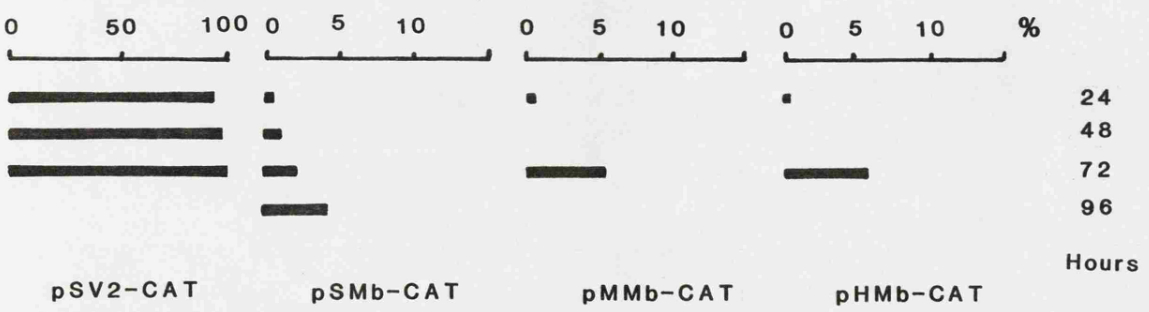
Figure 7.5

Analysis of CAT enzymatic activity in myoblasts transfected with pSMb-CAT, pMMb-CAT and pHMb-CAT

Top Myoblasts were transfected with equal amounts of the three myoglobin-CAT gene fusions and with pAT-CATp.l., salmon-sperm DNA and pSV2-CAT as controls. Pre-fusion myoblasts (24 hour samples) and fusing and post-fusion myoblasts (48-96 hour samples) were assayed for CAT activity. Low serum medium was added to the 48-96 hour samples 26 hours after the addition of the DNA precipitate. Assays were carried out for 4 hours using 50µl of cell extract for each sample. Autoradiography was for 2 days.

Bottom Graph showing the percentage of acetylation of chloramphenicol in the cell extracts determined by scintillation counting of the radioactive spots.


M.T.
pAT-CATp.I.
pSV2-CAT
pSMb-CAT
pMMb-CAT
pHMb-CAT
CM
CM
24 96 24 96 24 48 72 24 48 72 96
24 72 24 72 Hours


0 50 100 0 5 10 0 5 10 0 5 10 %
24
48
72
96
Hours
pSV2-CAT
pSMb-CAT
pMMb-CAT
pHMb-CAT

7.8 Replating of mouse myoblasts after transfection

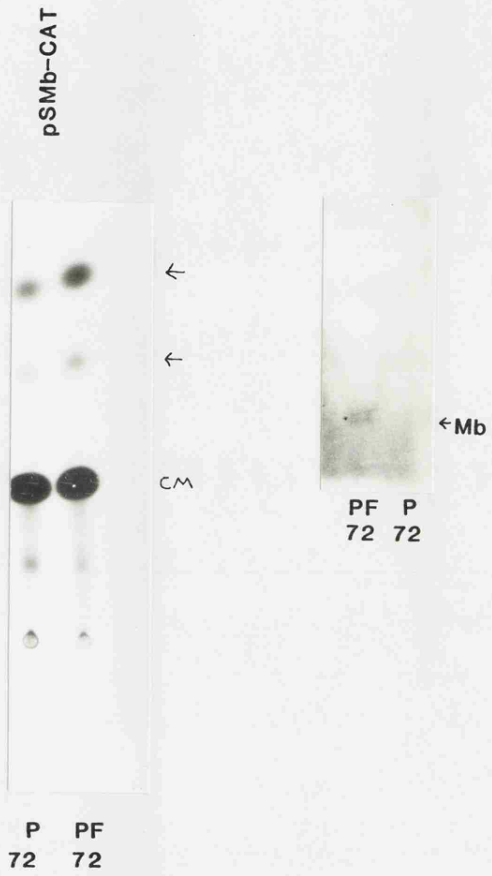
Another transfection procedure was carried out in order to try to assay parallel myoblast and myotube cultures, therefore providing an alternative to assay over a time course. pSMb-CAT was transfected into clone 1 myoblasts which were subsequently replated at high and low cell densities after the transfection procedure. Three 10 cm dishes of myoblasts were transfected as usual. After recovery from the glycerol shock, myoblasts were trypsinised from the culture dishes and pooled and split into dishes to give myoblasts at a high cell density (one dish only) and myoblasts at low cell density (equal number of cells distributed between 5 dishes). When the myoblasts had settled in the high density dish, the medium above the myoblasts was immediately replaced with low serum, fusion-stimulating medium. After 72 hours in culture (fusion time: 53 hours) the resulting cell morphology was mainly myotube fibres. The myoblasts replated at a low cell density had subsequently proliferated and the medium was changed frequently in an attempt to prevent precocious fusion in the cultures to have pure populations of proliferating myoblasts for the comparison. For the assay of CAT activity in these extracts, equivalent volumes of each extract were assayed so that the same initial number of transfected myoblasts were being analysed. Fig. 7.6 shows the result of the CAT assays; there is an increase in CAT activity in the fusing myoblast cell extract compared to pre-fusion extract (a relative increase of 2.6). Since myoblasts have proliferated, this estimate of increase in CAT activity is likely to be conservative if any insertion of the gene fusion had occurred and since the overall backround level of CAT activity is increased due to assay of a greater number of cells. The level of CAT activity in the pre-fusion extracts is about 2.5 fold above normal background levels observed as determined by scintillation counting. This

Figure 7.6

Assay of CAT enzymatic activity in myoblasts and myotubes in parallel

Myoblasts (3×10^6) were transfected with pSMb-CAT. After recovery from the glycerol shock, myoblasts were replated at high and low cell densities. The cells at high density were refed with low serum medium after replating. CAT assays (see left panel) were carried out 72 hours after the addition of the DNA precipitate on extracts from low density myoblasts (pre-fusion:P) and high density myoblasts which had been in low serum medium for 46 hours (post-fusion:PF). The positions of the chloramphenicol (CM) and mono-acetylated derivatives (arrow heads) are shown. The post-fusion cell extract converted 2.6 times more chloramphenicol than the pre-fusion cell extract. Autoradigraphy was for three days.

The right hand panel shows two lanes of a Northern blot on which RNAs isolated from pre- and post-fusion cell extracts were hybridised to the human myoglobin gene exon 2 probe (316 bp, see Table 3.1). Myoglobin gene transcripts (Mb) are indicated by an arrow. It cannot be determined from this blot whether a low level of myoglobin transcripts are present in the pre-fusion (P) RNA sample due to high background radioactivity levels. Autoradiography was for four days.


pSMb-CAT
CM
P 72
PF 72
Mb
PF 72
P 72

increased level of CAT activity in myoblast cell extracts could also be due to some precocious differentiaton in confluent clone 1 cultures which are difficult to maintain in a completely proliferative state (Chapter 5), although it was not possible to detect myoglobin mRNA in myoblast cultures on Northern blots.

If time had permitted, this transfection procedure could have been modified to transfect myoblasts in suspension and then replate at the appropriate cell density.

7.9 Summary and comment on experimental procedures

Transfection of myoglobin-CAT gene fusions and assay of CAT activity after differentiation of myoblasts over a transient expression period have demonstrated that CAT is expressed from the gene fusion when transfected into mouse myoblasts. Furthermore, the level of CAT activity appears to increase over a differentiation period. There is a coincidence of appearance of CAT activity and endogenous myoglobin gene transcripts. The expression of CAT driven by a myoglobin promoter in transient expression experiments appears to be tissue-specific because CAT activity is not detectable in mouse Ltk^- fibroblast cell extracts when fibroblasts are transfected with a myoglobin-CAT construct.

Some of the information for this tissue-specificity and stage specific expression would seem to be contained within the ~1 kb of myoglobin 5'-flanking region present in the myoglobin-CAT gene fusions. Both seal and human myoglobin-CAT gene fusions are expressed in differentiating mouse myoblasts which indicates that the regulatory signals controlling myoglobin tissue-specific and stage-specific expression have been conserved during many years of independent evolution in these species.

These preliminary transient expression results with cloned

myoglobin promoters are therefore encouraging, especially in view of the fact that not all transfections with transferred muscle-specific genes have resulted in expression of the particular gene fusion. For example, no expression could be detected from a transferred acetylcholine receptor gene (J. Merlie, personal communication). However, there are a number of experiments or modifications to experiments described above which must be carried out before it can be convincingly demonstrated that expression of the myoglobin gene fusion reflects that of the endogenous myoglobin gene during myoblast differentiation.

First and foremost, S1 nuclease analysis should be carried out to show that the initiation start site in the myoglobin-CAT gene fusions is identical to that used by the endogenous myoglobin gene. This experiment was attempted using the seal myoglobin-CAT RsaI fragment cloned into M13 (Section 6.6) to generate a single stranded hybridisation probe containing the expected site of transcrition initiation and ~500 bp of 5'-flanking DNA. The experiment was not successful and should be repeated before going on to other experiments.

An S1 nuclease analysis would also provide information on the hybrid gene RNA levels during differentiation over the transient expression period. This might indicate whether the observed increase in CAT activity is due to a build up in CAT protein levels from a previously established pool of hybrid gene transcripts in myoblasts or whether the increase in CAT activity reflects an increase in mRNA synthesis during differentiation.

Another major question raised by these results is whether the increase in CAT activity during differentiation is due to a specific response by the muscle-specific gene promoter to a "muscle-activator" or whether the observed effect is a non-specific global activation also operating at least on RSV as a result of terminal differentiation of the

cells. The restricted expression of the myoglobin-CAT genes in muscle versus non-muscle cells and the coincidence with the appearance of the endogenous myoglobin gene transcripts tends to suggest that the transfected myoglobin-CAT gene fusions are being regulated specifically and appropriately but the activation of the RSV promoter during myoblast differentiation in identical transfection experiments would tend to suggest that the activation is non-specific. Other researchers have noticed this effect with virus promoters, and the induction of expression is not consistent. The choice of this "control" plasmid was therefore unfortunate and it appears that the RSV LTR promoter is not a suitable control for transfection into differentiating myoblasts. During the myoblast to myotube transition there will be many changes in cellular "regulatory" factors, and it is not necessarily true that the same factor which activates the myoglobin promoter during myogenesis also activates the LTR promoter. A better control to test the specificity of the activation might be a gene expressed during terminal differentiation of another cell-type, for example the β-globin gene. A gene expressed constitutively in muscle cells during pre- and post-differentiation, for example a 5' β-actin gene fusion, would provide another control for this system.

In the CAT assays described in this Chapter the level of CAT activity appears to be low although the difference between activity and no activity is clearly visible. It is difficult to determine whether levels of CAT activity observed with myoglobin-CAT fusions are very low for a transferred muscle-specific gene, since in many reported transfections with muscle gene-CAT fusions, quantitations of CAT activity levels reported in the literature are restricted to the degree of induction during differentiation.

It is possible that a low level of expression might result from

lack of an important control signal in the myoglobin-CAT gene fusion. It has been reported for example that a troponin C gene requires sequences within introns 1-3 in order to be expressed (B. Paterson, personal communication) and an embryonic myosin heavy chain gene requires a sequence maybe as far away as 1.5 kb before the CAP site for high level expresion of a transfected mini-gene (V. Madhavi, personal communication). It would therefore be wise to include other portions of the myoglobin gene in further constructs.

Another reason for low level expression might be that much of the plasmid DNA is degraded by the time myoblasts become competent to express it. Southern blot analysis of plasmid DNA isolated from transfected cell extracts revealed that towards the end of the transient expression time course, plasmid DNA could not be detected in the extracts (not shown). The use of a cell line which differentiates rapidly, for example the MM14 mouse myoblast line of Chamberlain, Jaynes and Hauschka (1984) might be useful in that these myoblasts would be competent to express the myoglobin muscle gene fusion before the DNA becomes substantially degraded. Other useful cell lines might be the quiescent L6 and C2 myoblast cell lines isolated by Whalen et al. (1986). In these cell lines fusion may be rapidly induced in quiescent myoblasts with a serum-free medium, allowing the processes of growth and differentiation to be separated. The control over manipulation of growth and differentiation of these cell lines would obviously allow more control in transfection experiments using these myoblasts.

The system described here may not be useful at present to determine whether the seal myoglobin promoter is a more efficient promoter than the human or mouse myoglobin promoters. Although there is no information on the levels of mRNA from the transfected gene fusions, the preliminary experiments would indicate that all three promoters drive

similar levels of CAT gene expression as determined by the similar levels of CAT activity observed in transfected cell extracts. It therefore appears that in the particular seal myoglobin-CAT gene fusion constructed, sequences in the 5'-flanking region, up to ~1 kb, are not responsible for the observed increase in seal myoglobin mRNA observed *in vivo*. It might be that in the particular gene fusion a putative sequence (for example an enhancer sequence) specifying high level transcription has been deleted. Speculating further, it could be that mouse myoblasts lack a positive activator (or contain a negative regulator) which normally specifies increased transcriptional activity of the seal myoglobin promoter in seal myotubes or alternatively, levels of seal myoglobin mRNA are attained during muscle fibre maturation and therefore would not be seen in embryonic myoblast cell cultures which do not mature successfully *in vitro*.

As the *in vitro* system stands, the difference between myoglobin-CAT gene fusion expression and no expression is quite clear and therefore this system could be used at present to determine the effects of various promoter deletion mutations on expression of transfected myoglobin-CAT gene fusions. Although a number of deletion mutation strategies were planned, there was no time to carry out these experiments. In conclusion therefore, this system should be useful to study the molecular basis for myoglobin gene regulation during myogenesis.

CHAPTER 8

DISCUSSION

8.1 Structure of mammalian myoglobin genes

The recent cloning and characterisation of the mouse myoglobin gene (Chapter 3) has enabled further definition of the organisation of mammalian myoglobin genes.

Myoglobin of both cardiac and skeletal muscle is encoded by a single gene in the mouse. This is also true of the human and grey seal and therefore a single gene encoding myoglobin is likely to be the rule in mammals. In contrast to this, other members of the globin gene family are found as members of multigene families providing a variety of specialised gene products: for example, the β-like globin gene cluster in humans, encoding a number of embryonic, foetal and adult haemoglobins and the family of globin genes specifying leghaemoglobin in higher plants (see Jeffreys et al., 1983). The single myoglobin gene also stands in stark contrast to the large number of genes encoding muscle-specific protein isoforms (see Buckingham and Minty, 1983). The MHC multigene family for example, consists of ten members and has specific skeletal and cardiac isoforms. The role played by myoglobin in different muscles, muscle fibre types and during muscle development does not require specialised forms of myoglobin and its presence is determined by the physiological requirements of specific muscle types.

All the mammalian myoglobin genes characterised so far show the common three exon-two intron organisation found in vertebrate α- and β-globin genes. The introns of mammalian myoglobin genes, however, are much longer than their haemoglobin counterparts and this is also likely to be a stable feature of mammalian myoglobin genes. It appears that the exact size of the introns is not rigidly determined since mouse myoglobin

gene intron 2 is relatively short (1.5 kb versus 3.5 kb in human, 3.2 kb in seal) indicating a certain amount of fluidity in intron length. Similarly, intron 1, of the human gene is 5.8 kb in size, whereas seal and mouse are 4.8 kb and 4.4 kb in size respectively.

Myoglobin genes also have elongated 3'-nontranslated regions compared to α- and β-globin genes, although once again the length of this region shows interspecies variability: with the mouse 3'-nontranslated region being ~100 bp shorter than its human and seal counterparts. The sequence of neither introns nor 3'-nontranslated regions show any evidence for conservation between mammalian myoglobin genes. It is not known whether the elongated introns and 3'-nontranslated regions of myoglobin genes represent a functional adaption of globin family members incorporated into the myogenic developmental program. It would be interesting to determine at which point in evolution these long non-coding regions were established or the non-coding regions became shorter in haemoglobin genes.

An analysis of myoglobin gene promoter regions revealed that the "core" promoter region of myoglobin genes is atypical for a globin gene promoter. The only feature they share with globin promoters (and other specialised eukaryotic promoters) is a conserved TATA box in all three myoglobin genes located about 30 bp before the myoglobin gene CAP site.

Humans and mice contain eight fold less myoglobin mRNA in their skeletal muscle than seals. Sequence analysis of the myoglobin gene 5'-flanking regions does not reveal any specific sequence changes between the species which might account for the very obvious differences in the levels of myoglobin mRNA detected in human, seal and mouse skeletal muscle. The initial characterisation of seal and human myoglobin genes revealed that substitutions in the promoter region of the genes were more apparent in the purine-rich region termed the GAGA box. This lead to the

proposal that sequence differences in this region might determine in some way differences in the levels of myoglobin mRNA observed between these two species. However, the statistical analysis of observed base substitutions in the aligned human and seal myoglobin gene sequences (Section 3.5) demonstrated that this sequence is neither significantly diverged nor conserved between the two genes but is evolving at the same rate as DNA in the noncoding regions of the gene, presumably at or close to the neutral rate. Also this region is absent from the mouse myoglobin gene and therefore would not appear to play a role in myoglobin gene expression although the molecular basis for the adaption of the seal myoglobin to high level expression clearly awaits a functional test.

The sequence comparison between mammalian myoglobin genes led to the identification of a long highly conserved domain in the 5'-flanking region located about 120 bp before the mRNA CAP site of the genes. The significant sequence conservation in this region (for at least 80MY) suggests it may serve some functional role in myoglobin gene expression and will be discussed in Section 8.3.

8.2 Induction of myoglobin gene expression during myoblast differentiation in vitro

The early appearance of myoglobin in embryogenesis can apparently be modelled in cultures of differentiating embryonic myoblasts in vitro; both rat and mouse permanent myoblast cell lines synthesise myoglobin mRNA on fusion and differentiation to form myotubes. The induction of contractile protein genes and other muscle-specific genes also occurs during myoblast differentiation in vitro and as was found for the myoglobin gene, transcripts accumulate before the formation of visible myotubes. It has been demonstrated for a number of contractile protein genes that regulation of gene activation is primarily at the level of

transcription (see Section 1.4). The finding that a myoglobin-CAT gene fusion may be activated in differentiating myoblasts suggests that part of the regulation of myoglobin gene expression during myogenesis is at the level of transcription (see Section 8.5).

In contrast to the results described in this thesis and by Weller *et al.* (1986) using permanent rodent myoblast cultures, Gunning (personal communication) was unable to detect myoglobin gene transcripts in fusing cultures of cloned human skeletal muscle satellite myoblasts over a differentiation period of 15 days. Satellite myoblasts are isolated from adult skeletal muscles, but on differentiation to form myotubes *in vitro*, they recapitulate embryonic muscle development and express muscle gene isoforms found at the embryonic stage (see Strohman *et al.*, 1986). Gunning believes that differentiating satellite myoblasts more closely resemble the normal events and phenotypes which occur during skeletal muscle myogenesis *in vivo* and therefore permanent rodent myoblast cell lines may carry a regulatory mutation involving myoglobin gene expression. However, this seems unlikely for a number of reasons. Firstly, myoglobin gene transcripts are detectable extremely early in mouse embryogenesis (14 days of gestation, Weller *et al.*, 1986) at a time when myoblast fusion is occuring. Since contractile protein gene transcripts are detectable at this time, it would be expected that the myoglobin gene is activated along with the contractile protein genes during myoblast differentiation *in vitro*. Secondly, there are other data suggesting activation of the myoglobin gene during myoblast fusion in another cell system. For example, Kagen and Freedman (1973) have demonstrated that myoglobin can be detected by immunological assays very early in development in the chick embryo and this early appearance in embryogenesis is reflected by the synthesis of myoglobin in fusing chick primary myoblast cultures *in vitro* (Kagen and Freedman, 1974). Kagen and

Freedman (1974) detected a large increase in myoglobin concentrations in differentiating chick myoblasts after a period of 72 hours in culture, visible fusion to form myotubes occurring at 48 hours in this system. The appearance of myoglobin in these chick primary myoblast cultures therefore occurs somewhat after fusion. They demonstrated however, that the increase in myoglobin synthesis around the time of fusion was not dependant upon the fusion event since myoglobin is synthesised in the absence of fusion in calcium depleted medium.

Interestingly, their studies with myoblasts isolated from embryonic future pectoralis muscle (white, fast) and embryonic future thigh muscle (red, slow) showed differential myoglobin synthesis. Myoglobin cannot be detected in chicken pectoralis muscle *in vivo* by their immunological technique but is detectable in differentiating cultures of primary chicken pectoralis derived myoblasts although at a lower concentration than found in differentiating primary myoblasts derived from thigh muscle. Heywood, Havaranis and Herrmann (1973) however, were unable to detect myoglobin either in chicken pectoralis muscle *in vivo* or in differentiating myoblasts derived from this muscle *in vitro*. These experiments demonstrate that the early appearance of myoglobin in chick embryogenesis may be modelled in cultures of differentiating chick myoblasts *in vitro* and furthermore, the differential levels of myoglobin observed physiologically in different muscle types may also be modelled *in vitro* to a certain extent.

The lack of myoglobin gene expression in adult satellite myoblasts is therefore puzzling. It has recently been suggested that slow muscle satellite myoblasts may be different from myoblasts derived from fast muscle and it appears that adult satellite cells have diverged along a commitment pathway for fast and slow muscle respectively (Strohman *et al.*, 1986); if clones of fast muscle satellite myoblasts were used, it

seems likely that the myoglobin gene would not be expressed at significant levels.

Another similarity of the behaviour of the myoglobin gene during myogenesis with that of the contractile protein genes was obtained from a preliminary study of the methylation patterns around the myoglobin gene during myoblast fusion (J. Shaw, M. Price and A.J. Jeffreys, unpublished observations). DNA methylation has been implicated as one of the control processes responsible for the expression of specific genes during development (see Cedar, 1984). Considerable evidence has been obtained for a large number of cell-specific genes correlating expression of the genes with the degree of undermethylation of specific gene sequences. For example, genes encoding ovalbumin, conalbumin, and globin genes have sequences which are hypomethylated in cells that express them and hypermethylated in cells that do not.

Shani, Admon and Yaffe (1984) determined the methylation patterns around two contractile protein genes, α-skeletal actin and MLC2, in myoblasts and myotubes using methylation-sensitive restriction enzyme analysis. Their results demonstrated that no change in the methylation pattern around the genes occurred during the myoblast to myotube transition, and that both genes contained a number of unmethylated sites in both myoblast and myotube DNA. It also appeared that α-skeletal actin and MLC2 genes were less methylated in DNA from tissues which do not express either gene. We observed a similar result with the myoglobin gene. No change in the methylation pattern of various sequences in the 5'-flanking region of the myoglobin gene could be detected in myoblast DNA and myotube DNA using a number of methylation-sensitive restriction enzymes. The myoglobin gene sequences also appeared to be less methylated in fibroblast DNA which does not express the myoglobin gene. Therefore, in the case of the myoglobin gene and the contractile protein

genes, no correlation between undermethylation of the particular bases tested and activation of gene expression was observed.

Yisraeli et al. (1986) have recently extended the studies of methylation state of the α-skeletal actin gene. Using gene transfer experiments, they have demonstrated that the introduction of an α-skeletal actin gene fusion, methylated in vitro resulted in site-specific demethylation of the transferred gene only in cultures of myoblasts. This demethylation process mimics a similar process which occurs in the endogenous α-skeletal actin gene during normal muscle cell development.

The expression of the myoglobin gene concomittantly with the contractile protein genes during myogenesis and the similarity in their methylation states suggest that the myoglobin gene and contractile protein genes may share some common regulatory mechanisms. This is discussed below.

8.3 Sequence conservation in the 5'-flanking region of muscle-specific genes

There are a large number of examples in the literature of 5'-flanking sequences which are highly conserved in evolution between genes suggesting that these sequences are functionally significant in initiation and/or regulation of gene transcription. The sequences described are generally relatively small oligonucleotide sequences (for example, Schmidt, Yamada and de Crombrugghe 1984) but in some cases quite large regions of conservation are apparent (for example, Hu et al., 1986, Daubas et al., 1985).

The myoglobin gene contains a relatively long domain of sequence conservation in the 5'-flanking region. How may the functionality of these conserved regions be determined? One approach is to compare the particular conserved sequence with sequences which have been shown to be

essential for expression *in vitro* or which are conserved in other genes expressed in the same cell type. It might be expected that coordinate gene expression in response to "cell-specific" signals or particular extrinsic factors might involve shared regulatory elements. Are common sequences generally found in eukaryotic genes which are expressed in the same cell-type or are activated in response to common extrinsic factors? For a number of genes it would appear that homologous sequences signifying common *cis*-acting gene regulatory elements do exist. For example, three fat cell specific genes, adipocyte P2, glycerol-3-phosphate dehydrogenase and adipsin, which are coordinately activated during adipocyte cell differentiation, contain a 13 bp region of homology present in multiple copies in the 5'-flanking region of each gene. The three genes are members of three distinct gene families (Hunt *et al.*, 1986). Similarly, genes expressed specifically in the exocrine pancreas, amylase, chymotrypsin and trypsin I possess upstream elements that direct their differential cell-type expression (Boulet, Erwin and Ritter, 1986). The amylase gene and chymotrypsin gene upstream sequences have been mapped in detail. They lie 41 to 235 and 93 to 275 bp before the mRNA CAP site respectively, and exhibit the properties of cell-type specific enhancers. Sequences in the 5'-flanking region of seven other pancreatic exocrine genes show homology and are in a similar location to the amylase and chymotrypsin enhancer regions. Homology is not detected in endocrine genes or in a number of non-pancreatic genes tested. It is noteworthy that the pancreatic exocrine genes demonstrate diverse expression profiles during development.

The coordinate induction of transcription of the heat-shock genes in *Drosophila*, in response to a rise in temperature, appears to be mediated by a conserved heat-shock regulatory sequence found in the 5'-flanking region of heat-shock genes promoters. The homologous region

is found in an area which may adopt a hair-pin loop configuration (Pelham, 1982) and the heat-shock consensus sequence appears to be bound by a specific heat-shock transcription factor (Dynan and Tjian, 1985).

The vitellogenin and apoVLDL II yolk protein genes of the chicken are specifically transcribed in the liver upon oestrogenisation. Both genes have a number of common sequence motifs in regions of the genes which exhibit oestrogen induced changes in chromatin structure (Van Let Schip et al., 1986). A distinct homology to some of these sequences in vitellogenin genes in Xenopus laevis is also found.

Genes activated during development of a specific cell-type or in response to the same external stimuli therefore sometimes contain common sequences within their 5'-flanking region which may be implicated in their common regulation. There are however exceptions. A search for homologous sequences in a number of egg white proteins which are specifically induced in the ovary by four classes of steroid hormones was negative (Renkawitz et al., 1982). In another example, prolactin and growth hormone genes are structurally related and co-expressed within the same cell types during development which raises the possibility that they may be regulated by common factors. Cell type-specific enhancers have been described in the 5'-flanking regions of both genes but are structurally distinct cis-acting sequences under the regulation of discrete trans-acting factors (Nelson et al., 1986). Similarly, no common sequences are found in the α- and β-globin genes which are specifically expressed in erythroid cells (Charnay et al., 1984)

During muscle cell differentiation, a large number of unrelated genes are activated around the time of myoblast fusion (Section 1.4). It is therefore of interest to determine whether common sequences may be found in muscle-specific genes which would imply a relatively simple induction process in which coordinate control is mediated by a single

regulatory factor.

The myoglobin gene upstream conserved domain was compared to sequences in muscle specific genes which have been shown to be highly conserved in a number of species, show similarities from one gene to another and which have been shown to be required for regulated gene expression *in vitro*. A list of these published sequences which were compared to the myoglobin gene conserved sequence is given in Table 8.1.

Muscle-specific gene sequences homologous to the myoglobin gene conserved sequence (and flanking this region) were not detected by dot-matrix comparisons and by eye. The variety of sequences depicted in Table 8.1 demonstrate that while striking homologies can be observed between the same gene in different species which is probably indicative of a functional role in gene specific expression, a single sequence found in all muscle specific genes which could be considered as conferring a "muscle context" cannot be found. Some sets of muscle-specific genes may show common regulation which might explain why some consensus sequences appear to be present in a selected set of muscle-specific genes (Table 8.1). The sequences depicted in Table 8.1 demonstrate that the promoters of muscle-specific genes are complex. A variety of sequences in the same gene are shown to be essential for expression *in vitro*, and in other genes long conserved regions are found implying a number of functional sequences within the promoter region. L. Kedes (personal communication) has observed regions which seem to confer both positive and negative regulation on the human α-cardiac actin gene.

A simple model of common promoter sequence elements recognised by a specific factor appears to be too simplistic a concept to explain coordinate gene activation during myoblast differentiation and a more complex process is likely to be operating.

In view of some of the recent work carried out by Gunning

1 Troponin 1 gene consensus sequence

GCCCGAcaccCAAATAT

aga agacaagg

gggg

A B C

2 Sarcomeric actin CAAT box

GCCCGACACACCCAAATATGGC rat

GCCCAACACACCCAAATATGGC chicken

3 Human α-cardiac actin: CC $(A)_r$ GG boxes

CCCTATTTGG -197 chicken

CCCTATTTGG -240 human

CCTTAGATGG -165 chicken

CCTTACATGG -215 human

TGTGGCTTATTGTCCCCAAGGCT

4 α-skeletal actin hair-pin loop sequences

TCCTTCTTTGGTCAGTGCAGGAGACCCGGG

CCCAGGCTGAGAACCAGCCGAAGGA

Rat and chicken α-skeletal actin gene conserved sequences

-230 CTCTCCATATACGGCCCGGTCCGGTTCAGG CTACCTGGGCCAGGGCCA

CCTTCTTTGGTCAGTGCAGGAGACCC GTGCCCAACACCCAAATATGGCTTGGGAAGGGC -80

5 Creatine kinase repeat sequence

ACCTGGGGXXXGCCCAX

c ct c g c

6 MLC 2 and α-skeletal actin gene conserved sequences

-165 CCCTGCCCTTGGCTCCA + AGAATGGC

-242 GCTCCCTATATGGCCAT

7 Xenopus α-cardiac actin enhancer like sequence

-323 TACTCCATTTGCAGACCCCTG

8 Sequences required for regulated expression of chicken α-skeletal actin

-200 GGGGGCCCGGCAGACGCTCCTTATACGGCCCGGCCTCGCTCACCTGGGCCGCGGCC

AGGAGCGCCTTCTTTGGGCAGCGCGGGGCCGGGGCC -107

Table 8.1

Conserved sequence elements in the 5'-flanking region of muscle-specific genes

1 The 17 bp muscle gene consensus sequence is from Baldwin, Kittler and Emerson (1986) and is found 329 bp (quail fast troponin 1 gene), 100 bp (chicken and rat α-skeletal actin genes), 256 bp (chicken MLC 3 gene) and 411 bp (rat cardiac MHC gene) before the respective CAP sites. Two homologous regions (A and C) flank a more variable central core (B). The most prevalent nucleotide is shown in uppercase letters and the variant nucleotides in lower case.

2 The CAAAT box (underlined) is highly conserved in evolution in sarcomeric actin genes and homology extends over 20 bp around the CAAT box in the chicken and rat α-skeletal actin genes with only one nucleotide difference (Ordahl and Cooper, 1983)

3 $CC(A)_nGG$ boxes, shown here conserved at two positions (out of the five) between human and chicken α-cardiac actin genes were first described in the human α-cardiac actin gene (-240 to -40 bp, Minty and Kedes, 1986) and have been found in multiple copies in the 5'-flanking regions of the Xenopus α-cardiac actin gene (3 copies) α-skeletal actin genes (rat, chiken, 3 copies), β-actin genes (rat, chicken, 2 copies) and once in the rat α-cardiac MHC gene, chicken cardiac MLC2 gene, rat skeletal MLC2 gene and chicken cardiac Tn T gene. Upstream of the $CC(A)_nGG$ boxes in the human cardiac actin gene, sequences homologous to enhancer core sequences are found. Sequences showing a high degree of homology to the SV40 enhancer sequence are underlined. This region containing enhancer homology requires the $CC(A)_nGG$ in order to increase gene expression.

4 The 5'-flanking region of rat, mouse and chicken α-skeletal actin genes are conserved from the CAP sites to 300 bp upstream and sequences which could form hair-pin loop structures are found in this region. One of the potential stem loop structures (out of four) from the mouse α-skeletal actin gene is shown (Hu et al., 1986). Nudel et al. (1985) have also reported significant sequence conservation over 4 regions in the 5'-flanking regions of the rat and chicken α-skeletal actin genes from -230 to -80 bp, particularly around the CAAT box (underlined). The dots beneath the rat sequence indicate identical nucleotides in the chicken sequence.

5 A number of inverted and direct repeat sequences are found in the 5'-flanking region of the mouse muscle-specific creatine kinase gene (Jaynes et al., 1986). The consensus sequence of a small repeat found 3 times in the 5'-flanking region of the creatine kinase gene (before -240 bp), the rat and chicken α-skeletal actin genes (before -300 bp) and 1 time in the rat myosin heavy chain gene (before -200 bp) is shown. The nucleotides shown are invariant and those that differ are shown. Sequences containing these small repeats are required for in vitro expression of the creatine kinase gene.

6 Chicken and mouse α-cardiac actin genes contain a sequence homologous to the E1A enhancer core sequence (underlined). This sequence is also found at -100 in the mouse $MLC1_f$ gene (see Garner et al., 1986). Immediately upstream of this enhancer like sequence is the 17 nt sequence shown, conserved between α-actin and $MLC1_f$ genes and therefore

cont...

cont...

might confer muscle-specific gene expression to the enhancer should it function as one. The mouse α-cardiac gene contains a $(TG)_{24}$ in the same position as the SV40 enhancer like sequence in the human cardiac gene and therefore this gene may contain two non-overlapping enhancer regions. The 17 nt sequence shown, found at 242 bp before the CAP site is identical at all 17 positions in mouse and chicken α-skeletal actin genes.

7 The enhancer-like sequence found in the Xenopus α-cardiac actin gens is shown (underlined) and is homologous to the octamer sequence identified as a transcription factor binding site in the 5'-flanking region of the immunoglobulin heavy and K-light chain genes (Mohun, Garret and Gurdon, 1986). This sequence is also similar to the immunoglobulin enhancer sequence located in an intron in immunoglobulin heavy and light chain genes and to the enhancer-like sequence found in the 5'-flanking region of the U1 and U2 SnRNA genes in Xenopus.

8 The chicken α-skeletal actin gene sequence (-200 to -107 bp) is the border which has been defined as being necessary for the regulated expression of the α-cardiac actin gene (Bergsma et al., 1986). This region contains some of the hair-pin loop structures described in 4. Downstream of -107 bp, sequences involved in the transcription of the α-cardiac actin gene have been defined, including the CAAT and TATA boxes.

(personal communication), it is perhaps not too suprising that a single sequence in the muscle-specific genes is not detected. He analysed the kinetics of appearance of over thirty different muscle gene transcripts in differentiating human satellite myoblasts, and his data were consistent with a proposal that each muscle-specific gene is likely to have its own myogenic regulatory program. This is necessary to explain the different kinetics of appearance of most of the transcripts, the different levels of transcript accumulation and the up and down modulation of some transcript levels observed in culture. The complexity of muscle gene regulation is likely to be reflected in the complexity of the muscle gene promoters. This is observed in the rat α-fetoprotein gene for example, where over 7 kb of 5'-flanking region are required for its regulated expression *in vitro*. This region encompasses a promoter, a negative regulatory element and a complex cell-specific enhancer (Muglia and Rothman-Denes, 1986).

It is unlikely that the *cis*-acting sequences of each muscle-specific gene are recognised by a specific "myogenic" factor but more likely that a small number of muscle activators (or repressors) exist which may recognise different sub-sets of genes and bind each sub-set with different affinities to produce a wide range of combinatorial effects. It is interesting to note that it has recently been demonstrated using gel retardation techniques and foot-printing assays that the same factor binds to a similar sequence in the rat α-skeletal actin and MLC 2 genes (R. Aft-Konigsberg, personal communication).

8.4 Are all muscle-specific genes activated at the time of myoblast fusion?

A model to explain gene expression during muscle fibre development

proposed that all muscle-specific genes are activated at the time of myotube formation and later changes observed in phenotypes in different fibres is due to a selective modulation of gene expression. Evidence for this model came from the observation that both fast and slow isoforms of the contractile proteins are detectable in both future fast and slow myotubes formed in vitro and yet only one form predominates in adult fast or slow muscle fibres (see Hastings and Emerson, 1982).

The sequential appearance of specific MHC isoforms during development would argue against this model and suggest that there are likely to be multiple activation events occurring during muscle development. The clustering of the MHC genes within the same chromosomal domain may hold some answers to their particular mode of regulation during development.

In the case of myoglobin gene expression in skeletal muscle, regulation at two distinct stages is observed: first there is a low level activation of the gene during myoblast differentiation which would tend to support the total activation model since it is not thought that myoglobin is required early in skeletal muscle development (Longo, Koos and Power, 1973). In the second stage, there is a selective activation of myoglobin gene expression which gradually increases myoglobin mRNA pools during late pre-natal and post-natal periods. This second "activation" is possibly linked to the development of slow muscle fibres where myoglobin is predominantly found. This general activation followed by selective activation at a later stage of development may be a feature of the regulation of genes involved in muscle metabolism and may reflect a regulation suited to the metabolic demands of developing muscle fibres.

8.5 An in vitro system to study myoglobin gene expression

While interesting results may be obtained from a comparison of the

primary sequence of eukaryotic genes, the definitive role of conserved sequences may only be demonstrated by functional tests of the sequence in question in *in vitro* expression systems. The proposal that conserved sequences are involved in the regulation of gene expression has to some extent been borne out in functional studies with for example, insulin (Walker *et al.*, 1983), α-cardiac actin (Minty and Kedes, 1986) and α-skeletal actin (Bergsma *et al.*, 1986) genes, where conserved gene sequences have been shown to be required for gene expression *in vitro*.

The aim of this research was to develop an *in vitro* system in which some of the mechanisms responsible for the regulation of the muscle-specific myoglobin gene could be analysed and particulary to define the role of the long conserved region found in the 5'-flanking region of myoglobin genes. The induction of myoglobin gene expression during myoblast fusion *in vitro* (Chapter 4) provided a system in which the early activation of the myoglobin gene observed *in vivo* could be studied at the molecular level using gene transfer experiments. Contractile protein genes and several other muscle-specific genes are transcriptionally activated upon myoblast differentiation *in vitro*, and although a simple model of gene regulation under these circumstances is unlikely (Section 8.3) there may be some regulatory similarities in the initial induction of myoglobin and other muscle-specific genes. Myoglobin-CAT gene fusions introduced into differentiating myoblasts demonstrated that the gene fusions are expressed, although at a relatively low level and furthermore, expression increases 5 to 12 fold during myoblast fusion, coincidently with the appearance of endogenous myoglobin gene transcripts. The expression of the myoglobin-CAT gene fusions is cell-type specific. Therefore, sequences in the 5'-flanking region of myoglobin genes are able to confer tissue specificity and differentiation-linked inducibility on the myoglobin gene (although other

sequences outside of this region may be required). This is also observed for a number of contractile protein genes and a non-contractile protein gene: sequences in various regions in the 5'-flanking region are sufficient to direct the regulated expression of the genes (Melloul et al., 1984, Konieczny and Emerson, 1985, Minty and Kedes, 1986, Jaynes et al., 1986). This system will therefore be useful to determine the sequences required for this cell-type specific and developmental stage-specific gene expression and to define the role of the long conserved domain found in the 5'-flanking region of myoglobin genes using promoter deletion analysis.

Gene transfer experiments using the myoglobin promoter linked to a virus enhancer may be useful to define the core promoter of myoglobin genes. Interestingly, transfection experiments with an embryonic MHC gene, (V. Madahvi, personal communication) established that replacement of an upstream sequence of the MHC gene with the SV40 enhancer did not overule the tissue-specific activation of the MHC gene but merely acted as a "replacement" for a region of the MHC gene required for high level expression.

The in vitro system will be useful to determine the effects of a variety of physiological factors on myoglobin gene expression. For example, if myoglobin protein levels can be assayed, the effect of haem on myoglobin synthesis, known to induce the synthesis of haemoglobin may be analysed as well as the mechanism by which increased oxygen levels cause an increase in myoglobin mRNA levels (L. Kedes, personal communication).

The in vitro system as it stands may not be useful in determining the possible transcriptional differences between the different myoglobin promoters. All three promoters apperared to drive similar levels of CAT gene transcription and further experiments must be carried out in order

to establish whether the seal myoglobin gene promoter is transcriptionally more active due to primary sequence of the promoter region, activating factors which bind to the seal myoglobin promoter region more readily, decreased turnover of seal myoglobin mRNA or that high levels of seal myoglobin mRNA are controlled at a later stage of myogenesis (see Summary of Chapter 7). It is important to continue experiments in this direction since the myoglobin gene provides an excellent example to investigate the way in which a gene may become adapted to high level expression during evolution in response to a particular physiological requirement.

The differential expression of the myoglobin gene in foetal skeletal and cardiac muscle during development will probably require the use of animal models to determine the role of physiological signals influencing myoglobin gene expression in the two muscle types in vivo. It might be interesting however, to compare the levels of CAT gene expression from myoglobin-CAT gene fusions transfected into primary cultures of cardiac myocytes and skeletal myoblasts.

The second stage of myoglobin gene activation will almost certainly require the use of animal models or cultures of myoblasts able to differentiate and mature in vitro. The increase in myoglobin gene expression which occurs after stimulation of a fast fibre with the impulse frequency of a slow fibre is one way in which the increase in expression of the myoglobin gene in response to enforced slow fibre development may be investigated. For a full understanding of the molecular mechanisms involved in myoglobin gene expression during devlopment, experiments involving transgenic mice would be invaluable.

The single myoglobin gene exhibits diverse regulation both within the individual and between species. Differential gene expression is observed in different muscle types during development, in different fibre

types, between different species and in response to a number of functional demands made on muscle. The work carried out for this thesis has provided a base from which the different types of regulation exhibited by this muscle specific globin gene may be investigated.

BIBLIOGRAPHY

Aaji, C. and Borst, P. 1972. The gel electrophoresis of DNA. Biochem. Biophys. Acta. 269:192-200.

Adams, S. 1983. A molecular analysis of the β-globin gene cluster in the house mouse. Ph.D thesis. University of Leicester. England.

Adelstein, R.S. and Eisenberg, E. 1980. Regulation and kinetics of the actin-myosin-ATP interaction. Ann. Rev. Biochem. 49:921-956.

Affara, N.A., Robert, B., Jaquet, M., Buckingham, M.E. and Gros, F. 1980a. Changes in gene expression during myogenic differentiation. I. Regulation of messenger RNA sequences expressed during myotube formation. J. Mol. Biol. 140:441-458.

Affara, N.A., Daubas, P., Weydert, A. and Gros, F. 1980b. Changes in gene expression during myogenic differentiation. II. Identification of the proteins encoded by myotube-specific complementary DNA sequences. J. Mol. Biol. 140:459-470.

Antoine, M. and Niessing, J. 1984. Intron-less globin genes in the insect Chironomus thummi thummi. Nature 310:795-798.

Antonini, E. 1965. Interrelationship between structure and function of haemoglobin and myoglobin. Physiol. Rev. 45:123-170.

Appelby, 1974. In: The biology of nitrogen fixation. (ed. A. Quipsel) pp.521-554, North Holland, Amsterdam.

Askmark, H., Carlson, M. and Roxin, L-E. 1984. Myoglobin in rat hind limb muscles after denervation and during reinnervation. Muscle and Nerve. 7:656-661.

Aviv, H. and Leder, P. 1972. Purification of biologically active globin mRNA by chromatography on oligothymidilic acid-cellulose. Proc. Natl. Acad. Sci. USA. 69:1408-1412.

Bains, W., Ponte, P., Blau, H. and Kedes, L. 1984. Cardiac actin is the major actin gene product in skeletal muscle cell differentiation in vitro. Mol. Cell Biol. 4:1449-1453.

Baldwin, A.S., Kittler, E.L.W. and Emerson, C.P. 1985. Structure, evolution and regulation of a fast skeletal muscle troponin I gene. Proc. Natl. Acad. Sci. USA. 82:8080-8084.

Banerji, J., Rusconi, S. and Schaffner, W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27: 299-308.

Barany, M. 1967. ATPase activity of myosin correlated with speed of muscle shortening. J. Gen. Physiol. 50:197-216.

Barrie, P.A. 1982. A study on the β-globin gene cluster in man and the primates. Ph.D thesis, University of Leicester. England.

Bennet, M.R. and Pettigrew, A.G. 1974. The formation of synapses in striated muscle during development. J. Physiol. 241:515-545.

Bergsma, D.J., Grichnik, J.M., Gossett, L.M.A. and Schwartz, R.J. 1986. Delimination and characterisation of cis-acting DNA sequences required for the regulated expression and transcriptional control of the chicken skeletal α-actin gene. Mol. Cell. Biol. 6:2462-2475.

Betz. W.J., Caldwell, J.H. and Ribchester, R.R. 1979. The size of motor units during post-natal development of rat lumbrical muscle. J. Physiol. 297:463-478.

Biggin, M.D., Gibson, T.J. and Hong, H.F. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA. 80:3963-3965.

Birnboim, H.C. and Doly, J. 1979. A rapid alkaline procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.

Blake, C.C.F. 1978. Do genes-in-pieces imply proteins-in-pieces? Nature 273:267.

Blanchetot, A., Wilson, V., Wood, D. and Jeffreys, A.J. 1983. The seal myoglobin gene: an unusually long globin gene. Nature 301:732-734.

Blanchetot, A., Price, M. and Jeffreys, A.J. 1986. The mouse myoglobin gene: characterisation and sequence comparison with other mammalian myoglobin genes. Eur. J. Biochem. 159:469-474.

Blau, H.M., Pavlath, G.K., Hardeman, E.C., Chiu, C.-P., Siberstein, L., Webster, S.G., Miller, S.C. and Webster, C. 1985. Plasticity of the differentiated state. Science 230:758-766.

Bouche, J.P. 1981. The effect of spermidine on endonuclease inhibition by agarose contaminates. Anal. Biochem. 115:42-45.

Boulet, A.M., Erwin, C.R. and Rutter, W.J. 1986. Cell-specific enhances in the rat exocrine pancreas. Proc. Natl. Acad. Sci. USA 83:3599-3603.

Breathnach, R. and Chambon, P. 1981. Organistion and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. 50:349-383.

Breitbart, R.E., Nguyen, H.T., Medford, R.M., Destree, A.T., Mahdavi, V. and Nadal-Ginard, B. 1985. Intricate combinatorial patterns of exon splicing generate multiple regulated troponin T isoforms from a single gene. Cell 41:67-82.

Brisson, N. and Verma, D.P.S. 1982. Soybean leghemoglobin gene family : Normal, pseudo and truncated genes. Proc. Natl. Acad. Sci. USA 79:4055-4059.

Brooke, M.H. and Kaiser, K.K. 1974. The use and abuse of muscle histochemistry. Annals N.Y. Acad. Sci. 228:121-144.

Brown, D.D. 1984. The role of stable complexes that repress and activate eukaryotic genes. Cell 37:359-365.

Buckingham, M.E. 1977. Muscle protein synthesis and its control during differentiation of skeletal muscle cells in vitro. In: International Review of Biochemistry. Biochemistry of Cell Differentiation II, (ed. J.Paul) volume 15, p.269-322. University Park press, Baltimore.

Buckingham, M.E. 1985. Actin and myosin multigene families: their expression during the formation of skeletal muscle. Essays in Biochem. 20:77-109.

Buckingham, M.E. and Minty, A.J. 1983. Contractile protein genes. pp. In: N. Maclean, S. Gregory and A.J. Minty (ed.) Eukaryotic genes : their structure activity and regulation. Butterworth.

Buller, A.J., Eccles, J.C. and Eccles, R.M. 1960. Interactions between motorneurones and muscles in respect of the characteristic speeds of their responses. J. Physiol. 150:417-439.

Burke, R.E., Levine, D.n., Zajac III, F.E., Tsairis, P. and Engel, W.K. 1971. Mammalian motor units : physiological-histochemical correlation in three types of cat gastronemius. Science 174:709-712.

Butler, P.J. and Jones, D.R. 1982. The comparative physiology of diving in vertebrates. Adv. Comp. Physiol. Biochem. 8:179-368.

Butler-Browne, G.S., Bugaisky, L.B., Cuenoud, S., Schwartz, K. and Whalen, R.G. 1982. Dennervation of newborn rat muscle does not block the appearance of adult fast myosin. Nature 299:830-833.

Butler-Browne, G.S. and Whalen, R.G. 1984. Myosin isozyme transitions occurring during the postnatal development of the rat soleus muscle. Dev. Biol. 102:324-334.

Caravatti, M., Minty, A.J., Robert, B., Montarras, D., Weydert, A., Cohen, A., Daubas, P. and Buckingham, M.E. 1982. Regulation of muscle gene expression: The accumulation of messenger RNAs coding for muscle specific proteins during myogenesis in a mouse cell line. J. Mol. Biol. 160:59-76.

Carmon, Y., Czosnek, H., Shani, M. and Yaffe, D. 1982. DNase I sensitivity of genes expressed during myogenesis. Nucleic Acids Res. 10:3085-3097.

Cedar, H. 1984. DNA methylation and gene expression. In DNA methylation. (ed. A. Razin, H. Cedar and D.A. Riggs). pp.147-164. New York, Springer Verlag.

Chada, K., Magram, J. and Constantini, F. 1986. An embryonic pattern of expression of a human foetal globin gene in transgenic mice. Nature 319:685-689.

Chamberlain, J.S., Jaynes, J.B. and Hauschka, S.D. 1984. Regulation of creatine kinase induction in differentiating mouse myoblasts. Mol. Cell Biol. 5:484-492.

Charnay, P. Treisman, R., Mellon, P., Chao, M. Axel, R. and Maniatis, T. 1984. Differences in the human α- and β-glboin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell 38:251-263.

Charnay, P., Mellon, P. and Maniatis, T. 1985. Linker scanning mutagenesis of the 5'-flanking region of the mouse β-major globin gene: sequence requirements for transcription in erythroid and nonerythroid cells. Mol. Cell Biol. 5:1498-1511.

Chiquet, M., Eppenberger, H.M., Cerny, L. and Turner, D.C. 1981. Muscle morphogenesis: evidence for an organising function of exogenous fibronectin. Dev. Biol. 88:220-235

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294-5302

Chiu, C-P. and Blau, H.M. 1984. 5-azacytidine permits gene activation in a previously noninducible cell type. Cell 40:417-424.

Christian, C.N., Nelson, P.G., Peacock, J. and Nirenberg, M. 1977. Synapse formation between two clonal cell lines. Science 196:995-998.

Collins, F.S. and Weissmann, S.M. 1984. The molecular genetics of human haemoglobin. Prog. Nucl. Acids Res. 31:312-462.

Compton, R.S., Merrifield, P.A. and Konigsberg, I.R. 1986. Mitogens inhibit contractile protein synthesis in the absence of cell proliferation. In: Molecular Biology of Muscle Development. UCLA Symposia on Molecular and Cellular Biology. Volume 29 (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.61-75. Alan R. Liss, Inc. New York.

Couch, C.B. and Strittmatter, W.J. 1983. Rat myoblast fusion requires metalloendoprotease activity. Cell 32:257-265.

Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weis, M.L., Venta, P.J. and Tashian, R.E. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes which are expressed differentially during ontogeny. Nature 298:297-300.

Daubas, P., Robert, B., Garner, I. and Buckingham, M. 1985. A comparison between mammalian and avian fast skeletal muscle alkali myosin light chain genes : regulatory implications. Nucleic Acid Res. 13: 4623-4644.

Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M. and Darnell-Jr, J.E. 1981. Transcriptional control in the production of liver specific mRNAs. Cell 23:731-739.

Devlin, R.B. and Emerson, C.P. Jr. 1978. Coordinate regulation of contractile protein synthesis during myoblast differentiation. Cell 13:599-611.

Dhoot, G.K. and Perry, S.V. 1979. Distribution of polymorphic forms of troponin components and tropomyosin in skeletal muscle. Nature 278:714-718.

Dhoot, G.K. and Perry, S.V. 1980. The components of the troponin complex and development in skeletal muscle. Exp. Cell. Res. 127:75-87.

Dierks, P., Van Ooyen, A.V., Mantei, N. and Weissmann, C. 1981. DNA sequences preceeding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5' terminus. Proc. Natl. Acad. Sci. USA. 78:1411-1415.

Dierks, P., Van Ooyen, A.V., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32:695-706.

Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. 112:295-298.

Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.C. 1981. Rapid synthesis of oligodeoxyribonucleotides IV. Efficient, mechanised synthesis of heptadecaoxyribonucleotides by an improved solid phase phosphotriester route. Nucleic Acids Res. 9:1691-1706.

Dynan, W.S. 1986. Promoters for house-keeping genes. Trends in Genet. 8:196-197.

Dynan, W.S. and Tjian, R. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316:774-778

Edwards, Y.H., Lloyd, J.C., McMillan, S.L. and Benham, F.J. 1985a. Human glyceraldehyde-3-phosphate dehydrogenase: mRNA levels and enzyme activity in developing muscle. Mol. Cell Biol. 5:2147-2149.

Edwards, Y.H., Parker, M., Povey, S., West, L.F., Parrington, J.N. and Solomon, E. 1985b. Human myosin heavy chain genes assigned to chromosome 17 using a human cDNA clone as probe. Ann. Hum. Genet. 49:101-109.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. 1980. The structure and evolution of the human β-globin gene family. Cell 21:653-668.

Eichelberger, L., Fletcher, E.S., Geiling, E.M.K. and Vos, B.J. 1939. Muscle and blood haemoglobin in the dolphin. Science 90:443.

Eisenberg, B.R. 1985. Adaptability of ultrastructure in the mammalian muscle. J. Exp. Biol. 115:55-68.

Emerson, C.P. and Beckner, S.K. 1975. Activation of myosin synthesis in fusing and mononucleated myoblasts. J. Mol. Biol. 93:431-447.

Endo, M. 1977. Calcium release from the sarcoplasmic reticulum. Physiol. Rev. 57:71-108.

Endo, T. and Nadal-Ginard, B. 1986. Transcriptional and posttranscriptional control of c-*myc* during myogenesis: its mRNA remains inducible in differentiated cells and does not suppress the differentiated phenotype. Mol. Cell Biol. 6:1412-1421.

Ewton, D.Z. and Florini, J.R. 1981. Effects of the somatomedins and insulin on myoblast differentiation *in vitro*. Dev. Biol. 86:31-39.

Felsani, A., Maione, R., Ricci, L. and Amati,P. 1985. Polyoma virus as a tool for the study of the control of myogenic functions. Cold Spring Harbour Symp. Quant. Biol. 50:753-757.

Flavell, R.A. and Grosveld, F.G. 1983. Globin genes: their structure and expression. In Eukaryotic genes: their structure, activity and regulation. (ed. N. Maclean, S. Gregory and A.J. Minty). pp299-314, Butterworth.

Freshney, R.I. 1983 Culture of animal cells: a manual of basic technique. Alan R. Liss, Inc., New York.

Garner, I., Minty, A., Alonso, S., Barton, P. and Buckingham, M.E. 1986. A 5'-duplication of the α-cardiac actin gene in BALB/c mice is associated with abnormal levels of α-cardiac and α-skeletal actin mRNA in adult cardiac tissue. EMBO J. 5:2559-2567.

Garrels, J.I. 1979. Changes in protein synthesis during myogenesis in a clonal cell line. Dev. Biol. 73:134-152.

Gauthier, G.F., Lowey, S. and Hobbs, A.W. 1978. Fast and slow myosin in developing muscle fibres. Nature 274:25-29.

Gazdar, A.F., Chick, W.L., Oie, H.K., Sims, H.L., King, D.L., Weir, G.C. and Lauris, V. 1980. Continous, clonal, insulin and somatostatin secreting cell lines established from a transplantable rat islet cell tumour. Proc. Natl. Acad. Sci. USA. 77:3519-3523.

Gorman, C.M., Moffat, L.F. and Howard, B.H. 1982. Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. Mol. Cell Biol. 2:1044-1051.

Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H. 1982. The Rous-sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eucaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci. USA. 79:6777-6781.

Gorman, C., Laimons, L., Merlino, G.T., Gruss, P., Khoury, G. and Howard, B. 1984. A novel system using the expression of chloramphenicol acetyl transferase in eukaryotic cells allows the quantitative study of promoter elements. In: Eukaryotic gene expression. (ed. A. Kumar) pp129-139, Plenum press. New York, London.

Gorman, C.M., Rigby, P.W.J. and Lane, D.P. 1985. Negative regulation of viral enhances in undifferentiated embryonic stem cells. Cell 42:519-526.

Graham, F.L. and van der Eb, A.J. 1973. A new technique for the assay of infectivity of Human Adenovirus 5 DNA. Virology 52:456-467.

Grichnik, J., Bergsma, D.J. and Schwartz, R.J. 1986. Tissue restricted and stage specific transcription is maintained within 411 nucleotides flanking the 5-end of the chicken skeletal actin gene. Nucleic Acids Res. 14:1683-1701.

Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. 1982. DNA sequences necessary for transcription of the rabbit β-globin gene in vivo. Nature 295:120-126.

Grunstein, M. and Hogness, D.S. 1975. Colony hybridisation: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA. 72:3961-3965.

Gruss, P. and Khoury, G. 1981. Expression of SV40-rat preproinsulin recombinants in monkey kidney cells: Use of preproinsulin RNA processing signals. Proc. Natl. Acad. Sci. USA. 78:133-137.

Hamada, H., Petrino, M.G. and Kagunaga, T. 1982. A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA. 79:6465-6469

Hamer, D. 1977. In Recombinant Molecules: Impact on science and society. (ed. R.G. Beers, Jr and E.G. Bassett) pp.317-335. Raven, New York.

Hamer, D.H. and Leder, P. 1979. Expression of the chromosomal mouse β-maj-globin gene cloned in SV40. Nature 281:35-40.

Harris, A.J. 1981. Embryonic growth and innervation of rat skeletal muscles. Proc. Roy. Soc. B 293:20-31.

Harris, S. 1985. A comparitive study of β-globin pseudogenes in man and and the primates. Ph.D Thesis, University of Leicester, England.

Hastings, K.E.M. and Emerson, C.P. Jr. 1982. cDNA clone analysis of six co-regulated mRNAs encoding skeletal muscle contractile proteins. Proc. Natl. Acad. Sci. USA. 79:1553-1557.

Hentschel, C.C. 1982 Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S1 nuclease. Nature 295:714-716.

Hesse, J.E., Nickol, J.M., Lieber, M.R. and Felsenfeld, G. 1986. Regulated gene expression in transfected primary chicken erythrocytes. Proc. Natl. Acad. Sci. USA 83:4312-4316.

Heywood, S.M., Havarani, A.S. and Herrman, H. 1973. Myoglobin synthesis in cell cultures of red and white muscle. J. Cell Physiol. 82: 319-322.

Hickey, R., Skoultchi, A., Gunning, P. and Kedes, L. 1986. Regulation of a human cardiac gene introduced into rat L6 myoblasts suggests a defect in their myogenic program. Mol. Cell Biol. 6:3287-3290.

Hitchcock, S.E., Huxley, H.E. and Szent-Gyorgi, V. 1973. Calcium sensitive binding of troponin to actin-tropomyosin: a two site model for troponin action. J. Mol. Biol. 80:825-836.

Hochachka, P.W. 1985. Fuels and pathways as designed systems for support of muscle work. J. Exp. Biol. 115:149-164.

Holloszy, J.O. and Booth, F.W. 1976. Biochemical adaptions to endurance exercise in muscle. Ann. Rev. Physiol. 38:273-291.

Holtzer, H. 1978. Cell lineages, stem cells and the "quantal" cell cycle concept. In Stem cells and tissue homeostasis (ed. Lord, Potten and Cole) pp.1., Cambridge University Press.

Hu, C-H.M., Sharp, S.B. and Davidson, N. 1986. The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein coding region. Mol. Cell Biol. 6:15-25.

Hunt, T.L., Hurst-Calderone, S. and Dayhoff, M.O. 1978. In: Atlas of protein sequence and structure. pp.229-257 (ed. M.O. Dayhoff) National Biomedical Research Foundation, Washington.

Hunt, C.R., Ro, J.H-S., Dobson, D.E., Min, H.T., Spiegelman, B.M. 1986. Adipocyte P2 gene: Developmental expression and homology of 5'-flanking sequences among fat cell specific genes. Proc. Natl.Acad. Sci. USA 83:3786-

Huxley, H.E. 1973. Muscle contraction and cell motility. Nature 244: 445-449.

Irving, L. 1939. Respiration in diving mammals. Physiol. Rev. 19: 112-134.

Ish-Horowicz, D. and Burke, J.F. 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. 9:2989-2998

Jaynes, J.B., Chamberlain, J.S., Buskin, J.N., Johnson, J.E. and Hauschka, S.D. 1986. Transcriptional regulation of the muscle creative kinase gene and regulated expression in transfected mouse myoblasts. Mol. Cell Biol. 6:2855-2864.

Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M. and Williams, J.G. 1980. Linkage of adult α- and β-globin genes in X. laevis and gene duplication by tetraploidization. Cell 21:555-564.

Jeffreys, A.J., Harris, S., Barrie, P.A., Wood, D., Blanchetot, A. and Adams, S.M. 1983. Evolution of gene families : the globin genes in Evolution from Molecules to Men. (ed. D.S. Bendall), pp.175-195. Cambridge University Press.

Jeffreys, A.J., Wilson, V., Blanchetot, A., Weller, P., Geurts van Kessel, A., Spurr, N., Solomon, E. and Goodfellow, P. 1984. The human myoglobin gene: a third dispersed globin locus in the human genome. Nucleic Acids Res. 12:3235-3243.

Jeffreys, A.J., Wilson, V. and Thein, S.L. 1985. Hypervariable "mini-satellite" regions in human DNA. Nature 314:67-73.

Jolesz, F. and Sreter, F.A. 1981. Development, innervation and activity-induced changes in skeletal muscle. Ann. Rev. Physiol. 43:531-552.

Jones, K.A., Yamamoto, K.R. and Tjian, R. 1985. Two distinct transcriptional factors bind to the HSV thymidine kinase promoter in vitro. Cell 42:559-572.

Kagen, L.J. and Christian, C.L. 1966. Immunologic measurements of myoglobin in human adult and fetal skeletal muscle. Ann. J. Physiol. 211:656-660.

Kagen, L.A. and Freedman, A. 1973. Embryonic synthesis of myoglobin in vivo estimated by radioimmunoassay. Dev. Biol. 31:295-300

Kagen, L.A. and Freedman, A. 1974. Studies in the effects of acetylcholine, epinephrine, dibutyryl cyclic adenosine monophosphate, theophylline, and calcium on the synthesis of myoglobin in muscle cell cultures. Exptl. Cell Res. 88:135-142.

Karlsson, S. and Nienhuis, A.W. 1985. Developmental regulation of human globin genes. Ann. Rev. Biochem. 54:1071-1108.

Kelly, A.M. and Rubenstein, N.A. 1986. Muscle histogenesis and muscle diversity. In: Molecular Biology of Muscle Development. UCLA Symposia on Molecular and Cellular Biology. Volume 29 (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.77-84. Alan R. Liss, Inc. New York.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. 1958. A three dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature 181:662-666.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, R.D., Phillips, D.C. and Shore, V.C. 1960. Structure of myoglobin : a three dimensional Fourier synthesis at 2A resolution. Nature 185:422-427.

Kendrew, J.C., Watson, H.C., Strandberg, B.E., Dickerson, R.E., Phillips, D.C. and Shore, V.C. 1961. The amino acid sequence of sperm whale myoglobin: A partial determination by X-ray methods and its correlation with chemical data. Nature 190:666-670.

Kendrew, J.C. 1963. Myoglobin and the structure of proteins. Science 139:1259-1266.

Khazaie, K., Gounan, F., Antoniou, M., de Boer, E. and Grosveld, F. 1986. The β-globin gene promoter generates 5'-truncated transcripts in the embryonic/fetal erythroid environment. Nucleic Acids Res. 14: 7199-7212.

Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA. 78:454-458

Kirby, K.S. 1957. A new method for the isolation of deoxyribonucleic acids: evidence on the nature of bonds between deoxyribonucleic acid and protein. Biochem. J. 66:495-504.

Konieczny, S.F. and Emerson, C.-P. Jr. 1984. 5-azacytidine induction of stable mesoclemnal stem cell lineages from 10T1/2 cells: evidence for regulatory genes controlling determination. Cell 38:791-800.

Konieczny, S.F. and Emerson, C.-P. Jr. 1985. Differentiation, not determination, regulates muscle gene activation: transfection of troponin I genes into multipotential and muscle lineages of 10T1/2 cells. Mol. Cell Biol. 5:2423-2432.

Konigsberg, I.R. 1977. The role of the environment in the control of myogenesis in vitro. In. Pathogenesis of Human Muscular Dystrophies (ed. L.P. Rowland) p.779-798. Excerpta Medica, Amsterdam.

Konkel, D.A., Maizel, J.V. and Leder, P. 1979. The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell 18:865-873.

Krayev, A.S., Kramerov, D.A., Skryabin, D.P., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res. 8:1201-1215.

Kushner, S.R. 1978. An improved method for transformation of Escherichia coli with ColE1 derived plasmids. In: Genetic Engineering. (ed. H.W. Boyer and S. Nicosia). pp.17-23. North Holland Biomedical Press,Amsterdam,

Landsmann, J., Dennis, E.S., Higgins, T.J.V., Appelby, C.A., Kortt, A.A. and Peacock, W.J. 1986. Common evolutionary origin of legume and non-legume plant haemoglobins. Nature 324:166-168.

Lassar, A.B., Paterson, B.M. and Weintraub, H. 1986. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. Cell 47:649-656.

Lawrie, R.A. 1950. Some observations on factors affecting myoglobin concentrations in muscle. J. Agric. Sci. 40:356-366.

Lenfant, C., Johansen, K. and Torrance, J.D. 1970. Gas transport and oxygen storage capacity in some pinnipeds and the sea otter. Resp. Physiol. 9:277-286.

Linkhart, T.A., Clegg, C.H. and Hauschka, S.D. 1981. Myogenic diffentiation in permanent clonal mouse myoblast cell lines: Regulation by macromolecular growth factors in the culture medium. Dev. Biol. 86:19-30.

Lompre, A.M., Nadal-Ginard, B. and Mahdavi, V. 1984. Expression of the cardiac ventricular α- and β-myosin heavy chain genes is developmentally and hormonally regulated. J. Biol. Chem. 259:6437-6446.

Longo, L.D., Koos, B.J. and Power, G.G. 1973. Fetal myoglobin: quantitative determination and importance for oxygenation. Am. J. Physiol. 224:1032-1036.

Lui, C.C., Yansura, D. and Levinson, A.D. 1982. Direct expression of hepatitis B surface antigen in monkey cells from an SV40 vector. DNA 1:213-221.

Lusky, M. and Botchan, M. 1981. Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature 293:79-81.

Mace, H.A.F., Pelham, H.R.B. and Travers, A.A. 1983 Association of an S1 nuclease-sensitive structure with short direct repeats 5' of Drosophila heat shock genes. Nature 304:555-557.

Magram, J., Chada, K. and Constantini, F. 1985. Developmental regulation of a cloned adult β-globin gene in transgenic mice. Nature 315:338-340.

Mahdavi, V., Strehler, E.E., Periasamy, M., Wieczorek, D., Izumo, S., Grund, S., Strehler, M.-A. and Nadal-Ginard, B. 1986. Sarcomeric myosin heavy chain gene family : organisation and pattern of expression. In: Molecular Biology of Muscle Development. UCLA Symposia on Molecular and Cellular Biology. Volume 29. (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.345-361. Alan, R. Liss. Inc., New York.

Maniatis, T., Fritsch, E.F. and Sambrocke, J. 1982. Molecular cloning, a Laboratory Manual. Cold Spring Harbour Laboratory.

Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L. 1980. DNA dependant transcription of adenovirus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. USA. 77:3855-3859

Mantei, N., Boll, W. and Weissmann, C. 1979. Rabbit β-globin mRNA production in mouse L cells transformed with cloned rabbit β-globin chromosomal DNA. Nature 281:40-46.

Mathis, D., Oudet, P. and Chambon, P. 1980. Structure of transcribing chromatin. In Progress in Nucleic Acid Research and Molecular Biology (ed. W.E. Cohn) pp.1. New York Academic Press.

Mayer, Y., Czosnek, H., Zeelon, P.E., Yaffe, D. and Nudel, U. 1984. Expression of genes coding for skeletal muscle and cardiac actins in the heart. Nucleic Acids Res. 12:1087-1100.

McKeon, C., Schmidt, A. and de Crombrugghe, B. 1984. A sequence conserved in both the chicken and mouse α21 collagen promoter contains sites sensitive to S1 nuclease. J. Biol. Chem. 259:6636-6640.

Medford, R.M., Nguyen, H.T. and Nadal-Ginard, B. 1983. Transcription and cell cycle mediated regulation of myosin heavy chain gene expression during muscle cell differentiation. J. Biol. Chem. 258:11063-11073.

Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. 1981. Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. Cell 27:279-288.

Melloul, D., Aloni, B., Calvo, J., Yaffe, D. and Nudel, U. 1984. Developmentally regulated expression of chimeric genes containing muscle actin DNA sequences in transfected myogenic cells. The EMBO J. 5:983-990.

Merlie, J.P., Sobel, A., Changeux, J.P. and Gros, F. 1975. Synthesis of the acetylcholine receptor during differentiation of cultured embryonic muscle cells. Proc. Natl. Acad. Sci. USA. 72:4028-4032.

Merril, G.F., Hauschka, S.D. and McKnight, S.L. 1984. tk enzyme expression in differentiating muscle cells is regulated through an internal segment of the cellular tk gene. Mol. Cell Biol. 4:1777-1784.

Messing, J., Grea, R. and Seeburg, P.H. 1981. A system for shot-gun DNA sequencing. Nucleic Acids Res. 9:309-321

Messing, J. and Vieira, J. 1982. A new pair of M13 vectors for selecting either strand of double-digest restriction endonuclease fragments. Gene 19:269-276

Messing, J. 1983. New M13 vectors for cloning. Meth. Enz. 101:20-89.

Miata, Y. and Yoshioka, K. 1980. Selective elimination of muscle nerve terminals in the rat soleus muscle during development. J. Physiol. 309:631-646.

Mill, P.J. 1972. Respiration in the invertebrates. MacMillan Press Ltd. London.

Millikan, G.A. 1939. Muscle haemoglobin. Physiol. Rev. 19:503-534.

Minty, A. and Newmark, P. 1980. Gene regulation: New, old and remote controls. Nature 288:210-211.

Minty, A.J., Caravatti, M., Robert, B., Cohen, A., Daubas, P., Weydert, A., Gros, F. and Buckingham, M.E. 1981. Mouse actin messenger RNAS : construction and characterisation of a recombinant plasmid molecule containing a complementary DNA transcript of mouse α-actin mRNA. J. Biol. Chem. 256:1008-1014.

Minty, A.J., Alonso, S., Caravatti, M. and Buckingham, M.E. 1982. A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac actin mRNA. Cell 30:185-192.

Minty, A. and Kedes, L. 1986. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells : presence of an evolutionarily conserved repeat motif. Mol. Cell Biol. 6:2137-2145.

Minty, A., Blau, H. and Kedes, L. 1986. Two-level regulation of cardiac actin gene transcription : muscle-specific modulating factors can accumulate before gene activation. Mol. Cell Biol. 6:2137-2148.

Mohun, T.J., Garrett, N. and Gurdon, J.B. 1986. Upstream sequences required for tissue-specific activation of the cardiac actin gene in Xenopus laevis embryos. The EMBO J. 5:3185-3193.

Morton, H.J. 1970. A survey of commercially available tissue culture media. In Vitro. 6:89-108.

Moschonas, N., de Boer, E. and Flavell, R.A. 1982. The DNA sequence of the 5'-flanking region of the human β-globin gene: evolutionary conservation and polymorphic differences. Nucleic Acids Res. 10:2109-2120

Muglia, L. and Rothman-Denes, L.B. 1986. Cell type specific negative regulatory element in the control region of the rat α-fetoprotein gene. Proc. Natl. Acad. Sci. USA. 83:7653-7657.

Mulligan, R.C., Howard, B.H. and Berg, P. 1979. Synthesis of rabbit β-globin in cultured monkey kidney cells following infection with a SV40 β-globin recombinant genome. Nature 277:108-114.

Musin, B.S. 1969. Myoglobin content in skeletal and heart muscles of dogs at different age periods. Byulleten Eksperimental noi Biologii i Meditsiny, 67:10-12.

Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. and Ogata, K. 1984. Alternative transcription and two modes of splicing result in two myosin light chains from one gene. Nature 308:333-338.

Nadal-Ginard, B. 1978. Commitment, fusion and biochemcial differentiation of a myogenic cell line in the absence of DNA synthesis. Cell 15:855-864.

Nadal-Ginard, B., Breitbart, R.E., Strehler, E.E., Ruiz-Opazo, N., Periasamy, M. and Mahdavi, V. 1986. Alternative splicing: A common mechanism for the generation of contractile protein diversity. In: Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Volume 29. (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp. 387-410. Alan R. Liss, Inc., New York.

Nelson, C., Crenshaw III, E.B., Franco, R., Lira, S., Albert, V.R., Evans, R.M. and Rosenfeld, M.G. 1986. Discrete cis-active genomic sequences dictate the pituitary cell type-specific expression of rat prolactin and growth hormone genes. Nature 322:557-562.

Nemeth, P., Hoffer, H.W. and Pette, D. 1979. Metabolic heterogeneity of muscle fibres classified by myosin ATPase. Histochem. 63:191-201.

Nevo, E. 1979. Adaptive convergence and divergence of subterranean mammals. Annu. Rev. Ecol. Syst. 10:269-308.

Nguyen, H.T., Medford, R.M. and Nadal-Ginard, B. 1983. Reversibility of muscle differentiation in the absence of commitment: Analysis of a myogenic cell line temperature-sensitive for commitment. Cell 34:281-293.

Nishioka, Y. and Leder, P. 1979. The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. Cell 18:875-882.

Nobbs, C.C., Watson, H.C. and Kendrew, J.C. 1966. Structure of deoxymyoglobin: a crystallographic study. Nature 209:339-341

Nudel, U., Greenberg, D., Ordahl, C.P., Saxel, O., Neuman, S. and Yaffe, D. 1985. Developmentally regulated expression of a chicken muscle-specific gene in stably transfected rat myogenic cells. Proc. Natl. Acad. Sci. USA. 82:3106-3109.

Obitana, T., Kawashima, M., Kitani, S., Saitoh, O., Masaki, T., Bader, D.M. and Fischman, D.A. 1986 In: Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Volume 29. (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp. 143-152. Alan R. Liss, Inc., New York.

Ok-Ryun, C. and Engel, J.D. 1986. A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. Nature 323, 731-734.

Ontell, M. 1982. The growth and metabolism of developing muscle. In: Biochemical Development of the Foetus and Neonate, (ed. C.T. Jones) pp.213-241. Elsevier/North Holland Biochemical Press, Amsterdam.

Ordahl, C.P. and Cooper, T.A. 1983. Strong homology in the promoter and 3'-untranslated regions of chick and rat α-actin genes. Nature 303:348-349.

Ott, M-O., Sperling, L., Herbomel, P., Yaniv, M. and Weiss, M.C. 1984. Tissue-specific expression is conferred by a sequence from the 5' end of the rat albumin gene. The EMBO J. 3:2505-2510

Ozawa, E., Saito, K., Hagiwara, A.S., Saito, K. and Atsumi, A. 1986. Inorganic iron supports muscle growth as iron-bound transferrin does. In: Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Volume 29. (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp. 143-152. Alan R. Liss, Inc., New York.

Paterson, B.M., Eldridge, J., Seiler-Tuyns, A. 1986. Regulation of actin gene expression : analysis of potential regualtory regions. In: Molecular Biology of Muscle Development. UCLA Symposia on Molecular and Cellular Biology. Volume 29. (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.521-529. Alan R. Liss. Inc., New York.

Pattengale, P.K. and Holloszy, J.O. 1967. Augmentation of skeletal muscle myoglobin by a program of treadmill running. Am. J. Physiol. 213:783-785.

Pearson, M.L. 1980. Muscle differentiation in cell culture : A problem in somatic cell and molecular genetics. In: The molecular genetics of development. pp.361-417. (ed. T. Leighton and W.F. Loomis) Academic Press.

Pelham. H.R.B. 1982. A regulatory upstream promoter element in the Drosophila Hsp70 heat-shock gene. Cell 30:517-528

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. 1980. The evolution of genes: the chicken preproinsulin gene. Cell 20:555-566.

Perriard, J.C., Achtnich, U., Cerny, L., Eppenberger, H.M., Grove, B.K., Hossle, H.P. and Schafer, B. 1986. In: Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Volume 29. (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp. 143-152. Alan R. Liss, Inc., New York.

Perutz, M.F., Kendrew, J.C. and Watson, H.C. 1965. Structure and function of haemoglobin II. Some relations between polypeptide configuration and amino acid sequence. J. Mol. Biol. 13:669-678.

Peter, J.B., Barnard, R.J., Edgerton, V.R., Gillespie, C.A. and Stempel, K.E. 1972. Metabolic profiles of three fibre types of skeletal muscle in guinea pigs and rabbits. Biochemistry 11:2627-2634.

Pette, D. 1980. Plasticity of muscle. Berlin, Walter de Gruyter.

Pette, D. 1985. Metabolic heterogeneity of muscle fibres. J. Exp. Biol. 115:179-189.

Proudfoot, N.J., Gil, A. and Maniatis, T. 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 31:553-564.

Proudfoot, N.J. 1979. Eukaryotic promoters? Nature 279:376.

Quinn, L.S., Holtzer, H. and Nameroff, M. 1985. Generation of chick skeletal muscle cells in groups of 16 from stem cells. Nature 313:692-694.

Quinn, L.S. and Nameroff, M. 1986. Evidence for a myogenic stem cell. In Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Vol. 29 (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.35-45. Alan R. Liss, Inc. New York.

Rash, J.E. and Staehlin, L.A. 1974. Freeze-cleave demonstration of gap junctions between skeletal myogenic cells in vivo. Dev. Biol. 36:455-461.

Redfern, P.A. 1970. Neuromuscular transmission in newborn rats. J. Physiol. 209:701-709.

Renkawitz, R., Beug, H., Graf, T., Matthias, P., Grez, M. and Schutz, G. 1982. Expression of a chicken lysozyme recombinant gene is regulated by progesterone and dexamethasone after microinjection into oviduct cells. Cell 31:167-176.

Robert, B., Daubas, P., Akimenko, M.-A., Cohen, A., Garner, I., Guenet, J.-L. and Buckingham, M. 1984. A single locus in the mouse encodes both myosin light chains 1 and 3, a second locus corresponds to a related pseudogene. Cell 39:129-140.

Rogers, B.L. and Saunders, G.F. 1985. Transcriptional enhancers play a major role in gene expression. Bioessays 4:62-65.

Romero-Herrera, A.E. and Lehmann, H. 1971. Primary structure of human myoglobin. Nature 232:149-152

Romero-Herrera, A.E., Lehman, H., Joysey, K.A. and Friday, A.E. 1978. On the evolution of myoglobin. Phil. Trans. Royal Soc. Lon. 283: 61-163.

Romero-Herrera, A.E., Lehmann, H., Joysey, K.A. and Friday, A.E. 1973. Molecular evolution of myoglobin and the fossil record : a phylogenetic synthesis. Nature 246:389-395.

Rosenmann, M. and Morrison, P. 1969. Seasonal augmentation of myoglobin in the snow-shoe hare. J. Biol. Chem. 240:3353-3356

Rozek, C.E. and Davidson, N. 1983. Drosophila has one myosin heavy chain gene with three developmentally regulated transcripts. Cell 32:23-34.

Rugh, R. 1968. The mouse : Its reproduction and development. Burgess Minneapolis, Minnesota.

Ruiz-Opazo, N., Weinberger, J. and Nadal-Ginard, B. 1985. Comparison of α-tropomyosin sequences from smooth and striated muscle. Nature 315:67-70.

Rusche, J.R. and Howard-Flanders, P. 1985. Hexamine cobalt chloride promotes intermolecular ligation of blunt end DNA fragments by T4 DNA ligase. nucleic Acids Res. 13:1997-2008.

Salmons, S. and Sreter, F.A. 1976. Significance if impulse activity in the transformation of skeletal muscle type. Nature 263:30-34

Sanger, F., Coulson, A.R., Barrell, B.G.M., Smith, A.J.H. and Roe, B.A. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143:161-178.

Sarver, N., Gruss, P., Law, M.-F., Khoury, G. and Howky, P.M. 1981. Bovine papilloma virus deoxyribonucleic acid: A novel eucaryotic cloning vector. Mol. Cell. Biol. 1:486-496.

Sassone-Corsi, P. and Borrelli, E. 1986. Transcriptional regulation by trans-acting factors. Trends in Genet. 2:215-219.

Schaffner, W., Gross, K., Telford, J. and Birnsteil, M. 1976. Molecular analysis of the histone gene cluster of Psammechinus miliaris : II. The arrangement of the five coding and spacer sequences. Cell 8:471-478.

Schmidt, A., Yamada, Y and de Crombrugghe, B. 1984. DNA sequence comparison of the regulatory signals at the 5'-end of the mouse and chick α2 Type I collagen genes. J. Biol. Chem. 259:7411-7415.

Schneiderman, L.J. 1962. Myoglobin in the human foetus. Nature 194: 191-192.

Scholander, P.F. 1940. Experimental investigations on the respiratory functions in diving mammals and birds. Hvalradets Skrifter Norske Viderskaps-Akad. Oslo 22:5-131.

Schweighoffer, F., Maire, P., Tuil, D., Gaultron, S., Dalgelen, D., Bachner, L. and Kahn, A. 1986. In vivo developmental modifications of the expression of genes encoding muscle specific enzymes in the rat. J. Biol. Chem. 26:10271-10289.

Seiler-Tuyns, A., Eldridge, J.D. and Paterson, B.M. 1984. Expression and regulation of chicken actin genes introduced into mouse myogenic and nonmyogenic cells. Proc. Natl. Acad. Sci. USA. 81:2980-2984.

Serfling, E., Jasin, M. and Schaffner, W. 1985. Enhancers and eukaryotic gene transcription. Trends in Genet. 1:224-230.

Shani, M. 1986. Tissue-specific and developmentally regulated expression of a chimeric actin-globin gene in transgenic mice. Mol. Cell Biol. 6:2624-2631.

Shani, M., Admon, S. and Yaffe, D. 1984. The methylation state of two muscle-specific genes : restriction enzyme analysis did not detect a correlation with expression. Nucleic Acids Res. 12:7225-7234.

Shen, Y-M., Hirschhom, R.R., Mercer, W.E., Surmacz, E., Tsutsui, Y., Soprano, K. and Baserga, R. 1982. Gene transfer : DNA microinjection compared with DNA transfection with a very high efficiency. Mol. Cell Biol. 2:1145-1154.

Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal crossovers. Science 191:528-535

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-617.

Spandidos, D.A. and Paul, J. 1982. Transfer of human globin genes to erythroleukemic mouse cells. The EMBO J. 1:15-20.

Spandidos, D.A. and Wilkie, N.M. 1983. Expression of exogenous DNA in mammalian cells. In: Transcription and Translation. (ed. B.D. Hames and S.J. Higgins). pp 1-48. The practical approach series. IRL Press. Oxford, Washington D.C.

Spector, T. 1978. Refinement of the Coomassie blue method of protein quantitation. Anal. Biochem. 86:142-146.

Staden, R. 1980. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. 8:3673-3694.

Stein, J.P. 1986. Tissue-specific expression of a processed calmodulin gene. In Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Vol. 29 (ed. C. Emerson, D. Fischman, B., Nadal-Ginard and M.A.Q. Siddiqui) pp.35-45. Alan R. Liss, Inc. New York.

Strickland, N.C. and Goldspink, G. 1973. A possible indicator muscle for the fibre content and growth characteristics of porcine muscle. Anim. Prod. 16:135-146.

Strittmatter, W.J. and Couch, C.B. 1986. Myoblast metalloendoprotease: role in myoblast fusion. In Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Vol. 29 (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.167-178. Alan R. Liss, Inc. New York.

Strohman, R.C., Kardami, E., Moss, P., Micou-Eastwood, J. and Spector, D. 1986. Factors and conditions regulating muscle maturation in vitro. In Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Vol. 29 (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.153-165. Alan R. Liss, Inc. New York.

Takano, T. 1977. Structure of myoglobin refined at 2.0A resolution. J. Mol. Biol. 110:537-568.

Taylor, S.M. and Jones, P.A. 1979. Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. Cell 17:771-779.

Thomas, P.S. 1980. Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA. 77:5201-5205.

Tipler, T.D., Edwards, Y.H. and Hopkinson, D.A. 1978. Developmental changes in protein profiles of human cardiac and skeletal muscle. Ann. Hum. Genet. Lond. 41:409-418.

Toyota, N. and Shimada, Y. 1983. Isoform variants of troponin in skeletal and cardiac muscle cells cultured with and without nerves. Cell 33:297-304.

Turner, D.C., Gmur, G., Sieguist, M., Burckhardt, E. and Eppenberger, H.M. 1976. Differentiation in cultures derived from embryonic chicken muscle. Dev. Biol. 48:258-283.

Twigg, A.J. and Sherratt, 1980. Trans-complementable copy number mutants of plasmid ColE1. Nature 238:216-218.

Vandekerckhove, J. and Weber, K. 1978. At least six different actins are expressed in a higher mammal : an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. J. Mol. Biol. 126:783-802.

Vandekerckhove, J. and Weber, K. 1979. The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle, and rabbit slow skeletal muscle. Differentiation 14:123-133.

Van den Berg, J., Van Ooyen, A.V., Mantei, N., Schambock, A., Grosveld, G., Flavell, R.A. and Weissmann, C. 1978. Comparison of cloned rabbit and mouse β-globin genes showing evolutionary divergence of two homologous pairs of introns. Nature 276:37-44

Vanin, E.F. 1984. Processed psuedogenes : characteristics and evolution. Biochem. Biophys. Acta. 782:231-241.

Van het Schip, F., Strijker, R., Samallo, J., Gruber, M. and Geert, A.B. 1986. Conserved sequence motifs upstream from the co-ordinately expressed vitellogenin and apo VLDL II genes of chicken. Nucleic Acids Res. 14:8669-8680.

Vaughan, B.E. and Pace, N. 1956. Changes in myoglobin content of the high altitude acclimatized rat. Amer. J. Physiol. 185:549-556.

Wakabayashi, S., Matsubara, H. and Webster, D.A. 1986. Primary sequence of a dimeric bacterial haemoglobin from Vitroscilla. Nature 322:481-483.

Wakelam, M.J.O. 1985. The fusion of myoblasts. Biochem. J. 228:1-12.

Walker, M.D., Edlund, T., Boulet, A.M. and Rutter, W.J. 1983. Cell-specific expression controlled by the 5'-flanking region of insulin and chymotrypsin genes. Nature 306:557-561.

Walsh, F.S. and Phillips, E. 1981. Specific changes in cellular glycoproteins and surface proteins during myogenesis in clonal muscle cells. Dev. Biol. 81:229-237.

Weil, P.A., Donal, S.L., Segall, J. and Koeder, R.G. 1979. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18:469-484.

Weller, P., Jeffreys, A.J., Wilson, V. and Blanchetot, A. 1984. Organisation of the human myoglobin gene. The EMBO J. 3:439-446.

Weller, P.A., Price, M., Isenberg, H., Edwards, Y.H. and Jeffreys, A.J. 1986. Myoglobin expression : early induction and subsequent modulation of myoglobin and myoglobin mRNA during myogenesis. Mol. Cell Biol. 6:4539-4547.

Whalen, R.G., Sell, S.M., Butler-Browne, G.S., Schwartz, K., Bouveret, P. and Pinset-Harstrom, I. 1981. Three myosin heavy chain isoenzymes appear sequentially in rat muscle development. Nature 292:805-809.

Whalen, R.G. 1980. Contractile protein isozymes in muscle development: The embryonic phenotype. In: Plasticity of muscle (ed. D. Pette) Berin, Walter de Gruyter, pp.177-190.

Whalen, R.G. 1985. Myosin isoenzymes as molecular markers for muscle physiology. J. Exp. Biol. 115:43-54

Whalen, R.G., Butler-Browne, G.S., Pinset, C, Toutant, M., Watkins S.C., Ajioka, C.L., McCormick, D. and Riley, G.P. 1986. Control of myosin isoform expression during skeletal muscle development in rodents. In Molecular Biology of Muscle Development UCLA Symposia on Molecular and Cellular Biology. Vol. 29 (ed. C. Emerson, D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui) pp.237-251. Alan R. Liss, Inc. New York.

White, C.T., Hardies, S.C., Hutchinson, C.A. and Edgell, M.H. 1984. The diagonal-transverse homology search algorithm for locating similarities between two sequences. Nucleic Acids Res. 12:751-766.

Wickens, M.P. and Laskey, R.A. 1981. Expression of cloned genes in cell-free systems and in microinjected Xenopus oocytes. In Genetic Engineering Vol. 1 (ed. R. Williamson) pp.103-167. Academic Press.

Wieringa, B., Meyer, F., Reiser, J. and Weissmann, C. 1983. Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β-globin gene. Nature 301:38-43.

Wigler, M., Pellicer, A., Silverstein, S. and Axel, R. 1978. Biochemical transfer of single-copy eucaryotic genes using total cellular DNA as donor. Cell 14:725-731.

Wilderman, A.G., Sassone-Corsi, P., Gundstrom, T.,Zenke, M. and Chambon, P. 1984. Stimulation of in vitro transcription from the SV40 early promoter by the enhancer involves a specific trans-acting factor. The EMBO J. 3 :3129-3133

Wittenberg, J.B. 1970. Myoglobin-facilitated oxygen diffusion : role of myoglobin in oxygen entry into muscle. Physiol. Rev. 50:559-636.

Wittenberg, B.A., Wittenberg, J.B. and Caldwell, P.R.B. 1975. Role of myoglobin in the oxygen supply to red skeletal muscle. J. Biol. Chem. 250:9038-9043.

Wood, D., Blanchetot, A. and Jeffreys, A.J. 1982. Molecular cloning of seal myoglobin mRNA. Nucleic Acids Res. 10:7133-7144.

Wright, S., de Boer, E., Grosveld, F.G. and Flavell, R.A. 1983. Regulated expression of the human β-globin gene family in murine erythrolenkemic cell hybrids. Nature 305:333-336

Wright, S., Rosenthal, A., Flavell, R. and Grosveld, F. 1984. DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell 38:265-273.

Wright, W.E. 1984. The induction of muscle genes in neural cells. J. Cell Biol. 98:427-435.

Yaffe, D. 1968. Retention of differentiation potentialities during prolonged cultivation of myogenic cells. Proc. Natl. Acad. Sci. USA. 61:477-483.

Yaffe, D. 1969. Cellular aspects of muscle differentiation in vitro. Curr. Top. Dev. Biol. 4:37.

Yanish-Perron, C., Vieira, J. and Messing, J. 1985. Improved M13 phage cloning vectors and host strains : nucleotide sequences of M13mp18 and pUC19 vectors. Gene 33:103-119.

Yisraeli, J., Adelstein, R.S., Melloul, D., Nudel, U., Yaffe, D. and Cedar, H. 1986. Muscle-specific activation of a methylated chimeric actin gene. Cell 46:409-416.

Zak, R. 1974. Development and proliferate capacity of cardiac muscle cells. Circ. Res. 34/35 (Suppl. II) : 11-17.

Zak, R. 1981. Contractile function as a determinant of muscle growth, In: Cell and muscle motility.Vol. 1, pp.1-33. (ed. R.M. Dowben and J.W. Shay) Plenum Press.

Analysis of the structure and expression of mammalian myoglobin genes

Melanie Price

The myoglobin gene is a distant member of the globin gene superfamily, expressed in muscle rather than erythroid cells. In vertebrate muscle, myoglobin is the principle haemoprotein and serves to facilitate the diffusion of oxygen from the vascular system to the muscle mitochondria.

A single gene encodes myoglobin of both cardiac and skeletal muscle in the mouse. The characterisation of the mouse myoglobin gene established that it has the same three exon/two intron structure found in the α- and β-globin genes but differs from haemoglobin genes in having very long non-coding DNA regions. These features are shared by human and seal myoglobin genes.

Sequence comparison of human, seal and mouse myoglobin genes revealed a highly conserved upstream domain, approximately 120 bp before the myoglobin gene CAP site and extending over 200 bp. Although the position before the CAP site suggests it may play some regulatory role, this sequence has no homology to conserved 5'-flanking sequences described for contractile protein genes.

As expected from the early appearance of low levels of myoglobin mRNA in embryonic skeletal muscle, myoglobin gene expression is induced in both rat and mouse embryonic myoblasts during differentiation *in vitro*, concomitant with the contractile protein genes.

Myoglobin-CAT gene fusions containing ~1 kb of 5'-myoglobin DNA, including the conserved domain, are expressed on introduction into mouse myoblasts and furthermore, expression increases during myoblast differentiation over a transient expression period. This should provide a good *in vitro* system to study myoglobin gene expression at the molecular level during myogenesis.

UNIVERSITY LIBRARY LEICESTER

18 JUN 1987