

ABSTRACT

THE ISOLATION OF A MAMMALIAN MYOGLOBIN GENE

D. WOOD

Myoglobin is the major haemoprotein of vertebrate muscle where it facilitates oxygen diffusion from the blood to the mitochondria and acts as an oxygen store. Elevated levels of myoglobin are found in diving mammals and birds.

Myoglobins and haemoglobins form part of the globin superfamily and are related evolutionarily. The ancestral globin gene underwent a duplication between 500 and 800 million years ago to give rise to monomeric myoglobins and the haemoglobin lines. The globin superfamily of proteins has been extensively characterised and sperm whale myoglobin was the first protein whose three-dimensional structure was determined. In addition the haemoglobin genes represent the best characterised multigene family. However the myoglobin gene has not been investigated. In order to understand the relationship of myoglobin genes in relation to other globin genes from an evolutionary and structural aspect, a mammalian myoglobin gene has been isolated.

After unsuccessful attempts to clone myoglobin messenger RNA from human muscle, grey seal muscle containing high levels of myoglobin was used to prepare polyadenylated mRNA. cDNA was prepared by reverse transcription of polyadenylated mRNA and cloned into a plasmid vector. 4% of cDNA recombinants contained myoglobin cDNA inserts. One clone, shown to contain only 3' non-translated mRNA sequences, was used to determine that grey seal myoglobin is coded by a single gene which is transcribed to give a 1400 nucleotide mRNA considerably longer than related haemoglobin mRNAs. This cDNA clone was also used to isolate a myoglobin genomic clone from grey seal DNA cloned into a λ bacteriophage vector. The 3' end of this gene has been completely sequenced and the gene is interrupted by an intervening sequence at codon 105, a position precisely homologous to that in α - and β -globin genes. The 3' exon of the grey seal myoglobin gene is remarkably long and comprises codons 106-153, which give an amino acid sequence identical to that of the harbour seal myoglobin, and 548 bp of untranslated sequence. This untranslated sequence contains a typical polyadenylation signal and is much longer than the corresponding sequences of α - and β -globin genes.

THE ISOLATION OF A MAMMALIAN MYOGLOBIN GENE

by

David Wood

Thesis submitted for the degree of Doctor of Philosophy

in the University of Leicester

1984

UMI Number: U356097

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 U356097

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

Summary

Myoglobin is the major haemoprotein of vertebrate muscle where it facilitates oxygen diffusion from the blood to the mitochondria and acts as an oxygen store. Elevated levels of myoglobin are found in diving mammals and birds.

Myoglobins and haemoglobins form part of the globin superfamily and are related evolutionarily. The ancestral globin gene underwent a duplication between 500 and 800 million years ago to give rise to monomeric myoglobins and the haemoglobin lines. The globin superfamily of proteins has been extensively characterised and sperm whale myoglobin was the first protein whose three-dimensional structure was determined. In addition the haemoglobin genes represent the best characterised multigene family. However the myoglobin gene has not been investigated. In order to understand the relationship of myoglobin genes in relation to the other globin genes from an evolutionary and structural aspect, a mammalian myoglobin gene has been isolated.

After unsuccessful attempts to clone myoglobin messenger RNA from human muscle, grey seal muscle containing high levels of myoglobin was used to prepare polyadenylated mRNA. cDNA was prepared by reverse transcription of polyadenylated mRNA and cloned into a plasmid vector. 4% of cDNA recombinants contained myoglobin cDNA inserts. One clone, shown to contain only 3' non-translated mRNA sequences, was used to determine that grey seal myoglobin is coded by a single gene which is transcribed to give a 1400 nucleotide mRNA considerably longer than related haemoglobin mRNAs. This cDNA clone was also used to isolate a myoglobin genomic clone

from grey seal DNA cloned into a λ bacteriophage vector. The 3' end of this gene has been completely sequenced and the gene is interrupted by an intervening sequence at codon 105, a position precisely homologous to that in α - and β -globin genes. The 3' exon of the grey seal myoglobin gene is remarkably long and comprises codons 106-153, which give an amino acid sequence identical to that of the harbour seal myoglobin, and 548 bp of untranslated sequence. This untranslated sequence contains a typical polyadenylation signal and is much longer than the corresponding sequences of α - and β -globin genes.

Acknowledgements

All the work in this thesis was carried out in the Genetics Department of Leicester University and I am grateful to all the members, past and present who have given me help and advice.

In particular I should like to thank Dr Alain Blanchetot for his help with some of the cloning experiments; Ted Murray for initial work on cDNA synthesis; Victoria Wilson for expert technical help and Sheila Mackley for typing this thesis.

Finally, I should like to thank my supervisor, Dr Alec Jeffreys for his invaluable help, understanding and encouragement throughout the project.

Publications

Some of this work has been published.

"Molecular cloning of seal myoglobin mRNA" by D. Wood, A. Blanchetot and A.J. Jeffreys (1982), Nucleic Acids Research, 10, 7133-7144.

"The seal myoglobin gene : an unusually long globin gene" by A. Blanchetot, V. Wilson, D. Wood and A.J. Jeffreys (1983), Nature, 301, 732-734.

Abbreviations used

DEAE-cellulose	diethylaminoethyl-cellulose
DEPC	diethylpyrocarbonate
DTNB	dithionitrobenzene
DMSO	dimethylsulphoxide
EDTA	ethylenediaminetetra-acetic acid
IMS	industrial methylated spirit
PMSF	phenylmethylsulphonyl fluoride
s.d.s.	sodium dodecyl sulphate
SSC	saline sodium citrate
TEMED	NNN'N' tetramethyl ethylenediamine
DNA	deoxyribonucleic acid
cDNA	complementary DNA
RNA	ribonucleic acid
mRNA	messenger RNA
rRNA	ribosomal RNA
tRNA	transfer RNA
poly(A) ⁺ RNA	polyadenylated RNA
poly(A) ⁻ RNA	non-polyadenylated RNA
C-terminal	carboxy terminal
N-terminal	amino terminal
Mb	myoglobin
LC	light chain

Abbreviations Continued

bp	base pairs
kb	kilobase pairs
kd	kilodaltons
nt	nucleotides
PAGE	polyacrylamide gel electrophoresis

Contents

	Page
Chapter 1	
INTRODUCTION	1
1. MYOGLOBIN	1
(i) Historical Aspects	1
(ii) Structure	2
(iii) Function	4
(iv) Myoglobin Content of Foetal Muscle	6
(v) Evolution	7
2. MUSCLE AND MYOGENESIS	12
3. THE MOLECULAR BASIS OF MUSCLE CONTRACTION	12
4. THE CONTRACTILE PROTEINS OF MUSCLE	13
(i) Actin	13
(ii) Myosin	13
(iii) Tropomyosin	13
(iv) Troponin	14
(v) Minor Proteins	14
5. PROTEIN SYNTHESIS IN MUSCLE	14
6. ACTIN GENES	16
7. MYOSIN GENES	21
8. TROPOMYOSIN AND TROPONIN GENES	23
9. THE GLOBIN GENES	25
10. INTERVENING SEQUENCES IN GLOBIN GENES	28
11. LEGHAEMOGLOBIN GENES	33
12. OBJECT OF THE RESEARCH	36
Chapter 2	
MATERIALS AND METHODS	37

	Page
Chapter 3	DETECTION OF HUMAN MYOGLOBIN mRNA 76
1.	PURIFICATION OF HUMAN MYOGLOBIN 76
2.	ISOLATION AND ANALYSIS OF POLYADENYLATED RNA 76
	FROM HUMAN SKELETAL MUSCLE
3.	<u>IN VITRO</u> TRANSLATION OF HUMAN SKELETAL MUSCLE RNA 79
4.	IDENTIFICATION OF MYOGLOBIN IN <u>IN VITRO</u> TRANSLATION 80
	PROFILES BY IMMUNOPRECIPITATION
Chapter 4	THE CONSTRUCTION AND ANALYSIS OF A cDNA LIBRARY 81
	FROM ADULT HUMAN SKELETAL MUSCLE POLYADENYLATED RNA
1.	FIRST STRAND SYNTHESIS 81
2.	SECOND STRAND SYNTHESIS 82
3.	HOMOPOLYMER TAILING OF cDNA AND pAT153 84
4.	TRANSFORMATION OF <u>E.coli</u> HB101 85
5.	SCREENING THE HUMAN cDNA CLONE BANK 86
6.	POSITIVE SELECTION 87
7.	MAPPING AND SEQUENCING pLG667 cDNA INSERT 89
Chapter 5	ISOLATION AND CHARACTERISATION OF POLYADENYLATED 90
	RNA FROM GREY SEAL SKELETAL MUSCLE
1.	PURIFICATION OF GREY SEAL MYOGLOBIN 90
2.	ISOLATION OF POLYADENYLATED RNA 90
3.	<u>IN VITRO</u> TRANSLATION OF GREY SEAL POLYADENYLATED 91
	RNA
4.	IDENTIFICATION OF MYOGLOBIN IN <u>IN VITRO</u> TRANSLATIONS 92
	BY PURIFICATION FROM ISO-ELECTRIC FOCUSING
	POLYACRYLAMIDE GELS

	Page
Chapter 6	
THE CONSTRUCTION AND ANALYSIS OF A cDNA LIBRARY	93
FROM GREY SEAL SKELETAL MUSCLE POLYADENYLATED RNA	
1. CONSTRUCTION OF A GREY SEAL cDNA LIBRARY	93
2. IDENTIFICATION OF MYOGLOBIN cDNA CLONES	94
3. DNA SEQUENCE ANALYSIS OF MYOGLOBIN cDNA CLONES	96
4. ABUNDANCE OF MYOGLOBIN mRNA IN SEAL MUSCLE	97
POLYADENYLATED RNA	
5. DETECTION OF THE GREY SEAL MYOGLOBIN GENES BY	98
SOUTHERN BLOT ANALYSIS	
Chapter 7	
CLONING THE GREY SEAL MYOGLOBIN GENE	99
1. CONSTRUCTION OF A GENOMIC LIBRARY	99
2. SUBCLONING AND SEQUENCING	103
Chapter 8	
DISCUSSION	105
1. INTRODUCTION	105
2. cDNA CLONING	107
3. ISOLATION OF A MYOGLOBIN cDNA	112
4. ISOLATION OF A MYOGLOBIN GENE	116
5. PROSPECTS	118
6. CONCLUDING REMARKS	120
Bibliography	121

Chapter 1

INTRODUCTION

1. MYOGLOBIN

(i) Historical Aspects

In 1803 Xavier Bichat reported on the properties of a red pigment of muscle indicating its distinctiveness from the pigment of the circulating haemoglobin. He observed the increase in muscle pigmentation with age and that many red blooded animals, such as frogs, may have almost completely white skeletal muscle although rich in blood vessels. Kolliker, in 1850, asserted that the red colour of muscle arose from its own unique pigment which became bright red when oxygenated and darkened in the presence of a reducing agent. In 1897 Morner studied the spectrophotometric differences between the muscle pigment from dogs and canine haemoglobin. Morner suggested the name myochrome for this muscle pigment. However, in 1921, Gunther proposed the name myoglobin in order to emphasize the similarities of the pigment with haemoglobin. Gunther demonstrated quantitatively the deficiency of myoglobin in the skeletal tissue of the foetus and also described several myoglobinuric states in man (Kagen, 1973).

Whipple developed techniques for the measurement of myoglobin in muscle and in 1926 demonstrated the increase in canine muscle myoglobin with age and after prolonged exercise. In 1928 Whipple and Woodruff demonstrated the lowered content of myoglobin in the muscle of inactive patients with chronic wasting diseases (Kagen, 1973).

The distinctiveness of myoglobin and haemoglobin was finally demonstrated by Ray and Paff in 1930 using reflection spectrophotometry

from muscle in situ (Ray and Paff, 1930). In 1932, Theorell succeeded in crystallising myoglobin from horse cardiac muscle. He demonstrated the lower molecular weight of myoglobin compared to haemoglobin and the difference in oxygen binding of the two pigments (Theorell, 1932).

(ii) Structure

Myoglobin is a globular protein composed of a folded polypeptide portion, globin, and a prosthetic group, haem, which contains iron in its ferrous state (FeII). The disc-like water-insoluble haem is inserted into the hydrophobic region of the globin polypeptide. Myoglobin, as well as the other respiratory proteins, has the property of reversibly binding molecular oxygen. This reversible binding of oxygen takes place at the iron within the haem group. In myoglobin and haemoglobin, the oxygen-containing forms maintain the iron in the ferrous state. This special and critical property depends on the linkage between the haem and the globin, and is lost if the globin is denatured. Ferrimyoglobin (metmyoglobin) does not bind oxygen.

Myoglobin is the simplest haemoprotein capable of reversibly binding molecular oxygen, containing one polypeptide chain and one haem group per molecule. Haemoglobins are usually composed of several subunits, each containing a globin polypeptide chain with an associated haem group. Mammalian haemoglobins are made up of four such subunits, two identical α -like chains and two identical β -like chains.

Myoglobins have been isolated from the skeletal muscle, and sometimes the cardiac muscle, of a large number of animal species. The physical structure of myoglobin was elucidated by Kendrew and his colleagues in a number of pioneering studies (Kendrew et al., 1958, 1960, 1961, Kendrew, 1963; Nobbs et al., 1966). X-ray crystallographic tech-

niques were used to determine the three-dimensional structure of sperm whale myoglobin. A recent redetermination of the structure has been published by Takano (1977). From these studies the picture of myoglobin as a spherical globular protein has emerged. Most of the polypeptide chain is arranged as an alpha helix, with the majority of charged polar groups lying near the surface and the nonpolar groups directed inward. The haem moiety lies in a pocket lined with nonpolar groups. The protein consists of eight connected pieces of alpha helix with 121 of its 153 residues in these helical regions.

The amino acid composition and the amino acid sequence have been determined for myoglobins of several species (Romero-Herrera et al., 1978). All mammalian myoglobins so far sequenced are 153 amino acids in length, with a molecular weight of around 17 kd, compared with 146 amino acids for β -like globins and 141 for α -like globins.

Molluscan myoglobin may be dimeric with a molecular weight of around 34 kd, (Geraci et al., 1977; Read, 1967) or may occur in the monomeric form as in higher animals. This dimeric myoglobin has a sigmoidal oxygen dissociation curve similar in appearance to that for haemoglobin (Terwilliger and Read, 1971). Monomeric myoglobin shows a characteristic hyperbolic oxygen dissociation curve and has a greater affinity for oxygen at low oxygen tensions as first described by Theorell (1934).

The amino acid sequence of human skeletal muscle myoglobin is identical to that of cardiac muscle myoglobin. Similarly, the amino acid content of human adult and foetal myoglobin are also identical (Schneiderman, 1962).

Only five human myoglobin variants have been described (Boyer et al., 1963; Boulton et al., 1969, 1970, 1971, a, b, c). All five variants result from single amino acid substitutions and have no apparent effect on function and do not appear to be associated with disease. Study of these variants in heterozygotes, who show equal amounts of variant and normal myoglobin, suggests that the myoglobin gene is present as a single copy (Boyer et al., 1963).

(iii) Function

Myoglobin has been found in the muscle of animals from many different phyla. In molluscs it is found in buccal, adductor, radular and heart muscle (Rossi-Fanelli et al., 1958; Manwell, 1963; Koppenheffer and Read, 1969; Geraci et al., 1977). In higher animals myoglobin forms the major haemoprotein of cardiac and red skeletal muscle. The concentration of myoglobin is higher in muscles which carry out sustained or periodic work with a fairly slow period (Millikan, 1939). In birds myoglobin is found in the breast muscle of active forms such as geese and ducks but completely absent in the white breast muscle of the nonflying chicken. The remaining skeletal muscle of chicken has the same concentration of myoglobin as that of duck and geese (Kagen and Linder, 1968).

In fish the mass of the body musculature is white, but the lateral band muscles used for slow sustained movements are deeply red (Barets, 1961; Tekeuchi, 1959). Gill muscle is also red, requiring myoglobin to sustain the constant movement of water over the gills surface.

The colour of red muscle is due to the presence of both myoglobin and cytochromes. Lawrie (1952, 1953a,b), described the correlation between the redness of muscle and myoglobin content. Muscles richest in myoglobin are those which show constant slow work of long duration. Thus there is three times the level of myoglobin in the pectoral muscle of the ocean

going Manx shearwater, with relatively slow wing movements, than in the same muscle of the land based pigeon. The myoglobin content of particular skeletal muscles seems related to its capacity for oxidative metabolism. Cardiac muscle is rich in myoglobin and depends largely on oxidative metabolism whereas muscle fibres with less myoglobin have greater glycolytic activity (Kagen, 1973).

Although much is known about the structure of myoglobin, little is known of its formation and destruction. Myoglobin in the muscle fibre sarcoplasm is probably in solution. During prolonged tonic contractions of red muscle, the blood supply to the tissue is diminished and some of the oxygen requirement is provided from that stored in the myoglobin. Thus myoglobin is found in higher amounts in animals regularly deprived of access to air for short periods of time. The myoglobin content of muscle of the surface diving dolphins which dives for about two minutes is around 7 mg/g wet weight (Lehmann, 1974) whereas land mammals such as man and cattle have between 3 and 5 mg/g wet weight myoglobin in skeletal muscle. Deeper diving seals have around seven times the amount of myoglobin as the skeletal muscle of man (Robinson, 1939) while the very deeply diving sperm whale has up to ten times the myoglobin content of man (Wittenberg, 1970). During the prolonged dives of sperm whales, the arterioles to the skeletal muscles constrict, diverting blood to the vital organs, while the muscles derive their oxygen from the rich myoglobin stores.

Myoglobin can combine with oxygen at lower oxygen tensions than haemoglobin. The mechanics of the action of myoglobin and its function were reviewed by Millikan as early as 1939. One of the prime functions of myoglobin is in facilitated oxygen diffusion. Myoglobin augments the flow of oxygen toward the mitochondria and enables the muscle to obtain a

higher level of sustained work than could occur in its absence. Wittenberg (1970) and Wittenberg et al. (1975) described how myoglobin facilitates oxygen diffusion in muscle tissue. They suggest that a steady state of oxygen supply and steep gradients of free and myoglobin-bound oxygen are necessary for myoglobin to enhance the flux of oxygen within the muscle fibre.

The role of myoglobin as a storage mechanism within the cell is of considerable importance. Such a role is supported by the rich myoglobin content of muscle of deep-diving mammals. It seems likely that myoglobin acts as a short-term store and releases its oxygen during prolonged muscular contractions.

(iv) Myoglobin Content of Foetal Muscle

As early as 1803 Bichat observed that foetal skeletal muscle is deficient in myoglobin. The influence of maturation on the myoglobin content of muscle is a good example of muscle differentiation. Kagen and Christian (1966) determined that human adult skeletal muscle contained 4-5 mg of myoglobin per gram wet weight, and muscle of children between 1 and 17 months of age had approximately one third of this value. The skeletal muscle of newborn infants and fetuses was found to contain less than one per cent of the myoglobin content of adult skeletal muscle. Similar findings have been found for a range of mammals and birds (Lawrie, 1950; Musin, 1969; Froning et al., 1968; Kagen and Linder, 1968).

The developmental appearance of a number of muscle proteins, including myoglobin, in human skeletal and cardiac muscle has been studied by Tipler et al. (1978). They observed that myoglobin was not detectable in skeletal muscle tissue until after 24 weeks gestation, although it is readily detectable in cardiac muscle at 20 weeks gestation. It seems

therefore that the oxygen storage and transport functions of myoglobin are required by foetal cardiac muscle earlier in development than in the relatively quiescent foetal skeletal muscle. These workers also observed no differences between adult and foetal myoglobin in size or isoelectric point.

(v) Evolution

Early studies in molecular evolution concentrated on the comparison of amino acids of related proteins (Dayhoff, 1972). It was established that homologous protein sequences of different species varied in a phylogenetically consistent manner. It has been found that the more closely related the species the greater the similarity in amino acid sequence. Detailed molecular phylogenies have been deduced which tend to reflect phylogenetic relationships already established from taxonomic and palaeontological studies. Individual protein sequences appear to evolve at a constant rate regardless of the particular lineage being studied (Wilson et al., 1977). Thus the idea of an evolutionary molecular clock has emerged, where the rate of evolution has been found to vary greatly for different proteins but to be quite constant for a given protein independent of lineage.

The molecular evolution of myoglobin at the protein level has been studied extensively (Romero-Herrera et al., 1973; Goodman et al., 1975; Holmquist et al., 1976; Romero-Herrera et al., 1977; Romero-Herrera et al., 1978; Czelusniak et al., 1982). The approach used in these studies has been to compare the differences in the amino acid sequences of the myoglobin from many species and, using fossil evidence to estimate times of divergence of these species, construct cladograms using the fewest possible number of amino acid substitutions. Romero-Herrera et al. (1973) compared the amino acid sequences of myoglobin of eighteen living

species. They constructed a cladogram and derived a possible ancestral myoglobin chain.

In order to study the evolution of myoglobin it is essential to consider the function of the protein. Myoglobin resembles haemoglobin at the haem contacts, but not at those sites responsible for haem-haem interaction in haemoglobin (Romero-Herrera and Lehmann, 1971; Perutz et al., 1965). The internal hydrophobic residues are similar in all globins. Of the 18 amino acid sequences studied by Romero-Herrera et al., (1973) mutations had been fixed at only sixty seven sites out of 153. Sixty of these involve the 120 surface residues and only seven of the thirty three internal residues. Thirteen residues participate in maintaining the shape of the myoglobin monomer (Englander and Staley, 1969). These residues tend to differ in the globin families, but amongst the myoglobins they are highly conserved. An invariant leucine residue in myoglobin (G5 Leu) becomes a phenylalanine residue in haemoglobins, which is involved in maintaining a more upright position for the haem plate in deoxyhaemoglobin (Morimoto et al., 1971). Interestingly the abnormal haemoglobin "Heathrow" has a leucine (β G5) at this position (White et al., 1973) and shows a very high oxygen affinity, which suggests that this residue is important in the high oxygen affinity of myoglobin. In myoglobin H23 Tyr is in a helical part of the protein but in a non-helical part (HC2) in haemoglobin (Perutz, 1965). This invariant residue is involved in the transition from the deoxyhaemoglobin to the oxyhaemoglobin tetramer (Perutz, 1970). In the myoglobins an external flat plane is formed by the C helix and the CD interhelical bend with invariant residues at positions C3, C6, CD2, CD5 and CD8. This region may be highly conserved to allow interaction with mitochondria during the facilit-

ation of oxygen diffusion (Scholander, 1965; Wittenberg, 1965). In haemoglobin residues C3 and C6 are involved in maintaining contacts between monomer subunits of the molecule (Perutz et al., 1968) and the conservation of different residues in this part of the myoglobin protein may guarantee that it remains monomeric. The interface residues in haemoglobin are hydrophobic in the haemoglobins, however the corresponding residues in myoglobin are hydrophilic. Ohno (1970) proposes that the C terminal residues in myoglobin have been lost in the haemoglobins in order to allow them to form tetrameric molecules. Romero-Herrera et al. (1978) suggest that the 21 bases of the mRNA coding for seven invariant myoglobin residues at position 133-139 may form a hairpin structure with the bases coding for the C-terminal amino acids in myoglobin at position 146-152. If this secondary structure exists in myoglobin mRNA it may provide stability or function as a recognition site ^{for} translation. This study should be able to answer this question in the myoglobin genes being studied.

Goodman et al. (1975) compared the amino acid sequences of 55 globins including monomeric haemoglobins and myoglobins. They deduced that during the evolutionary transition from monomer to tetramer the residues which acquired cooperative functions changed more rapidly than the other positions, and attributed these changes to Darwinian natural selection to produce more optimal function. In their study a maximum parsimony method was used to reconstruct a genealogy of the globin chains. Their genealogy shows that the separation of mammalian myoglobin branch from the α - β haemoglobin branches occurred after the separation of lamprey globin (Li and Riggs, 1970). They consider that homotetramers preceded heterotetramers in haemoglobin evolution and that the ancestral homo-

tetramer had β -like chains. The haemoglobins of the primitive vertebrates the lampreys are monomeric when oxygenated, but form dimers in the deoxygenated state (Anderson and Gibson, 1971) as well as transitory tetramers (Behlke and Schelee, 1970). Goodman et al. suggest that a similar haemoglobin existed in the earliest vertebrates which released oxygen more readily at reduced pH and in its aggregated state. In order to facilitate more efficient delivery of oxygen to the tissues of fast-moving animals it was necessary to evolve a heterotetrameric haemoglobin. After the β - α duplication, positive selection could act on the α locus while stabilising selection was stronger at the β locus. Once the β_4 type haemoglobin was replaced with the $\alpha_2 \beta_2$ tetramer positive selection for a more specialised β chain intensified. Once this functionally superior tetrameric haemoglobin emerged, more intense stabilising selection occurred at the α and β loci. Goodman et al. (1975) present detailed evidence for this interpretation and go on to suggest that natural Darwinian selection continued to shape the finer adaptations of the protein particularly at the external sites involved in interactions with other molecules.

Subsequent studies of globin and myoglobin evolution utilised more and more available sequence data (Holmquist et al., 1976; Romero-Herrera et al., 1977). The most recent study by Czelusniak et al. (1982) used amino acid sequence data from 195 globin chains of 87 vertebrate and 19 non-vertebrate species. In addition they used nucleotide sequence data to construct a genealogy of 40 α - and β -haemoglobin genes and pseudogenes of eight vertebrate species. They conclude that myoglobin, β -globin, embryonic and adult α -globins of living birds and mammals had their origins early in vertebrate evolution. Rates of amino acid substitutions in the diverging globin lineages were exceptionally

fast. However late in the vertebrate phylogeny amino acid substitutions occurred at a much slower rate. The early period of fast evolution involved amino acid substitutions at positions which acquired new or altered functions. For example after the α - β duplication the fastest rates occurred at residues responsible for subunit cooperativity i.e. $\alpha_1\beta_2$ contacts, Bohr effect sites and 2,3-diphosphoglycerate binding sites. Later in evolution change at these sites occurred at a much lower rate.

Czelusniak et al. suggest that around 700 million years ago molluscan myoglobin diverged from the remainder of the globin lineage and that 500 million years ago agnathan myoglobin diverged from gnathostome haemoglobin. Lamprey myoglobin is cladistically closer to lamprey haemoglobin than any other globin. Around 450-500 million years ago they suggest that a gene duplication occurred giving rise to lineages leading to gnathostome myoglobins and haemoglobins. Further gene duplications occurred in the haemoglobin lineage giving rise to embryonic foetal and adult haemoglobins (see Section 9, The Globin Genes). This study places the myoglobin-haemoglobin gene duplication much closer to the α - β -globin gene duplication than previously suggested (Dayhoff, 1972; Efstratiadis et al., 1980) and it remains to be seen whether this estimate is accurate. The sequence analysis of myoglobin genes from different species should help to estimate more accurately the timing of these events.

The hagfish (Eptatretus stoutii) has several loci for monomeric haemoglobin (Ohno and Morrison, 1966) and exhibits gene duplication with the retention of equivalent function and apparently no divergence. Haemolysates from 12 hagfish revealed five haemoglobin phenotypes with four to six distinct haemoglobins. Each haemoglobin was shown to be monomeric with molecular weights of around 18 kd. Ohno and Morrison postulate that these monomers are controlled by genes at four loci.

2. MUSCLE AND MYOGENESIS

The muscular tissue of man composes around 40% of his total body weight. Muscle converts chemical energy into mechanical work.

The general scheme of muscle development is well understood (reviewed by Fischman, 1972). A mononucleated, mitotically dividing cell population accumulates at sites destined to form muscle. Early in myogenesis certain cells elongate and assume a bipolar, spindle shape. These mononucleated cells show no DNA synthesis or mitotic activity and are called myoblasts.

The next stage in myogenesis is the formation of multinucleated syncytia, termed myotubes, by the cytoplasmic fusion of myoblasts. After cell fusion the bulk synthesis of actin and myosin and other contractile proteins begins, although it appears that myoblast fusion is not an absolute requirement for such synthesis (Buckingham, 1977). Subsequent development of the muscle cell involves the assembly of myofibrils, mitochondrial proliferation, glycogen deposition and innervation.

3. THE MOLECULAR BASIS OF MUSCLE CONTRACTION

Myofibrils have a banded appearance along their length, as a result of the arrangement of partially overlapping arrays of protein filaments constituting the contractile apparatus. The two principal proteins of the contractile apparatus are actin and myosin. Myosin constitutes around 38% of total muscle fibre protein while actin accounts for around 18%. Actin and myosin are arranged into separate filaments which slide past each other during contraction. The lengths of the thick myosin-containing filaments and the thin actin-containing filaments remain constant during contraction. When a muscle is activated an interaction takes place between the myosin and actin molecules which enables the myosin to split ATP at a rapid rate, providing energy required for contractions (Reviewed by Huxley, 1973).

4. THE CONTRACTILE PROTEINS OF MUSCLE

(i) Actin

Actin is a highly conserved protein found in all eukaryotic cells and functionally involved in cell motility, mitosis, muscle contraction and the maintenance of cytoskeletal structure. The α -actin of muscle is a globular protein which in the absence of sodium chloride forms the monomeric G-actin, while under physiological salt conditions actin forms the double stranded polymer, F-actin. The actin polypeptide has a molecular weight of around 43 kd (Lewis *et al.*, 1963).

(ii) Myosin

Myosin is a large and complex molecule composed of two heavy chains of approximately 200 kd each, which are wound together in a double stranded α -helical arrangement. In addition at least four smaller polypeptides, myosin light chains, are located at the head portion of the myosin molecule. Myosin from fast skeletal muscle contains two distinct light chains (21 kd and 16 kd), released on alkali treatment, which seem to be involved in the regulation of the myosin ATPase. Two other light chains (18 kd), released on treatment with dithionitrobenzene (DTNB), seem to be involved in calcium binding (Kendrick-Jones *et al.*, 1976).

(iii) Tropomyosin

Tropomyosin is a major regulatory protein in muscle, involved in the calcium ion dependent interaction of actin and myosin. It is closely associated with actin in muscle fibres. The protein is dimeric, composed of two types of subunit : α -tropomyosin (34 kd) and β -tropomyosin (36 kd) (Laki, 1971).

Different muscle types have different proportions of the two forms, fast muscle having an α to β ratio of 3.8 : 1. The two forms of tropomyosin have similar amino acid compositions.

(iv) Troponin

Troponin acts as a calcium binding protein in muscle cells and acts with tropomyosin on the actin filament to activate muscle contraction. The protein is found in only certain muscle types being absent from smooth muscle and some invertebrate muscles. There appear to be three troponin polypeptides, troponin T (37 kd), troponin I (24 kd) and troponin C (19 kd), (Maruyama, 1971).

(v) Minor Proteins

In addition to the above muscle proteins a number of other minor proteins are found in myofibrils. α -actinin is universally distributed with a molecular weight of 95 kd and is found associated with actin. In skeletal muscle a number of high molecular weight proteins have been identified all of which copurify with myosin, these include the C protein (140 kd) and the B and F proteins (150-180 kd) (Laki, 1971).

5. PROTEIN SYNTHESIS IN MUSCLE

The differentiation of muscle fibres from myoblasts is characterised by striking changes in cellular function and morphology. These changes are accompanied by pronounced changes in the pattern of protein synthesis. Most significantly there is a several hundred-fold increase in the synthesis of the muscle filament proteins actin and myosin, (Devlin and Emerson, 1978).

The changes in muscle protein synthesis during muscle cell differentiation has been extensively reviewed by Buckingham (1977).

The regulation of protein synthesis during muscle differentiation may be at the level of transcription, nuclear or cytoplasmic processing or translation of the relevant mRNA molecules. Expression of the major muscle proteins myosin, troponin and tropomyosin appears to be coordinately regulated.

Early experiments concentrated on the regulation of myosin protein synthesis. Heywood and his colleagues suggested that an mRNA selectivity operates during the initiation steps of protein synthesis (Heywood, 1973; Rourke and Heywood, 1972). Initiation factor 3 (IF3) was found to be responsible for this messenger selectivity. Subsequent studies using both myosin and myoglobin mRNAs and initiation factors from muscle suggests that muscle contains a number of messenger recognition factors (Thompson et al., 1973; Heywood and Kennedy, 1974; Heywood et al., 1974). Initiation factor 3 derived from red muscle is involved in the translation of both myosin and myoglobin mRNAs, while IF3 derived from white muscle is only effective in the translation of myosin mRNA.

Recently Affara et al. (1980 a and b) monitored changes during myogenic differentiation in a number of gene products at a number of levels : the nuclear RNA, the cytoplasmic mRNA and the pattern of polypeptides synthesised. These workers used three cell lines representing different stages in myogenic development; an undifferentiated pluripotent mouse embryonal carcinoma cell line, a committed mouse teratocarcinoma-derived myoblast cell line and a multinucleate terminally differentiated myotube cell line. By following the hybridisation of labelled, single-copy mouse DNA with nuclear RNA and poly(A)+ polysomal RNA from these cell lines, they estimate that there is a 30% increase in nuclear RNA complexity in passing from the pluripotent embryonal carcinoma cell line to the myoblast and myotube stages. In addition the embryonal carcinoma nuclear RNA sequences remain as a subset of the myoblast and myotube stage nuclear RNA. With each stage of differentiation a new group of mRNA sequences enters the polysomes. Complementary DNA (cDNA) probes enriched in the new sequences appearing at the myoblast and myotube stages were prepared.

By hybridisation of these cDNA probes with nuclear RNA from the three stages of differentiation, these workers demonstrated that the expression of these new polysomal sequences is controlled at the level of transcription. In addition, gene sequences specific to the myotube stage were resistant to attack by DNase I in the chromatin of the embryonal carcinoma and myoblast cells, whereas gene sequences expressed at all stages of development are readily digested.

In the second part of their study Affara et al. examined the changes in the pattern of protein synthesis during the terminal differentiation of a mouse myogenic cell line. The levels of α -actin and myosin light chains (LC1, LC2 and LC3), tropomyosin and troponin greatly increases after cell fusion.

Thus, a number of approaches have been used to analyse the control of gene expression during myogenesis. These approaches suggest that control may be at the level of transcription and translation. With recent advances in the field of recombinant DNA technology it has become possible to clone specific DNA sequences containing the genes for many of these contractile proteins. These cloned DNA sequences have been used to study the gene structure and arrangement of actin and myosin genes in a number of species. It should be possible to use these cloned DNAs to study the expression of these muscle genes in both in vivo and in vitro systems.

6. ACTIN GENES

Actin is a highly conserved, ubiquitous protein having been found in all eukaryotic cells from protists to mammals. It forms part of the cytoskeleton and is involved in cell mobility and muscle contraction.

The genes encoding actins form a complex gene family which show differential gene expression during development. There are three forms of actin in mammalian cells, β -actin and γ -actin are cytoplasmic forms found in all cells while α -actin is found only in muscle cells.

α -actin is not observed in myoblast cells prior to cell fusion, but its synthesis is substantially increased during muscle cell differentiation. Amino acid sequencing of cytoplasmic and α -actins show that muscle actins have a different N-terminal sequence from that of cytoplasmic actins. The three different forms of actin appear to be products of different mRNA species. In addition α -actins from smooth and striated muscle have slightly different primary sequences (Firtel, 1981).

Actins have been studied in lower organisms particularly yeast and Dictyostelium. Only one form of actin is found in yeast, consistent with the presence of only one actin gene. In Dictyostelium actin synthesis is developmentally regulated. Several actins have been observed in Dictyostelium, the primary sequences of these actins being more similar to mammalian cytoplasmic actins than to mammalian muscle actins. Dictyostelium actins appear to be encoded by approximately 17 genes (Kindle and Firtel, 1978; McKeown et al., 1978; McKeown and Firtel, 1981).

The number of actin genes seems to be species dependent as do the number and position of intervening sequences within actin genes. The Dictyostelium actin gene appears to have no intervening sequences (Firtel et al., 1979), while the yeast actin gene has one intervening sequence interrupting the fourth codon (Gallwitz and Sures, 1980; Ng and Abelson, 1980). Drosophila has six actin genes with at least one intervening sequence (Fyrberg et al., 1980). Three forms of actin have been more similar to mammalian cytoplasmic actins than to mammalian muscle actins.

Dictyostelium actins appear to be encoded for by approximately 17 genes (Kindle and Firtel, 1978; McKeon et al., 1978; McKeon and Firtel, 1981).

In Drosophila melanogaster three forms of actin have been identified (Actins I, II and III). Actin I is found only in muscle tissue and appears during myogenesis. Tobin et al. (1980) isolated a genomic clone from D. melanogaster using a Dictyostelium cDNA plasmid, and established that the Drosophila genome contained a number of actin coding sequences which they localised at dispersed sites on polytene chromosomes. Fyrberg et al. (1980) concluded that the Drosophila genome contained six different actin genes, each identified in Drosophila melanogaster (Actin I, II and III). Actin I is found only in muscle tissue and appears during myogenesis. Tobin et al. (1980) isolated a genomic clone from D. melanogaster using a Dictyostelium cDNA plasmid, and established that in the Drosophila genome the actin coding sequences are widely dispersed in the polytene chromosomes. The six Drosophila actin genes each code for a different message (Fyrberg et al., 1980). One of these clones λ DmAZ, contains a gene which codes for a non-muscle actin with a single intervening sequence of 1.65 kb. This clone was used to isolate six genomic clones each containing a different Drosophila actin genes (Fyrberg et al., 1981). These genes encode several different proteins, but none comparable in primary structure to vertebrate muscle actins. The position of intervening sequences within these genes is not conserved, one gene is split at codon 13, a second at codon 3 and a third at codon 307. The sizes of these intervening sequences are 0.63 kb, 1.65 kb and 0.357 kb respectively. The variability in the positions and sizes of intervening sequences is in sharp contrast to the situation observed in globin genes (see below).

Fyrberg et al. (1983) have shown that the six Drosophila actin mRNAs accumulate in a stage - and tissue-specific manner. They used the unique 3' untranslated portions of the cloned actin genes to monitor levels of their respective mRNAs in whole organisms and dissected body parts. Two genes act5C and act42A are expressed in undifferentiated cells and probably encode cytoplasmic actins. Two other genes act57A and act87E are expressed mainly in larval, pupal and adult intersegmental muscles, while act88F is expressed in adult thorax muscle and act79B in thorax and leg muscles. These workers postulate that these genes come under regulation of different regulatory molecules. The genes for the cytoplasmic forms may respond to levels of mitogenic compounds. The other genes may respond to molecules synthesised early in muscle cell differentiation.

Sea urchins have around eleven different actin genes, each with several intervening sequences (Durica et al., 1980). Although some actin genes share intervening sequences in common positions, no obvious pattern of intervening sequence position is observable in this gene family, unlike the case in the globin gene family. This difference may reflect some fundamental difference in gene stability within these two gene families. A study of the actin gene family may help to elucidate the function and evolutionary properties of intervening sequences.

The observation that the lower eukaryotes and Drosophila synthesise only cytoplasmic forms of actin, suggests that vertebrate actins evolved after insects separated from the phylogenetic line that gave rise to vertebrates. The report of Storti et al., 1978 seems to contradict Fyrberg's findings that Drosophila only encode cytoplasmic actins. However, it seems that some of these genes code for actin proteins utilised in striated muscle (Davidson et al., 1982).

The α -actin genes of chick have been analysed and there appears to be 4-7 different actin genes. Schwartz et al. (1980) constructed a cDNA probe complementary to the α -actin mRNA, and found that the muscle specific α -actin sequence had only a 70% homology with the non-muscle β - and γ -actin sequences. Ordahl et al. (1980) also constructed cDNA probes complementary to chick α -actin mRNA. One of these probes was used to isolate actin genes from a genomic library of chick DNA. Clones containing α -actin sequences were isolated, and one of these was subsequently sequenced (Fornwald et al., 1982). The gene is interrupted by six very short intervening sequences between 105 and 225 bp long. The position of these six intervening sequences is identical to those found in the rat α -actin gene.

Katcoff et al. (1980) isolated cDNA recombinant plasmids containing sequences complementary to rat non-muscle actin mRNA, muscle actin mRNA and mRNA coding for myosin light chain 2. The rat skeletal muscle α -actin cDNA clones were further analysed (Shani et al., 1981) and one of these was used to isolate the rat skeletal muscle actin gene (Nudel et al., 1982) which was found to contain six intervening sequences.

Minty et al. (1981) constructed a recombinant plasmid molecule containing a cDNA complementary to mouse α -actin mRNA. This recombinant was subsequently used to isolate a cDNA plasmid complementary to a foetal skeletal muscle actin mRNA.

Humphries et al. (1981) used the mouse actin probe to determine that the human genome contains approximately 20 actin genes. These workers isolated 12 actin recombinants from a human genomic library which represent 9 different genes, which probability calculations indicate were picked

from a pool of at least 20 different actin genes. Engel et al. (1981) have used cloned actin genes from Drosophila and chicken to isolate actin containing sequences from a human genomic library, and have reached a similar estimate for the number of human actin genes.

The discordant pattern of intervening sequences in actin genes may be due to the insertion of intervening sequences during evolution or the elimination of intervening sequences from a more complex ancestral pattern. The origin and evolutionary significance of the pattern of intervening sequences in actin genes remains to be investigated in more detail, and may help elucidate the role of these sequences in eukaryotic genes.

7. MYOSIN GENES

Myosin genes for both heavy and light chain myosins have been cloned from a number of different species.

Mutants of the soil nematode Caenorhabditis elegans that show uncoordinated body movements (unc-54) have been analysed (Brenner, 1974). The unc-54 loci specifies a major myosin heavy chain found in the body wall musculature. A second myosin heavy chain is found in the body wall musculature as well as two others in the pharynx. MacLeod et al. (1981) have cloned the unc-54 gene from genomic libraries of the nematode. The gene contains seven introns which do not separate functional domains. Other unc genes in C. elegans code for other contractile proteins, unc-15 for paramyosin and unc-92 for an actin polypeptide.

Bernstein et al. (1983) report the isolation and characterisation of a Drosophila muscle myosin heavy chain gene. The myosin heavy chain protein is present in larval leg and thoracic muscle and is encoded by a

single gene. Chromosomal mapping showed that this gene does not map close to any of the six actin genes but maps to a region containing a cluster of flight muscle mutations.

Recombinant cDNA clones containing sequences of two different myosin heavy chain genes from chicken embryonic skeletal muscle have been constructed by Umeda et al. (1981). A myosin heavy chain gene has also been constructed by cDNA cloning from rat skeletal muscle by Medford et al. (1980). These workers have found that myosin heavy chain genes in rat are tightly linked. In addition a myosin heavy chain gene appears to have 32 intervening sequences which do not separate functional domains, and which are not conserved in position in adult myosin heavy chain genes. There appears to be some 17-20 myosin heavy chain genes in rat. Nudel et al. (1980) have also isolated myosin heavy chain DNA sequences from a rat genomic library and conclude that there are several rat myosin heavy chain genes.

The amino acid sequences of chick myosin light chains have been determined (Matsuda et al., 1981). Skeletal muscle myosin contains two alkali light chains LC_1 and LC_3 . Sequence homology between these and other tissue specific light chains suggest these proteins are related and may have evolved from one ancestral gene. The nucleotide sequence of cloned cDNAs for chick skeletal myosin alkali light chains LC_1 and LC_3 have been determined (Nabeshima et al., 1982) and the sequence encoding the C-terminal 141 amino acid residues and the entire 3'-untranslated region of the LC_1 and LC_3 mRNAs is identical. In contrast, the 5' 108 nucleotides of LC_1 and the 5' 78 nucleotides of LC_3 are different. Nabeshima et al. (1984) have isolated a genomic clone of myosin alkali

light chain which is interrupted by eight intervening sequences. Exons 1 and 4 are specific for LC_1 , while exons 2 and 3 are specific for LC_3 . Exons 5-9 are shared by both LC_1 and LC_3 . Exon 1 contains the 5'-non coding sequence and N-terminal 42 amino acid residues of LC_1 , while exon 2 contains the 5'-non coding sequence and ATG initiation codon of LC_3 . The precursor RNAs for LC_1 and LC_3 are initiated at different sites and are subsequently spliced by different pathways to produce the two mature mRNAs from one gene. The mechanism of this differential splicing is unknown but it seems likely that the secondary or tertiary structure of the precursor mRNA has an important role in selecting the splice site.

DNA sequences corresponding to rat myosin light chain 2 (Katcoff et al., 1980) and mouse myosin light chains 1 and 3 (Robert et al., 1982) have been isolated. One mouse myosin light chain gene appears to have several introns.

8. TROPOMYOSIN AND TROPONIN GENES

Tropomyosins are a family of related proteins widely distributed in nature. Skeletal muscle contains two forms of tropomyosin (α and β). MacLeod (1981) constructed recombinant cDNA plasmids containing sequences complementary to chicken skeletal muscle α -tropomyosin. These plasmids were subsequently sequenced (MacLeod, 1982) and two distinct α -tropomyosin mRNAs were identified specifying a major and a minor α -tropomyosin.

Helfman et al. (1983) used an immunological screen to identify clones encoding chicken tropomyosin. They constructed a cDNA library of around 9000 clones from chicken smooth muscle mRNA using the plasmid expression

vector pUC8, to enable transcription using the lacZ gene. Colonies were screened with rabbit anti-tropomyosin antibody. Two colonies were detected by this screening procedure. The purified plasmids from these colonies selected a particular chicken mRNA which in in vitro translations gave rise to tropomyosin. Direct sequence analysis of the cDNA insert of these clones (600 and 900 base pairs) confirmed that these clones encode portions of the tropomyosin protein.

Hastings and Emerson (1982) used a cDNA cloning approach to investigate muscle gene regulation during differentiation of cultured embryonic myoblasts of quail (Coturnix coturnix). cDNA clones corresponding to α -actin, α -tropomyosin, myosin heavy chain, myosin light chain 2, troponin-C and troponin-I were isolated. In general, sequence analyses of these clones suggest that embryonic quail muscle cultures express adult genes for these proteins rather than a special set of genes encoding embryonic isoforms. Three forms of troponin-I are known, a fast, a slow and a cardiac form. The N-terminus of these proteins is greatly diverged while the C-terminus show some homologies. The fast form of troponin I has been cloned from embryonic quail muscle by Emerson. This gene contains five intervening sequences which appear to separate the gene into exons coding for functional domains. One exon codes for a troponin-C binding domain, while a second codes for an actin binding domain. Troponin-C has a fast and a slow/cardiac form. The fast form troponin-C gene has been cloned and contains a number of intervening sequences. The cardiac and slow forms of troponin-C are encoded by the same gene.

9. THE GLOBIN GENES

The globin gene family represents the best characterised multigene family in higher organisms. Globins are widespread and include tetrameric haemoglobins of higher vertebrates, monomeric haemoglobins of proto-chordates, invertebrate globins, monomeric myoglobins and monomeric leg-haemoglobins found in the root nodules of nitrogen fixing plants.

The globins of man are coded for by two unlinked sets of genes. The α -globin related family is located on chromosome 16 (Deisseroth et al., 1977) and consists of a ζ -globin gene expressed early in embryogenesis and two very similar α -globin genes expressed in the foetus and adult. The β -globin related family is located on the short arm of chromosome 11 (Deisseroth et al., 1978; Jeffreys et al., 1979; Gusella et al., 1979) and consists of an embryonic ξ -globin, two very similar foetal globin genes ($^G\gamma$ and $^A\gamma$), a minor adult δ -globin gene and the major adult β -globin gene. In addition there are an unknown number of myoglobin genes. All of these globin genes show significant amino acid sequence homology and have arisen by a number of gene duplications followed by subsequent sequence divergence. By assuming that globin genes evolve independently and in a clock-like manner it is possible to deduce the timing of these gene duplications (Dayhoff, 1972; Efstratiadis et al., 1980). The most ancient duplication occurred between 500 and 800 million years ago and gave rise to the ancestors of haemoglobin and myoglobin genes. The $\alpha\beta$ -globin gene duplication occurred about 500 million years ago, early in vertebrate evolution. The β -globin gene family evolved more recently with a foetal-adult duplication about 200 million years ago, an ξ - γ duplication around 100 million years ago and the δ - β duplication around 40 million years ago.

The α and β -globin genes have been studied extensively in mammals, birds and amphibians as well as in some plants.

The α -globin gene cluster on human chromosome 16 is arranged within 25 kb of DNA in the order 5'- ξ_2 - $\psi\xi$ - α_2 - α_1 -3'. The β -globin gene cluster on human chromosome 11 is arranged in a 45 kb region of DNA in the order 5'- ξ - γ^G - γ^A - $\psi\beta_1$ - δ - β -3'. All genes are orientated in the same direction and are separated by stretches of intergenic DNA. Only 8% of the DNA in these clusters codes for globin mRNA, and 8% comprises the globin intervening sequences. The function of the remaining intergenic DNA is unknown, but consists of single copy DNA and repetitive elements (Adams et al., 1980; Baralle et al., 1980; Coggins et al., 1980; Fritsch et al., 1980).

The overall arrangement of the globin gene cluster has been shown to be indistinguishable in man, gorilla and baboon (Barrie et al., 1981). These sequences may have been under some selective constraint during the 20-40 million years of recent primate evolution. One possible explanation for the retention of this arrangement may be the existence of chromatin domains; regulation of gene activity may be achieved by modulating the packing conformation of these domains (Van der Ploeg et al., 1980).

New World monkeys and prosimians have different arrangements of the β -globin gene cluster (Barrie et al., 1981). Analysis of the lemur β -globin gene cluster shows that it is the shortest so far described in mammals and is similar in arrangement to that found in the rabbit (Lacy et al., 1979). This suggests that a simple cluster of β -globin genes was established before the radiation of the mammals at least 85 million years ago.

The β -globin cluster of rabbit is arranged in the transcriptional order 5'- β^4 - β^3 - β^2 - β^1 -3' (Hardison et al., 1979). The β^1 gene encodes the adult β -globin chain and the β^3 and β^4 genes encode embryonic and/or foetal globin chains. The β^2 gene is a pseudogene containing frameshift mutations and premature termination codons that render it incapable of encoding a functional β -globin chain (Lacy and Maniatis, 1980).

The β -globin cluster of mouse has also been characterised (Jahn et al., 1980) with the genes arranged in the same transcriptional orientation, 5'- ξ^3 - β^h - β^{h1} - β^{h2} - β^{h3} - β^{maj} - β^{min} -3', in 70 kb of DNA, β^{h2} and β^{h3} are pseudogenes.

α -globin genes have been characterised in man. (Orkin, 1978; Lauer et al., 1980) and the apes (Zimmer et al., 1980) and in the mouse (Leder et al., 1981).

Unlinked clusters of α - and β -globin genes have been characterised in the chicken (Hughes et al., 1979; Engel and Dodgson, 1980). Unlike the mammalian β -globin clusters, the chicken embryonic genes (ρ and ξ) are located at the 5' and 3' ends of the cluster while the hatching (β^H) and adult (β) genes are found at the centre of the cluster in the order, 5'- ρ - β^H - β - ξ -3'.

The adult globin genes in Xenopus laevis have been characterised (Jeffreys et al., 1980; Patient et al., 1980). The major adult α' and β' globin genes are closely linked in the order 5'- α' - β' -3'. This suggests that the initial $\alpha\beta$ globin gene duplication was a tandem duplication and occurred around 500 million years ago, and that these tandem duplicates have remained closely linked in amphibia. These genes have subsequently become unlinked in the lineage leading to birds and mammals. This unlinking may have occurred by some transposition event, a chromosomal translocation or a chromosome duplication to give two unlinked sets of α - β clusters, followed by silencing of linked α - or β -

globin genes. X.laevis has a second $\alpha\beta$ -globin gene cluster coding for minor adult globin chains. This cluster appears to have arisen by tetraploidisation, as the related species X.tropicalis has a single $\alpha\beta$ -globin cluster. X.tropicalis represents a contemporary equivalent of the ancestral species that underwent the tetraploidisation event.

10. INTERVENING SEQUENCES IN GLOBIN GENES

All active vertebrate globin genes studied contain two intervening sequences. In each case the intervening sequences occur at precisely homologous positions in the genes. The human β -globin genes are interrupted between codons 30 and 31 and codons 104 and 105, by intervening sequences of 122-130 bp and 850-904 bp respectively (Efstratiadis et al., 1980). The human α -globin genes, the mouse α - and β -globin genes, the rabbit β -globin genes, the chicken β - and ξ -globin genes and the Xenopus α - and β -globin genes contain intervening sequences at homologous positions (Jeffreys and Flavell, 1977; Dodgson et al., 1979; Hardison et al., 1979; Konkell et al., 1979; Jahn et al., 1980; Lauer et al., 1980; Dolan et al., 1981; Patient et al., 1980). The discontinuous nature of globin genes must have been in existence for at least 500 million years and must have preceded the α - β gene duplication.

A number of sequences similar to sequences coding for globin polypeptides have been found in many species. These sequences however contain changes in the coding sequence including insertions, deletions and changes creating stop codons and missense reading such that no globin polypeptide can be made from these sequences. Such sequences have been termed pseudogenes and have been found in many gene families apart from the globin genes. Low stringency hybridisations have been used to search for

globin related sequences in the β -globin cluster of man. A $\psi\beta 1$ globin gene was detected between the γ^A - and δ -globin genes (Fritsch et al., 1980; Efstratiadis et al., 1980). This pseudogene sequence is present in many primates (Barrie et al., 1981; Barrie, Harris and Jeffreys, personal communication). This pseudogene as well as the $\psi\alpha 1$ and $\psi\zeta$ in man (Proudfoot and Maniatis, 1980), the rabbit $\psi\beta 2$ gene (Lacy and Maniatis, 1980), the goat $\psi\beta^x$ and $\psi\beta^z$ genes (Cleary et al., 1980, 1981) the mouse $\beta h 3$ pseudogene (Jahn et al., 1980) and the lemur $\psi\delta$ gene (Jeffreys et al., 1982) all contain intervening sequences at homologous positions to functional globin genes. One remarkable pseudogene, however, the mouse $\psi\alpha 3$ gene has had both intervening sequences precisely removed during evolution (Nishioka et al., 1980; Vanin et al., 1980). Sufficient base changes have occurred in the $\psi\alpha 3$ gene to distort the translational sequence such that it is unable to encode a functional globin polypeptide. The mechanism of intron loss in the mouse $\psi\alpha 3$ gene is unknown. The possibility that a cDNA copy of the mouse α -globin mRNA has illegitimately recombined into the chromosome seems to be ruled out because homology between the $\psi\alpha 3$ gene and the functional α -globin gene extends beyond the 5' cap site of the mRNA molecule although the possibility of an abnormally long germ line transcript cannot be ruled out. It is possible that the mature mRNA or cDNA may act as a template for gene conversion. In this model the mRNA (or cDNA) hybridises to the gene at a growing fork during DNA replication and the introns are looped out and removed by nicking and exonuclease activities. A third model postulates a retroviral origin of the $\psi\alpha 3$ gene (Flavell, 1982). This model suggests that the mouse α -globin mRNA became incorporated into a retroviral RNA which in turn became integrated into the mouse genome in its DNA form.

It is known that retroviruses can incorporate cellular genes into their RNA (Bishop, 1981) sometimes with the loss of introns (Goff et al., 1980). Further evidence for this mechanism for the origin of the mouse $\Psi\alpha 3$ gene has been provided by Lueders et al. (1982) who show that the $\Psi\alpha 3$ gene is flanked by sequences homologous to retroviral-like mouse intracisternal A particle RNA. These repeats are however invert repeats and occur at a large distance from the $\Psi\alpha 3$ gene.

Intronless pseudogenes have been reported to occur in human β -tubulin genes (Wilde et al., 1982a,b; Gwo-Shu Lee et al., 1983). These workers used the 3' untranslated region of a human β -tubulin cDNA to isolate four distinct regions from a genomic library. One of these regions contained a functional β -tubulin with three intervening sequences all with correct consensus splice sequences. This gene codes for two mRNAs, due to the presence of two polyA addition sites, of 1800 nucleotides and 2600 nucleotides. The three other regions represent intronless pseudogenes, two of these seem to have been derived by integration into the host germ line of cDNA copies of the 1800 nucleotide mRNA and the other by integration of a cDNA copy of the 2600 nucleotide mRNA. They calculate that these integration events occurred 4, 10 and 13 million years ago. The three pseudogenes have a polyA tract and extend to the 5' cap sites and are therefore full length cDNA copies. All three pseudogenes are flanked by short direct repeats of different length with no relationship to each other. All three contain genetic lesions precluding translation to give a functional polypeptide. The mechanism of insertion of these cDNA copies is unknown, however human placental tissue does contain an RNA-dependent DNA polymerase (Nelson et al., 1981) which could reverse transcribe mRNAs expressed in the germ line. Gwo-Shu Lee et al., (1983) predict that in multi-gene families where germ line expression occurs, a significant portion of

sequences will be accounted for by pseudogenes generated via a RNA intermediate.

Similar pseudogenes have been reported for the rat α -tubulin gene (Lemishka and Sharp, 1982) the human metallothionein gene (Karin and Richards, 1982) and a human β -actin gene (Hamada et al., 1982).

In the human immunoglobulin gene family a ψV_k gene has been shown to have introns but with altered splicing signals as well as premature termination codons (Bentley and Rabbitts, 1980). In addition a human immunoglobulin light chain pseudogene has been shown to have lost its J-C intron (Hollis et al., 1982).

Split genes are transcribed into colinear precursors from which the intron transcripts are removed by splicing. Sequence analysis has shown that intron-exon junctions show similarities (Breathnach and Chambon, 1981). If an enzyme system exists for intron removal as in mouse $\psi \alpha 3$ it is interesting to speculate as to why all introns have not been removed. It may be that the enzyme system is very inefficient or that introns could have some function. It is interesting that normal globin genes whose introns have been removed are not expressed (Hamer and Leder, 1979).

Many split genes have arisen by the duplication of already split genes. Gene duplication by unequal crossing over may explain how some split genes may have arisen (Sakano et al., 1979a,b). These workers explain how split immunoglobulin light and heavy chain genes evolved by multiple duplications of an unsplit DNA segment coding for a polypeptide with a similar size to that of the present homologous domains of the heavy chains.

It was suggested by Gilbert (1978) that interrupted genes present a way to facilitate the evolutionary development of more complex proteins. Two previously separate functions might be brought together by recomb-

ination within intervening sequences to place their coding sequences within the same transcriptional unit. Thus a novel genetic function may be generated by the shuffling of exons at the DNA or RNA level. This mechanism has led to the suggestion that exons might encode discrete domains within proteins (Blake, 1979). In the case of the globins exons do seem to encode discrete functional or structural regions in the globin polypeptide (Craik et al., 1980, 1981). The central exon codes for the haem-binding domain of globin, and the other two exon products seem to be required to maintain a stable haemprotein complex. Similarly the four exons of the chicken lysozyme gene each specify different functional regions of the protein (Jeffreys, 1981; Gō, 1982).

Doolittle (1978) suggests that the discontinuous structure of eukaryotic genes is the primitive form and that prokaryotic genes have lost their intervening sequences during evolution. He suggests that in the common ancestor of eukaryotes and prokaryotes replication, transcription and translation was unfaithful, and that genetic information in these cells would be redundant and not tightly organised as in Escherichia coli. Intervening sequences would ensure that transcripts of exons, reiterated but often incorrectly transcribed and replicated, would at least occasionally assemble correctly to code for functional proteins. As the transcriptional and replicational processes became more faithful, evolutionary pressure to eliminate these sequences would increase providing there was a premium for rapid replication in this system. Thus prokaryotic genomes became more streamlined and could only evolve at a point mutational level while 'higher' forms retained the genetic plasticity inherent in the genomes of their primitive ancestors.

Darnell (1978) and Crick (1979) have suggested that genes containing intervening sequences may once have been continuous but subsequently

acquired introns by the insertion of transposable elements. It is necessary to invoke the pre-existence of splicing mechanisms in order to render such events non-lethal. The transposable character of these insertion elements would be then lost in evolution. In this model eukaryotic genes were at one time uninterrupted and strictly colinear with the polypeptide they encoded. There is no direct evidence for intervening sequences being acquired by eukaryotic nuclear genes during evolution. It is possible that the irregular organisation of intervening sequences in actin genes may be due to the insertion of intervening sequences (see Section 6). However, generally it seems that the number of introns in a gene decreases during evolution.

It has been postulated that inserted DNA sequences act as controlling elements in Zea mays (Nevers and Sadler, 1977). The introns of the cytochrome b gene in yeast mitochondrial DNA may have arisen by insertion. In some yeast strains three of the five introns including the mRNA maturase sequence are absent (Nobrega and Tzagoloff, 1980). It is possible that these three dispensible introns were once transposons and that the maturase has evolved from a transposase originally used to excise DNA, not RNA (Borst and Grivell, 1981).

11. LEGHAEMOGLOBIN GENES

Leghaemoglobin is a myoglobin like monomeric haemoprotein found only in the root modules of legumes symbiotically associated with Rhizobian. The leghaemoglobin protein is encoded by the plant genome (Sidloi et al., 1978). Leghaemoglobin is structurally and functionally related to myoglobin and haemoglobin (Appleby, 1974) and until recently was thought to have arisen by convergent evolution. Soybean (Glycine max) root modules contain four major leghaemoglobins (Fuchsman and Appleby, 1979)

as well as some minor forms. The differences in the amino acid sequences of the various leghaemoglobins are small. Computer analysis of the amino acid sequences of globins including several leghaemoglobins suggest that leghaemoglobins and globins have a common evolutionary origin (Hunt et al., 1978).

Recently a number of leghaemoglobin genes have been isolated from soybean DNA (Sullivan et al., 1981; Jensen et al., 1981; Hyldig-Nielsen et al., 1982; Brisson and Verma, 1982). The leghaemoglobin genes are interrupted by three intervening sequences at codons 32, 68-69 and 103-104. The positions of the first and third intervening sequences are homologous to the two intervening sequences found in all other known globin genes. The second, central, intervening sequence is unique to leghaemoglobin genes.

Leghaemoglobin has never been detected in any plant other than legumes which appeared about 200 million years ago, (Hyldig-Nielsen et al., 1982). The presence of leghaemoglobin in legumes is enigmatic as estimates of its time of divergence, assuming a constant rate of globin evolution, suggest leghaemoglobin diverged around 1500 million years ago (Dayhoff, 1972). It is possible that the leghaemoglobin gene is a result of convergent evolution involving the recombination of two haem binding exons present in the plant genome with unknown function, with two other plant gene exons according to the mechanism of Blake (1981).

It is interesting to speculate on other possible mechanisms whereby legumes may have obtained their leghaemoglobin gene. There are many similarities between rhizobial and agrobacterial infections of plant cells. Agrobacterium tumefaciens is known to be capable of inserting bacterial genes into the plant genome (Zambryski et al., 1980). It is possible therefore that a Rhizobium globin-like gene was transferred into the

genome of a primitive legume host, from which leghaemoglobin evolved. This mechanism is unlikely however as intervening sequences have never been detected in prokaryotes. It is possible however that an animal globin sequence became incorporated into a Rhizobium plasmid, similar to the Ti plasmids of *Agrobacterium*, and subsequently became incorporated into the legume genome. The animal globin gene sequence may have entered the Rhizobium cell by a simple process of transformation.

Another possible mechanism for the presence of this plant globin gene is that it was translocated there by some viral agent. For example, an insect borne plant pathogenic virus such as rhabdo-virus (Francki and Randles, 1980) may have been the vector for an insect globin gene to be transferred to a legume. It would be interesting, therefore, to examine the globin genes of insects and determine whether their structure resembled that of leghaemoglobin. If these genes have three intervening sequences it would seem likely that the central intervening sequence has been eliminated in the lineage leading to the vertebrates sometime before the α - β gene duplication. Another possible alternative for this gene transfer may have been through the agency of some retroviral like RNA. It is possible that an unprocessed globin gene RNA, complete with introns, was transferred as part of the retroviral RNA and became incorporated into the legume genome.

The existence of three intervening sequences in the leghaemoglobin gene seems incompatible with the hypothesis that exons encode functional protein domains, as the central exon in the leghaemoglobin gene is split into two halves. However, Gō (1981) has analysed the β -globin polypeptide and is able to discern four distinct modules or sectors. She suggests that globin genes may have originally consisted of four exons, and predicts an additional intervening sequence at almost the exact position that the extra leghaemoglobin gene intervening sequence is found.

Brisson and Verma (1982) have demonstrated that leghaemoglobin genes in the soybean genome, represent a moderately complex family of sequences which include functional genes linked to pseudogenes.

12. OBJECT OF THE RESEARCH

Haemoglobin genes represent the best characterised vertebrate multigene family and have proved useful in our understanding of gene structure and expression. However, nothing is yet known of another member of the globin gene family, the myoglobin gene. Comparisons of haemoglobin and myoglobin genes may provide information concerning the early events in globin gene evolution as well as enable us to study the properties of genes incorporated within radically different (erythroid and myogenic) developmental programmes. Analysis of the myoglobin gene in relation to the expression of the muscle contractile protein genes would be facilitated by the molecular cloning of a myoglobin gene.

The work in this thesis is concerned with the molecular cloning of a mammalian myoglobin gene. Initial, unsuccessful, attempts were made to isolate a cDNA clone complementary to human skeletal muscle myoglobin mRNA. The skeletal muscle from adult grey seal (Halichoerus grypus) was therefore used as a rich source of myoglobin mRNA, to synthesise a myoglobin cDNA clone. This clone was used to estimate the number of myoglobin genes in the grey seal genome and the size of the myoglobin mRNA, as well as to isolate a myoglobin gene from a genomic library of grey seal DNA cloned into bacteriophage λ .

Chapter 2

MATERIALS AND METHODS

1. TISSUE

Human adult skeletal muscle from post-mortems and foetal limb skeletal muscle was supplied by Dr Yvonne Edwards (University College, London). Human adult skeletal muscle from amputations of calf muscle (67 year old male) was supplied by Dr Simon Walker (Pathology Department, Leicester). Grey seal (Halichoerus grypus) skeletal muscle from juvenile animals was supplied by Dr John Prime (British Antarctic Survey, Cambridge).

2. RECOMBINANT PLASMIDS

pBG1 DNA was a gift from Professor C. Weissmann (Zurich). pαA91 was supplied by Dr Margaret Buckingham (Paris).

3. ENZYMES, ANTIBIOTICS, CHEMICALS AND REAGENTS

Proteinase K, bovine pancreatic ribonuclease A, lysozyme, S1-nuclease, lysostaphin, protein-A, deoxyribonucleoside triphosphates, dithiothreitol, spermidine trichloride, putrescine dichloride, salmon sperm DNA (sodium salt), bovine serum albumin, dimethyldichlorosilane, piperidine, ampicillin, diethylpyrocarbonate, L-methionine and N,N,N',N'-tetramethylethylenediamine were obtained from Sigma, London, England. Tri-iso-propylnaphthalene sulphonic acid, glyoxal and hydrazine were obtained from Eastman Kodak, Rochester, N.Y., U.S.A.; restriction endonuclease EcoRI, calf intestinal phosphatase and calf liver tRNA from the Boehringer Corporation, London, England; DNA Polymerase I, restriction

endonucleases AccI, AluI, AvaII, BamHI, BglIII, DdeI, EcoRI, HaeIII, HhaI, HindIII, HinfI, HpaII, PstI, Sau96I, Sau3A, and Taq Y1, and urea from Bethesda Research Laboratories Inc., Rockville, Maryland, U.S.A; deoxy-ribonuclease I from Worthington Biochemical Corporation, Freehold, New Jersey, U.S.A; avian myeloblastosis virus reverse transcriptase from Dr J.W. Beard, Life Sciences Incorporated, St. Petersburg, Florida, U.S.A; T4 polynucleotide kinase and terminal deoxynucleotidyl transferase from PL Biochemicals Inc., Milwaukee, Wisconsin, U.S.A; ficoll 400 and Sephadex G-50 from Pharmacia Fine Chemicals, Sweden; polyvinylpyrrolidone and phenol (AR) from Fisons, Loughborough, England; oligo d(T) cellulose Type 3, from Collaborative Research Incorporated, Waltham, Massachusetts, U.S.A; anti-human myoglobin antiserum, anti-human transferrin and agarose from Miles Laboratories Ltd., Stoke Poges, England; acrylamide and bis-acrylamide from Bio-Rad Laboratories Ltd., Watford, England; dimethyl sulphate from Aldrich Chemical Co. Ltd., Gillingham, England; Aquasol-2 from New England Nuclear, Massachusetts, U.S.A. Deoxyadenosine, deoxycytidine and deoxythymidine 5'- α - 32 P triphosphates, specific activity 2-3000 Ci/mMole, deoxy 8- 3 H guanosine 5'-triphosphate, specific activity 12 Ci/mMol, adenosine 5'- γ - 32 P triphosphate, specific activity 3000 Ci/mMole, L- 35 S-methionine, specific activity 1100 Ci/mMol and rabbit reticulocyte lysate were obtained from Amersham International Ltd., Amersham, England. All other chemicals used were analytical grade.

4. MEDIA

The following liquid medium was used: Luria Broth (10 g Difco Bacto Tryptone, 5 g Difco Bacto Yeast Extract, 5 g NaCl per litre of distilled water).

Agar plates were prepared by solidifying liquid medium with 15 g Davis New Zealand Agar per litre; 6 g agar per litre was used to prepare

soft agar overlays.

BBL Agar, used for phage assays and initial growth, contained 10 g Trypticase (Baltimore Biological Laboratories) and 5 g NaCl per litre of distilled water solidified with 15 g or 6 g agar as above. Soft agar overlays were always supplemented with 10 mM MgCl_2 .

5. AGAROSE GEL ELECTROPHORESIS

Horizontal agarose slab gels were used: generally 2-3% for RNA test gels, 0.4% for restriction endonuclease test gels and 0.8% for restriction endonuclease analysis by "Southern blotting". All gels were run in 40 mM Tris-acetate buffer pH7.7 containing 1 mM EDTA and 0.5 $\mu\text{g}/\text{ml}$ ethidium bromide, (Aaij and Borst, 1972). RNA and DNA samples were mixed with 0.5 volume of a 2% suspension of agarose beads in 20 mM EDTA containing 10% glycerol and a small amount of bromophenol blue as an electrophoresis dye marker (Schaffner et al., 1976). Samples were loaded into 5 mm x 7 mm deep slots and gels were run at room temperature at 6 V/cm.

To run DNA in the single stranded form, samples were denatured by the addition of 0.1 volume of 1.5 M NaOH, 0.1 M EDTA 5 minutes before loading. Samples were loaded and left for 5 minutes before running.

6. PREPARATIVE GEL ELECTROPHORESIS

Preparation of samples, gel loading and electrophoresis were as described by Jeffreys et al. (1980). The amount of native DNA loaded was adjusted so that no more than 0.5 μg of DNA entered each 1 mm^2 of gel surface; more than this amount caused overloading in this system.

DNA in gels was located by the fluorescence of bound ethidium bromide in the DNA under long wave length ultra violet light.

DNA was recovered from gel slices by inserting them into dialysis bags together with a small amount of diluted electrophoresis buffer (8 mM Tris-Acetate) as described by Smith (1980), the bags placed in shallow buffer just sufficient to maintain electrical contact and the DNA electro-eluted out of the gel into the dialysis bag. The DNA was rinsed off the dialysis membrane with diluted electrophoresis buffer. The DNA was recovered from solution by phenol extraction followed by ethanol precipitation.

7. PHOTOGRAPHY

Gels were photographed using a Nikon F camera and Ilford Ilfodata HS23 Type J500 P 35 mm film. DNA was visualised by bound ethidium bromide fluorescence, using a short wave length ultra violet light transilluminator.

8. GENERAL TECHNIQUES USED IN DNA AND RNA PREPARATION AND RECOVERY

(i) Phenol Extraction

DNA and RNA solutions were deproteinised by mixing with 0.5 volumes of phenol : chloroform : isoamyl alcohol : 8-hydroxyquinoline (100 : 100 : 4 : 0.1, W : V : V : W) saturated with 10 mM Tris-HCl, pH7.5, and briefly centrifuged to separate the phases. The upper aqueous phase containing the DNA or RNA was removed and the phenol layer re-extracted with an equal volume of 10 mM Tris-HCl, pH7.5.

The phenol used was AR grade and not redistilled.

(ii) Ethanol Precipitation

DNA and RNA were precipitated from solution by the addition of 0.1 volumes of 2 M (or 3 M) sodium acetate, pH5.6, and 2 (or 2.5) volumes of ethanol and chilling for 5 minutes in an industrial methylated spirit (I.M.S.) - dry ice bath. DNA precipitates were pelleted by centrifugation at $16300 \times g$ for 5 minutes at -15°C , and then rinsed with 70% ethanol, centrifuged in an Eppendorf centrifuge for 2 minutes and the 70% ethanol removed.

(iii) Methoxyethanol/Phosphate Extraction Of Carbohydrate

In order to remove carbohydrate and further clean DNA preparations, DNA solutions were mixed with equal volumes of 2.5 M potassium phosphate, pH8.0, and 2-methoxyethanol and the resulting turbid solution centrifuged (Jeffreys et al., 1980). The upper phase was then dialysed overnight against 100 μM Tris-HCl, 10 μM EDTA pH7.5.

(iv) Butanol Concentration

DNA solutions were concentrated by extraction of water with butan-2-ol. Solutions were mixed with butan-2-ol and briefly centrifuged. The top phase was discarded and the lower phase extracted 3 times with diethyl ether to remove remaining butan-2-ol. Traces of diethyl ether were removed by gentle passage of air over the solution.

9. PREPARATION OF DNA FROM TISSUE

Tissue was stored at -80°C . The tissue was thawed in 5 volumes of ice-cold 150 mM NaCl, 100 mM EDTA pH8.0, and homogenised in this solution using a Sorvall Omnimixer. The homogenate was lysed by adding an equal volume of 8% tri-isopropyl-naphthalene-sulphonic acid, 2% sodium dodecyl sulphate, 12% butan-2-ol dissolved in water. The lysate was then mixed with 0.25 volumes of 5 M sodium perchlorate and phenol extracted without

re-extracting the phenol layer. Mixing was gentle to avoid shearing high molecular weight DNA. Nucleic acids were precipitated by adding 2 volumes of ethanol and the precipitate rinsed with 70% ethanol. The DNA was then dissolved in 0.1 x TNE (TNE = 50 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, pH7.5) at 0-4°C and the solution incubated at 37°C for 20 minutes with 100 µg/ml heat treated pancreatic ribonuclease followed by 100 µg/ml proteinase K in 1 x TNE, 1% sodium dodecyl sulphate, at 37°C for 20 minutes. The DNA was phenol extracted and precipitated with 0.1 volumes of 2 M sodium acetate, pH5.6, and 2 volumes of IMS without chilling or centrifugation. The precipitate was redissolved in a small volume of 10 mM Tris-HCl, pH7.5, and methoxyethanol/phosphate extracted to remove carbohydrates (Kirby, 1957). DNA was precipitated from the upper phase using sodium acetate and IMS as before. The pellet was rinsed 3 times with 70% IMS and redissolved in 10 mM Tris-HCl, 1 mM EDTA, pH7.5, and then dialysed against 2 l of 2 mM Tris-HCl, 0.1 mM EDTA, pH7.5, overnight at 4°C.

DNA concentrations were determined by measurement of optical density of DNA solutions at 260 nm using a Unicam SP500 Spectrophotometer.

DNA quality was tested by agarose gel electrophoresis.

Low molecular weight denatured salmon sperm DNA, used as carrier in filter hybridisations, was prepared by shearing salmon sperm DNA to a length of about 0.4 kb by heating to 100°C for 20 minutes in 0.3 M NaOH, 20 mM EDTA. The DNA solution was then neutralised by adding 1 M Tris-HCl, pH7.5 to 40 mM and then concentrated HCl to pH7.8. The DNA was then phenol extracted and precipitated with IMS, vacuum dried and redissolved in water.

10. PREPARATION OF TOTAL RNA FROM TISSUE

Tissue was stored at -80°C . 500 g was fragmented with a hammer and placed in liquid nitrogen for 60 minutes. These pieces were then shattered and homogenised in three volumes of ice-cold 75 mM NaCl, 50 mM EDTA, pH8.0, 0.1% diethyl pyrocarbonate in a Waring Blendor which had been chilled overnight at -20°C . The homogenate was lysed by adding sodium dodecyl sulphate to a final concentration of 1% and shaking on ice for 10 minutes. The lysate was then phenol extracted without re-extracting the phenol layer. Nucleic acids were precipitated by adding 2 volumes of IMS and the precipitate rinsed with 70% IMS. The nucleic acid was then dissolved in water treated with 0.1% diethylpyrocarbonate. Nucleic acids were further precipitated by the addition of 0.1 volumes of 2 M sodium acetate pH5.6 and 2 volumes of ethanol and rinsed with 70% ethanol. The nucleic acids were then dissolved in a small volume of water.

In order to remove DNA the RNA was pelleted in a caesium chloride gradient as described by Gilsin et al., 1974 and Ullrich et al., 1973. 10 g of caesium chloride were added to 2.2 ml 0.5 M EDTA, 0.22 ml 1 M Tris-HCl, pH7.5 and 5.3 ml nucleic acid solution making final concentrations of 5.7 M caesium chloride, 0.1 M EDTA, 0.02 M Tris-HCl, pH7.5. Samples were centrifuged for 22 hours at $160000 \times g$ at 25°C . The RNA pellet was redissolved in a small volume of water and dialysed against water at 4° overnight.

RNA concentrations were determined by measurement of optical density of RNA solutions at 260 nm using a Unicam SP500 Spectrophotometer.

RNA quality was tested by agarose gel electrophoresis.

11. PREPARATION OF POLYADENYLATED RNA

Polyadenylated RNA was purified on oligo (dT) cellulose columns as described by Aviv and Leder (1972). Total RNA was loaded onto columns of 0.3 ml bed volume, in binding buffer (0.5 M NaCl, 10 mM Tris-HCl, pH7.5) unbound non-polyadenylated RNA was washed through the column with 15-20 bed volumes of binding buffer. Bound polyadenylated RNA was eluted from the column with 10 bed volumes of elution buffer (10 mM Tris-HCl, pH7.5).

The unbound RNA fractions were loaded twice more onto the same column and the bound fractions containing polyadenylated RNA were pooled. Generally this polyadenylated RNA was then heated at 65° for 5 minutes, cooled on ice, made up to binding buffer conditions and reloaded onto the same column and the bound polyadenylated RNA eluted with elution buffer. The polyadenylated RNA was recovered by adding 0.1 volumes of 2 M sodium acetate, pH5.6, and 2 volumes of 100% ethanol, chilling at -80°C for 5 minutes and spinning at 10,000 rpm for 10 minutes. RNA pellets were rinsed with 70% ethanol and redissolved in a small volume of water.

Polyadenylated RNA concentrations were determined by measurement of optical density of RNA solutions at 260 nm using a Unicam SP500 Spectrophotometer.

RNA quality was tested by agarose gel electrophoresis.

12. IN VITRO TRANSLATION OF RNA

Translation products of total and polyadenylated RNA were analysed in an in vitro translation system using rabbit reticulocyte lysate. Assays were carried out at 30°C for 60 minutes at a final radioactive concentration of 1 mCi/ml. Typical assays contained 8 µl rabbit reticulo-

cyte lysate and 2 μ l labelled amino acid (L-[35 S]-methionine).

Reactions were initiated by adding 1 μ l of RNA containing 0.1-0.3 μ g of polyadenylated RNA or 1-3 μ g of total RNA. Zero message controls were carried out by adding 1 μ l of distilled water instead of RNA.

After incubation in vitro, translations were terminated by adding L-methionine to a final concentration of 10 mM and heat treated pancreatic RNase to a final concentration of 0.1 mg/ml and incubating at 37°C for 15 minutes.

13. ANALYSIS OF IN VITRO TRANSLATION PRODUCTS BY SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS

Labelled translation products of in vitro translations were analysed by sds-polyacrylamide gel electrophoresis as described by Laemmli (1970). The constitution of buffer and solutions is given in Table 2:1. Gels were composed of a 7.2% stacking gel and a 15% or 17% separating gel. Freshly prepared ammonium persulphate was always used. Electrophoresis was carried out using a Raven slab gel apparatus (IN/96). Samples were boiled for 2 minutes in 2 volumes of sample buffer. Electrophoresis was carried out at 120 V per gel until the bromophenol blue was 5 mm from the bottom of the gel. Gels were stained for one hour in 400 ml 10% v/v acetic acid, 50% v/v methanol, 0.1% Coomassie brilliant blue. Diffusion destaining was carried out by shaking the gel in 400 ml 10% v/v acetic acid, 5% v/v methanol overnight with one change of solution. Gels were photographed on a white light transilluminator.

Fluorographs were prepared by dehydrating the gels in two changes of dimethyl sulphoxide (DMSO), followed by impregnation with 1,4-bis-2-(4 methyl-5-phenyloxazolyl)-benzene (dimethyl POPOP) exactly as described by

Table 2:1

Solutions and buffers used in electrophoresis

A. Separating Gel Buffer:

0.75 M Tris-HCl, pH8.8, 0.2% w/v sodium dodecyl sulphate.

B. Stacking Gel Buffer:

0.025 M Tris-HCl, pH6.8, 0.2% w/v sodium dodecyl sulphate.

C. Acrylamide Solution:

44% w/v acrylamide, 0.3% w/v N, N'-methylene-bis-acrylamide.

D. Electrophoresis Buffer:

0.125 M Tris, 0.192 M glycine, 0.1% w/v sodium dodecyl sulphate, pH8.3.

E. Sample Buffer:

0.625 M Tris-HCl, pH6.8, 20% v/v glycerol, 4% w/v sodium dodecyl sulphate, 5% v/v 2-mercaptoethanol, 0.001% w/v bromophenol blue.

F. Gel Composition (ml):

Separating Gel	15%	17%
Buffer A	13.5	13.5
C	9.2	10.4
distilled water	3.6	2.8
ammonium persulphate	0.95	0.95
(10 mg/ml)		

TEMED	0.075	0.075
-------	-------	-------

Stacking gel, 10 ml Buffer B, 3.3 ml C, 7.2 ml distilled water, 0.5 ml ammonium persulphate, 0.04 ml TEMED.

The gel apparatus was plugged with the same gel composition as the separating gel.

Bonner and Lasky (1974). The gels were dried onto Whatman 3 MM Chromatography Paper in a Bio-rad gel drying unit and exposed in an X-ray cassette with a Kodak X-Omat XRP5 X-ray film at -80°C . Generally films were prefogged to increase sensitivity.

14. IMMUNOPRECIPITATION OF SPECIFIC IN VITRO TRANSLATION PRODUCTS USING PROTEIN-A PRECIPITATION

Myoglobin was immunoprecipitated from in vitro translations of human skeletal muscle polyadenylated RNA, using anti-human myoglobin antiserum raised in goats and protein-A precipitation. 2 μl of a 24 μl in vitro translation mixture were removed as inputs and 2 x 10 μl taken for immunoprecipitation. To one 10 μl aliquot 1 μl of anti-human myoglobin antiserum was added, to the second 10 μl aliquot 1 μl of anti-human transferrin antiserum was added. This was repeated for 24 μl in vitro translation products of total nucleic acid and zero RNA controls.

24 μl of immunoprecipitation buffer were then added to all tubes (immunoprecipitation buffer : 50 mM Tris-HCl, pH7.5, 100 mM sodium chloride, 0.1% w/v N.P.40, 1 mM PMSF). The tubes were left at 0°C overnight. 10 μl of Protein-A were added and the tubes left at room temperature for one hour. The tubes were centrifuged in an Eppendorf centrifuge for 2 minutes, supernatants saved, and the pellets washed 4 times in immunoprecipitation buffer. The pellets were then taken up in 30 μl of polyacrylamide gel sample buffer (see Table 2:1) boiled for 2 minutes and centrifuged for 2 minutes in an Eppendorf centrifuge. Sample buffer was also added to aliquots of the supernatant after precipitation and all samples loaded onto a 15% polyacrylamide gel.

15. PREPARATION OF MYOGLOBIN BY ISO-ELECTRIC FOCUSING AND POLY-
ACRYLAMIDE GEL ELECTROPHORESIS

Adult human skeletal muscle was homogenised in an equal volume of water and centrifuged at 12,000 g for 10 minutes. The supernatant was removed and centrifuged at 5,000 g for one hour. The supernatant was dialysed overnight against water at 0°C and concentrated by surrounding the dialysis tubing in solid polyethylene glycol 6000. This concentrate was loaded onto a LKB Ampholine PAG plate pH5.5-8.5 (LKB 1804-103) which had been prefocussed for one hour at 20 W, 1.5 kV. Samples were focussed for 2-3 hours until all the pigmented proteins could be resolved. Human haemoglobin markers were also focussed. The major non-haemoglobin red pigmented band (myoglobin) was excised and protein extracted by mashing the gel slice in 20 mM sodium chloride in 4 changes of solution. The solution was centrifuged and concentrated by placing it in dialysis tubing surrounded by solid polyethylene glycol 6000. The concentrated solution was dialysed overnight against 20 mM sodium chloride, 2 mM Tris-HCl, pH7.5.

Further purification of human apo myoglobin was achieved by electrophoresing the iso-electrically focussed protein in a 15% sds-polyacrylamide gel. The region of gel containing myoglobin was excised and mashed in 3 volumes 60 mM Tris-HCl pH7.0, 0.2% sodium dodecyl sulphate, 10 mM 2-mercaptoethanol. The protein was allowed to elute from the gel for 3 hours at 4°C. The gel pieces were pelleted by centrifugation and the supernatant stored. More buffer was added to the gel pellet and remaining protein allowed to elute overnight at 4°C. The gel pieces were pelleted and both supernatants pooled, dialysed against several changes of water, concentrated to a small volume with solid polyethylene glycol and stored at -20°C.

Seal myoglobin was prepared from skeletal muscle by iso-electric focussing as described above.

16. SEPARATION OF MYOGLOBIN FROM IN VITRO TRANSLATIONS OF SEAL

POLYADENYLATED RNA AND POSITIVELY SELECTED RNA BY ISO-ELECTRIC

FOCUSSING

0.9 µg of seal polyadenylated RNA in a 30 µl reaction volume or 1 µl of positively selected RNA (Materials and Methods Section 32) in a 10 µl reaction volume were in vitro translated in the rabbit reticulocyte system.

After in vitro translation 1/5 of each reaction mix was removed as input and the remaining mixture mixed with 15 µl purified grey seal myoglobin. This mixture was then loaded onto a LKB Ampholine PAG plate, pH5.5-8.5 (LKB 1804-103). After electrofocussing the myoglobin band was excised and either the myoglobin recovered as described above and loaded onto a 17% sds-polyacrylamide gel or the gel slice boiled for 5 minutes in polyacrylamide gel sample buffer (Table 2:1) and loaded directly into the sample slots of a 17% sds-polyacrylamide gel. Samples were electrophoresed and fluorographed as described above.

17. PREPARATION OF SINGLE STRANDED COMPLEMENTARY DNA FROM POLYADENYLATED

RNA

Avian Myeloblastosis virus (AMV) reverse transcriptase was used to synthesise complementary DNA (cDNA) copies of polyadenylated RNA in the presence of an oligo (dT) primer. Preliminary experiments were carried out in order to optimise reaction conditions with regard to substrate and enzyme concentration. All reactions were carried out in 50 mM Tris-HCl (pH8.3), 60 mM KCl, 6.0 mM MgCl₂, 10 mM dithiothreitol, 0.6 mM dATP, 0.06 mM dCTP, 0.6 mM dGTP, 0.6 mM dTTP, 100 µg/ml actinomycin D and 7.5 µg/ml oligo (dT)₁₀ primer. Optimising experiments were in 10 µl reaction volume.

The reaction was monitored by the incorporation of α - ^{32}P -dCTP (2-3000 Ci/mMole) at a final radioactive concentration of 0.1 mCi/ml. Reactions were incubated at 42°C for 2 hours and terminated by the addition of an equal volume of 1% sodium dodecyl sulphate, 0.1 M EDTA, pH8.0. Nucleic acid was ethanol precipitated twice in the presence of 10 μg of alkali denatured salmon sperm DNA. Pellets were redissolved in 20 μl 10 mM Tris-HCl, pH7.5, prior to Cerenkov counting. In order to analyse radioactive cDNA synthesis, samples were alkali denatured to remove the RNA template and electrophoresed in 1.5% agarose gels. The gels were dried down onto a glass plate with hair-dryers and autoradiographed for 10-12 hours using Kodak X-Omat R film, XRP5.

Bulk preparations of single stranded cDNA were direct scale ups of the above reaction using optimal substrate and enzyme conditions. However no carrier salmon sperm DNA was used in precipitations and the cDNA/mRNA hybrid molecules were separated from unincorporated α - ^{32}P -dCTP by Sephadex G50 column chromatography. Enough Sephadex G50 was pre-swollen in 10 mM Tris-HCl, pH7.5, 1 mM EDTA, pH8.0, to produce a bed volume of 1.5 or 3.0 ml. Columns were set up in siliconised Pasteur pipettes plugged with glass wool. Columns were washed with 5 times the bed volume of 10 mM Tris-HCl, 1 mM EDTA, pH7.5. The reaction mix, after incubation, was applied to the column and eluted in 150 μl fractions with 10 mM Tris-HCl, 1 mM EDTA, pH7.5.

Generally 40 μg of polyadenylated RNA was reverse transcribed using 550 units of reverse transcriptase and 15 μg oligo (dT)₁₀ in 2 ml reaction volume in the presence of 40 μCi α - ^{32}P -dCTP. The ratio of labelled to unlabelled dCTP was 1 : 7 500 (this ratio was changed to 1 : 750 for later experiments). This ratio was selected in order to reduce the radiolysis of cDNA by the decay of ^{32}P . The

quality of cDNA synthesised was assayed by the electrophoresis in 1.5% agarose gels of alkali-denatured samples of the reaction mix. Gels were dried down and autoradiographed.

18. PREPARATION OF DOUBLE STRANDED COMPLEMENTARY DNA

The RNA of cDNA/mRNA hybrid molecules was hydrolysed by incubation at 67°C for 15 minutes in 0.2 M sodium hydroxide. After incubation an equal volume of 1 M Tris-HCl, pH7.5, was added and the mixture neutralised by the addition of 1 M hydrochloric acid. Single stranded cDNA was ethanol precipitated twice, redissolved in water and Cerenkov counted in order to estimate yields.

Optimising experiments for second strand synthesis were carried out in 10 µl reaction volumes in 50 mM Tris-HCl, pH8.3, 60 mM KCl, 60 mM MgCl₂, 0.6 mM dATP, 0.6 mM dCTP, 0.6mM dGTP, 0.6 mM dTTP. No radioactive α-³²P-dCTP was incorporated into the second strand. Reactions were incubated at 42°C for 2 hrs and terminated by the addition of an equal volume of 1% sodium dodecyl sulphate, 0.1 M EDTA, pH8.0. Double stranded cDNA was recovered by ethanol precipitation.

The extent of the reaction was monitored by assessing the proportion of S1-nuclease resistant reaction products. 0.5 µg of single stranded cDNA was incubated with 550 units of reverse transcriptase in 0.3 ml reaction volume. This mixture was incubated for 2 hours at 42°C. The reaction was stopped by adding EDTA to a final concentration of 20 mM and the mixture extracted with phenol/chloroform mix with two re-extractions. The aqueous phases were ethanol precipitated and double stranded cDNA further purified by passage over a Sephadex G50 column as described above. The excluded fractions were pooled and made to a final concentration of 0.025 M sodium chloride, 0.3 M

sodium acetate, 4.5 mM zinc chloride, pH4.5, S1-nuclease was added to a final concentration of 1500 units/ml. An aliquot of this mixture was removed and 1 μ g of alkali denatured high molecular weight human DNA and 1 μ g of double stranded λ DNA digested with EcoRI were added to provide internal controls for S1-nuclease activity. Both mixtures were incubated at 37°C for 15 minutes. The main digest tube was then frozen at -20°C. The mixture of double stranded cDNA, single stranded human DNA and double stranded λ DNA x EcoRI was electrophoresed in a 1.5% agarose gel to assay for S1-nuclease digestion. After checking S1-nuclease digestion in the tested aliquot the main digest was thawed, extracted 3 times with phenol/chloroform and ethanol precipitated. The dried pellet was redissolved in a small volume of water and Cerenkov counted to estimate the yield of double stranded cDNA.

The size of double stranded cDNA was determined by electrophoresis in 1.5% agarose gels followed by autoradiography.

In order to check that the S1-nuclease had removed the 'hairpin' structures at the end of the double stranded cDNA molecules, samples were mixed with 1 μ g of high molecular weight human DNA and denatured in 0.2 M sodium hydroxide for 5 minutes at room temperature. This mixture was neutralised, made to S1-nuclease reaction conditions, λ DNA x EcoRI added and incubated with 5 units of S1-nuclease as described above. The mixture was electrophoresed against input double stranded cDNA, double stranded cDNA treated with S1-nuclease and alkali-denatured double stranded cDNA. After electrophoresis the gel was photographed, dried down and autoradiographed. The input track of this autoradiograph was scanned with a densitometer in order to estimate the size range of double stranded cDNA molecules.

19. ADDITION OF HOMOPOLYMER TAILS TO DOUBLE STRANDED cDNA MOLECULES

After S1-nuclease treatment and phenol extraction the double stranded cDNA was passed over a Sephadex G50 column. Terminal deoxy-nucleotidyl transferase was used to catalyse the addition of deoxy-cytidine residues to the 3' ends of double stranded cDNA. In preliminary experiments 3.0 ng of double stranded cDNA in 1.5 mCi/ml α -³²P-dCTP (2000-3000 Ci/mmole) 0.1 M potassium cacodylate, pH6.9, 1 mM dithiothreitol, 1 mM dCTP were used in a 10 μ l reaction volume. 1 μ l of 10 mM CoCl₂ was added to the above mix and the mixture prewarmed at 37°C for 2 minutes. 1.7 units of terminal transferase were then added and the mixture incubated at 37°C. 1 μ l aliquots were removed at intervals and immediately quenched with 10 μ l of 10% sodium dodecyl sulphate, 0.5 M sodium chloride. 229 μ l of H₂O were added to this mixture along with 5 μ l 0.5 M EDTA, pH8.0, 50 μ g high molecular weight salmon sperm DNA and 500 μ l 8% trichloroacetic acid (TCA). Samples were left on ice for 10 minutes and filtered through GFC Whatman glass fibre filters. The filters were rinsed with 5 x 5 mls of 8% TCA and 3 x 3 mls 100% Industrial Methylated Spirits. The filters were then dried and Cerenkov counted.

In bulk preparations of "C"-tailed double stranded cDNA 46 ng of DNA were incubated in a reaction mix as described above suitably scaled up. At 20, 30, 40 and 60 minutes 1/4 of the reaction mixture was removed and quenched with 10% sodium dodecyl sulphate, 0.5 M NaCl to stop the reaction. These aliquots were then pooled and passed over a Sephadex G50 column. The excluded fractions were pooled and concentrated to approximately 20 μ l with butan-2-ol. Residual butan-2-ol was removed with diethyl ether and the solution Cerenkov counted to estimate the length of tails added.

20. ADDITION OF HOMOPOLYMER TAILS TO LINEARISED PLASMID pAT153

pAT153 was linearised with the restriction endonuclease PstI. Oligo (dG) homopolymer tails were added to the 3' termini of the linearised plasmid by incubating 5 µg of DNA in a final volume of 100 µl in 140 mM cacodylic acid 30 mM Trizma base (neutralised to pH6.9 with potassium hydroxide) 1 mM dithiothreitol, 18 µCi deoxy 8-³H guanosine 5'-triphosphate. Cobalt chloride was added to a final concentration of 1 mM, the reaction mixture prewarmed at 37°C for 2 minutes, 40 units of terminal transferase were added and the reaction incubated at 37°C. At 30, 60 and 90 minutes 2 µl aliquots were removed for gel analysis and 30 µl quenched with an equal volume of sodium dodecyl sulphate, 1% 0.1 M EDTA, pH8.0. These large aliquots were pooled, phenol extracted and loaded onto a Sephadex G-50 column. The excluded fractions were pooled and concentrated to 20 µl with butan-2-ol. This sample was electrophoresed in a 0.4% agarose gel and the DNA recovered from the gel slice as described in Materials and Methods Section 6.

The 2 µl aliquots were electrophoresed in a 0.4% agarose gel, the gel slice excised and boiled in 1 ml of water and counted in 4 ml Aquasol-2 in a liquid scintillation counter in order to estimate the incorporation of ³H-dGTP into the DNA.

21. ANNEALING G-TAILED VECTOR DNA AND C-TAILED DOUBLE STRANDED DNA

Equimolar amounts of C-tailed double stranded cDNA and G-tailed pAT153 were annealed in 10 mM Tris-HCl pH7.5, 5 mM EDTA and 0.1 M NaCl for 10 minutes at 65°C, followed by 2 hours at 42°C. The reaction was allowed to cool to room temperature overnight and kept on ice until ready for use.

22. TRANSFORMATION

An overnight culture of *E.coli* HB101 (recA, hsdR_k⁻ hsdM_k⁻ pro⁻ leu⁻ thi⁻ lacY⁻ endA⁻ rpsL 20 ara 14 galK₂ Xyl-5 mtl-1 supE44 trp⁻) in Luria broth with thymine at 20 µg/ml was diluted one hundred fold into identical medium and grown to A₆₀₀ = 0.2. The culture was then rediluted as before and again grown to A₆₀₀ = 0.2. 10 ml aliquots of cells were then pelleted by centrifugation at 12000xg for 10 minutes at 4°C and resuspended in 5 ml of ice-cold 50 mM CaCl₂ and kept on ice for 15 minutes. Cells were repelleted and suspended in 500 µl of 50 mM CaCl₂ and kept on ice. Typically, 100 ng of annealed DNA were diluted to 50 µl with water and 34 µl of this mixed with 100 µl of competent cells and 10 µl of transformation buffer (50 mM CaCl₂, 50 mM MgCl₂, 50 mM Tris-HCl, pH7.5). Volumes were made up to 160 µl with water. Mixtures were kept on ice for 30 minutes and then at 37°C for 2 minutes and then at room temperature for 10 minutes. 1 ml of Luria broth with thymine at 20 µg/ml was added to each mixture and tubes incubated at 42°C for 2 minutes and at 37°C for 60-90 minutes with shaking. The contents of each tube were then plated on Luria agar plates supplemented with thymine at 20 µg/ml and tetracycline at 7.5 µg/ml. Controls were transformations with the linearised tailed vector and with pAT153.

Plates were incubated overnight at 37°C.

23. SCREENING OF TRANSFORMANTS

The method used was a modification of the procedure of Grunstein and Hogness (1975). Colonies from each transformation experiment were transferred using sterile toothpicks onto two Luria agar plates containing 20 µg/ml thymine and 7.5 µg/ml tetracycline overlaid with 9 cm diameter Schleicher and Schull gridded nitrocellulose filters

(Schleicher and Schüll, pore size 0.45 μ M). One plate acted as a master while the other was a replica. Both masters and replicas were incubated at 37°C overnight. The master plate was then stored at 4°C.

Filters were removed from the replica plates and treated in the following manner by placing them on Whatman 3 MM chromatography paper flooded with the appropriate solution. At no time was liquid allowed onto the upper surface to prevent any merging of colonies.

Filters were first treated with 1.5 M NaCl, 0.5 M NaOH for 10 minutes and then twice with 1.0 M Tris-HCl pH7.5 each time for 2 minutes. This was followed by 1.5 M NaCl, 0.5 M Tris-HCl pH7.5 for 4 minutes after which filters were blotted dry and baked, colony side up, for 5 hours at 80°C. The filters were then hybridised as described below using as probe a single stranded 32 P-labelled cDNA prepared by reverse transcription from polyadenylated RNA. Positively hybridising colonies were detected by autoradiography.

24. PREPARATION OF 32 P-LABELLED cDNA BY REVERSE TRANSCRIPTION

The method used was a modification of the procedure of Marotta et al. (1974). 0.2 μ g of polyadenylated RNA was incubated with 5 units of avian myeloblastosis virus reverse transcriptase for 3 hours at 42°C in the presence of 50 mM Tris-HCl, pH8.3, 60 mM KCl, 6 mM MgCl₂, 10 mM dithiothreitol, 0.6 mM dATP, dGTP and dTTP, 4 μ M dCTP, 10 μ g/ml of oligo (dT), 0.25 mg/ml of actinomycin D and 10 μ Ci of α - 32 P dCTP. Recovery was exactly the same as that described below for the recovery of nick translated DNA using high molecular weight salmon sperm DNA as carrier and redissolving the precipitated nucleic acid in 200 μ l 10 mM Tris-HCl, pH7.5. The cDNA was rendered single stranded for hybridisation by RNA

hydrolysis. EDTA was added to 20 mM and NaOH to 0.2 M. This mixture was incubated for 15 minutes at 67°C before being diluted with an equal vol of 1 M Tris-HCl, pH7.5 and then neutralised to pH7.0 with 1 M HCl.

25. SMALL SCALE PREPARATION OF PLASMID DNA

Two methods were used to make DNA from small cultures of E.coli HB101 harbouring recombinant plasmids.

(a) This method was a modification of that described by Klein et al. (1980).

E.coli HB101 harbouring recombinant plasmids was grown in 10 mls of Luria broth supplemented with 20 µg/ml thymine and 7.5 µg/ml tetracycline at 37°C. These cultures were allowed to grow to A₅₅₀ 0.9-1.0 at which stage they were amplified by adding chloramphenicol to a concentration of 100 µg/ml. The cultures were then grown overnight at 37°C. The cells were pelleted and resuspended in 0.5 ml 50 mM Tris-HCl pH8.0 and transferred to a 1.5 ml Eppendorf tube. 50 µl of a 10 mg/ml solution of lysozyme in 10 mM Tris-HCl were added and the mixture left at room temperature for 15'. An equal volume of phenol/chloroform mix was then added, shaken and the mixture centrifuged for 15' in an Eppendorf centrifuge. The aqueous phase was removed and re-extracted with 0.3 ml of phenol/chloroform mix. Plasmid DNA was recovered by ethanol precipitation.

(b) The method used was a modification of that described by Birnboim and Doly (1979). 2 ml cultures of E.coli HB101 harbouring recombinant plasmids were grown in Luria broth supplemented with 20 µg/ml thymine and 7.5 µg/ml tetracycline at 37°C overnight. 1.5 ml of these cultures were transferred to 1.5 ml Eppendorf tubes, the cells pelleted and resuspended in 100 µl of lysis buffer (25 mM Tris-HCl (pH8.0), 10 mM EDTA (pH8.0), 50 mM sucrose, 1 mg/ml lysozyme) this mixture was left on ice

for 10 minutes. 200 μ l of 0.2 M NaOH, 1% sodium dodecyl sulphate was added and left on ice for 5 minutes. 150 μ l of 3 M potassium acetate, pH4.8, were added and the mixture left on ice for 10 minutes. Tubes were centrifuged in an Eppendorf centrifuge for 5 minutes at room temperature, 0.4 ml supernatant removed and 1 ml 100% ethanol added. The tubes were placed at -80°C to allow the DNA to precipitate and then spun for 2 minutes in an Eppendorf centrifuge. DNA pellets were re-dissolved in 100 μ l of 0.2 M sodium acetate, pH5.6, and 300 μ l of 100% ethanol added. DNA was allowed to precipitate at -80°C , the DNA pelleted in an Eppendorf centrifuge, rinsed with 1 ml 70% ethanol, vacuum dried and redissolved in a small volume of water.

26. LARGE SCALE PREPARATION OF PLASMID DNA

This method was found suitable for E.coli HB101 harbouring pAT153 and recombinant plasmids. 10 mls of culture supplemented with 20 $\mu\text{g/ml}$ thymine and 7.5 $\mu\text{g/ml}$ tetracycline were grown overnight at 37°C . This 10 ml culture was added to two 500 ml baffled flasks of Luria broth supplemented with 20 $\mu\text{g/ml}$ thymine and 7.5 $\mu\text{g/ml}$ tetracycline and grown overnight at 37°C .

Cells were pelleted and resuspended in 40 ml of lysis solution (see above) and left for 5 minutes at 0°C . 80 ml of alkaline/sodium dodecyl sulphate solution (see above) were added and the mixture left at 0°C for 5 minutes. 60 ml of 3 M potassium acetate, pH4.8, were added and the mixture left at 0°C for 5 minutes. The mixture was then spun in a GS3 bottle at 8000 rpm for 10 minutes at 4°C . The supernatant was poured through a funnel plugged with polyallomer wool and nucleic acid precipitated by addition of 100 ml iso propanol followed by centrifugation, at 8000 rpm for 10 minutes at 4°C . The DNA pellet was rinsed with 70% ethanol, dried with

diethyl ether and redissolved in 20 ml of 10 mM Tris-HCl, 1 mM EDTA, pH7.5. 4 ml of 5 mg/ml ethidium bromide and 23.76 g A.R. CsCl were added. The solution was divided between two 50 Ti tubes, which were then sealed and centrifuged in a Beckman 50 Ti rotor at 39000 rpm for 40 hours at 15°C. After centrifugation the lower plasmid band was visualised with ultra violet light and removed using a syringe and tubing. Ethidium bromide was removed by several extractions with iso-propanol saturated with a solution of caesium chloride. Two volumes of water were added to the DNA and the DNA precipitated by adding two volumes of ethanol. The precipitate was pelleted, rinsed with 70% ethanol, vacuum dried and dissolved in 0.5 ml 10 mM Tris-HCl (pH7.5).

27. LABELLING DNA BY 'NICK TRANSLATION'

The procedure used was a modification of that of Rigby et al. (1977). 15 μ Ci of α 32 P-dCTP (2-3000 Ci/mMole) (= 1.5 μ l of a stabilised aqueous solution of the triethylammonium salt) were made up to a final volume of 25 μ l with 50 mM Tris-HCl, pH7.8, 5 mM MgCl₂, 10 mM 2-mercaptoethanol, 4 μ M dGTP, 4 μ M dATP and 4 μ M dTTP. 0.1 μ g of DNA, 16 pg of deoxyribonuclease I and 2 units of E.coli DNA polymerase I were added and the mix incubated for 1 hour at 15°C. The reaction was stopped by the addition of 25 μ l of 0.5% sodium dodecyl sulphate, 12.5 mM EDTA, 10 mM Tris-HCl, pH7.5, phenol extracted and centrifuged at top speed in an MSE benchcentrifuge for 2 minutes. After removing the aqueous upper phase the phenol layer was reextracted with 50 μ l 10 mM Tris-HCl, pH7.5. 50 μ g of high molecular weight salmon sperm DNA were added to the pooled aqueous phases and the DNA ethanol precipitated without centrifugation or chilling. The ethanol was removed and the DNA

precipitate rinsed with 70% ethanol before being redissolved in 200 μ l of 10 mM Tris-HCl, pH7.5. The DNA was reprecipitated and rinsed as before and then redissolved in 500 μ l of 10 mM Tris-HCl, pH7.5. Specific activities of 10^7 to 10^8 dpm/ μ g DNA were generally obtained.

28. RESTRICTION ENDONUCLEASE DIGESTION

Digestions were carried out according to manufacturers recommended conditions. For genomic mapping experiments 15 μ g of DNA was digested at a concentration of 100 μ g/ml. Completion of digestion was checked by removing a 0.5 μ g aliquot and electrophoresing this through a 0.4% agarose slab gel alongside molecular weight markers. Where digestion was incomplete a further aliquot of restriction endonuclease would be added and the incubation repeated. After digestion EDTA was added to a final concentration of 20 mM and the DNA phenol extracted, ethanol precipitated, vacuum dried and redissolved in 10 mM Tris-HCl, pH7.5. This solution was used for subsequent digestion with other restriction endonucleases when required.

For digestion of plasmid DNA 1-2 μ g of DNA were digested at a concentration of 100 μ g/ml and the digestion products electrophoresed in 0.5%, 1% or 2% agarose slab gels without recovery of the DNA from the digest.

29. TRANSFER OF DNA TO NITROCELLULOSE FILTERS

Digested DNA samples containing 15 μ g in 20 μ l 10 mM Tris-HCl, pH7.5 were denatured and loaded into 20 cm x 20 cm horizontal 0.8% agarose slab gels and the gels run until the bromophenol blue had run about 8 cm from the loading slots. Gels were then washed, with occasional, gentle rocking, in 20 x SSC (SSC = saline sodium citrate; 1 x SSC is 0.15 M

NaCl, 15 mM trisodium citrate, pH7.0 for 20 minutes at room temperature. The denatured DNA was transferred onto nitrocellulose filters (Sartorius, 0.45 μ m pore size) by following a modification of the procedure of Southern (1975). The transfer apparatus was constructed by placing a glass plate on supports in a tray flooded with 20 x SSC and covering it with 2 sheets of Whatman 3 MM chromatography paper soaked in 20 x SSC; these sheets also dipped into the bottom of the tray to act as wicks. The tray and paper were covered with a sheet of cling-film and a hole cut in this slightly smaller than the slab gel. The exposed paper was soaked with 20 x SSC and the gel placed so that all the exposed area was covered by the gel and the DNA in the gel all lay within the borders of the exposed area. Great care was taken to ensure no air-bubbles were trapped between the gel and paper. After flooding the gel surface with 3 x SSC a Sartorius filter soaked in 3 x SSC was placed on the gel, again avoiding air bubbles. A similar sized piece of Whatman 3 MM chromatography paper soaked in 3 x SSC was placed over the filter followed by three dry sheets. These were covered with three layers of Boots Disposable Nappies cut to fit the surface area. On top was placed a glass plate and a 2 kg weight, and the apparatus left at 4°C overnight. The filters were subsequently washed with 3 x SSC for 20 seconds and then baked for 5 hours at 80°C to bind the DNA onto the filter.

30. TRANSFER OF RNA TO NITROCELLULOSE FILTERS

2 μ g of polyadenylated RNA and 10 μ g of total RNA were vacuum dried and redissolved in 20 μ l RNA denaturation mix. (10 mM sodium phosphate, pH6.85, 16% v/v glyoxal deionised by mixing with Bio-rad AG 501 - x 8 (D) beads, 50% v/v dimethylsulphoxide) and incubated at 50°C for 1 hour.

2 μ l of 10 mM sodium phosphate, pH7.0, 0.05% bromophenol blue were added to each sample and the samples loaded onto 2.5% agarose gels made by dissolving 7.5 g of agarose in 300 ml 10 mM sodium phosphate (pH7.0). Gels were electrophoresed at 150 V for 2 hours in 10 mM sodium phosphate, pH7.0, during which time electrophoresis buffer was recirculated using a peristaltic pump.

After electrophoresis RNA was transferred to nitrocellulose filters as described for DNA samples, except that the nitrocellulose filter was soaked in 20 x SSC before placing on the gel.

31. FILTER HYBRIDISATION

DNA hybridisation and filter washing was at 65°C in a gently rocking water bath. Hybridisation solutions were thoroughly degassed under vacuum before use, as air bubbles coming out of solution give rise to intense black spots of background labelling (Jeffreys and Flavell, 1977). Flat bottomed perspex boxes with rubber gasket-sealed lids, with bottom interior dimensions 10 cm x 4 cm and 1.7 cm deep and with sloping sides were used as incubation chambers. Circular chambers of similar design with a bottom interior diameter of 9 cm were used for screening recombinant plaque lifts and Grunstein-Hogness filters.

The baked filters from DNA transfers were cut into 9 cm x 3 cm strips using a long wave length ultra violet top illuminator to view filter-bound ethidium bromide along the DNA tracks. The strips were sandwiched between two blank filters, 3.7 cm x 9.4 cm and always completely submerged in all solutions to prevent drying out which gives rise to high background labelling.

Filters were first incubated in 3 x SSC for 20 minutes and then treated for 1 hour with 0.2% ficoll, 0.2% polyvinylpyrrolidone, 0.2%

bovine serum albumin in 3 x SSC (Denhardt, 1966). This solution was then replaced with the same mixture supplemented with low molecular weight denatured salmon sperm DNA at 50 µg/ml and sodium dodecyl sulphate at 0.1% and incubated for 1 hour. The salmon sperm DNA was added to reduce non-specific binding of labelled probe.

The filters were then incubated for a further 1 hour in a similar solution containing 9% dextran sulphate (Sigma, approximate average molecular weight 500,000 Daltons); the dextran sulphate greatly increases the filter hybridisation kinetics (Wahl et al., 1979; Jeffreys et al., 1980). Strips were then transferred to 12 ml of an identical solution containing no more than 10 µg/ml of ³²P-labelled DNA probe; DNA probes were denatured at 100°C for 5 minutes before adding to the hybridisation solution.

After overnight hybridisation, strips were given four 10 minute washes in complete hybridisation mix preheated to 65°C with the labelled probe omitted and using an intermediate bath of the same solution between each transfer. This was followed by two 30 minute washes in 0.3 to 0.1 x SSC containing with 50 µg/ml low molecular weight denatured salmon sperm DNA and 0.1% sodium dodecyl sulphate preheated to 65°C. Finally strips were rinsed in cold 3 x SSC, blotted dry and autoradiographed at -70°C for 1-14 days using Kodak X-Omat R film, XRP5, in conjunction with an Ilford Tungstate intensifying screen.

The baked filters from RNA transfers were cut into 9 cm x 3 cm strips. The strips were sandwiched between two blank filters and always completely submerged in all solutions. DNA marker filters were hybridised as described above. RNA filters were prehybridised for 8-20 hours at 42°C in 50% v/v formamide, 5 x SSC, 50 mM sodium phosphate, pH6.5, 250 µg/ml alkali denatured salmon sperm DNA, 0.02% Ficoll (Sigma,

type 400), 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone (Sigma PVP-360). The nick translated probes were denatured by immersing in a 100°C bath for 5 minutes and added to the hybridisation mix (4 parts of above prehybridisation buffer mixed with one part 50% w/v dextran sulphate). After mixing the probe into the hybridisation mix thoroughly, the RNA filters were transferred to the hybridisation buffer and incubated at 42°C overnight.

After hybridisation the RNA filters were washed in 4 changes of 2 x SSC, 0.1% sodium dodecyl sulphate for 5 minutes each at room temperature, and then in 2 changes of 0.1 x SSC, 0.1% sodium dodecyl sulphate for 15 minutes each at 50°C.

The RNA filters were dried at room temperature for 60 minutes and autoradiographed at -70°C using Kodak X-Omat R film, XRP5, in conjunction with an Ilford Tungstate intensifying screen.

32. POSITIVE SELECTION OF mRNA BY FILTER BOUND cDNA CLONES (HYBRID RELEASED TRANSLATION)

In order to assign cloned cDNAs to specific proteins the technique of Hybrid Released Translation was used. Cloned cDNAs were bound to nitrocellulose filters and hybridised to polyadenylated RNA, the filters were washed and after elution from the filter the bound message translated in the rabbit reticulocyte system.

Initially, cloned cDNAs were analysed as pools of 10-15 individual clones and screened. Some of these pooled sets of clones were subsequently analysed individually by the same method.

Plasmid DNA made by one of the small scale preparations was made to 5 mM Tris-HCl, 10 mM sodium chloride, 1 mM EDTA, pH7.5 and incubated at 37°C for 20 minutes with 100 µg/ml heat treated pancreatic ribonuclease

followed by 100 $\mu\text{g}/\text{ml}$ proteinase-K in 50 mM Tris-HCl, 100 mM sodium chloride, 5 mM EDTA, pH7.5, 1% sodium dodecyl sulphate, at 37°C for 20 minutes. The DNA was phenol extracted and ethanol precipitated and 5-10 μg redissolved in 5 μl of denaturation buffer (1 M sodium chloride, 0.5M sodium hydroxide, 10 mM EDTA). The DNA was incubated at 65°C for 5 minutes, spun down and spotted onto Schleicher and Schüll 9 cm diameter nitrocellulose filters (0.45 μ pore size) in an ordered array. The spots were allowed to dry for 20 minutes at room temperature, and rinsed in 3 x SSC for 2 minutes before baking at 80°C for 5 hours.

Each spot of DNA was numbered on the reverse of the filter with a soft pencil and washed four times in water at room temperature. The filters were then washed 2 times for 2 minutes in water at 80°C to remove any unbound DNA. Each spot was then carefully cut from the filter with a scalpel and all the spots placed in the minimum volume of hybridisation mix required to completely cover all the filters. (Hybridisation mix : 50% v/v twice recrystallised formamide, 0.4 M sodium chloride, 10 mM PIPES - Na pH6.4, 4 mM EDTA, 0.6 mg/ml calf liver tRNA). Filters were incubated in siliconised glass pots at 41°C for 3 hours with continuous shaking. Filters were then transferred to fresh hybridisation mix for a further two hours at 41°C. After this prehybridisation, the solution was replaced by hybridisation mix supplemented with polyadenylated RNA at concentrations between 25-50 $\mu\text{g}/\text{ml}$ of hybridisation mix. (Generally 20-25 filters could be hybridised in 0.8 ml of hybridisation mix). The filters were then incubated overnight at 41°C with continuous shaking.

Control filters with bound p β G1DNA were hybridised separately in hybridisation mix supplemented with 50-100 ng of rabbit globin mRNA in 160 μl hybridisation volume.

After hybridisation, filters were washed twice in 10 ml of 1 x SSC, 0.5% sodium dodecyl sulphate for 3 minutes each at room temperature, followed by 3 washes in 10 ml of 0.1 x SSC, 0.1% sodium dodecyl sulphate for 3 minutes each at room temperature. The filters were then washed at 50°C in 10 ml of 0.1 x SSC 0.1% sodium dodecyl sulphate for 5 minutes, followed by two washes in 10 ml 10 mM Tris-HCl pH7.5, 1 mM EDTA for 3 minutes each at 50°C.

Filters were then individually placed in Eppendorf tubes, 150 µl of water added and incubated at 80°C for 2 minutes. The water was then removed and 100 µl of fresh water added and the incubation repeated. The water from these two elutions was pooled and RNA ethanol precipitated twice in the presence of 4 µg of calf liver tRNA. The RNA was re-dissolved in 2-3 µl of water ready for use in the in vitro translation system.

33. PREPARATION OF PARTIAL DIGESTS OF GREY SEAL DNA USING RESTRICTION

ENDONUCLEASE SAU3A1

The Sau3A partial digests of grey seal DNA were prepared by Dr Alec Jeffreys. 13 aliquots each of 40 µg of grey seal DNA at a concentration of 200 µg/ml in 10 mM MgCl₂, 50 µM NaCl, 0.5 M dithiothreitol, 10 mM Tris-HCl, pH7.4 were incubated at 37°C for 20 minutes with increasing amounts of restriction endonuclease Sau3A ranging from 1 to 20 units. 0.4 µg samples were removed from each 40 µg digests and tested for partial digestion by agarose gel electrophoresis. The digests were pooled and the DNA recovered by phenol extraction and ethanol precipitation.

The DNA was vacuum dried, redissolved in water and electrophoresed in a preparative 0.4% agarose gel flanked by λ DNA x HindIII molecular

weight markers. Sau3A partials between 13 and 21 kb were electrophoresed into a dialysis bag and recovered (Materials and Methods - 4).

34. GROWTH AND HARVESTING OF REPLACEMENT VECTOR

Phage lysates were prepared by a modification of the method of Blattner et al. (1977). A fresh plaque of bacteriophage λ 47.1 (Loenen and Brammar, 1980) was dispensed into 1 ml of phage buffer (6 mM Tris-HCl, pH7.2, 10 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.005% gelatin) and 0.1 ml of the phage suspension was added to 5×10^7 cells of E.coli C600 in Luria broth supplemented with 10 mM MgCl_2 . After adsorption for 20 minutes at room temperature the mixture was diluted into 400 ml of Luria broth supplemented with 10 mM MgCl_2 and 0.2% maltose in two 2 l baffled conical flasks. Cultures were incubated with vigorous shaking at 37°C overnight. Obvious lysis and bacterial debris were apparent the following day. Lysates were cleared by centrifugation at 13000 x g for 10 minutes at 4°C and the phage harvested by centrifugation at 51000 x g for 3 hours at 4°C. Each pellet was resuspended in 1 ml of phage buffer.

35. PREPARATION OF PHAGE DNA

Phage suspensions were lysed by adding sodium dodecyl sulphate to 1% and EDTA to 20 mM. This was followed by phenol extraction and two ethanol precipitations without chilling. Vacuum dried pellets were dissolved in 0.1 x TNE and then treated with heat treated ribonuclease and proteinase. The DNA was then phenol extracted and ethanol precipitated three times without chilling and without centrifugation. Precipitates were vacuum dried and redissolved in 50 μ l 10 mM Tris-HCl and then extracted with 2-methoxyethanol/phosphate. The upper phase was removed and the lower phase re-extracted by adding 30 μ l of water and an equal

volume of 2-methoxyethanol and potassium phosphate. The upper phases were pooled and precipitated with 0.5 ml of ethanol. The upper and lower phases were carefully removed leaving DNA at the interface. This was dissolved in 50 μ l 10 mM Tris-HCl pH7.5 and repeatedly ethanol precipitated without chilling or centrifugation until addition of ethanol no longer caused the liquid to turn cloudy. The precipitates were vacuum dried and redissolved in 100 μ l 10 mM Tris-HCl, pH7.5.

36. PREPARATION OF VECTOR ARMS

Vector arms were prepared by Dr Alec Jeffreys 60 μ g of λ 147.1 DNA were digested to completion with restriction endonuclease BamHI and the DNA recovered by phenol extraction and ethanol precipitation, vacuum dried and redissolved in 10 mM Tris-HCl pH7.5. The left and right arms were recovered by electrophoresis into a dialysis bag in a preparative 0.4% agarose gel and the DNA recovered as described earlier.

37. PREPARATION OF λ 147.1 GREY SEAL RECOMBINANTS

Recombinants were prepared by Dr Alec Jeffreys. The first stage was annealing of the cohesive termini of the arms of λ 147.1. 8 μ g of the left and right arms of λ 147.1 were incubated in 16 μ l 0.1 M Tris-HCl pH7.5 and 10 mM $MgCl_2$ at 42 $^{\circ}$ for 1 hour and stored at -20 $^{\circ}$ C.

For ligation, 1 μ g of annealed λ 147.1 arms were mixed with 0.44 μ g of size selected grey seal DNA fragments (i.e. equimolar amounts) and the DNA mixture incubated at 4 $^{\circ}$ C overnight in ligase buffer (66 mM Tris-HCl pH7.5, 6.6 mM $MgCl_2$, 10 mM dithiothreitol and 0.4 mM ATP) in a total volume of 20 μ l and in the presence of 5 units of T4 DNA ligase. 3 μ l aliquots were removed for testing. The remainder were stored at -20 $^{\circ}$ C. To each 3 μ l test sample was added 20 μ l of 66 mM Tris, 10 mM $MgCl_2$ and

the tubes then heated at 68°C for 3 minutes to melt annealed cohesive termini before being placed on ice. Each sample was divided into two, one half alkali denatured, and then both halves electrophoresed in a 0.3% agarose gel alongside unligated mixtures to test the effectiveness of the ligation.

38. IN VITRO PACKAGING

The method used was a modification of the procedure, of Enquist and Sternberg (1979) using E.coli λ lysogens BHB2688 [N205 recA⁻ (λ imm434 cIts b2 red3 Eam4 Sam 7)/ λ] and BHB2690 [N205 recA⁻ (λ imm434 cIts b2 red3 Dam15 Sam7)/ λ] (Hohn, 1979). Streaks of both strains were grown on nutrient agar for 2 days at 30°C. Overnight cultures from single colonies were grown by inoculating 2 ml Luria broth and incubating at 30°C.

(i) Preparation Of a Freeze-thaw Lysate (FTL) From Strain BHB2688

Three 500 ml cultures of BHB2688 in Luria broth in 2 litre baffled conical flasks were grown to $A_{600} = 0.3$ and induced by placing in a water-bath at 45°C for 15 minutes without shaking. The induced cultures were then grown at 37-38°C for 1 hour shaking hard. Cultures were cooled on ice and the cells harvested by centrifugation at 16000 x g for 10 min at 4°C. All the supernatant was drained off by pouring off the bulk and then letting the tube rest at an angle, on ice, with the pellet uppermost. After 5 minutes all drained liquid was removed. Each pellet (250 ml worth of cells) was resuspended in 0.5 ml 10% sucrose, 50 mM Tris-HCl, pH7.5. All the pellets were pooled into 2 Sorvall tubes and to each one was added 75 μ l of fresh lysozyme solution (2 mg/ml in 0.25 M Tris-HCl, pH7.5). These were gently and thoroughly mixed and then

quick-frozen in liquid nitrogen. At this stage the preparation could be stored at -80°C if necessary. The preparations were warmed briefly at room temperature and then at 4°C until completely thawed; they were then cooled on ice. To each tube was added 75 μl of buffer M1 (6 mM Tris-HCl, pH7.5, 30 mM spermidine trichloride, 60 mM putrescine dichloride, 18 mM MgCl_2 , 15 mM ATP neutralised with .880 ammonia, 28 mM 2-mercaptoethanol) with thorough, gentle, mixing. The tubes were then spun at $35000 \times g$ for 50 minutes at 4°C . The supernatant was removed and distributed into pre-cooled microfuge tubes in 50-100 μl aliquots. These were then quick-frozen in liquid nitrogen and stored at -80°C .

(ii) Preparation Of A Sonicated Extract (SE) From Strain BHB2690

One 500 ml culture of BHB2690 was grown at 30°C to $A_{600} = 0.3$, induced, grown for 1 hour at $37-38^{\circ}\text{C}$ and then cells harvested as described for the FTL above. Each pellet was suspended in 0.5 ml of buffer A (20 mM Tris-HCl, pH8.0, 3 mM MgCl_2 , 0.05% (v/v) 2-mercaptoethanol, 1 mM EDTA, pH7.0), the samples pooled in a plastic sonicating tube and diluted with 2.6 ml of buffer A. The cells were sonicated without foaming using 15 x 3 sec blasts of an ethanol sterilised microtip in an MSE 150 watt Ultrasonic Disintegrator keeping the suspension on ice. At this point the mixture was no longer viscous. The cell debris was pelleted by centrifugation at $6000 \times g$ for 6 minutes at 0°C . This produced a small pellet; a large pellet at this stage indicated intact bacterial cells and hence inadequate sonication. The supernatant was distributed into pre-cooled microfuge tubes in 50 μl aliquots, quick-frozen in liquid nitrogen and stored at -80°C .

(iii) IN VITRO PACKAGING OF λ 47.1/GREY SEAL RECOMBINANTS

Before the in vitro packaging extracts were used for packaging recombinant DNA, the efficiency of the system was tested using λ 47.1 and λ wt as DNA substrates. λ 47.1 was packaged into infectious phage particles at 7.5×10^6 pfu/ μ g whilst λ wt DNA was packaged at 3.7×10^7 pfu/ μ g.

For in vitro packaging, components were mixed in the following order : 0.15 μ g of DNA (in a volume of 3.6 μ l), 9 μ l of buffer A, 1.4 μ l of M1 buffer, 8.4 μ l of SE, and 14 μ l FTL. Mixtures were incubated at 25°C for 1 hour and then diluted on ice into 0.5 ml of phage buffer. 10 μ l of each were then plated in serial dilutions onto the following E.coli strains : WL87 (803, supE, supF, hsdR_k⁻, hsdM_k⁺, tonA, trpR⁻, metB), WL95, a P2 lysogen of WL87 for recombinant selection and ED8910 (803, supE, supF, recB21, recC22, hsdS). 240 μ l of each diluted preparation were also plated onto single plates of WL87 and ED8910. All soft agar overlays were supplemented with 10 mM MgCl₂ and 0.2% maltose. No more than 30 μ l of a packaging reaction or its diluted equivalent were used for a single plating. Controls were λ 47.1 ligated arms and intact λ 47.1 alone. Plates were incubated at 37°C overnight. Subsequent reactions were scaled up by 3.3 and after dilution into 0.5 ml of phage buffer were divided into four 155 μ l aliquots each of which was plated on a single 9 cm plate of WL95.

39. SELECTION OF PHAGE RECOMBINANTS CONTAINING GREY SEAL MYOGLOBIN

GENE

The method used was that of Benton and Davis (1977). The plates screened were those on which about half the diluted in vitro packaging reaction had been plated onto strain WL95. After phage growth the agar was hardened by cooling the plates at 4°C for 15 minutes. 88 mm diameter nitrocellulose filters (Sartorius 0.45 µm pore size) were placed directly onto the agar without trapping air bubbles. Phage were allowed to transfer onto the filters for 5 minutes during which time filters and plates were marked for future orientation. The filters were then gently peeled off and washed for 1 minute in 0.1 M NaOH, 1.5 M NaCl and then neutralised by washing for 1 minute in 0.2 M Tris-HCl pH7.5, 2 x SSC before being blotted dry and baked for 5 hours at 80°C. The filters were then hybridised with ³²P labelled pSM 178 DNA as described earlier. Washing after overnight hybridisation followed the standard pattern finishing with a final stringency wash of 0.3 x SSC. Filters were then dried and recombinant phage-probe hybrids detected by autoradiography for 1-4 days.

40. PURIFICATION OF POSITIVELY HYBRIDISING RECOMBINANT CLONES

Each positively-hybridising recombinant plaque or region was picked, resuspended in 0.5 ml of phage buffer, shaken with chloroform and 100 µl plated in serial dilutions onto strain ~~WL95~~10 genotype.

Plates were incubated overnight at 37°C. Plates with well separated plaques were then screened as before and positively hybridising clones and rescreened as described above.

41. PHAGE AMPLIFICATION

Positively-hybridising well separated plaques were dispersed into 0.5 ml phage buffer and 100 μ l used for serial dilutions on E₃₉₁₀ using Luria agar plates; the remaining 0.4 ml was used for a single plate. After overnight incubation at 37°C plates with near confluent lysis were overlaid with 3 ml of Luria broth supplemented with 10 mM MgCl₂ and left to stand for 10 minutes at room temperature. The liquid and soft agar overlays were then removed, gently mixed with the agar well macerated then cleared by centrifugation at 16000 x g for 5 minutes. One drop of chloroform was added to the supernatant which was stored at 4°C.

42. RESTRICTION ENDONUCLEASE MAPPING OF RECOMBINANT PHAGE

0.5 μ g of recombinant phage DNA was used for each digestion in a final volume of 10 μ l. Digested DNA was not recovered as was usual for larger DNA digests. For subsequent digestion with another restriction endonuclease 1 μ l of a 10 x conversion buffer was added directly to adjust the salt conditions to those required by the next restriction endonuclease. Digests were electrophoresed directly in 0.4% agarose gels without recovering the DNA.

43. PREPARATION OF LINEARISED AND PHOSPHATASED PLASMID pAT153

Linearised, phosphatased pAT153 was prepared by Dr Alec Jeffreys. 20 μ g of plasmid pAT153 was linearised by digestion with restriction endonuclease HindIII. The DNA was recovered in the usual way redissolving the pellet in 10 mM Tris-HCl, pH7.5. 15 μ g of linearised plasmid DNA in 150 μ l 10 mM Tris-HCl, pH7.5, was incubated with 0.15 units of calf intestinal phosphatase at 37°C for 1 hour. The DNA was extracted twice with phenol before being ethanol precipitated and redissolved in 10 mM Tris-HCl, pH7.5.

44. LIGATION OF FRAGMENTS INTO PLASMID pAT153

Recombinant phage DNA was digested with the restriction endonuclease HindIII and the DNA recovered in the usual way. Fragments produced by digestion with restriction endonuclease HindIII were ligated with plasmid pAT153 linearised with restriction endonuclease HindIII. The linearised plasmid had been phosphatased using calf intestinal phosphatase.

A total of 100 μ g of linearised, phosphatased plasmid DNA and recombinant phage fragment DNA were mixed in equimolar proportions in a total reaction volume of 25 μ l of ligase buffer and incubated with 1.5 units of the DNA ligase at 4°C overnight. Ligation was tested by agarose gel electrophoresis.

45. FRAGMENT PREPARATION AND END LABELLING FOR SEQUENCING

(i) Restriction Endonuclease Digestion

10 μ g of recombinant plasmid DNA was digested with an appropriate restriction endonuclease under conditions just compatible with complete digestion in a volume of 100 μ l. 5 μ l were checked on agarose gels for complete digestion. If digestion was incomplete more enzyme was added and the incubation repeated. DNA was recovered from completed digests by a phenol extraction followed by two ethanol precipitations. The DNA pellet was redissolved in 20 μ l 10 mM Tris-HCl pH7.5.

(ii) Labelling DNA Ends With Polynucleotide Kinase And γ -³²P ATP

10 μ g of restricted DNA in 50 μ l of 10 mM Tris-HCl pH7.5 were incubated at 37°C for 45 minutes with 0.1 units of calf intestinal phosphatase. After incubation the DNA was recovered by two phenol extractions followed by two ethanol precipitations. The DNA pellet was redissolved in 10 μ l 10 mM Tris-HCl, pH7.5 and incubated with 5 units of T4 polynucleotide kinase at 37°C for 30 minutes in the presence

of 50 mM Tris-HCl, pH7.6, 10 mM MgCl_2 , 50 mM dithiothreitol, 4 mM spermidine, 0.1 mM EDTA, and 25 μCi of $\gamma\text{-}^{32}\text{P}$ ATP, in a reaction volume of 15 μl . 0.7 μl were removed and electrophoresed in an agarose gel. The gel slice containing the labelled fragment was excised and Cerenkov counted in order to estimate incorporation. DNA was recovered from the reaction by two phenol extractions followed by two ethanol precipitations.

(iii) Labelling DNA Ends By 'Filling-in' With Reverse Transcriptase and $\alpha\text{-}^{32}\text{P}$ dNTPs

10 μg of DNA in a total volume of 50 μl was incubated with 16 units of reverse transcriptase at 37°C for 90 minutes in the presence of 100 mM Tris-HCl pH8.0, 1 mM 2-mercaptoethanol, 10 mM MgCl_2 , 10 μM dATP, 10 μM dGTP, 10 μM dTTP and 30 μCi $\alpha\text{-}^{32}\text{P}$ dCTP. Different mixes were used to suit the end being labelled, using different labelled and unlabelled dNTPs. After incubation DNA was recovered by phenol extraction followed by two ethanol precipitations.

(iv) Second Restriction Endonuclease Digestion

Labelled DNA was digested in a volume of 100 μl with an appropriate restriction endonuclease. 5 μl were checked against 5 μl of input for complete digestion. After digestion the DNA fragment(s) was recovered by electroelution from a preparative agarose gel as described earlier. The labelled DNA was then cleaned by using DEAE-cellulose columns as follows:

For each gel slice 2 DEAE-cellulose columns were prepared using autoclaved Whatman preswollen DE52-cellulose in 1/10 x agarose gel electrophoresis buffer. Columns were constructed by plugging a 1 ml plastic Eppendorf pipette tip with autoclaved polyallomer wool (aquarium filter wool), cutting off 2 mm from the end to enlarge the

orifice and loading with a 100 μ l - 200 μ l bed volume of DEAE-cellulose which was then rinsed with 1/5 x electrophoresis buffer. The column used was that with the fastest flow rate. A third Eppendorf pipette tip was plugged with polyallomer wool. The eluted DNA was first passed through the tip with polyallomer wool and then loaded onto a DEAE-cellulose column. The column was then washed with 0.5 ml of 1/5 x electrophoresis buffer and then the bound DNA eluted from the column with two 200 μ l aliquots of 1 M NaCl, 50 mM Tris-HCl pH7.5 1 mM EDTA. 1 ml of ethanol was added and the mixture chilled in a dry ice-IMS bath for 10 minutes. The precipitated DNA was pelleted by centrifugation for 10 minutes in an Eppendorf centrifuge the supernatant removed and the pellet redissolved in 250 μ l of 0.3 sodium acetate pH6.0. This was followed by 750 μ l ethanol, chilling and centrifugation as before. The pellet was then rinsed with 80% ethanol very gently to avoid dislodging it; and the tube centrifuged as before before removing the alcohol. The pellet was vacuum dried and redissolved in 55 μ l of distilled water ready for sequencing.

(v) DNA Sequencing

All the procedures used for DNA sequencing of 32 P end-labelled DNA fragments were exactly those described by Maxam and Gilbert (1980) in their published protocol.

46. CONTAINMENT

All recombinant DNA experiments described were performed under the Containment Category I conditions stipulated by the Genetic Manipulation Advisory Group.

Chapter 3

DETECTION OF HUMAN MYOGLOBIN mRNA

1. PURIFICATION OF HUMAN MYOGLOBIN

Myoglobin accounts for approximately 1% of the protein of adult human skeletal muscle. Myoglobin was purified from this muscle by iso-electric focussing and sds-polyacrylamide gel electrophoresis.

Homogenates of adult human skeletal muscle were loaded onto an ampholine polyacrylamide iso-electric focussing gel and focussed against adult haemoglobin markers. After focussing, the major non-haemoglobin red pigmented protein band (myoglobin) was excised and the native myoglobin recovered from the gel slice.

This protein was further purified by electrophoresis in a 15% sds-polyacrylamide gel and by staining marker protein tracts the region of gel containing the myoglobin protein excised, and the protein recovered. This protein was used in subsequent experiments as 'pure' human apo-myoglobin.

2. ISOLATION AND ANALYSIS OF POLYADENYLATED RNA FROM HUMAN SKELETAL MUSCLE

Initially a number of methods were used to isolate RNA from skeletal muscle. All of these methods involved the homogenisation of frozen muscle samples, lysis of cells with either detergent or a detergent based mixture, phenol extraction of protein and ethanol precipitation of nucleic acids. Initially homogenates were lysed in an equal volume of 8% tri-isopropyl-naphthalene-sulphonic acid, 2% sodium dodecyl sulphate and 12% butan-2-ol. Preparations using this lysis mixture were usually

badly degraded and subsequent preparations used a gentle lysis in 1% sodium dodecyl sulphate. Preparations using this method of lysis were generally not badly degraded as revealed by agarose gel electrophoresis where relatively high molecular weight RNAs including ribosomal RNA bands could be detected (Figure 3:1).

A number of sources of human skeletal muscle tissue were used to prepare RNA. Adult tissue was generally from post mortems or amputated calf muscle. Foetal samples were from abortuses and neonate samples from still born infants. It was found that RNA prepared from post mortem samples was often badly degraded (Figure 3:1). This was probably due to long delays of up to 24 hours between death and the freezing of tissue samples. In addition amputated calf muscle often gave yields of badly degraded RNA (Figure 3:1). This was possibly because the patients were suffering from diabetes and their muscle tissue may have been ischemic due to poor circulation.

The method finally adopted to prepare RNA from muscle samples generally gave good yields of relatively undegraded RNA (Figure 3:1). It was found necessary to make all solutions up with water treated with 0.1% diethylpyrocarbonate in order to inhibit ribonuclease activity from the lysed cells. Tissue was stored at -80°C until required. Before use the muscle was placed in liquid nitrogen for at least one hour, in order to facilitate easier homogenisation. All subsequent procedures were carried out on ice and as swiftly as possible, using precooled sterile equipment. After homogenisation the tissue was gently lysed, phenol extracted and nucleic acids ethanol precipitated. Typical yields of nucleic acid from adult human skeletal muscle are shown in Table 3:1.

RNA was separated from DNA by centrifugation in caesium chloride. Total nucleic acid samples were made up to 5.7 M caesium chloride and centrifuged in a Swing-out rotor at $160000 \times g$ for 22 hours at 25°C .

Figure 3:1 Agarose gel electrophoresis of skeletal muscle nucleic acid preparations

Nucleic acid preparations from various sources of skeletal muscle were electrophoresed in 2% agarose gels. Preparations of nucleic acid from adult skeletal muscle using the 'harsh' lysis procedure (A lanes 1 and 2) show an intense high molecular weight DNA band (D) and a smear of RNA, a low molecular weight tRNA band (t) is visible, but no bands corresponding to 28S and 19S ribosomal RNA can be seen. On alkali treatment of the nucleic acid preparation most of the RNA is degraded to low molecular weight oligonucleotides (lanes 3 and 4).

In preparations of nucleic acid from the same muscle using the 'gentle' lysis procedure (B) higher molecular weight RNA bands can be visualised (lanes 1 and 2) which are degraded on alkali treatment (lanes 3 and 4).

Preparations of nucleic acid from human adult amputated calf muscle often showed considerable RNA degradation (C lanes 1) with no high molecular weight RNA visible. The smear of low molecular weight RNA treated with alkali shows further degradation (lane 2). Similarly some adult skeletal muscle post mortem samples showed no high molecular weight RNA bands (D lanes 1 and 2).

Human adult skeletal muscle nucleic acid prepared from post mortem samples by gentle lysis in the presence of diethylpyrocarbonate (E lanes 1 and 2) showed a high molecular weight DNA band (D) and 28S and 18S ribosomal RNA bands (28 and 18) as well as a smear of RNA with fainter RNA bands superimposed on it and a tRNA band. On alkali treatment (lanes 3 and 4) this RNA was degraded into small molecular weight oligonucleotides.

O denotes the origin of electrophoresis.

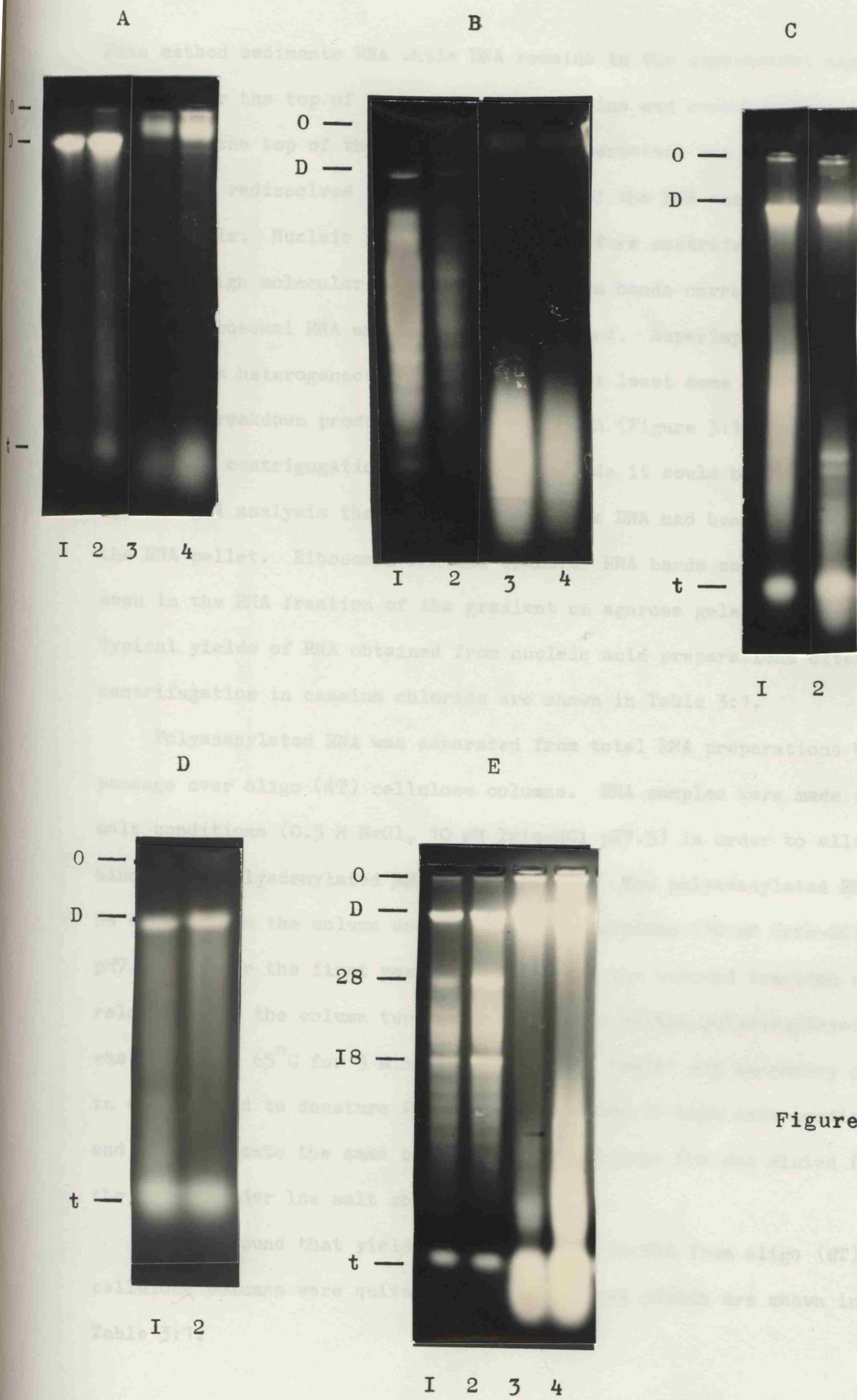


Figure 3: I

This method sediments RNA while DNA remains in the supernatant and forms a band near the top of the gradient. Proteins and carbohydrates form a 'skin' at the top of the gradient. The supernatant was removed and the RNA pellet redissolved in water. Quality of the RNA was checked on agarose gels. Nucleic acid preparations before centrifugation clearly showed a high molecular weight DNA band, two bands corresponding to 28S and 18S ribosomal RNA and a small 5S RNA band. Superimposed on these bands was a heterogeneous smear of RNAs. At least some of this RNA was probably breakdown products of ribosomal RNA (Figure 3:1).

After centrifugation in caesium chloride it could be shown by agarose gel analysis that all or most of the DNA had been separated from the RNA pellet. Ribosomal RNA and transfer RNA bands could be clearly seen in the RNA fraction of the gradient on agarose gels. Typical yields of RNA obtained from nucleic acid preparations after centrifugation in caesium chloride are shown in Table 3:1.

Polyadenylated RNA was separated from total RNA preparations by passage over oligo (dT) cellulose columns. RNA samples were made to high salt conditions (0.5 M NaCl, 10 mM Tris-HCl pH7.5) in order to allow the binding of polyadenylated RNA to the column. The polyadenylated RNA could be eluted from the column under low salt conditions (10 mM Tris-HCl pH7.5). After the first passage and elution the unbound fraction was reloaded onto the column two more times. The eluted polyadenylated RNA was heated at 65°C for 5 minutes in order to 'melt' any secondary structure in the RNA and to denature RNA aggregates, made to high salt conditions and reloaded onto the same column. Polyadenylated RNA was eluted from the column under low salt conditions.

It was found that yields of polyadenylated RNA from oligo (dT) cellulose columns were quite variable. Typical yields are shown in Table 3:1.

Table 3:1 Typical yields of nucleic acid prepared from human skeletal muscle

<u>Nucleic acid fraction</u>	<u>Yield in mg/kg tissue</u>	<u>% of total nucleic acid</u>
Total nucleic acid	150	100
Total RNA (after CsCl centrifugation)	75	50
polyA ⁺ mRNA (after oligo (dT) cellulose chromatography)	0.4	0.3
polyA ⁻ RNA (after oligo (dT) cellulose chromatography)	~ 70	~ 47

The quality of polyadenylated RNA was checked on agarose gels. A range of molecular weights of RNA could be detected. Little 5S RNA was present, whereas both ribosomal RNA bands could be detected (Figure 3:2). This ribosomal RNA contamination of polyadenylated RNA did not seem to inhibit subsequent in vitro translations or cDNA synthesis.

3. IN VITRO TRANSLATION OF HUMAN SKELETAL MUSCLE RNA

RNA and polyadenylated RNA preparations from adult human skeletal muscle were translated in vitro in a rabbit reticulocyte lysate system. After incubation, non-radioactive methionine and RNase were added to prevent further incorporation of radioactive methionine and to chase label into mature polypeptides. The products of the in vitro translation were analysed by electrophoresis in sds-polyacrylamide gels. Gels of high concentration (15% and 17%) were used in order to obtain good resolution between globin (16.4 kd) and myoglobin (17.2 kd).

A wide range of different molecular weight proteins up to 200 kd could be detected in the in vitro translation profiles on sds-polyacrylamide gels. The smallest detectable protein in in vitro translations of adult human polyadenylated RNA was of approximately 17 kd (Figure 3:3). Smaller proteins could not be detected in this system as they would be obscured by the large globin band from the lysate.

When human myoglobin was electrophoresed and the stain profile compared with the in vitro translation product profile it was observed that the 17 kd protein comigrated with purified human myoglobin.

Figure 3:2 Agarose gel electrophoresis of skeletal muscle
polyadenylated RNA

Polyadenylated RNA from human adult skeletal muscle was electrophoresed in 2% agarose gels. Lane 1 shows polyadenylated RNA with some ribosomal 28S and 18S RNA contamination, but no DNA or tRNA. On alkali treatment the RNA is degraded to low molecular weight oligonucleotides (lane 2).

O denotes the origin of electrophoresis and nt = nucleotides.

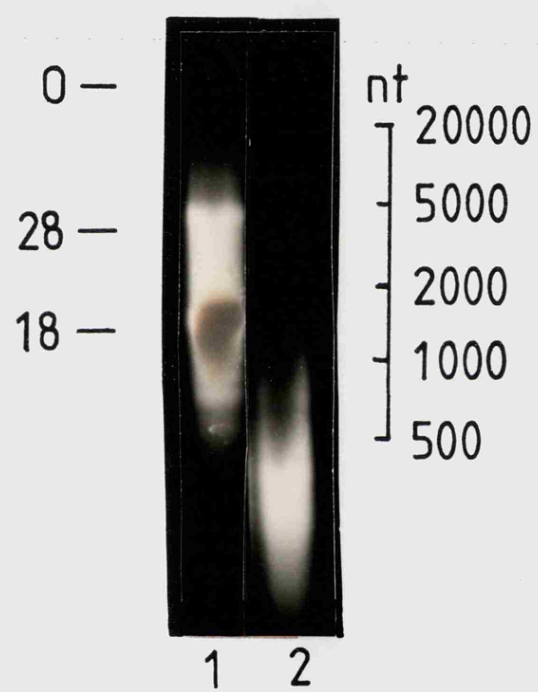


Figure 3:2

Figure 3:3 In vitro translation profile of human skeletal muscle
polyadenylated RNA

0.1 µg of rabbit globin mRNA (lane 2) and 0.1 µg of human skeletal muscle polyadenylated RNA (lane 3) were translated in the rabbit reticulocyte lysate system and the translation products electrophoresed in 17% sds-polyacrylamide gels. A range of translation products could be seen including a number of small proteins between 17 and 20 kd. Myoglobin migrates as a 17 kd protein.

The 42 kd protein which appears in the zero RNA control (Lane 1) is present in the reticulocyte lysate. This protein comigrates with actin in sds-polyacrylamide gels but has been shown to have a different isoelectric point to actin and is probably binding unincorporated ³⁵S-methionine.

O denotes the origin of electrophoresis.

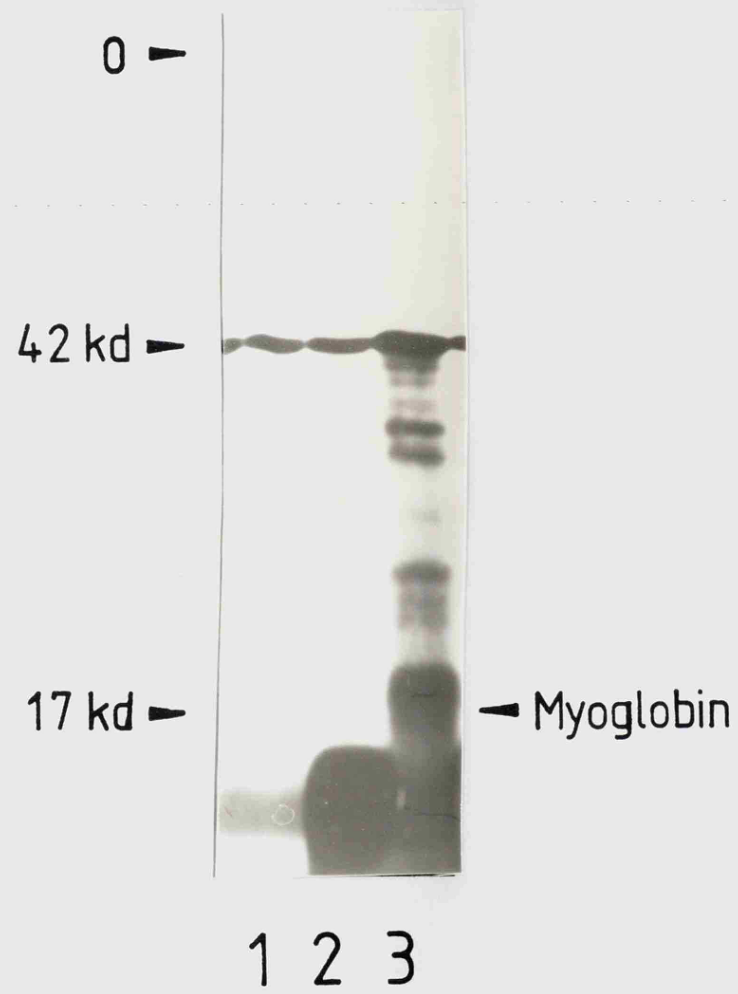


Figure 3:3

4. IDENTIFICATION OF MYOGLOBIN IN IN VITRO TRANSLATION PROFILES BY IMMUNOPRECIPITATION

Human myoglobin is a protein of 17.2 kd molecular weight. In order to determine whether the 17 kd band which comigrates with purified human myoglobin is myoglobin, immunoprecipitation with antihuman myoglobin antiserum was performed.

Adult human skeletal muscle polyadenylated RNA was translated in vitro in the rabbit reticulocyte lysate system. The translation mixture was immunoprecipitated with anti-human myoglobin antiserum raised in goats. As a control, a similar translation mixture was immunoprecipitated with anti-human transferrin antiserum. After incubation at 0°C overnight Protein-A was added to the mixture in order to precipitate any antigen-antibody complexes. Precipitated products were analysed on sds-polyacrylamide gels. Using anti-human myoglobin antiserum, a protein corresponding to the small 17 kd protein in in vitro translations could be specifically immunoprecipitated. With the anti-human transferrin antiserum no protein could be specifically immunoprecipitated from in vitro translations of adult human skeletal muscle polyadenylated RNA (Figure 3:4).

Figure 3:4 Immunoprecipitation of myoglobin from in vitro translations of total nucleic acid and polyadenylated RNA from human skeletal muscle

The in vitro translation profiles of polyadenylated RNA (A) total nucleic acid (B) and a zero RNA control (C) and immunoprecipitations from these in vitro translations are shown. Anti-human myoglobin (m) or antihuman transferrin (t) antisera were added to the in vitro translations of the zero RNA (zero) total nucleic acid (total) and polyadenylated RNA (polyA) and any antigen-antibody complexes precipitated using protein-A. The pellets and supernatants from these immunoprecipitation were electrophoresed in a 17% sds-polyacrylamide gel.

Myoglobin was present in the in vitro translations of total nucleic acid and polyA RNA as well as the supernatants of the immunoprecipitated samples of these translations. No myoglobin was present in the zero RNA controls. Myoglobin (17 kd) was specifically immunoprecipitated from the polyA translation to which anti-human myoglobin antiserum had been added but not from the zero RNA or total nucleic acid in vitro translations. The anti-human transferrin antiserum did not precipitate any protein from either the total nucleic or polyA RNA translations. The scale is shown in kilodaltons (kd).

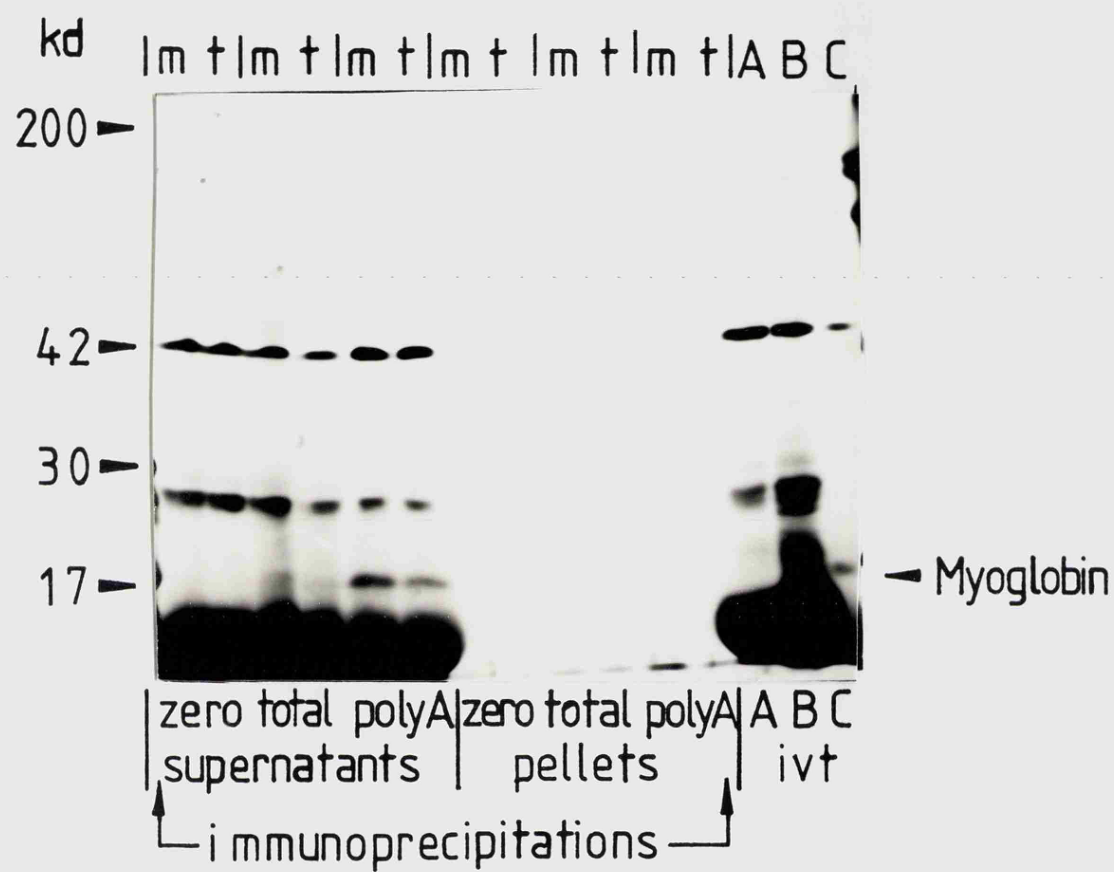


Figure 3:4

Chapter 4

THE CONSTRUCTION AND ANALYSIS OF A cDNA LIBRARY FROM ADULT HUMAN SKELETAL MUSCLE POLYADENYLATED RNA

Adult human skeletal muscle polyadenylated RNA was used to construct a cDNA library in E.coli HB101. A schematic representation of the steps involved in the construction of a cDNA library are shown in Figure 4:1 .

1. FIRST STRAND SYNTHESIS

Single stranded cDNA was made from polyadenylated RNA using an oligo (dT)₁₀ primer and AMV reverse transcriptase. Use of an oligo (dT) primer should ensure that cDNA synthesis is initiated near the 3' terminus of the RNA in the poly(A) tract. Synthesis of cDNA was monitored by following the incorporation of α -³²P-dCTP into the synthesised cDNA strand.

Reaction conditions were optimised using a standard reaction of 10 μ l volume containing 1 μ Ci of α -³²P-dCTP (2-3000 Ci/mMole). Initial experiments showed that when 1 μ g rabbit globin mRNA was reverse transcribed in a 10 μ l reaction reverse transcriptase activities of greater than 3 units per 10 μ l reaction resulted in decreased yields of cDNA. This is presumably due to ribonuclease contamination of the reverse transcriptase, as described by Buell et al. (1979). Therefore, optimising experiments were carried out at a reverse transcriptase concentration of 2.75 units per 10 μ l reaction volume. In order to determine the optimal concentration of polyadenylated RNA required, varying amounts of polyadenylated RNA were reverse transcribed in a standard 10 μ l reaction. The yields of single stranded cDNA generally obtained are shown in Table (4:1).

In order to estimate the size of cDNA synthesised, the cDNA/polyadenylated RNA hybrids were hydrolysed in alkali and the single stranded

Figure 4:1 Strategy for the synthesis of a cDNA copy of polyadenylated mRNA and insertion into a plasmid vector by homopolymer tailing

Polyadenylated mRNA is copied by reverse transcriptase (r.t.) and treated with alkali to form a single stranded cDNA. Reverse transcriptase is used to synthesise the second strand and S1 nuclease (S1) used to remove the 'hairpin' loop. The double stranded cDNA is then tailed with oligo (dC) using terminal transferase (t.t.).

The plasmid vector (pAT153) is linearised with Pst1, tailed with oligo (dG) using terminal transferase and annealed with the oligo (dC) tailed cDNA, thus recreating the Pst1 target sites. The annealed DNA is then used to transform E.coli HB101 and tetracycline transformants selected (Tc^R).

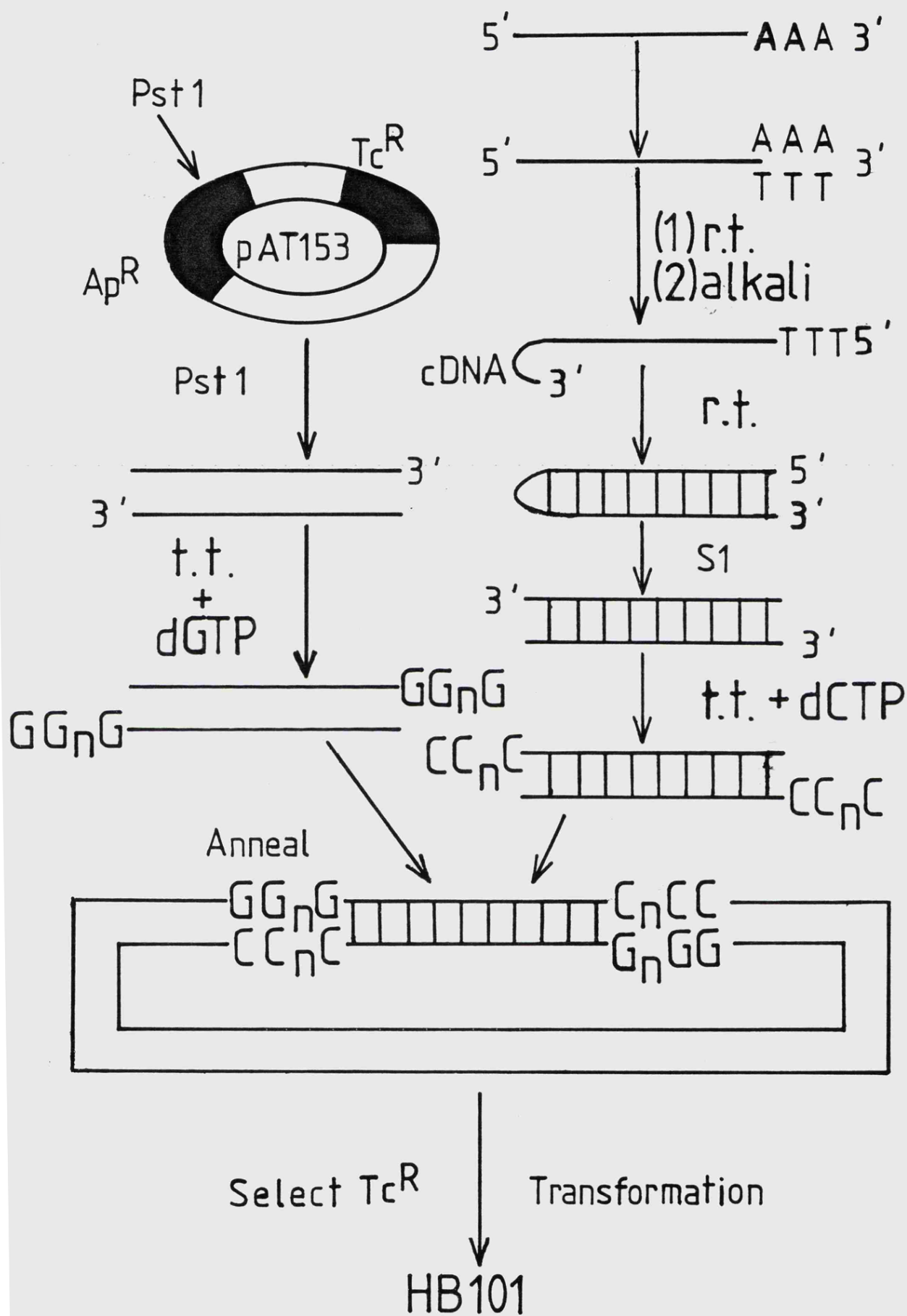


Figure 4:1

Table 4:1

Yields of cDNA transcribed from poly(A)⁺ RNA with reverse transcriptase
at a concentration of 2.75 units/10 μ l reaction volume

<u>μg polyA⁺</u>	p moles dCTP	p moles dNTP	<u>ng cDNA</u>	<u>ng cDNA/μg polyA⁺</u>
	<u>incorporated</u>	<u>incorporated</u>		
0.10 [*]	7.3	29.2	9.7	97.0
0.05	7.9	31.6	10.4	208.0
0.10	22.6	90.4	29.8	298.0
0.20	14.3	57.2	18.9	94.5
0.40	24.5	98.0	32.3	80.8
0.80	16.5	66.0	21.8	27.3
1.60	18.4	73.6	24.3	15.2
3.20	16.4	65.6	21.7	6.8

* globin mRNA control

cDNA electrophoresed in agarose gels, which were subsequently dried down and autoradiographed. From these gels it was determined that cDNAs between 3000 nucleotides and less than 100 nucleotides could be synthesised, with an average size of approximately 300 nucleotides (Figure 4:2). It is clear that the majority of cDNA molecules are shorter than 1000 nucleotides and only a small proportion are as long as 3000 nucleotides in length.

Bulk preparations of single stranded cDNA were direct scale-ups of the optimised reaction conditions. 40 μ g of polyadenylated RNA were incubated with 550 units of reverse transcriptase and 15 μ g oligo (dT) in a 2 ml reaction volume, at a radioactive concentration of 20 μ Ci α - 32 P-dCTP per ml of reaction volume (2-3000 Ci/mMole). The ratio of labelled to unlabelled dCTP was reduced to 1 : 7500, to minimise radio-lysis of single stranded cDNA.

As previous experiments had shown that unincorporated α - 32 P-dCTP could be ethanol precipitated to a certain degree, the cDNA/polyadenylated RNA hybrids were recovered by chromatography on Sephadex G50 column into 10 mM Tris-HCl pH7.5. After alkali hydrolysis of the RNA it was determined that 1.5 μ g of single stranded cDNA was obtained from this reaction.

2. SECOND STRAND SYNTHESIS

Synthesis of the second cDNA strand was achieved using reverse transcriptase in the "self-primed" reaction. This reaction relies on a region of self complementarity at the 3' end of a cDNA molecule which results in the formation of a "hairpin" structure (Leis and Hurwitz, 1972). Addition of reverse transcriptase allows the hairpin to act as primer for elongation, resulting in the synthesis of a double stranded copy of the cDNA with one "closed" end (containing sequences derived from

Figure 4:2 Agarose gel electrophoresis of single stranded cDNA

cDNA/mRNA hybrids were hydrolysed in alkali and the cDNA electrophoresed in 2% agarose against single stranded DNA markers (unlabelled). The gels were photographed, dried down and autoradiographed. Lane 1 has twice the amount of labelled cDNA as lane 2. The size range of cDNAs is from around 100 bases to around 3000 bases with an average size of around 300 bases.

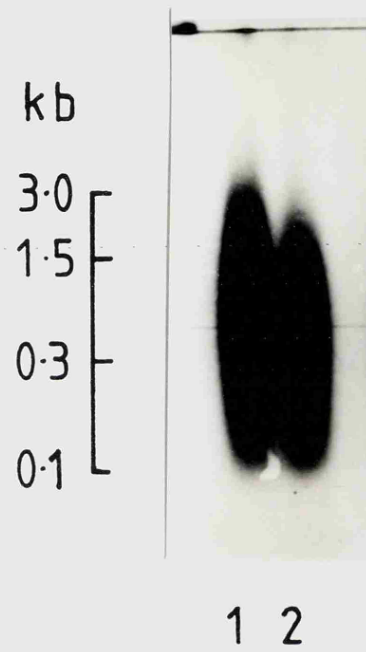


Figure 4:2

the 5' end of the mRNA) and one "open" end (containing 3' derived mRNA sequences).

The second strand synthesis reaction was monitored by determining the degree of protection of the initial ^{32}P -single stranded cDNA molecules against S1-nuclease which is specific for single stranded DNA. After second strand synthesis (see Materials and Methods) cDNA molecules were recovered by Sephadex G50 column chromatography to remove any deoxy-nucleotide triphosphates as concentrations of $1\text{ }\mu\text{M}$ deoxynucleotide triphosphates inhibit S1-nuclease activity (Wickens, 1978).

Optimisation experiments indicated that 0.7 units of S1-nuclease in a $10\text{ }\mu\text{l}$ reaction would totally digest $1\text{ }\mu\text{g}$ of heat-denatured human genomic DNA but would not digest $1\text{ }\mu\text{g}$ of λ DNA cleaved with EcoRI (double stranded) Figure (4:3).

$1\text{ }\mu\text{g}$ of single stranded cDNA was incubated with reverse transcriptase and recovered by Sephadex G50 column chromatography, and then treated with S1-nuclease. 250 ng of double stranded cDNA were recovered from this reaction.

In order to estimate the amount of single stranded cDNA converted to double stranded form the degree of S1-nuclease resistance was determined. An aliquot of the S1-nuclease treated double stranded cDNA was mixed with $1\text{ }\mu\text{g}$ of human DNA as an internal control, alkali denatured and neutralised with hydrochloric acid. This DNA was mixed with $1\text{ }\mu\text{g}$ of λ DNA x EcoRI and incubated at 37°C for 15 minutes with 5 units of S1-nuclease. Samples were analysed on a 2% agarose gel which was photographed, dried down and autoradiographed. Gel analysis revealed that the human DNA was completely degraded by S1-nuclease but that the λ DNA x EcoRI was completely resistant to S1-nuclease attack. The autoradiograph (Figure 4:4) revealed

Figure 4:3 Agarose gel electrophoresis of S1 treated DNA

1 μ g of heat denatured human DNA was mixed with 1 μ g of λ DNA cleaved with EcoR1 and electrophoresed in a 1.5% agarose gel (lanes 3-7). Similar DNA samples were treated with varying amounts of S1 nuclease : 1.5 units, 1.0 units, 0.7 units, and 0.3 units (lanes 8, 9, 10 and 11 respectively). Lanes 1 and 2 contain λ x HindIII and pBR322 x Sau3A DNA markers respectively.

0 = origin of electrophoresis.

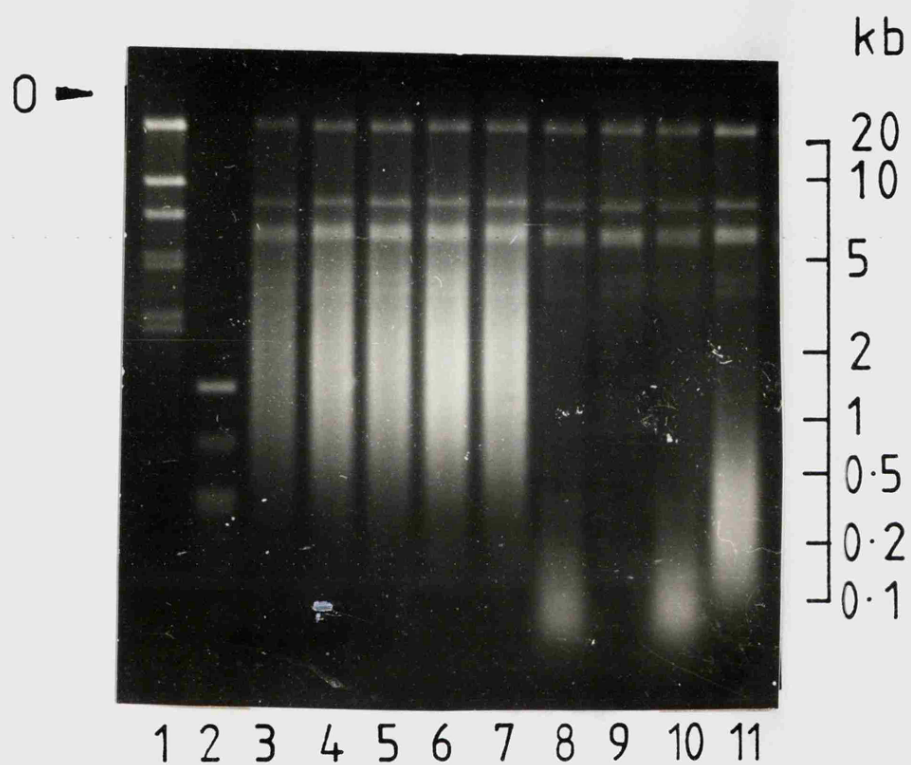


Figure 4:3

Figure 4:4 Agarose gel electrophoresis of S1 treated cDNA

³²P-labelled single stranded cDNA (lane 1) was treated with S1 nuclease (lane 2) and degraded to small oligonucleotides. Double stranded cDNA, after S1 nuclease treatment to remove hairpins, was alkali denatured and treated with S1 nuclease and degraded to small oligonucleotides (lane 3). Non alkali denatured double stranded cDNA remained resistant to S1 nuclease (lane 4). Lane 5 is input double stranded cDNA.

0 denotes the origin of electrophoresis.

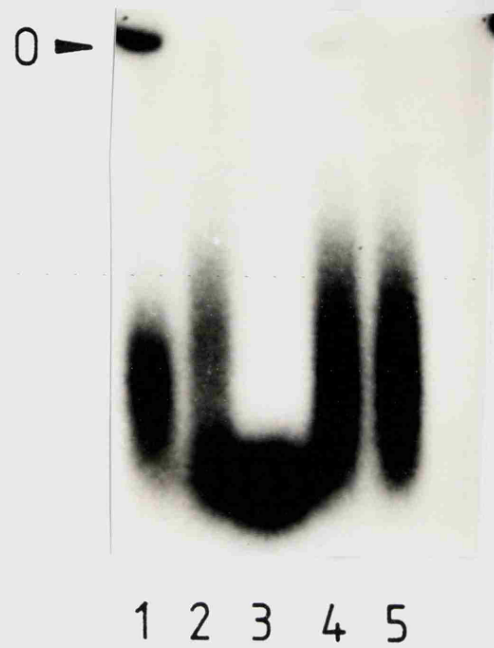


Figure 4:4

that the S1-nuclease treated double stranded cDNA remained resistant. The alkali denatured double stranded cDNA was sensitive to S1-nuclease. This implies that the hairpin structures had been removed by S1-nuclease treatment, as double stranded cDNA with hairpins would immediately 'snap back' after denaturation and neutralisation and remain S1-nuclease resistant.

3. HOMOPOLYMER TAILING OF cDNA AND pAT153

Terminal deoxynucleotidyl transferase, from calf thymus, catalyses the addition of nucleotides to a suitable 3' terminus in the absence of a template. Terminal transferase will synthesise a homopolymer tract or "tail" of defined length at both 3' termini of the double stranded cDNA. If a homopolymer tail of the complementary nucleotide is added to the termini of a linearised plasmid then stable recombinants can be formed by annealing the double stranded cDNA to the plasmid (Jackson *et al.*, 1972; Lobban and Kaiser, 1973).

In order to insert double stranded cDNA into the PstI site of pAT153, poly dG - poly dC homopolymer tailing was used. Adding a tail of "C" residues to the cDNA and "G" residues to the plasmid DNA, the PstI site of the plasmid is reconstructed which allows the excision of the cDNA insert using PstI.

(i) "C" tailing double stranded cDNA

Optimisation experiments were carried out as described in Materials and Methods. Approximately 40 ng of double stranded cDNA were tailed, the incorporation of "C" residues being monitored by the incorporation of α -³²P-dCTP. An average of 60 "C" residues were added to each 3' terminus of the double stranded cDNA (Figure 4:5).

Figure 4:5 Homopolymer tailing of double stranded cDNA

Double stranded cDNA was incubated with terminal transferase and aliquots removed at 5, 10, 15, 30, 60 and 90 minutes, quenched with sds and precipitated with TCA in the presence of high molecular weight salmon sperm DNA. Samples were filtered through glass fibre filters which were dried and Cerenkov counted. The incorporation of α - ^{32}P -dCTP was determined and the average number of "C" residues per end of each cDNA molecule calculated. The number of "C" residues added increases with time for the first 60 minutes and then levels off at just over 100 residues per end.

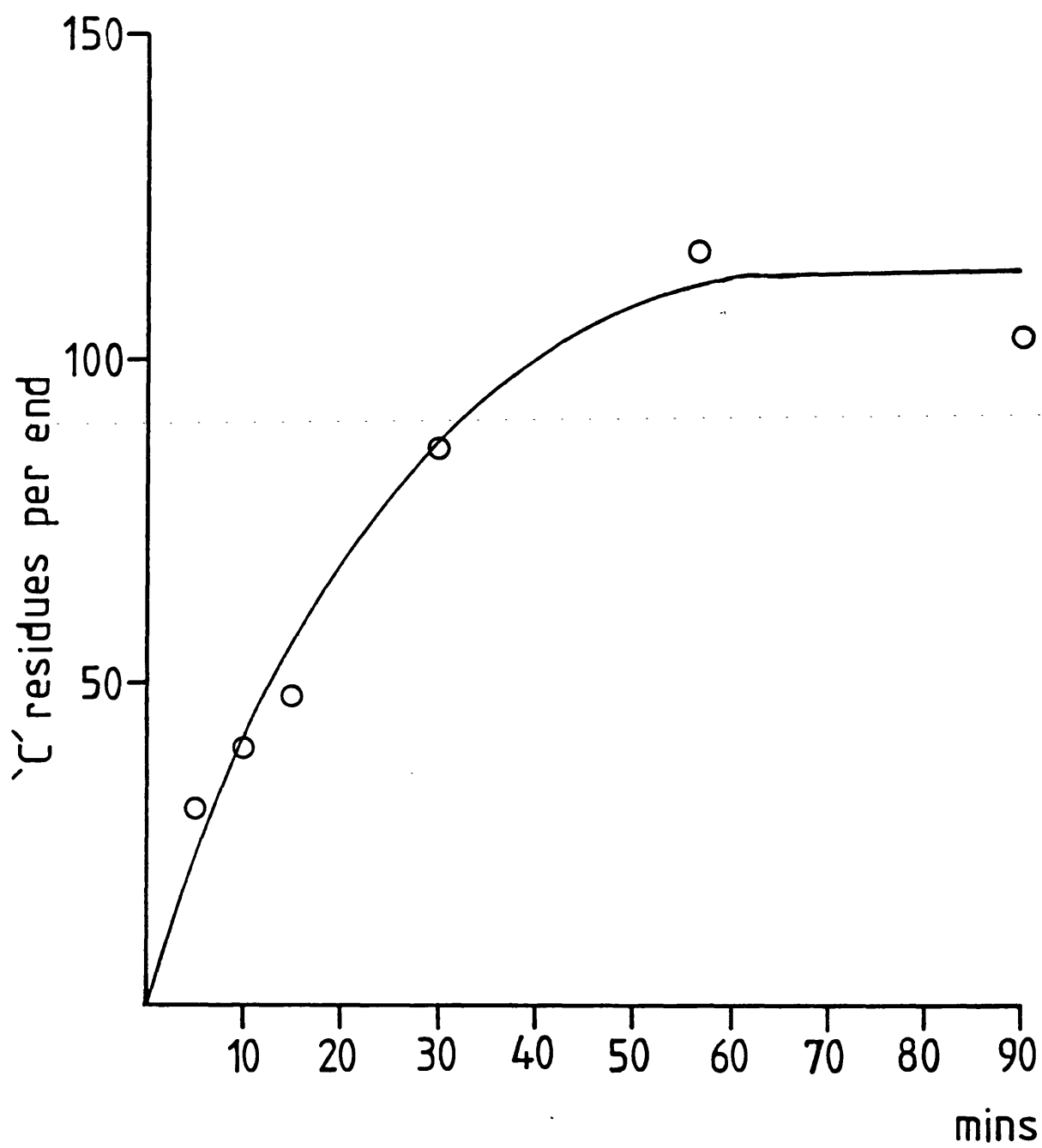


Figure 4:5

(ii) "G" tailing pAT153 DNA

pAT153 was linearised with the restriction endonuclease PstI and "G" tailed as described in Materials and Methods. Incorporation of "G" residues was monitored by the incorporation of ^3H -dGTP. An average of 10 "G" residues were added to each 3' terminus of plasmid DNA.

The tailed double stranded cDNA and tailed plasmid DNA were then annealed as described in Materials and Methods.

4. TRANSFORMATION OF *E.coli* HB101

pAT153 is a derivative of pBR322 (Twigg and Sherrat, 1980) which has lost a segment of DNA required for mobilisation of the plasmid but retaining the genes encoding ampicillin and tetracycline resistance. The PstI site of pAT153 lies in the ampicillin resistance gene between the sequences encoding amino acids 182 and 183 of the β -lactamase (Sutcliffe, 1978). All tailed cDNAs were introduced into pAT153 at this PstI site.

Calcium chloride treatment of *E.coli* cells was used to render them competent for DNA transformation (Mandel and Higa, 1976). In order to minimise the possibility of rearrangements arising during the bacterial amplification of the inserted DNA the recombination-deficient strain HB101 was used for DNA transformation (Boyer and Roulland-Dussoix, 1969).

In addition to the large scale transformation using the recombinant DNA, a series of small-scale control transformations were performed in parallel. Transformants were selected on tetracycline at 7.5 $\mu\text{g}/\text{ml}$. In order to estimate the efficiency of transformation, closed-circular pAT153 DNA was used to transform HB101. In addition a "mock" recombinant DNA sample was used as a control, this sample was "G" tailed pAT153 DNA which had gone through the annealing reaction in the absence of tailed cDNA. The relative transformation efficiencies of closed circular DNA, recombinant DNA and "mock" recombinant DNA are shown in Table (4:2).

Table 4:2 Relative transformation efficiencies of vector DNA, recombinant DNA and "mock"
recombinant DNA (-tailed cDNA)

<u>DNA</u>	<u>Amount/transformation (μg)</u>	<u>Number of transformants</u>	<u>Number of transformants/μg DNA</u>
pAT153	0.01	2×10^4	2×10^6
recombinant	0.1	2×10^2	2×10^3
"mock" recombinant	0.1	0	0

A total of approximately 200 transformants were obtained from transformations using recombinant DNA. 20 of these transformants were tested for resistance to 25 µg/ml ampicillin. Insertion of cDNAs into the PstI site of pAT153 might be expected to inactivate the β-lactamase gene and render all transformants sensitive to ampicillin. However only 7 of the 20 transformants tested were found to be ampicillin sensitive. This has been reported by Villa-Komaroff et al. (1978). The PstI site in the β-lactamase gene of pAT153 interrupts the gene between the sequences encoding amino acids 182 and 183 (Sutcliffe, 1978). Thus the active site of the β-lactamase must lie within the N terminal 182 amino acids of the protein.

5. SCREENING THE HUMAN cDNA CLONE BANK

The clone bank of approximately 200 clones was screened using the in situ hybridisation method devised by Grunstein and Hogness (1975). Bacterial colonies were replicated onto nitrocellulose filters, lysed in situ with alkali and hybridised to a radioactively labelled probe.

In order to identify clones specific to adult muscle mRNAs, duplicate filters were hybridised with ³²P labelled cDNA reverse transcribed from adult and foetal polyadenylated RNA. 19 week foetal polyadenylated RNA was used as no myoglobin protein is detectable in foetal muscle at this stage (Tipler et al., 1978). However, this does not discount the possibility that foetal muscle may contain repressed myoglobin mRNA. Adult and foetal probes were used in order to detect adult specific clones which might be expected to be strong candidates for clones containing a myoglobin cDNA sequence.

It is clear from Figure (4:6) that the colonies hybridised with different intensities. The intensity of each spot reflects the relative

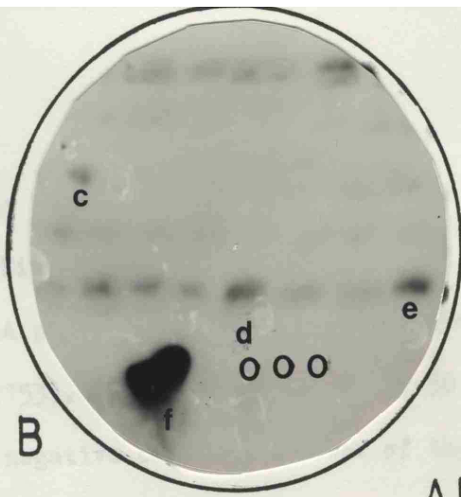
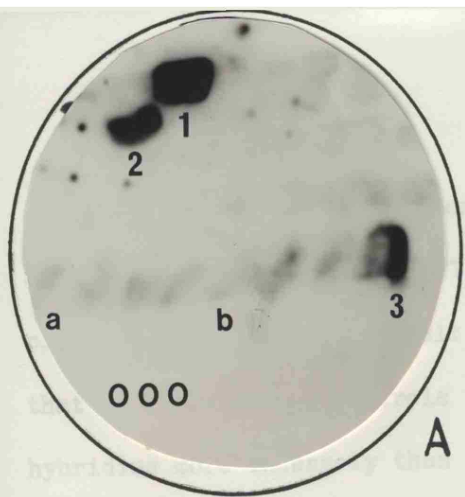
Figure 4:6 Colony hybridisation of human cDNA clone bank with adult and foetal probes

Approximately 200 cDNA clones were streaked onto 2 sets of four ungridded nitrocellulose filters: (A-A', B-B', C-C' and D-D'). One set (A, B, C, D) was hybridised with 32 P-labelled adult cDNA (ADULT), the other (A', B', C', D') with similarly labelled foetal cDNA (FOETAL). The probes were made by the reverse transcription of human skeletal muscle polyadenylated RNA from the relevant source.

Two sets of clones could be distinguished. One set (1-8) corresponds to eight strongly hybridising adult specific clones. The position of the equivalent clones in the foetal screen are designated 1'-8'. The second set corresponds to 13 adult clones (a-m) which are also detected by the foetal probe (a'-m'). The letters or numbers used to indicate these clones are placed directly beneath the relevant clone.

It should be noted that the positions of the clones on the two sets of filters are not perfectly aligned due to inconsistency in the streaking out procedure.

pAT153 controls are ringed.



ADULT

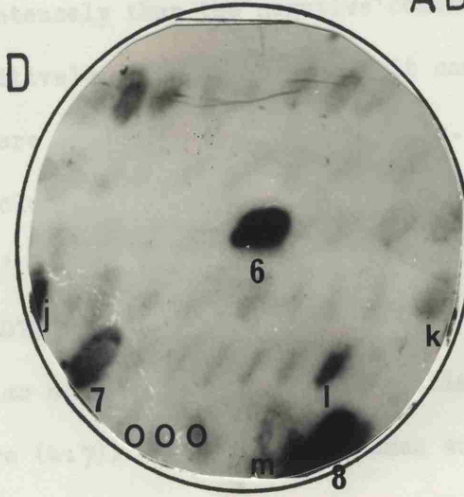
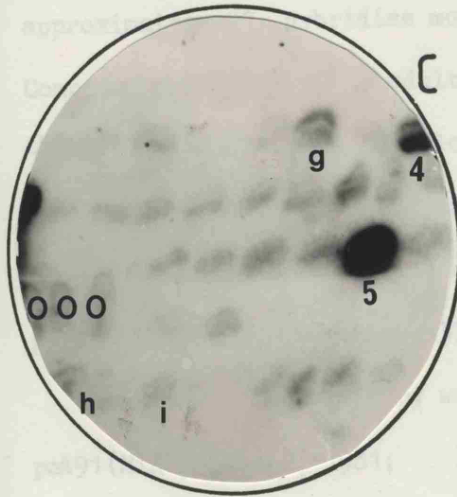
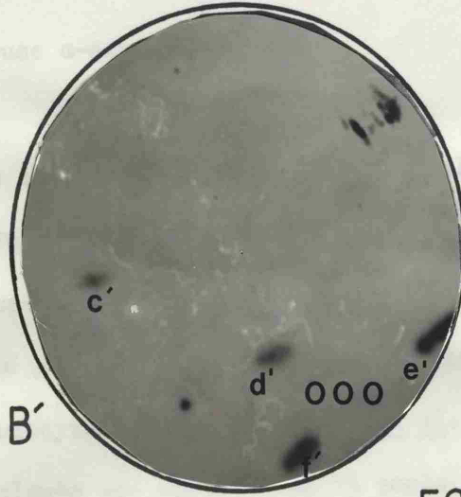
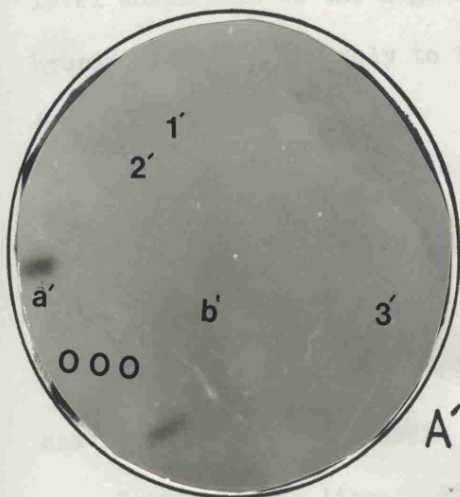
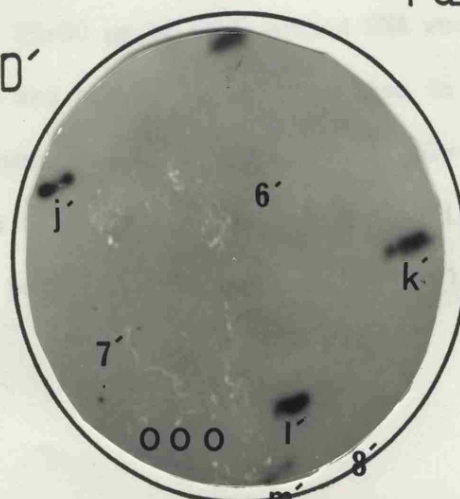
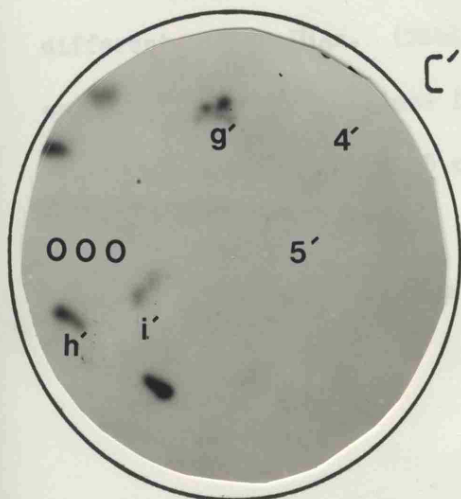


Figure 4:6



FÖETAL



abundance of the mRNA corresponding to the cloned sequence in the polyadenylated RNA (Williams, 1981). Figure (4:6) shows that most of the 200 clones screened with the adult cDNA probe hybridise at a level similar to that of the negative controls (pAT153). However approximately 50 clones hybridise more intensely than the negative controls. Fewer of the foetal clones approximately 20, hybridise more intensely than the negative controls. Comparing the foetal and adult positively hybridising clones it can be seen that most of the 'foetal' clones are detected by the adult probe. However 8 strongly hybridising adult clones are not detected by the foetal cDNA probe. These 'adult specific' clones have been designated pLG623, pLG648, pLG676, pLG714, pLG737, pLG745, pLG752 and pLG793.

The human cDNA library was also screened with a mouse α -actin probe pA91 (Minty *et al.*, 1981; Figure (4:7)) to detect any human actin cDNAs. It can be seen from Figure (4:8) that most clones hybridise at a level comparable to the negative controls. One clone however, pLG646, hybridised very intensely to the mouse α -actin probe.

6. POSITIVE SELECTION

Positive selection was used to identify plasmids containing mRNA sequences encoding translation products. The plasmid DNA was bound to nitrocellulose filters and hybridised to adult skeletal muscle polyadenylated RNA, unhybridised RNA was removed and the hybridised RNA recovered and translated in the rabbit reticulocyte system (Cleveland *et al.*, 1980).

Initially the library of 218 clones was screened as 15 pools of 14 different cloned DNAs. (Table 4:3) 25-30 μ g of each pooled DNA was spotted onto nitrocellulose filters and hybridised as described in Materials and Methods to 25 μ g of human adult skeletal muscle polyadenylated RNA. Figure 4:9 shows the in vitro translation profiles of positively selected RNA from these pools. Some pools e.g. 4 and 11 show

Figure 4:7 Partial restriction map of a mouse α -actin plasmid
(p α A91)

The map of p α A91 is adapted from Minty et al. (1981). Sites for BamH1 (B), EcoR1 (E), HindIII (H) and PstI (P) are shown. The cDNA insert of p α A91 contains around 90% of the coding sequence of α -actin (amino acids 30-374) and approximately 300 nucleotides of 3' noncoding sequence.

The 1.1 kb PstI fragment of p α A91 was isolated by preparative gel electrophoresis, labelled by nick translation and used as a probe for actin sequences in the human skeletal muscle cDNA library.

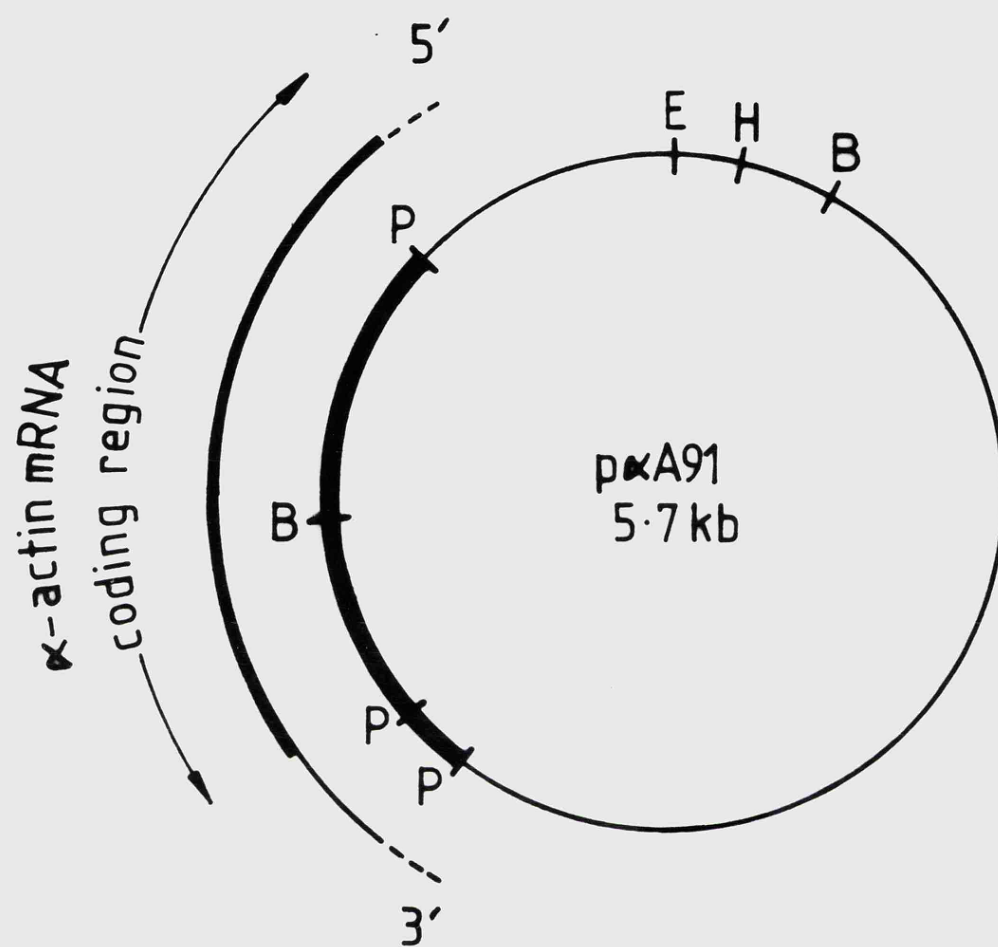
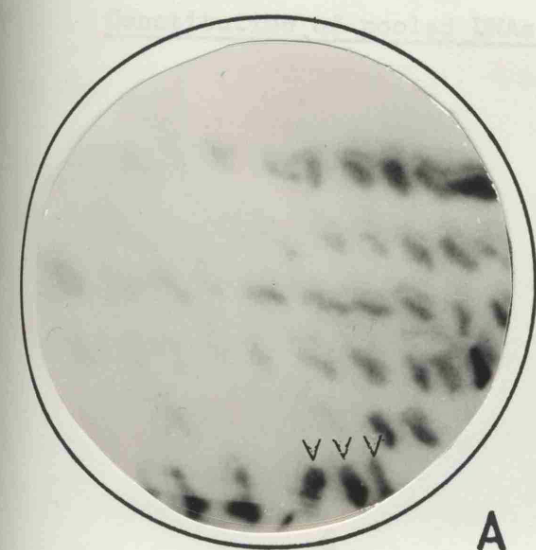


Figure 4:7

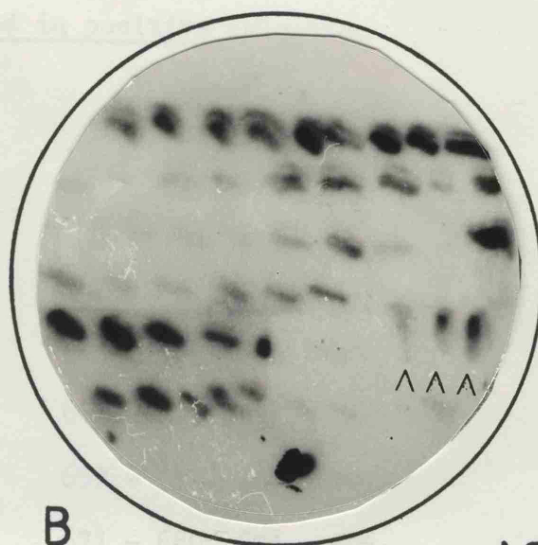
Figure 4:8 Colony hybridisation of human cDNA library with mouse
 α -actin probe

Approximately 200 cDNA clones spotted onto 4 nitrocellulose filters (A, B, C, D) were hybridised with a ^{32}P -labelled mouse α -actin probe derived from p α A91 (Figure 4:5). pAT153 controls are arrowed. X designates a strongly hybridising clone, pLG646, corresponding to a tentative human α -actin clone.

The high background level of hybridisation is probably due to contamination of the α -actin probe with plasmid DNA.

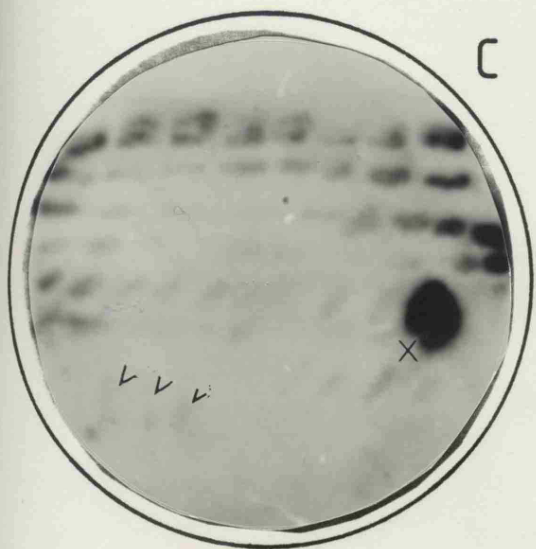


A

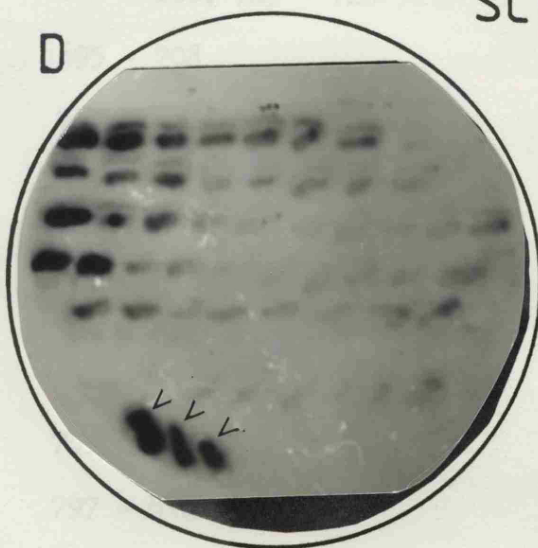


B

ACTIN
SCREEN



C



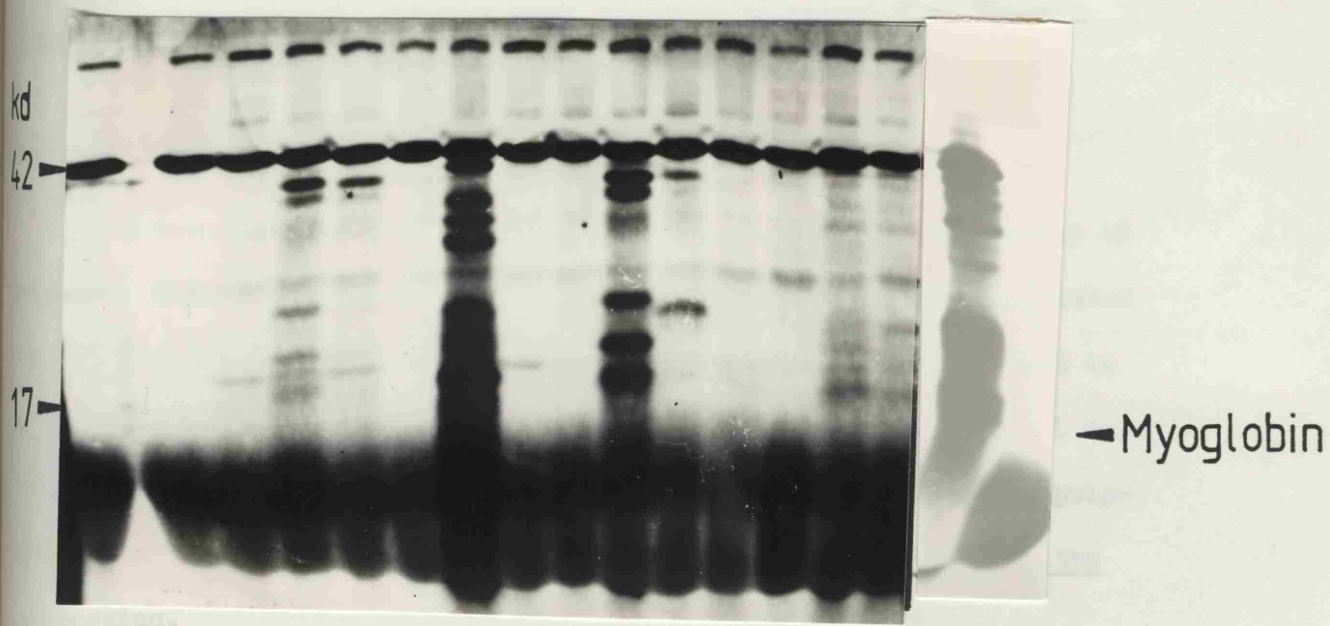
D

Figure 4:8

Table 4:3

Constitution of pooled DNAs used in positive selection

<u>Pool</u>	<u>pLG Number</u>
1	601 - 614
2	615 - 628
3	629 - 642
4	643 - 656
5	657 - 670
6	671 - 680, 691 - 694
7	681 - 690, 723 - 726
8	695 - 708
9	709 - 722
10	727 - 740
11	741 - 754
12	755 - 768
13	669 - 782
14	783 - 796
15	797 - 818



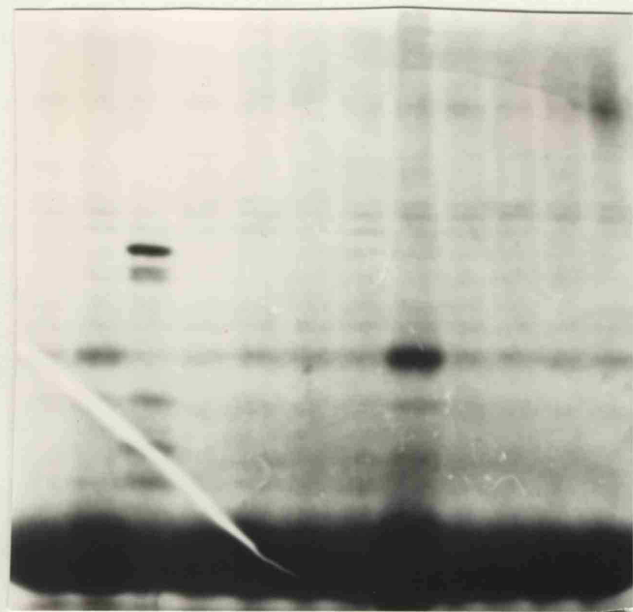
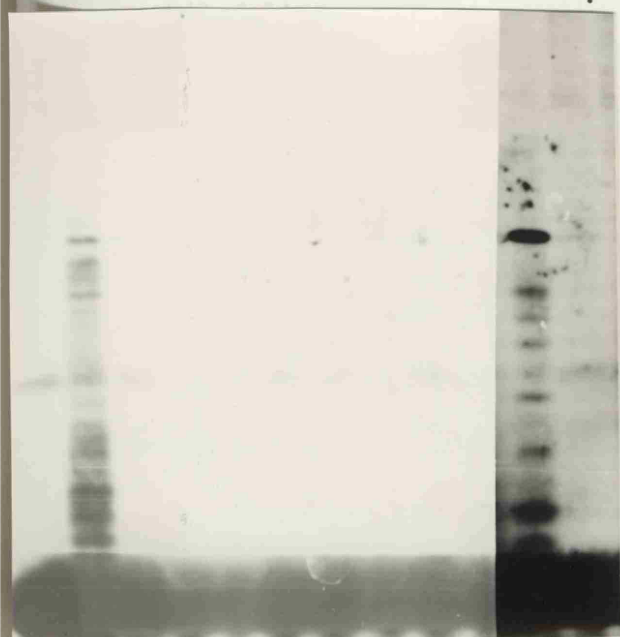
0 L 1 2 3 4 5 7 8 9 10 11 12 13 15 P G

pooled DNAs

Figure 4:9

pool 5

pool 13



G P 659 660 661 662 663 664 665 666 667 668
pLG

720 710 714 715 716 713 717
pLG
G 721 722 pβG1 O

Figure 4:9 Positive selection of mRNA by pooled and individual
recombinant cDNA clones

The human skeletal muscle cDNA library was pooled into 15 pools of around 14 clones and hybridised to human skeletal muscle polyadenylated RNA, the hybridised RNA was in vitro translated and electrophoresed in a 17% sds-polyacrylamide gel. O = blank filter only hybridised. L = rabbit reticulocyte lysate (zero RNA). P = human skeletal muscle polyadenylated RNA in vitro translation. G = rabbit globin mRNA in vitro translation.

The DNA of individual pools 5 and 13 were hybridised separately, along with a control p β G1 DNA containing rabbit β -globin cDNA which was hybridised to rabbit globin mRNA. Recombinant pLG667 positively selects all the mRNAs encoding the proteins selected by pool 5 and similarly pLG714 selects mRNAs selected by pool 13.

kd = kilodaltons.

no novel proteins compared to the zero DNA track. Other pools e.g. 5 and 9 show several proteins when compared to the zero DNA track. Some tracks share common proteins e.g. 1 and 7, indicating that these pools share at least one clone which selects the same mRNA from total polyadenylated RNA. Other pools e.g. 5 and 9 share a number of common proteins but also a number of different proteins.

Pool 5 and pool 13 seemed to have a protein which migrates in the region of the gel where myoglobin in the total polyadenylated RNA translation profile migrates. The individual DNAs from these two pools were therefore used in a positive selection experiment.

When this experiment was carried out it was found that all the proteins corresponding to pool 5 were selected by the single plasmid pLG667 and similarly the proteins corresponding to pool 13 were selected by the single plasmid pLG714. There are two possible explanations for this phenomenon. Firstly, pLG667 and pLG714 may contain a cDNA sequence which is common to a number of different muscle mRNAs which are selected and translated to give a number of proteins. Secondly, the cDNA contained in pLG667 and pLG714 may select only one mRNA which gives rise to a number of premature termination products or degradation products after in vitro translation of specific length which appear on sds-polyacrylamide gel electrophoresis as a number of differently sized bands.

pLG667 seemed to select an mRNA which could be translated in vitro to give a protein which comigrates with purified myoglobin in sds-polyacrylamide gels, in addition to a number of other proteins. In order to determine whether this cDNA contained any myoglobin DNA sequence the cDNA insert of pLG667 was sequenced as described by Maxam and Gilbert (1980).

7. MAPPING AND SEQUENCING pLG667 cDNA INSERT

pLG667 (Figure (4:10)) contains a cDNA insert at the PstI site of pAT153 of approximately 190 base pairs. As a result of the homopolymer tailing PstI sites have been regenerated at each end of the cDNA sequence. pLG667 contains one Taq Y1 site approximately 75 bp from the right hand PstI site and a HhaI site approximately 145 bp from this site.

pLG667 was cleaved with restriction endonuclease Taq Y1. The DNA fragments were treated with alkaline phosphatase and labelled with γ -³²P-ATP using polynucleotide kinase. The DNA was recovered and cleaved with the restriction endonuclease PstI. The DNA was electrophoresed in a 3% agarose gel and the two smallest fragments of approximately 115 bp and 75 bp were excised from the gel and the DNA recovered.

DNA sequencing chemistry was performed on these two fragments and the samples electrophoresed as described by Maxam and Gilbert (1980).

The preliminary sequence of the cDNA insert in pLG667 is shown in Figure (4:11).

It can be seen from this sequence that the homopolymer tails are of 30 and 16 base pairs. No sequence corresponding to a poly(A) tail or AATAAA box (polyadenylation signal) can be detected in this sequence, suggesting that this cDNA corresponds to mRNA upstream of the 3' end of the message.

Comparing the cDNA sequence to the predicted DNA sequence derived from the human myoglobin amino acid sequence (Dayhoff et al., 1972), no similarities could be detected. This suggests that the cDNA insert in pLG667 is not derived from the message corresponding to human myoglobin.

Figure 4:10 Sequencing strategy for the cDNA insert of pLG667

The cDNA insert of pLG667 was partially sequenced by the Maxam-Gilbert technique using the TaqY1 site (T) within the cDNA insert. The TaqY1 site was kinase-labelled and then cleaved with Pst1 (P). The 75 bp and 115 bp fragments were recovered from an agarose gel and sequencing chemistry performed.

pLG 667

100 bp

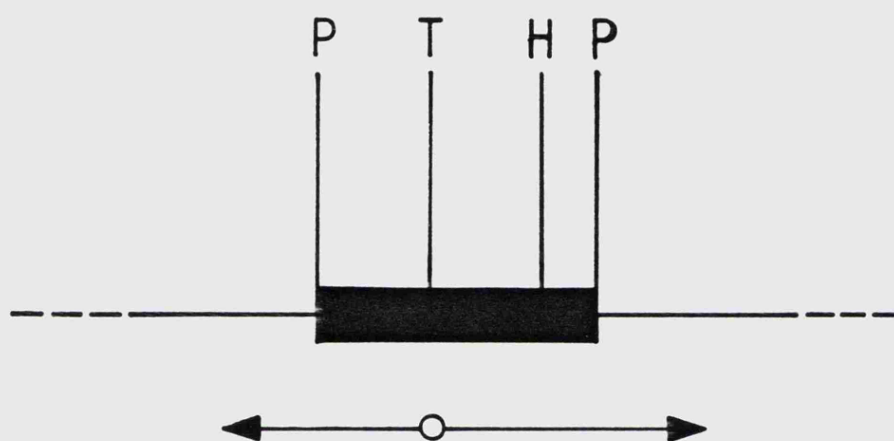


Figure 4:10

Figure 4:11 Preliminary sequence of the cDNA insert of pLG667

The cDNA insert contains 178 base pairs. The 5' terminus has a G.C tail of 30 base pairs and the 3' terminus has a G.C tail of 16 base pairs. The remaining sequence (132 bp) contains no polyadenylation signal or sequence encoding the human myoglobin amino acid sequence. The sequence around the TaqY1 site has not been determined and is designated by dashes (-). Ambiguous sequence is shown as N. Dots above the sequence occur every tenth base.

gggggggggggggggggggggggggggggggggCCGCCGACTCCATCTACAACAGCNNTCNNT
GTGT----TCGA-----AGGAAGGACCTNGNTATGCCAACACGTCGTGCTGGGTGCCAG
CGCGGTGATCTCTTTCTGCATGCGGTGAGCGGTCCTAGGGTTcccccccccccccccc

Figure 4:II

Chapter 5

ISOLATION AND CHARACTERISATION OF POLYADENYLATED RNA FROM GREY SEAL

SKELETAL MUSCLE

The red skeletal muscle of diving mammals such as whales and seals contains between 5 and 10 times the concentration of myoglobin protein than does the equivalent muscle in other mammals (Robinson, 1939; Wittenberg, 1970). One possible consequence of this phenomenon is that the skeletal muscle of diving mammals may contain a higher concentration of myoglobin mRNA than does human skeletal muscle. For this reason the skeletal muscle of the grey seal (Halichoerus grypus) was used as a source of polyadenylated RNA. The muscle tissue was obtained from Dr John Prime of the British Antarctic Survey, Cambridge. The muscle provided was skeletal red muscle although the exact muscle type, sex and age were unknown.

1. PURIFICATION OF GREY SEAL MYOGLOBIN

Homogenates of grey seal skeletal muscle were loaded onto an ampholine polyacrylamide iso-electric focussing gel and focussed. The major non-haemoglobin red pigmented protein band (myoglobin) was excised and the native myoglobin recovered from the gel slice (Figure 5:1).

2. ISOLATION OF POLYADENYLATED RNA

RNA was prepared from grey seal skeletal muscle tissue as described for human skeletal muscle. Samples were made brittle by chilling in liquid nitrogen, shattered, homogenised and lysed with sodium dodecyl sulphate. All procedures were carried out using solutions made up with water treated with 0.1% diethylpyrocarbonate to inhibit RNase activity.

Figure 5:1 Iso-electric focussing of grey seal myoglobin

Homogenates of grey seal skeletal muscle were loaded onto an LKB ampholine polyacrylamide I.E.F. gel (pH 5.5 - 8.5). Two major red pigmented bands could be seen, the myoglobin band which migrated at a pK of 7.4 was excised and the protein recovered. The other major band had a higher pK (about 7.8) and probably corresponds to myoglobin in a different oxidation state. Other faint red pigmented bands (pK 6.7 - 7.1) probably correspond to grey seal haemoglobins.

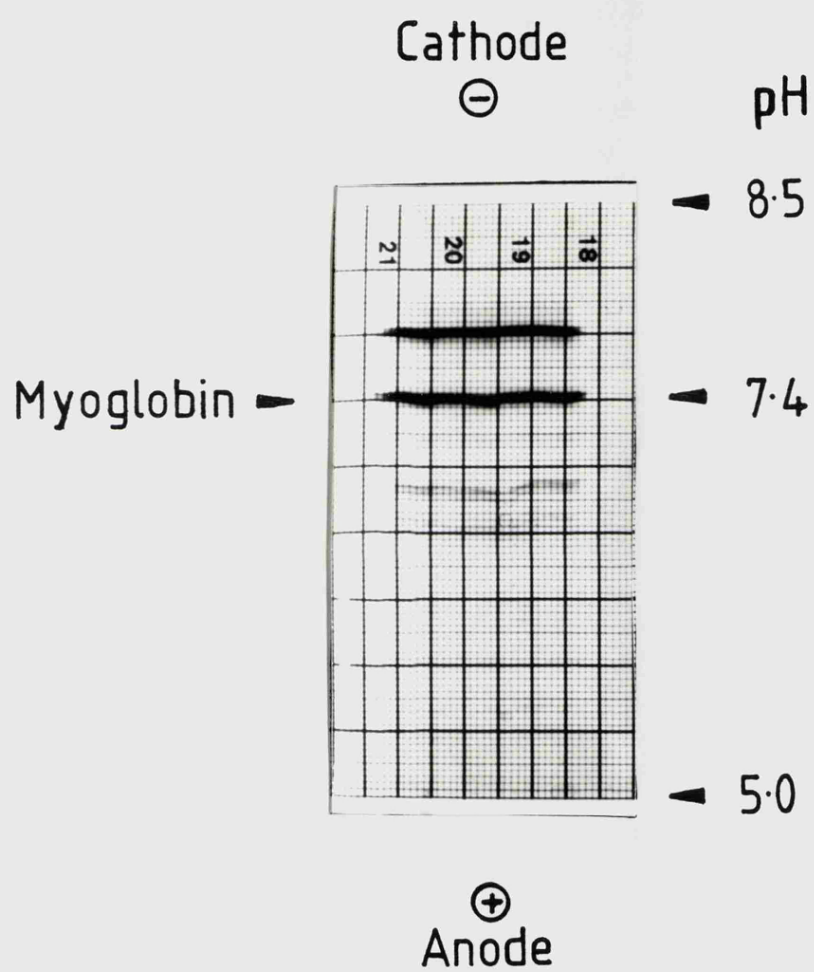


Figure 5:1

Typical yields of nucleic acid, RNA and polyadenylated RNA from grey seal skeletal muscle are shown in Table (5:1).

RNA prepared from grey seal skeletal muscle was generally undegraded and bands corresponding to 28S or 18S ribosomal RNA bands could be seen on agarose gel electrophoresis (Figure 5:2). The good quality of RNA prepared from this tissue can be accounted for by the fact that the samples were frozen immediately after the animals were killed. The subsequent procedures used to isolate RNA from this tissue have been developed in order to produce minimum degradation of the RNA.

3. IN VITRO TRANSLATION OF GREY SEAL POLYADENYLATED RNA

Polyadenylated RNA from adult human and grey seal skeletal muscle was translated in vitro in the rabbit reticulocyte system in the presence of ^{35}S methionine and the products electrophoresed on SDS-polyacrylamide gels (Figure 5:3a). It is clear from these that the human and seal profiles are very similar. In the seal polyadenylated RNA translation profile there is a protein of approximately 17 kd which comigrates with purified grey seal myoglobin. This protein is present in greater quantities in the translation profile of seal polyadenylated RNA than is the equivalent protein in translation profiles of human polyadenylated RNA. It seems therefore that grey seal skeletal muscle may be significantly enriched for myoglobin mRNA compared to human skeletal muscle. Alternatively the seal myoglobin mRNA may be translated more efficiently in the rabbit reticulocyte system.

Grey seal total RNA, polyadenylated RNA and non-polyadenylated RNA were translated in the rabbit reticulocyte system and the products

Table 5:1

Typical yields of nucleic acid prepared from grey seal skeletal muscle

<u>Nucleic acid fraction</u>	<u>Yield/^{mg}kg tissue</u>	<u>% of total nucleic acid</u>
Total nucleic acid	200	100
Total RNA (after CsCl centrifugation)	120	60
poly(A) ⁺ mRNA (after oligo (dT) cellulose chromatography)	0.7	0.35
poly(A) ⁻ RNA (after oligo (dT) cellulose chromatography)	~ 75	~ 37.5

Figure 5:2 Preparation of nucleic acid, RNA and polyadenylated mRNA
from grey seal skeletal muscle

1 µg of total nucleic acid was electrophoresed in a 2% agarose gel (lanes 1 and 4), DNA (D), 28S and 18S ribosomal RNA (28 and 18) and tRNA (t) bands could be clearly seen. After centrifugation in CsCl no DNA or tRNA bands were present (lane 2 - 1 µg). Lane 3 shows 10 µg of total nucleic acid with a wide range of molecular weights. Lane 5 shows 1 µg of seal polyadenylated mRNA with only faint 28S and 18S bands. Lane 5 shows 10 µg of total nucleic acid after alkali treatment.

O = origin of electrophoresis. nt = nucleotides.

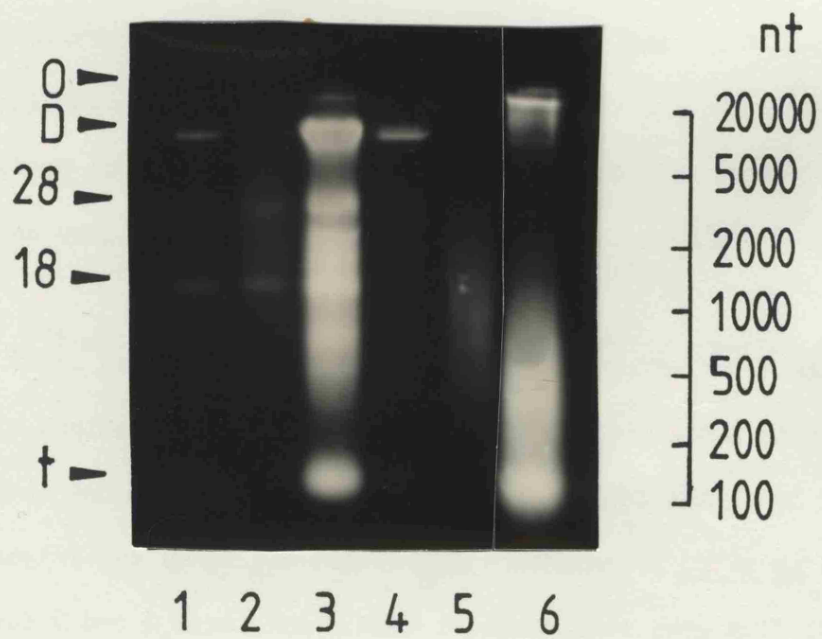


Figure 5:2

Figure 5:3 In vitro translation profiles of grey seal polyadenylated RNA

- (a) 0.1 μ g of rabbit globin mRNA (lane 2) human skeletal muscle polyadenylated RNA (lane 3) and grey seal skeletal muscle polyadenylated RNA (lane 4) were translated in the rabbit reticulocyte lysate system and electrophoresed in 17% sds-polyacrylamide gels. The myoglobin band migrated at 17 kd and the grey seal profile is greatly enhanced for this protein. Lane 1 is a zero RNA control.
- (b) Lane 1 corresponds to in vitro translation of 0.1 μ g of rabbit globin mRNA. Lanes 2 and 3 are in vitro translations of 3 and 1 μ g of total RNA from grey seal skeletal muscle. Lanes 4 and 5 correspond to in vitro translations of 0.3 and 0.1 μ g of non-polyadenylated RNA from grey seal skeletal muscle and lanes 6 and 7 correspond to 0.3 μ g and 0.1 μ g of polyadenylated RNA from grey seal skeletal muscle. The total RNA clearly has a low proportion of active mRNA. The polyadenylated RNA fraction is greatly enriched for mRNA species, including myoglobin mRNA, compared to the non-polyadenylated RNA fraction.

analysed on sds-polyacrylamide gels (Figure 5:3b). From these gels it is clear that the non-polyadenylated RNA fraction is depleted in myoglobin mRNA. This suggests that the majority of myoglobin mRNA molecules are polyadenylated.

4. IDENTIFICATION OF MYOGLOBIN IN IN VITRO TRANSLATIONS BY PURIFICATION FROM ISO-ELECTRIC FOCUSSING POLYACRYLAMIDE GELS

In order to confirm further that the 17 kd protein from in vitro translations of grey seal skeletal muscle RNA was myoglobin, its migration in iso-electric focussing gels was studied. In vitro translation products of grey seal skeletal muscle polyadenylated RNA were mixed with grey seal myoglobin. This mixture was focussed on an ampholine polyacrylamide iso-electric focussing gel (LKB Ampholine PAG Plate pH5.5-8.5). The band corresponding to grey seal myoglobin was excised and the protein recovered from the gel slice. This protein was analysed on a 17% sds-polyacrylamide gel. It was determined from this analysis that the purified myoglobin comigrated on both iso-electric focussing gels and sds-polyacrylamide gels with the radioactively labelled 17 kd in in vitro translations of grey seal skeletal muscle polyadenylated RNA (Figure 5:4). The fact that some of the 17 kd translation product copurifies with seal myoglobin on iso-electric focussing gels suggests that at least some of this protein is myoglobin. This 17 kd protein copurifies with native myoglobin on iso-electric focussing gels and must therefore be the native myoglobin associated with the haem group. The rabbit reticulocyte lysate is therefore a useful system for in vitro translation as it will translate messages from diverse sources and in the case of myoglobin, provides the haem groups from the lysate to produce the native myoglobin holoprotein.

Figure 5:4 Identification of myoglobin mRNA in seal muscle poly(A)⁺
RNA by iso-electric focussing

RNAs from grey seal muscle were translated in a rabbit reticulocyte lysate system in the presence of L-³⁵S methionine and the products analysed by sds-PAGE in 17% gels. '-' = no RNA, 'RNA' = 1 µg total muscle RNA, poly(A)⁺ RNA = 0.1 µg muscle polyadenylated RNA. The lysate from the translation of polyadenylated RNA was mixed with pure seal myoglobin and analysed by iso-electric focussing. Mb + = sds PAGE of IEF purified myoglobin mixed with lysate from poly(A)⁺ RNA translations showing co-purification of labelled myoglobin with unlabelled pure seal myoglobin (Mb). Mb - = sds PAGE of IEF purified myoglobin mixed with lysate from zero RNA translation showing no co-purification of any labelled product.

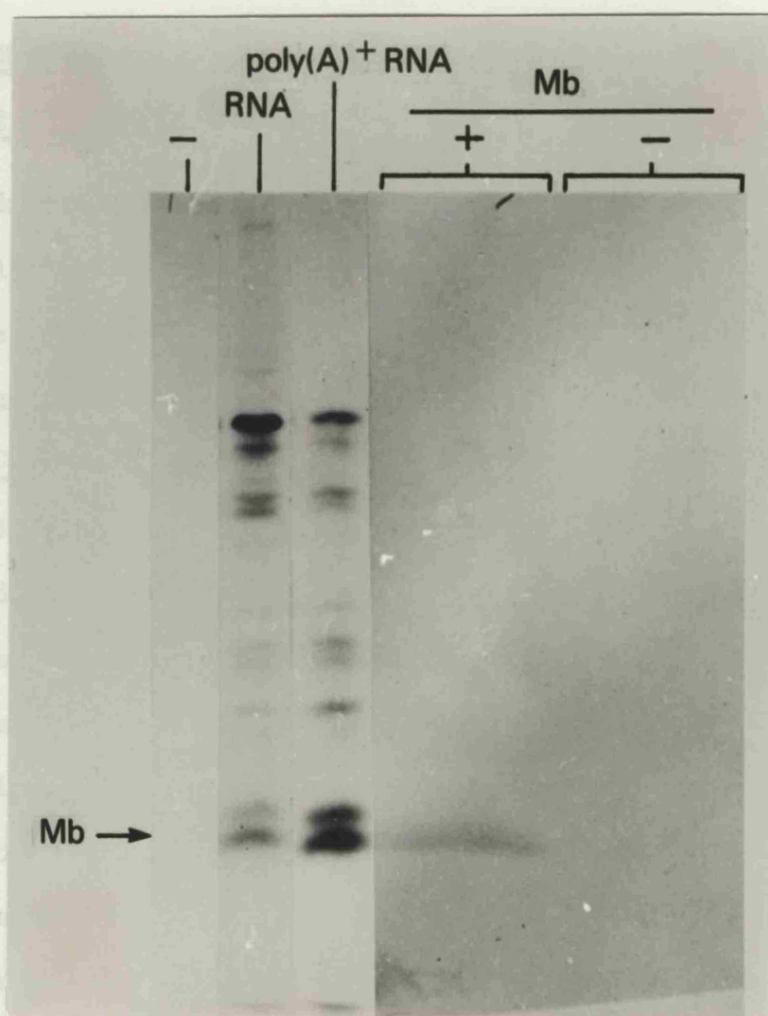


Figure 5:4

Chapter 6

THE CONSTRUCTION AND ANALYSIS OF A cDNA LIBRARY FROM GREY SEAL SKELETAL MUSCLE POLYADENYLATED RNA

Grey seal skeletal muscle polyadenylated RNA was used to construct a cDNA library in E.coli HB101. An identical cloning strategy was used in the construction of this library as was used for the construction of the human library as described in Chapter 4 (Figure 4:1).

1. CONSTRUCTION OF A GREY SEAL cDNA LIBRARY

Single-stranded cDNA was made by incubating 50 µg of seal muscle polyadenylated RNA in the presence of an oligo (dT)₁₀ primer and AMV reverse transcriptase. Approximately 2.5 µg of ³²P-labelled cDNA were recovered. Double-stranded cDNA was prepared by incubating 2 µg ³²P-labelled single-stranded cDNA with reverse transcriptase. Double-stranded cDNA was recovered and flush ended with S1-nuclease. 1 µg of S1-nuclease treated double-stranded cDNA was recovered. Agarose gel electrophoresis and autoradiography of the single-stranded cDNA showed a range of relatively short cDNA molecules from around 100 to 1000 nucleotides in length. This may have been the result of using partially degraded polyadenylated RNA from the original muscle sample as well as a relatively inefficient 'first strand' cDNA synthesis, (Figure 6:1). Oligo (dC) homopolymeric tails were added to the 3' termini of the double-stranded cDNAs using α-³²P-dCTP and terminal deoxynucleotidyl transferase. The average length of these 'C' tails was around 40 nucleotides. This 'tailed' double-stranded cDNA was mixed with an equimolar amount of PstI-cleaved, oligo (dG) tailed pAT153 DNA (average length of tails 8 nucleotides) in an annealing reaction. The annealed DNA was transformed into E.coli HB101 and recombinants selected in the presence of tetracycline.

Figure 6:1 Agarose gel electrophoresis of single and double stranded
cDNA from grey seal mRNA

cDNA/mRNA hybrids (lane 2) and double stranded cDNA (lanes 3 and 4) were electrophoresed against a ^{32}P labelled Sau3A digest of pBR322 DNA (lane 1) in a 2% agarose gel. The size range of single stranded cDNA was between 0.5 and 3.0 kilobases with an average size of around 0.5 kilobases. The size range of the double stranded cDNA was between 0.1 and 1.0 kilobases, with an average size of around 0.5 kilobases.

approximately 1000 clones were isolated. As transformants were obtained from 'cock' transformants using 'G' tailed pMT133 DNA which had been annealed in the absence of 'G' tailed DNA. This indicated that all 1000 transformants were likely to be recombinants.

The recombinant DNA was digested with *Eco*RI and separated on agarose gels. The DNA prepared by recombinant DNA was digested with *Eco*RI (Figure 6:2). Initial intensity, reflectance sequences corresponded to clones correspond to clones correspond to DNA prepared from genomic DNA.



1 2 3 4

In vitro translation of polyadenylated RNA shows 2-3 prominent labelled products in the molecular weight range 17-20 kd (Figure 5:1). The smallest of these co-electrophoresed previously with gray seal myoglobin (17 kd), when pure myoglobin was added to the translation products prior to electrophoresis and subsequent staining. The polyadenylated RNA was separated on a polyacrylamide gel and stained with ethidium bromide. The autoradiograph of the labelled RNA showed no degradation. By adding pure native myoglobin to an *in vitro* translation of polyadenylated RNA followed by separating the reaction mixture on an isoelectric focusing gel, and electrophoresing the purified myoglobin band on a polyacrylamide gel it was possible to show that only the 17 kd labelled component co-purified with seal myoglobin.

Figure 6:1

These results indicated that myoglobin cDNA should be moderately abundant in seal skeletal muscle. That at least some myoglobin cDNAs are

Approximately 1600 clones were isolated. No transformants were obtained from 'mock' transformations using 'G' tailed pAT153 DNA which had been annealed in the absence of 'C' tailed cDNA. This indicated that all 1600 transformants were likely to be recombinants.

The recombinant cDNA clones were replica-plated onto nitrocellulose filters and screened by hybridisation with ^{32}P -labelled single-stranded cDNA prepared by reverse transcription of grey seal polyadenylated RNA (Figure 6:2). Individual colonies gave a range of hybridisation intensities, reflecting the relative abundance of polyadenylated RNA sequences corresponding to each cDNA clone. Strongly hybridising cDNA clones correspond to relatively abundant mRNAs, while faintly hybridising clones correspond to mRNAs present at low abundance in the polyadenylated RNA prepared from grey seal skeletal muscle.

2. IDENTIFICATION OF MYOGLOBIN cDNA CLONES

In vitro translation of seal muscle polyadenylated RNA show 2-3 prominent labelled products in the molecular weight range 17-20 kd (Figure 5: 3). The smallest of these co-electrophoresed precisely with grey seal myoglobin (17 kd), when pure myoglobin was added to the translation products prior to electrophoresis and comparing the stained polyacrylamide gel with the autoradiograph of the labelled in vitro translation products. By adding pure native myoglobin to an in vitro translation of polyadenylated RNA followed by focussing the reticulocyte lysate on an isoelectric focussing gel, and electrophoresing the purified myoglobin band on a polyacrylamide gel it was possible to show that only the 17 kd labelled component co-purified with seal myoglobin.

These results indicated that myoglobin mRNA should be moderately abundant in seal skeletal muscle, that at least some myoglobin mRNAs are

Figure 6:2 Hybridisation analysis of seal muscle cDNA library with
 ³²P-labelled cDNA

Recombinant cDNA clones were replica plated onto nitrocellulose filters and screened by hybridisation with ³²P-labelled single stranded cDNA prepared from seal polyadenylated RNA and filters autoradiographed. Two such filters are shown, individual colonies giving a range of hybridisation intensities. Strongly hybridising clones correspond to relatively abundant mRNAs while faintly hybridising clones correspond to low abundance mRNAs.

O denotes control colonies containing vector pAT153 DNA.

polyadenylated and that an in vitro translation native synthesis system with base is produced.

Figure 6:2

Around 150 independently hybridizing cDNA clones were selected from the pool of cDNA clones (Figure 6:2) and recombinant

plasmid DNA. The clones were pooled into 9 pools of 16 or 17 clones each. The clones were then transfected into cells and the cells were grown in the presence of ³⁵S-methionine. The cells were then lysed and the RNA was extracted. The RNA was then hybridized with the cDNA probes. The results of the hybridization are shown in Figure 6:2.

In a recombinant cDNA library, the cDNA clones are inserted into a plasmid vector. The recombinant plasmids are then transfected into cells. The cells are grown in the presence of ³⁵S-methionine. The cells are then lysed and the RNA is extracted. The RNA is then hybridized with the cDNA probes. The results of the hybridization are shown in Figure 6:2.

The DNA of this pool were therefore sequenced individually and the recombinant cDNA clones were identified. The results of the sequencing are shown in Figure 6:2.

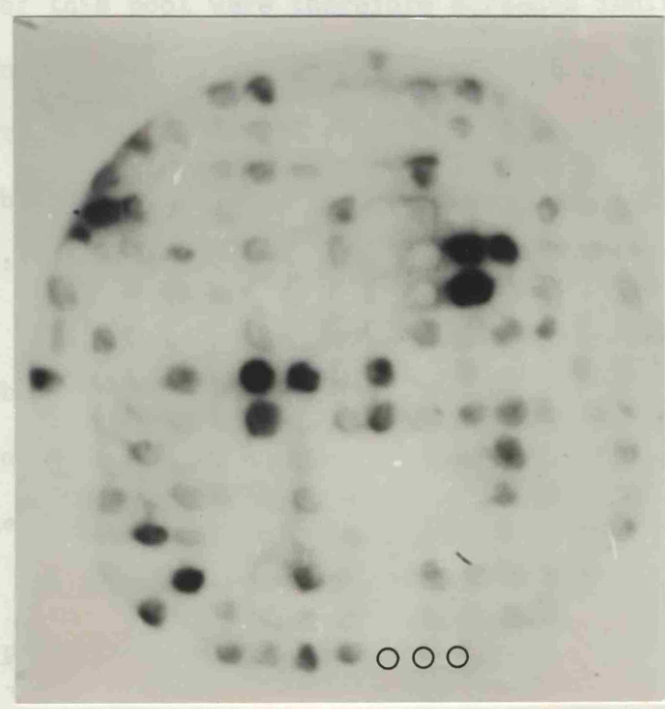
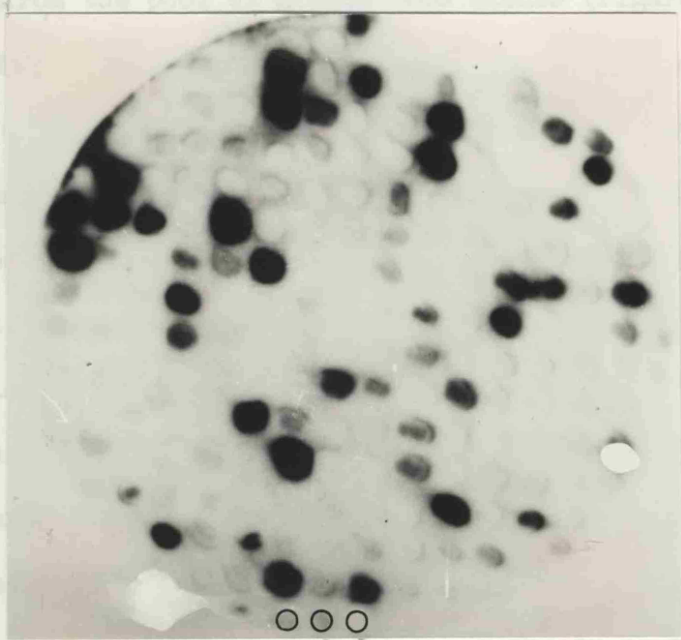
appeared to be a single cDNA clone. The cDNA clone was then sequenced and the results of the sequencing are shown in Figure 6:2.

spontaneous cDNA inserts (Figure 6:2). The cDNA clones were then sequenced and the results of the sequencing are shown in Figure 6:2.

pGH-75 by the presence of a specific cDNA clone. The cDNA clone was then sequenced and the results of the sequencing are shown in Figure 6:2.

of myoglobin. In each case at least 50% of the cDNA clones were found to be identical to the cDNA clone. In contrast, the cDNA clones were found to be identical to the cDNA clone.

labelled pGH-75. The cDNA clones were then sequenced and the results of the sequencing are shown in Figure 6:2.



polyadenylated and that on in vitro translation native myoglobin complexed with haem is produced.

Around 150 moderately to intensely hybridising cDNA clones were selected from the pool of around 1600 clones (Figure 6:2) and recombinant plasmid DNAs prepared for each clone. These were pooled into 9 pools of 16 or 17 different cloned DNAs (Table 6:1), bound to nitrocellulose filters and hybridised with seal muscle polyadenylated RNA. RNA which had hybridised with the cDNA clones was recovered and translated in vitro in the rabbit reticulocyte system. Labelled products were analysed by sds-polyacrylamide electrophoresis and autoradiography (Figure 6:3).

In addition a random selection of approximately 50 independent recombinant DNAs were screened individually by this positive selection procedure. Pool 8 appeared to contain a DNA or DNAs which selected RNA which in turn directed the synthesis of the 17 kd labelled apomyoglobin. The DNAs of this pool were therefore screened individually and the recombinant plasmids pSM168, pSM175, pSM178, pSM179, pSM180 and pSM182 appeared to select RNA which directed the synthesis of the 17 kd labelled apomyoglobin identifying these clones as containing myoglobin cDNA inserts (Figure 6:4). This identification was confirmed for pSM144, pSM174 and pSM178 by filter selection of RNA, in vitro translation and purification of myoglobin from the reticulocyte lysate by isoelectric focussing in the presence of pure unlabelled grey seal myoglobin. In each case at least 50% of the 17 kd labelled polypeptide co-purified with seal myoglobin. In contrast, RNA selected by non-myoglobin cDNA clones translated to give no labelled polypeptides which co-purified with myoglobin (Figure 6:5).

Table 6:1

Constitution of pooled DNAs used in positive selection

<u>Pool</u>	<u>pSM Number</u>
1	1-16
2	17-33
3	33-50
4	101-116
5	117-133
6	134-150
7	151-167
8	168-184
9	185-200

Figure 6:3 Screening pools of recombinant plasmid DNA by hybrid
release translation

16 different recombinant plasmid DNAs were pooled, bound to nitro-cellulose, hybridised to seal muscle polyadenylated RNA, the hybridised RNA translated in a reticulocyte system and the protein products analysed by sds-PAGE. Translation profiles for 3 pools are shown, lanes 1, 2 and 3. Lane 0 is a zero RNA control and lane pA shows the translation profile of grey seal skeletal muscle polyadenylated RNA. The 17 kd myoglobin (Mb) band is indicated and can be seen in polyadenylated RNA and pool 2 but not in pools 1 or 3.

Figure 6:3

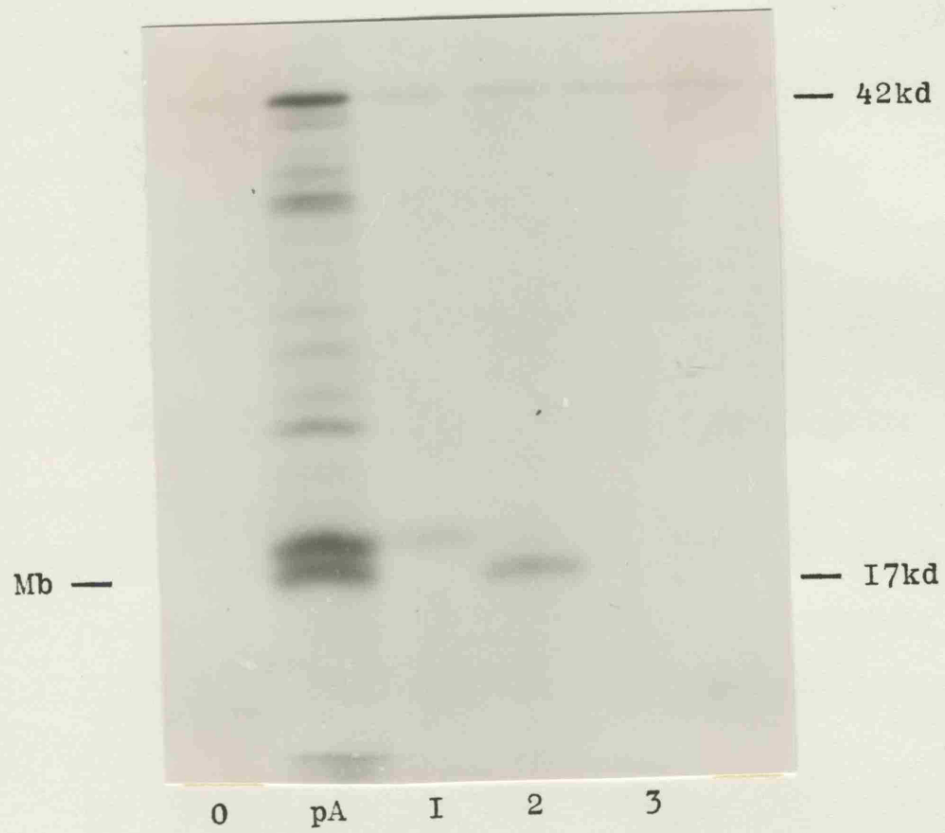


Figure 6:4 Identification of myoglobin cDNA clones by hybrid release translation

17 individual seal muscle cDNAs from pool 8 (Table 6:1) were screened by hybrid release translation. pA = translation profile of grey seal skeletal muscle polyadenylated RNA sharing 17 kd myoglobin band (Mb). Lane 1 is a zero RNA control in vitro translation. Lanes 2-18 show the in vitro translation profiles of cDNA plasmids pSM168-184 respectively. Lanes 2, 8, 9, 12, 13 and 14 corresponding to cDNA clones pSM168, 174, 175, 178, 179 and 180 show the 17 kd myoglobin band indicating that these clones contain myoglobin cDNA inserts which select myoglobin mRNA on hybridisation. Other clones pSM173 and 181 (lanes 7 and 15) appear to select mRNA which codes for a 42 kd polypeptide. These lanes also show other less intense bands which may correspond to breakdown products of the larger protein.

Figure 6:4

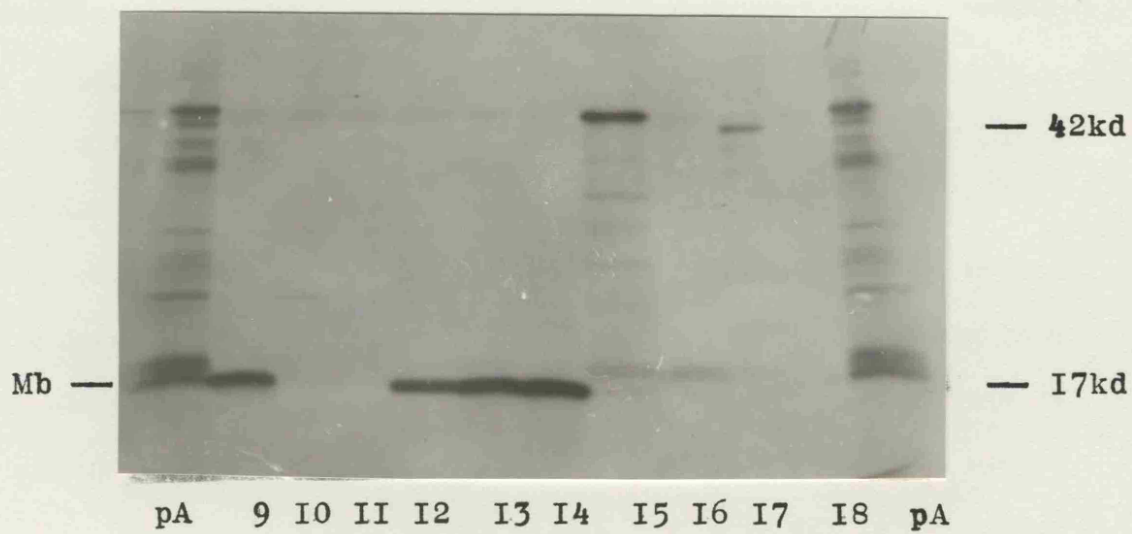
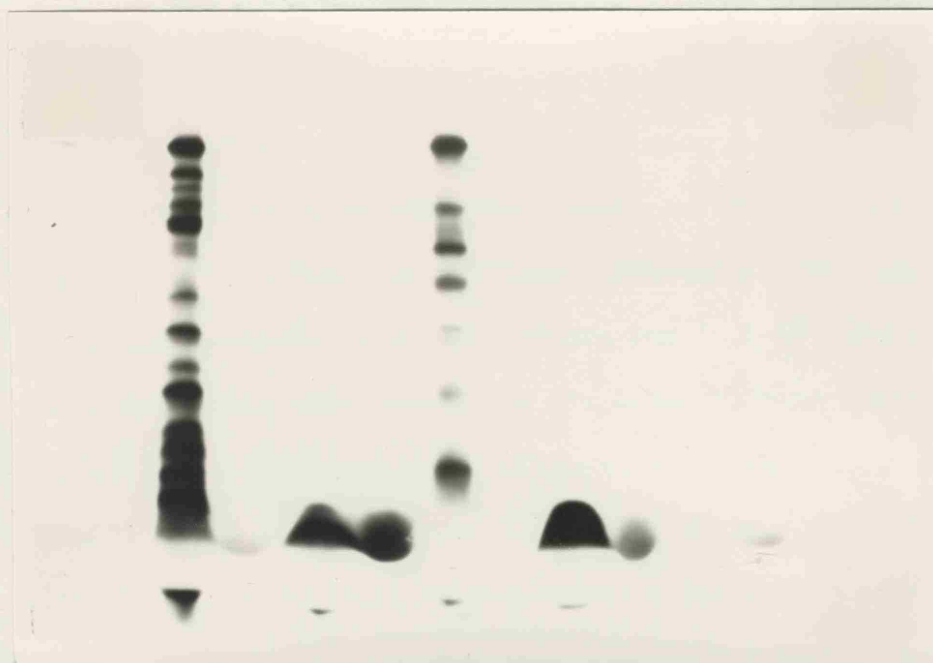


Figure 6:5 Identification of myoglobin cDNA clones by hybrid release translation and iso-electric focussing of labelled myoglobin

Myoglobin containing cDNA clones were bound to nitrocellulose, hybridised to seal muscle RNA and the hybridised RNA translated in a reticulocyte lysate. Some of this lysate was analysed by sds-PAGE (lanes marked 'a') the remainder of the lysate was focussed on iso-electric focussing gels with unlabelled myoglobin and the myoglobin band cut out of the gel. The protein from these gel slices were analysed by sds-PAGE (lanes marked 'b'). 0 = no DNA bound to filter. pA = grey seal polyadenylated RNA in vitro translation (a) and electrophoresis of translation products which co-purify with unlabelled myoglobin (b). 1, 3 and 4 = pSM144, 174 and 178 myoglobin cDNA clones, respectively, bound to nitrocellulose filters. 2 and 5 = pSM173 and pSM170 non-myoglobin cDNA clones respectively, bound to nitrocellulose filters. Mb indicates the migration of unlabelled grey seal myoglobin. Labelled myoglobin from in vitro translations of RNA hybridising to the myoglobin cDNA clones co-purifies with unlabelled seal myoglobin (lanes 1b, 3b, 4b) but no translation product from RNA hybridising to non-myoglobin cDNAs co-purifies with unlabelled seal myoglobin (lanes 2b and 5b). The unusual migration of myoglobins in lanes 1a and 1b and 3a and 3b can be accounted for by the sample containing the polyacrylamide gel slice which results in the wide appearance of the myoglobin band.

Figure 6:5



0 pA I 2 3 4 5
a b a b a b a b a b a b

(Proudfoot and Brownlee, 1976).

A second myoglobin cDNA clone (pM171) was also sequenced and found to contain a cDNA insert corresponding exactly in sequence to positions 50 and 50' of the pM178 cDNA insert (Figure 6:4), followed by a (GG)₁₁ sequence prior to the cDNA (GG)₁₁ tail. These two independently derived cDNA clones must represent cDNA reverse transcribed from the same mRNA species as their sequences are identical.

Comparison of the pM178 and pM171 myoglobin cDNA sequences with the published amino acid sequence of the closely related harbour seal myoglobin (Dayhoff, 1972) showed that myoglobin sequences were not present in this cDNA. This indicates that the 5' non-translated region of seal myoglobin

3. DNA SEQUENCE ANALYSIS OF MYOGLOBIN cDNA CLONES

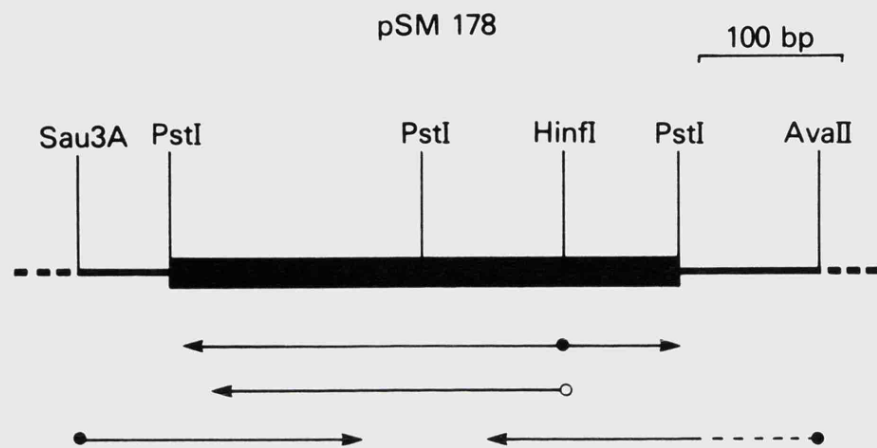
Six independent myoglobin cDNA clones were analysed by restriction endonuclease cleavage. As expected, the cDNA inserts were short, ranging from 150-355 bp in length, and showed certain similarities in restriction patterns including an internal Pst1 cleavage site in addition to the reconstituted Pst1 sites flanking the inserts.

The largest myoglobin cDNA insert, in the recombinant plasmid pSM178, was mapped in detail and completely sequenced. The sequencing strategy involved fragments which were kinase and fill-in labelled at the Hinf1 site within the cDNA insert and fragments fill-in labelled at the adjacent vector Sau3A and AvaII sites (Figure 6:6). The insert consisted of 301 base pairs of mRNA sequence flanked by (dG.dC)₁₆ and (dG.dC)₃₄ tails. One end of the molecule corresponded to the end of a mRNA sequence, with a (dA)₈ remnant of the poly(A) tail preceded at a distance of 17 base pairs by an AATAAA sequence commonly found towards the 3' end of mammalian mRNAs (Proudfoot and Brownlee, 1976).

A second myoglobin cDNA clone (pSM151) was also sequenced and found to contain a cDNA insert corresponding exactly in sequence to positions 60 and 309 of the pSM178 cDNA insert (Figure 6:6), followed by a (dA)₁₁ sequence prior to the oligo (dG.dC) tail. These two independently derived cDNA clones must represent cDNA reverse transcribed from the same mRNA species as their sequences are identical.

Comparison of the pSM178 grey seal myoglobin cDNA sequences with the published amino acid sequence of the closely related harbour seal myoglobin (Dayhoff, 1972) showed that myoglobin sequences were not present in this cDNA. This indicates that the 3' non-translated region of seal myoglobin

Figure 6:6



(a)

[illegible]

(b)

Figure 6:6 The myoglobin cDNA sequence of pSM178

The restriction endonuclease cleavage map of pSM178 is shown for Sau3A, Pst1, HinF1 and Ava11 (a). Restriction fragments were kinase-labelled at the 5' terminus (open circles) or fill-in labelled at the 3' terminus (closed circles). The extent of the sequence is shown by the length of the arrows, all the restriction sites of the insert are overlapped and the DNA sequence has been determined off both strands. The AATAAA polyadenylation signal and the remnant of the poly(A) tail are underlined, and the oligo (dG.dC) tails shown in lower case (b).

mRNA is at least 293 nucleotides in length in marked contrast to the short (51-135 nucleotide) 3' non-translated regions reported for a variety of avian and mammalian α - and β -globin mRNAs (Efstratiadis *et al.*, 1980; Liebhaber *et al.*, 1981; Roninson and Ingram, 1981; Hampe *et al.*, 1981). Myoglobin is 153 amino acid residues in length and therefore of similar size to α - and β -globins (141 and 146 amino acid residues respectively). This predicts that seal myoglobin mRNA should be substantially longer than globin mRNA. In order to test this prediction a Northern blot analysis of seal muscle polyadenylated RNA probed with ^{32}P -labelled pSM178 cDNA to detect myoglobin mRNA was carried out (Figure 6:7). A discrete hybridising component of approximately 1400 nucleotides in length was detected by this analysis. This is considerably longer than rabbit β -globin mRNA (750 nucleotides) detected in a parallel Northern blot hybridisation. This difference in size represents a major difference between haemoglobin and myoglobin mRNAs and suggests that seal myoglobin mRNA contains approximately 800 nucleotides of non-translated sequences in addition to 465 nucleotides of coding sequences.

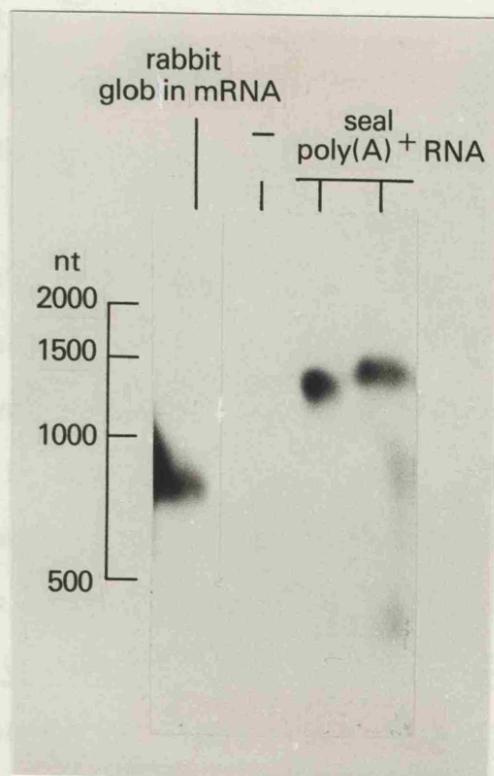
4. ABUNDANCE OF MYOGLOBIN mRNA IN SEAL MUSCLE POLYADENYLATED RNA

By relating the cDNA clones which positively select myoglobin mRNA from seal muscle polyadenylated RNA (Figure 6:4) to those clones which hybridise to cDNA made from seal muscle polyadenylated RNA (Figure 6:2) it is clear that myoglobin cDNA clones represent intensely hybridising clones. The approximate abundance of myoglobin mRNA was determined by hybridising a seal muscle cDNA library of around 800 clones with a purified myoglobin cDNA sequence. In order to avoid problems of cross-hybridisation between cDNA clones and the oligo (dG.dC) tails in pSM178, a 215 base pair *Sau96I*-*HinfI* cDNA fragment was isolated from the

Figure 6:7 Detection of grey seal myoglobin mRNA by use of myoglobin
cDNA clone pSM178

pSM178 DNA was digested with HhaI and the 642 bp fragment containing the cDNA insert isolated, labelled with ^{32}P by nick translation and used as a probe in a Northern blot hybridisation to denatured grey seal muscle polyadenylated RNA blotted onto nitrocellulose from a 2% agarose gel (seal poly(A)⁺ RNA). Also shown is the hybridisation of rabbit β -globin cDNA (from the plasmid p β G1) to rabbit globin mRNA and a zero message control hybridisation (-). RNA sizes were estimated by parallel electrophoresis of single stranded ^{32}P -labelled marker DNA (λ DNA x HindIII and pBR322 x Sau3A).

Figure 6:7



relatively short myoglobin cDNA of pSM151, which has a Sau96I site at the 5' end of the cDNA created by fusion of the oligo (dG.dC) tail to a CC dinucleotide at position 60-61 (Figure 6:6). This Sau96I site is obviously absent in the pSM178 cDNA insert. This pure myoglobin cDNA was labelled with ^{32}P by nick translation and hybridised to the seal cDNA library (Figure 6:8). Around 4% (34/780) of cDNA clones hybridised, including all those which had previously been identified as containing myoglobin cDNA inserts by their property of positively selecting myoglobin mRNA (Figure 6:4). In contrast, clones which had been shown to contain non-myoglobin cDNA inserts were not detected in the hybridisation with the purified myoglobin cDNA from pSM151. It appears therefore that myoglobin mRNA is one of the most abundant mRNA species in adult grey seal skeletal muscle and accounts for approximately 4% of all polyadenylated RNAs in this tissue.

5. DETECTION OF THE GREY SEAL MYOGLOBIN GENE BY SOUTHERN BLOT ANALYSIS

The myoglobin cDNA insert in pSM178 was labelled with ^{32}P and used to detect the myoglobin gene in Southern blot hybridisations of restriction endonuclease digested grey seal DNA (Figure 6:9). In each digest, a single major hybridising component was detected, and by analysing double digests of grey seal DNA, an unambiguous restriction endonuclease cleavage map could be constructed around this single myoglobin gene (Figure 6:10). At least one additional faintly hybridising component could also be detected in several digests of seal DNA, suggesting that the seal genome contains additional diverged sequence(s) containing at least the 3' end of a myoglobin gene.

Figure 6:8 Relative abundance of myoglobin mRNA

cDNA clones were replicated onto nitrocellulose filters and the colonies hybridised with ^{32}P -labelled myoglobin cDNA (isolated as a 215 bp Sau96I - HinfI DNA fragment from pSM151) and the filters autoradiographed (B). Colonies shown to contain myoglobin cDNA inserts by hybrid release translation (Figure 6:4) are marked 'M', non-myoglobin cDNA clones are marked '-' (A). Clones known to contain myoglobin cDNAs in this way also hybridise with the myoglobin cDNA from pSM178. Around 4% of all the colonies tested hybridised with the myoglobin cDNA insert and this was taken as an estimate of the relative abundance of myoglobin mRNA in grey seal muscle.

The filter labelled 'A' represents cDNA clones hybridised to ^{32}P -labelled single stranded cDNA prepared as described in Figure 6:2.

Figure 6:8

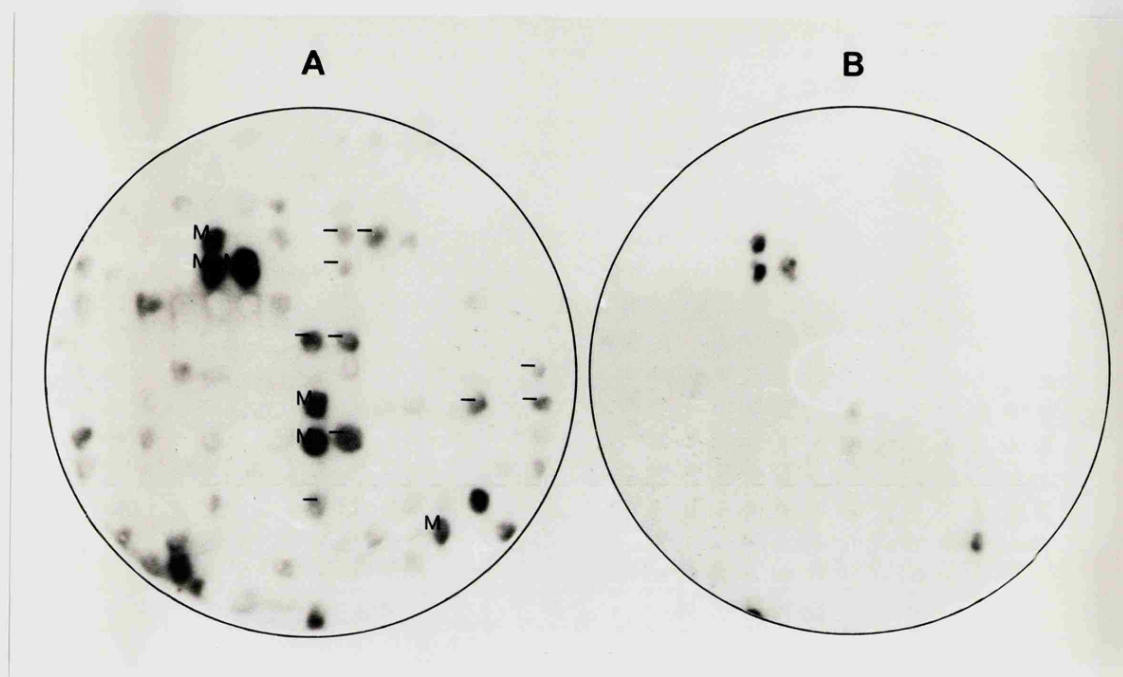


Figure 6:9 Seal myoglobin gene fragments detected by use of the myoglobin cDNA clone pSM178

A 642 bp HhaI fragment of pSM178 containing the entire myoglobin cDNA insert was isolated, labelled with ^{32}P by nick translation and used as a probe in Southern blot hybridisations (in 1 x SSC at 65°) against grey seal genomic DNA digested with various combinations of BamHI (B), BglII (Bg), EcoRI (E) and HindIII (H). DNA sizes were estimated by parallel electrophoresis of single stranded ^{32}P labelled marker DNA (λ DNA x HindIII and pBR322 DNA x Sau3A). The probe detects a 12.0 kb B, a 3.7 kb E, a 3.8 kb Bg and a 6.9 kb H fragment. The E and Bg fragments are contained within the H fragment, as H cleaves neither fragment. The B fragment overlaps the H fragment by 2.7 kb and the E fragment by 1.5 kb, positioning the E site within the 2.7 kb H - B fragment. The Bg fragment overlaps the B by 1.3 kb and the E fragment is trimmed by Bg thus positioning a Bg site within the 1.5 kb Bg - B fragment. The cleavage site map for these enzymes is presented in Figure 6:10.

Figure 6:10 Restriction endonuclease cleavage site map of the grey seal myoglobin gene

The physical map of restriction sites around the single myoglobin gene is shown. The 3' end of the gene is located in the shaded region
B = BamHI, H = HindIII, E = EcoRI, Bg = BglII.

Figure 6:9

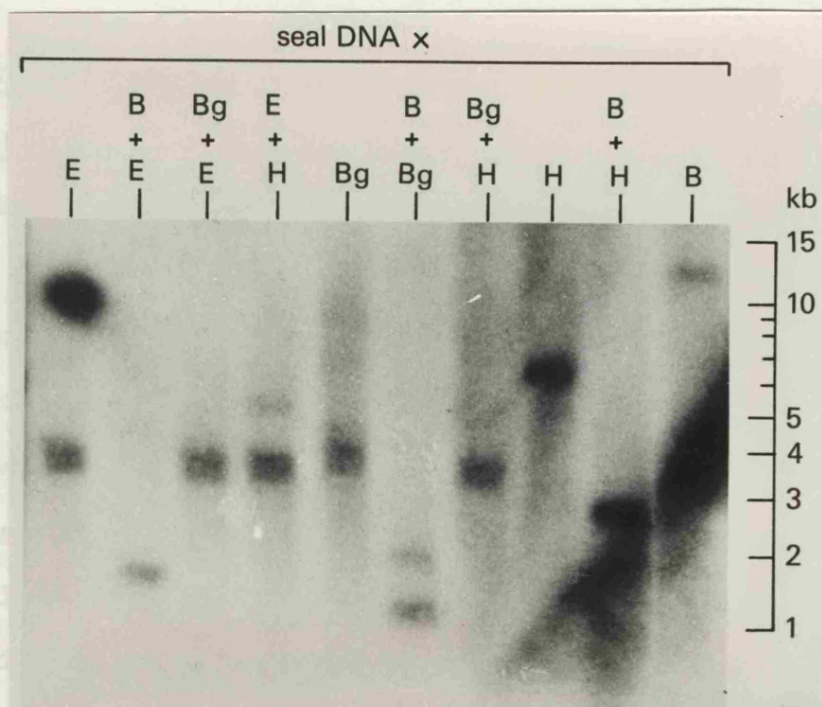
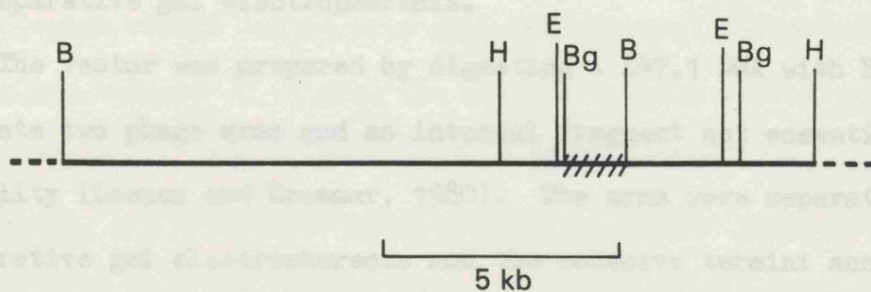


Figure 6:10



Chapter 7

CLONING THE GREY SEAL MYOGLOBIN GENE

1. CONSTRUCTION OF A GENOMIC LIBRARY

(i) Strategy

The strategy for constructing a genomic library of grey seal DNA was to ligate grey seal DNA fragments produced by partial digestion with *Sau3A* into the *Bam*HI sites of the replacement vector λ L47.1 (Loenen and Brammar, 1980).

Sau3A has a four base pair recognition sequence and therefore has a cutting site frequency of one per approximately 256 base pairs. By using partial digestion the number of possible ways of obtaining a fragment of specified length is increased as well as generating a random collection of large DNA fragments. By generating large average fragment sizes the number of phage plaques that have to be screened is reduced.

(ii) Preparation of Grey Seal DNA Fragments and Ligation into λ L47.1

Aliquots of high molecular weight grey seal DNA were digested with different amounts of *Sau3A*, pooled and fragments of 13 to 21 kb isolated by preparative gel electrophoresis.

The vector was prepared by digesting λ L47.1 DNA with *Bam*HI to generate two phage arms and an internal fragment not essential for phage viability (Loenen and Brammar, 1980). The arms were separated by preparative gel electrophoresis and the cohesive termini annealed. Equimolar amounts of arms and grey seal DNA were ligated using T4 DNA ligase. Typically 1 μ g of annealed λ L47.1 arms were ligated to 0.44 μ g of grey seal DNA fragments.

Using the equation of Clarke and Carbon (1976):

$$N = \frac{\ln (1-P)}{\ln (1-f)}$$

it is possible to calculate the number of different, independently derived phage recombinants needed to give a 99% probability of finding a given single copy sequence in the library. Where N = the number of plaques, p = probability and f = the fraction of the total genome represented by each fragment. f maybe calculated by dividing the average fragment size by the haploid genome size. Assuming a typical mammalian haploid genome size of 3×10^9 base pairs for the seal genome and an average fragment size of 15 kb, about 9×10^5 recombinants are needed. The system used in this work is therefore an efficient way of constructing a library of phage recombinants that can be screened on a small number of 9 cm plates without amplification.

Eight plates produced from two packaging reactions representing 460,000 plaque forming units were screened using the method of Benton and Davis (1977). The plaques were screened by hybridisation with the 642 bp fragment from a HhaI digest of pSM178, which contains part of the 3' untranslated region of grey seal myoglobin mRNA, labelled with ^{32}P by nick translation. Four positively hybridising regions were detected. With an average fragment size of 15 kb and a genome size of 3×10^9 bp, 460,000 clones would be expected to include about 2 clones of a given gene. This result therefore suggests that the seal genome contains only one or a few myoglobin genes.

(iv) Characterisation of Recombinant Phage

The four positively hybridising regions were picked as 2.5 mm plugs of agar using a glass tube. These were replated on E.coli ED8910 at low plaque density, rescreened with 2 cycles of rescreening. On rescreening only two strongly hybridising plaques could be detected (λ SM.1 and λ SM.19) and DNA made as described by Jeffreys et al., (1982).

DNA from each of these two recombinant phage was digested with restriction endonucleases both singly and in pairwise combinations and the resulting fragments separated by electrophoresis in the native form in neutral agarose gels. The DNA fragment patterns were then photographed and the DNA in the gels subjected to acid/alkali denaturation (Wahl et al., 1979) before transfer to nitrocellulose filters by Southern blotting. The filters were hybridised with labelled grey seal myoglobin cDNA (from pSM178) to identify fragments containing grey seal myoglobin gene sequences (Figure 7:1).

The maps of λ SM.1 and λ SM.19 are shown in Figure 7.2. Comparison between the genomic map of the 3' end of the myoglobin gene and the map of the recombinant DNAs shows that no detectable rearrangement has occurred in this region of DNA during the cloning process. The cloned DNA in λ SM.1 and λ SM.19 accounts for about 8.6 kb and 14.9 kb of the grey seal genome respectively.

The map of the recombinants λ SM.1 and λ SM.19 show a slight reduction in fragment sizes from those of the genomic map. However this can be attributed to differences in electrophoretic separation; in genomic mapping the DNA fragments were denatured in alkali before electrophoresis whereas in the characterisation of the recombinants the DNA fragments were denatured after electrophoresis. The estimated size of

Figure 7:1 Southern blot analysis of λ SM.1 DNA

λ SM.1 DNA was cleaved singly with BamHI (B), EcoRI, HindIII (H) and BglII (Bg) and electrophoresed in an agarose gel, denatured and transferred to nitrocellulose filters which were hybridised with labelled pSM178 DNA (Upper panel). Similarly double digests of λ SM.1 DNA were also analysed (Lower panel).

The scale is in kilobase pairs (kb).

Fragments run in the native form were as long as 1000 bp in the single stranded form (Barrell, 1980).

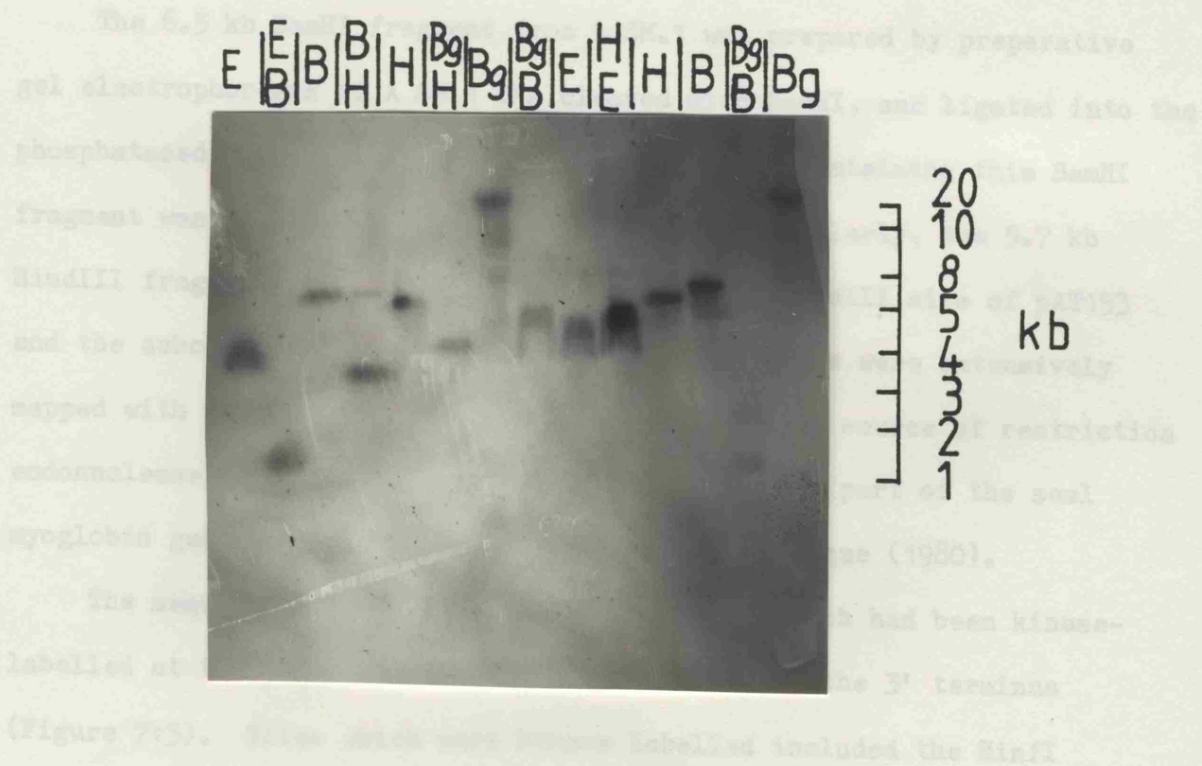
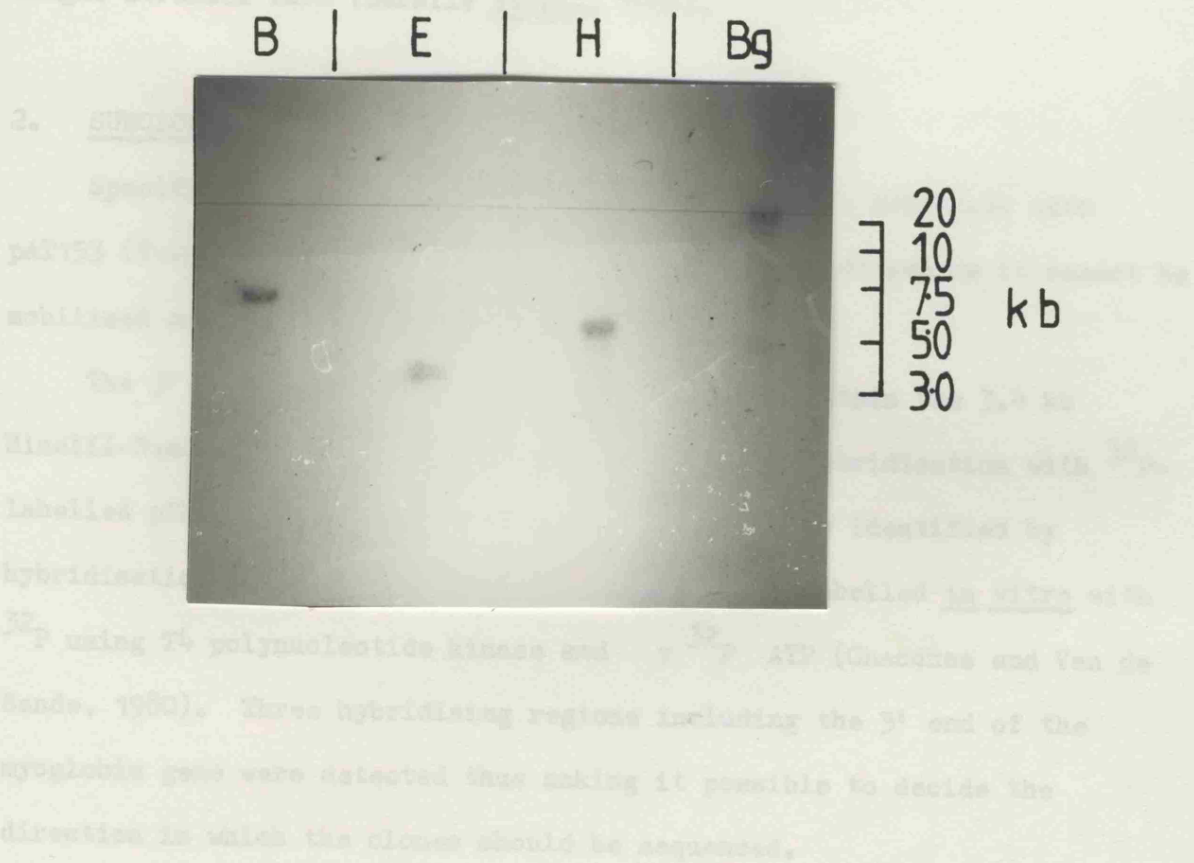


Figure 7:1

fragments run in the native form tends to be lower than those ran in the single stranded form (Baralle et al., 1980).

2. SUBCLONING AND SEQUENCING

Specific restriction endonuclease fragments were subcloned into pAT153 (Twigg and Sherratt, 1980). This vector was chosen as it cannot be mobilised and has a higher copy number than pBR322.

The 3' end of the myoglobin gene was located within the 3.4 kb HindIII-BamHI fragment of λ SM.1 by Southern blot hybridisation with ^{32}P -labelled pSM178 DNA. Additional coding regions were identified by hybridisation with seal muscle polyadenylated RNA labelled in vitro with ^{32}P using T4 polynucleotide kinase and γ ^{32}P ATP (Chaconas and Van de Sande, 1980). Three hybridising regions including the 3' end of the myoglobin gene were detected thus making it possible to decide the direction in which the clones should be sequenced.

The 6.5 kb BamHI fragment from λ SM.1 was prepared by preparative gel electrophoresis of λ SM.1 DNA cleaved with BamHI, and ligated into the phosphatased BamHI site of pAT153. The subclone containing this BamHI fragment was designated pSM1.17 (Figure 7:2). Similarly, the 5.7 kb HindIII fragment of λ SM.1 was ligated into the HindIII site of pAT153 and the subclone designated pSM1.9. These subclones were extensively mapped with restriction endonucleases and used as a source of restriction endonuclease fragments used in the determination of part of the seal myoglobin gene sequence by the Maxam-Gilbert technique (1980).

The sequencing strategy used DNA fragments which had been kinase-labelled at the 5' terminus or fill in labelled at the 3' terminus (Figure 7:3). Sites which were kinase labelled included the HinfI

Figure 7:2 Restriction endonuclease cleavage map of the 3' exon
of the grey seal myoglobin gene

A BamHI (B), BglII (Bg), EcoRI (E) and HindIII (H) restriction map was determined using λ SM.1 and λ SM.19 DNA. The leftmost BamHI site was created at the lefthand end of λ SM.1 during cloning and is absent from λ SM.19 and seal genomic DNA. Restriction fragments from λ SM.1 and λ SM.19 were subcloned into pAT153 and designated pSM1.17 and pSM1.9.

Figure 7:3 Sequencing strategy for the 3' exon of the grey seal
myoglobin gene

The sequencing strategy used restriction fragments isolated from pSM1.17 and pSM1.9. DNA fragments were kinase labelled at the 5' terminus (open circles) or fill-in labelled at the 3' terminus (closed circles) using reverse transcriptase. Myoglobin sequences are represented by the solid box and untranslated myoglobin mRNA sequence by the open box. Hf = HinfI E = EcoRI Bg = BglII S = Sau3A A = AvaII.

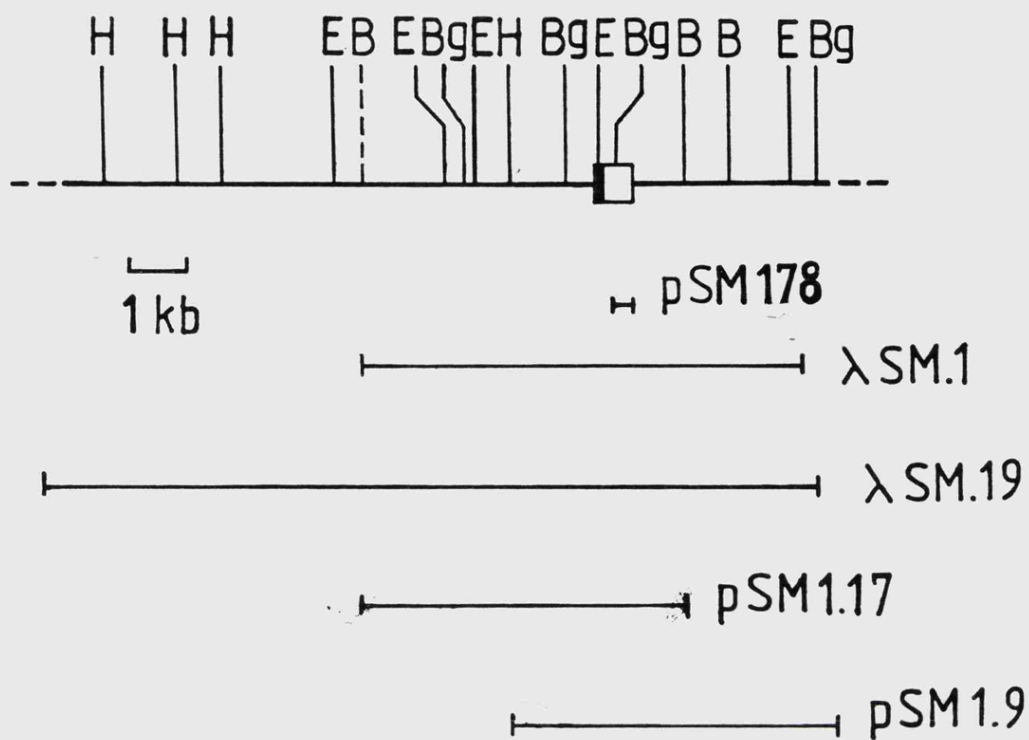


Figure 7:2

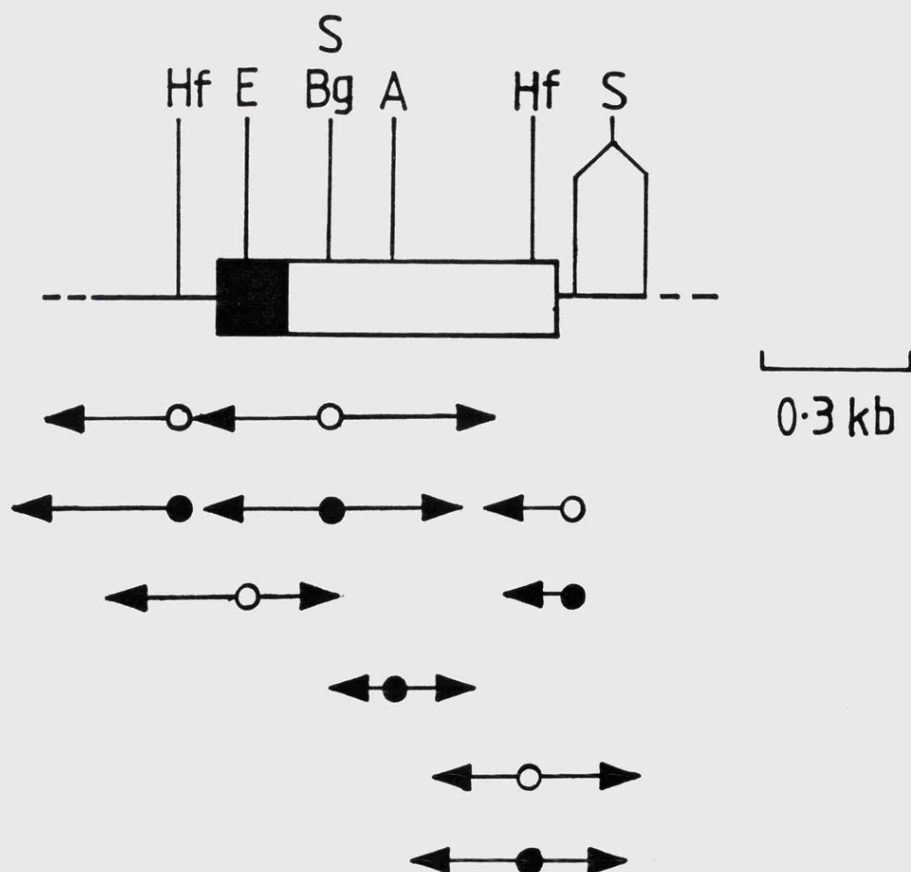


Figure 7:3

sites, the BglII site, the EcoRI site, and the Sau3A site. Sites which were fill-in labelled included the HinfI sites, the BglII site, the Sau3A site and the AvaII site. The DNA sequence obtained from the sequencing of these fragments is shown in Figure 7:4. The sequence is continuous and the majority of bases confirmed by sequencing both strands of DNA. As expected, this sequence extended upstream from the polyadenylation site and included a segment of myoglobin-coding DNA, covering codons 106-153 and predicting an amino acid sequence identical to that determined for the harbour seal.

Figure 7:4 DNA sequence of the 3' end of the grey seal myoglobin gene

Sequences present in mature myoglobin mRNA are shown in capital letters. The intron-exon junction (boxed) conforms to the GT-AG rule of Breathnach and Chambon (1981). The final 293 base pairs of the 3'-untranslated region are identical to the sequence of the myoglobin cDNA clone, pSM178 (Figure 6:6b). The AATAAA polyadenylation sequence is underlined and the poly(A) addition site arrowed. The 3' exon comprises codons 106-153 and the intron preceding this interrupts the coding sequence at homologous positions to the second intron of all α - and β -globin genes.

Dots above the sequence indicate every tenth base.

ccgactgggāaacttgggtāaaaccacttāatgggagcatttcccttccittttotgotgco
 agcatgggggācagtgtcctggggtccttgcāccgaggtgggācctgggttccācogtoggcaco
 tgtctcccāgggtggtagggāgagaaggaāttaatccacāgagacccctōgatagcaogo
 ttgggtgtgtāataagccctōgataaatggācogtggagtaāagagaaggtgggaotcagto
 tgattcccatōtggctcctggogtgtgcgāccctgtgaccāgoacccccctttcgtotoco

I10

phe ile ser glu ala ile ile his val leu his ser lys
 ccttgcagTC.ATC.TCA.GAA.GCC.ATC.ATC.CAT.GTC.CTG.CAC.AGC.AAG.

I20

I30

his pro ala glu phe gly ala asp ala gln ala ala met lys lys
 CAT.CCC.GCG.GAA.TTC.GGG.GCC.GAC.GCC.CAG.GCT.GCC.ATG.AAA.AAG.

I40

ala leu glu leu phe arg asn asp ile ala ala lys tyr lys glu
 GCC.CTG.GAG.CTG.TTC.CGG.AAT.GAC.ATC.GCT.GCC.AAA.TAC.AAG.GAG.

I50

leu gly phe his gly ter
 CTG.GGG.TTC.CAT.GGC.TAAGCACCTGGGCGCCCCCACTCCCACCCCCACCCACTCG
 GGCCCCGGGGCCAAGGGTCGGGTGGGCCTAGATCTCCTGTAGCCATGTAGAGTTTGCTTC
 TGAGTGTCTGCTTTGTTTGAGAGGTGGGTGGGAGAGGCTGAGGGGCTGGGCTAGGCTCTT
 CAGGGTGGCTTCGCTGCCTGGAGACTGGGGTCCTCGTGAGGGGTCATCACCCCGGGAACC
 AGGAGACATGTGACCCTTTGCTCACTGTGGCCCCGCAGTGTTAATTCGCCTCTCTCTTAA
 ACTCCAACCCAATTCCTCCAACCTCCAAACTCAACCAACACCTTCACCCTCACCTCTAATC
 ACAAATCCAAGCCATAAACCAAGGCCTGGCCATAAAGATGCTCTCCCATCACCCCCGCC
 CCAAAGTCTGTCTGCAGTGAGAAAGAGGTGTGGGCGGAGAGGGACAGCCACACGCATCCC
 GAGAGAGGCGACGGGGGTCTGGAACATAAGTGTGCCTTCTCAGCGAATGGAGTGACTCAC
 CTGGTTTTAATAAAAGACCTTGCAGCATCACagtcgctgggtgatcttcccttccoa
 gagcaagtgaaggggccccaggggaagaaatggctgggattgtgtgtgoagacagccctg
 gggggcgaggcagaggagaggaggaggaggaggaggaaacggggcaggaggaggagaggagao
 gggaagctgccttcccttagatcacccctcagtggoggggacagtgctgggcaogggggaag
 gagggggg

Figure 7:4

Chapter 8

DISCUSSION

1. INTRODUCTION

Myoglobin facilitates oxygen transport and acts as an oxygen store in red muscle. It is a monomeric protein of 153 amino acids and has been sequenced from a great number of species. The existence of this red haemoprotein in muscle has been known since the early 19th century when it was discovered to be distinct from the haemoglobin of blood. Advances in biochemical and molecular biological methods has enabled the study of this protein and its relationship to the haemoglobins to make great progress. This study is part of that progress.

The globins, including myoglobin, have been studied extensively at the protein level with regard to the evolution of a multigene family. In addition vertebrate globins are the most intensively studied family of proteins at the physical biochemical and molecular levels.

In 1958 Kendrew and his co-workers determined the three-dimensional structure of sperm whale myoglobin - the first protein structure to be determined. This established myoglobin as a classic protein to be studied at all levels including evolution, physiology and structure.

It is now possible to isolate and clone specific DNA sequences of single copy genes from higher organisms. This has allowed major advances in the study of gene evolution organisation and structure. It is apparent that little of the DNA in higher organisms encodes proteins.

There are large tracts of intergenic DNA, satellite DNAs and intervening sequences whose functions, if any, remains unknown. The discovery of intervening sequences in particular has led to intense speculation on the evolution of gene structure in eukaryotes. The possible genetic functions of non-coding DNA can be studied by analysing the molecular evolution of DNA. Extragenic regulatory elements in eukaryotes may be involved in changes in morphology during evolution (King and Wilson, 1975).

The vertebrate haemoglobin genes have been extensively studied using recombinant DNA techniques. Much knowledge has now accumulated on the arrangement, structure and organisation of globin genes in many species ranging from amphibians to man. In addition genes encoding the related monomeric leghaemoglobins have recently been cloned (Jensen *et al.*, 1981; Brisson and Verna, 1982; Wiborg *et al.*, 1982; Hyldig-Neilsen, 1982). These leghaemoglobin genes are unique in all the globin genes studied in that they are interrupted by three intervening sequences as opposed to the usual two.

It is surprising, therefore, that no attempt, until this study, has been made to clone a myoglobin gene. Myoglobin is an important member of the globin protein family. The myoglobins and haemoglobins are evolutionarily related, but are now incorporated within different developmental programmes - erythroid and myogenic. The first gene duplication in the evolutionary history of the globin gene family was that which gave rise to lineages leading to haemoglobins and myoglobins and may have occurred up to 10^9 years ago. Comparisons of haemoglobin and myoglobin genes could therefore provide information about early events in globin gene evolution, the organisation of intervening sequences, the chromosomal arrangement of haemoglobin and myoglobin genes and the properties of homologous genes incorporated into radically different developmental programmes.

2. cDNA CLONING

The aim of this project has been to clone a mammalian myoglobin gene. The approach used to achieve this aim was to construct a cDNA clone bank using skeletal muscle polyadenylated RNA and to identify myoglobin cDNA clones by their ability to select myoglobin mRNA from polyadenylated RNA, which could then be translated in vitro to give the myoglobin protein. Once identified these cDNA clones can be used to screen a genomic library to isolate the myoglobin gene.

The preparation of a cDNA clone bank involves a series of enzymic steps (Williams, 1981). These enzymic steps are technically demanding, time consuming and expensive. The justification for using this approach as opposed to screening a genomic bank is a) no direct screen for a myoglobin gene is possible without the use of a cDNA b) a cDNA clone bank contains fewer clones than a genomic bank. Eukaryotic mRNAs are present at varying abundances in different cell types and the frequency of occurrence of a particular clone in a cDNA bank is roughly proportional to this abundance. Myoglobin forms 1% of total protein in the myoglobinised muscle of land mammals and up to ten times this amount in the muscle of diving mammals. It is therefore a reasonable assumption that myoglobin mRNA is relatively abundant in these tissues. Using this tissue as a source of mRNA it should be possible to make a cDNA clone bank of a relatively small number of clones and have a reasonable chance of finding a myoglobin cDNA clone. c) every cDNA clone contains an mRNA sequence and therefore a positive hybridisation signal should mean a clone should contain the required sequence. Screening of genomic banks usually generates false positives due to the contamination of mRNA preparations with ribosomal RNA which can be copied into cDNA with low efficiency.

d) it is not necessary to obtain a full length cDNA copy of a particular mRNA in order to use this clone to screen a genomic bank for the gene required. Providing there is enough specificity in the cloned cDNA it should be relatively straightforward to isolate a specific genomic clone.

The construction of a cDNA clone bank involves several enzymic steps. At any of these steps the presence of contaminating exonuclease or endonuclease activities may lead to the cloning of only partial cDNA copies. It is therefore necessary to characterise each enzyme used before using them on valuable samples. It may be advisable to perform 'dummy' experiments using easily obtainable mRNA such as globin mRNA before using these enzymes with painstakingly prepared mRNA. Alternatively 'pilot' runs on small proportions of the prepared sample can be carried out in appropriately scaled down reactions for each stage of the cDNA construction. Generally, this 'pilot' run approach has been used in this study.

The first stages in the construction of a cDNA clone bank is the preparation of mRNA from a tissue in which the gene of interest is expressed. Between 1 and 2% of total cytoplasmic RNA is mRNA, a relatively large amount of mRNA needs to be prepared in order to construct a cDNA clone bank of reasonable size, particularly if subsequent screening procedures involves the use of mRNA to make cDNA probes or in positive or negative selection hybridisations.

The major concern in preparing total cell RNA from a tissue is the action of cellular ribonucleases in the degradation of RNA. The use of ribonuclease inhibitors, low temperatures and speed in the extraction procedure is strongly recommended. DNA can easily be removed from RNA preparations by pelleting the RNA in caesium chloride (Gilsin et al., 1974; Ullrich et al., 1973).

It is possible to make cDNA clones using total RNA preparations, however ribosomal RNA sequences can also be copied (Williams and Lloyd, 1979). It is therefore useful to purify mRNA before cloning. Most eukaryotic mRNAs contain a tract of around 100 (A) residues at their 3' termini (Kates, 1970; Lim and Cannelakis, 1970). Polyadenylated RNA can be prepared by separation on an affinity column of oligo(dT) cellulose (Aviv and Leder, 1972) or poly(U) sepharose (Palatnik et al., 1979).

Once polyadenylated RNA has been purified it is possible to prime the cDNA synthesis using oligo(dT) primers of 10-20 nucleotides. These primers hybridise to the poly(A) tract of the mRNAs and form a substrate for reverse transcriptase. Use of these primers should ensure that cDNA synthesis is initiated at the 3' end of the mRNA. E.coli poly(A) polymerase can be used to add poly(A) tracts to the 3' ends of non-polyadenylated RNAs such as ribosomal RNA. If some of the nucleotide sequence is known, specific synthetic primers can be used to make cDNAs of a particular mRNA such as rat insulin mRNA (Chan et al., 1979).

Most cDNA cloning has involved the insertion of a double stranded cDNA into a bacterial plasmid and this has been the method adopted in this study. An alternative method is to insert an mRNA-cDNA hybrid into a plasmid using complementary homopolymeric tails (Wood and Lee, 1976; Zain et al., 1979). Reverse transcriptase from Avian Myeloblastosis virus is used in the presence of a primer, to synthesise the first strand. The mRNA template is then removed by alkali hydrolysis and the second synthesised by the self priming synthesis reaction. A region of self complementarity at the 3' end of a cDNA molecule forms a 'hairpin' structure (Leis and Hurwitz, 1972). Synthesis of the second strand can be by reverse transcriptase, as in this study, or by DNA polymerase I,

the Klenow fragment or T⁴ DNA polymerase. The phosphodiester bonds within the 'hairpin' loop of a double stranded cDNA can be removed by S1 nuclease from Aspergillus oryzae.

S1 nuclease digestion invariably leads to the loss of sequences corresponding to the extreme 5'-terminal region of the mRNA. A procedure for the efficient cloning of the entire sequence of an mRNA is presented by Land et al. (1983). This method involves the priming of second strand cDNA synthesis by oligo(dG) hybridised to the 3' homopolymer tail of dCTP-tailed single stranded cDNA. Full length double stranded cDNA is purified from agarose gels and after a second tailing step with dCTP, the cDNA is annealed to dGTP tailed plasmid DNA.

Double stranded cDNA may be introduced into a plasmid by the use of homopolymeric tailing or restriction enzyme linkers. In this study C tails were added to the cDNA and G tails to the PstI cleaved vector pAT153 using terminal transferase. PolydG-polydC tailing was used as these tails are more stable than polydA-polydT hybrids and the PstI site of the vector is regenerated allowing the excision of the cDNA insert (plus tails).

Annealed cDNA and vector can be introduced into bacterial cells by transformation. The bacterium acquires any drug resistance carried by the plasmid. E.coli cells can be treated with calcium chloride to make them competent for DNA transformation (Mandel and Higa, 1976). In this study the E.coli strain HB101 has been used for transformation as it is recombination deficient (Boyer and Roulland-Dussoix, 1969).

Once transformants have been obtained it is necessary to screen them to identify a clone containing the sequence of interest. Transformants can be spotted onto nitrocellulose filters, lysed and hybridised to a radioactively labelled probe, (Grunstein and Hogness, 1975). The probe

used may be a labelled RNA or a labelled cDNA. The amount of probe hybridising to a clone is directly proportional to the abundance of the sequence in an mRNA population. In this study radioactively-labelled cDNAs of polyadenylated RNA were hybridised to cDNA clones in order to determine which clones were recombinant, plus the relative RNA abundance of individual cDNA clones.

Further screening can be carried out by the analysis of translation products directed by RNA complementary to the plasmid DNA. In this study a positive selection procedure has been adopted. Plasmid DNA was bound to nitrocellulose and hybridised to polyadenylated RNA. Hybridised RNA was eluted and translated in vitro. In this way it is possible to correlate recombinant plasmids with a specific translation product. An alternative, negative selection, procedure may be used. In this method - hybrid arrested translation - mRNA-DNA hybrids are formed and the mRNA translated in vitro. As mRNA-DNA hybrids do not translate in vitro it is possible to look for a translation product which disappears or is reduced in amount and to correlate a cDNA clone with that translation product. Hybrid arrested translation is useful in identifying high abundance mRNA sequences for which a high-resolution gel separation technique is available. However, it is not suitable for identifying medium or low abundance mRNA sequences (Williams, 1981).

Once a cDNA clone of interest has been identified it can be directly sequenced and compared to any known amino acid sequence and then used to isolate the gene of interest from a genomic library.

3. ISOLATION OF A MYOGLOBIN cDNA

Initial attempts at obtaining a myoglobin cDNA clone from humans proved unsuccessful. It is essential when attempting to make cDNA clones that the original mRNA preparation is of high quality. RNA preparations were made from human skeletal muscle. These muscle samples were often post mortems and amputations. It was clear from our preparations of RNA from these tissues that a large amount of degradation had occurred. However, these preparations proved a useful substrate in in vitro translation experiments. It was shown that polyadenylated RNA from human skeletal muscle could be translated in vitro to give a number of relatively high molecular weight protein products as visualised on SDS polyacrylamide gels. It is clear then that these preparations of polyadenylated RNA contained some full length copies of the more abundant muscle mRNAs. By the use of SDS polyacrylamide gels and immunoprecipitation it was shown that one of these translation products was the 17 kd protein myoglobin. It is clear from this result that at least some of the human skeletal muscle myoglobin mRNA is polyadenylated. It was essential for the strategy adopted that myoglobin could be recognised on SDS polyacrylamide gels of in vitro translations of skeletal muscle polyadenylated RNA.

A cDNA clone bank was made from preparations of human skeletal muscle polyadenylated RNA. This bank of around 200 clones was screened using the cDNA insert of a mouse α -actin containing plasmid. Actin mRNA should be a relatively abundant message in skeletal muscle and therefore actin containing cDNA clones should be readily detected with this probe and this was found to be the case. As no similar probe existed for obtaining a myoglobin cDNA a different screening procedure was adopted. This screen involved hybrid released translation or positive selection of myoglobin

mRNA. It was found by this screen that the cDNA clone bank did not appear to contain a myoglobin cDNA. It was concluded that the abundance of myoglobin mRNA in human skeletal muscle was too low for a myoglobin cDNA to be detected from a cDNA bank of this size.

Diving mammals such as whales and seals have 6-10 fold elevated levels of myoglobin protein in skeletal muscle compared with man and other terrestrial mammals. These elevated levels of myoglobin could be due to elevated levels of myoglobin mRNA in these tissues. Skeletal muscle from adult grey seal (Halichoerus grypus) was therefore used as a potentially rich source of myoglobin mRNA, to synthesise a myoglobin cDNA clone.

RNA was prepared from adult grey seal skeletal muscle. Electrophoresis of total RNA in 2% agarose gels indicated some degree of post-mortem degradation. The polyadenylated RNA showed a range of sizes up to 4000 nucleotides in length. In vitro translation of grey seal skeletal muscle polyadenylated RNA showed a range of translation products up to 200 kd in size. 2-3 prominent in vitro labelled translation products could be resolved in the molecular weight range, 17-20 kd. The smallest of these co-electrophoresed with grey seal myoglobin. The other small labelled polypeptides (18-20 kd) probably included myosin light chains 17-21 kd (Weeds and Lowey, 1971). The 17 kd labelled translation product was shown to copurify with pure native grey seal myoglobin on isoelectric focussing gels.

These results indicate that myoglobin mRNA should be moderately abundant in seal muscle, that at least some myoglobin mRNA molecules are polyadenylated and that translation of this message in a rabbit reticulocyte lysate produces native myoglobin complexed with haem.

A cDNA clone bank was made using grey seal skeletal muscle polyadenylated RNA. Agarose gel electrophoresis of this cDNA showed a relatively low size range of between 100 and 1000 nucleotides. This is probably due to the use of partially degraded polyadenylated RNA and an inefficient first strand synthesis.

Around 1600 cDNA clones were screened using the hybrid released translation screen. Several filter selected RNAs directed the synthesis of only the 17 kd labelled apomyoglobin, identifying the corresponding cDNA clones as containing myoglobin cDNA inserts.

These clones were analysed by restriction endonuclease cleavage and shown to contain short cDNA inserts of between 150 to 355 bp in length. The largest myoglobin cDNA insert of pSM178 was completely sequenced. The insert consisted of 301 bp of mRNA sequenced flanked by (dG.dC)₁₆ and (dG.dC)₃₄ tails. One end corresponded to the end of a mRNA sequence, with a (dA)₈ remnant of a poly(A) tail preceded at a distance of 17 bp by an AATAAA sequence commonly found towards the 3' end of mammalian mRNAs (Proudfoot and Brownlee, 1976).

The amino acid sequence of the myoglobin of the harbour seal (Phoca vitulina) has been published (Bradshaw and Gurd, 1969). Comparison of the pSM178 grey seal myoglobin cDNA sequence with this amino acid sequence showed that myoglobin sequences were not present in this cDNA. This indicates that the 3' non-translated region of grey seal myoglobin mRNA is at least 293 nucleotides in length. This is in marked contrast to the short 3' non-translated regions reported for a variety of avian and mammalian α - and β -globins which range between 51 and 135 nucleotides (Efstratiadis et al., 1980; Liehaber et al., 1981; Roninson and Ingram, 1981; Hampe et al.). This suggests that seal myoglobin mRNA should be substantially longer than globin mRNA as the polypeptides are similar in

length, myoglobin being 153 residues and α - and β -globin being 141 and 146 residues respectively. This prediction was confirmed by a 'Northern' blot analysis of seal polyadenylated RNA probed with labelled pSM178 cDNA to detect myoglobin mRNA. A discrete hybridising component of approximately 1400 nucleotides was detected which is considerably longer than rabbit β -globin mRNA (750 nucleotides) detected in a parallel Northern blot hybridisation. This is a major difference between haemoglobin and myoglobin mRNAs and suggests that seal myoglobin mRNA contains approximately 800 nucleotides of non-translated sequences in addition to 465 nucleotides of coding sequence. It is interesting to speculate on the significance of this difference between myoglobin and globin mRNA.

The approximate abundance of myoglobin mRNA in grey seal skeletal muscle was established by hybridising the seal muscle cDNA library against a purified myoglobin cDNA. A 215 bp *Sau96I*-*HinfI* cDNA fragment was isolated from the relatively short myoglobin cDNA of pSM151. This fragment eliminates the (dG,dC) tails which would cross-hybridise to all the clones in the cDNA library. 4% of cDNA clones hybridised to this purified myoglobin cDNA, including those which had previously been identified as containing myoglobin cDNA inserts. In addition, most of the clones which hybridised most strongly with labelled cDNA prepared from total polyadenylated RNA hybridised with myoglobin cDNA. It can be concluded, therefore, that myoglobin is one of the most abundant mRNA species in grey seal skeletal muscle and accounts for approximately 4% of all polyadenylated RNA in this tissue. This confirms the assumption that the high level of myoglobin in the muscle of diving mammals is correlated with high levels of myoglobin mRNA at least in the grey seal.

In order to begin to analyse the genomic organisation of myoglobin genes in grey seal the purified myoglobin cDNA was used to detect the

myoglobin gene in Southern blot hybridisations of seal genomic DNA. In each digest a single major hybridising component was detected and an unambiguous cleavage map can be constructed around this single myoglobin gene. Other faintly hybridising components could also be detected in several digests, suggesting that the seal genome may contain additional diverged sequences containing at least the 3' end of a myoglobin gene.

There is tentative genetic evidence for a single myoglobin gene in man (Boyer et al., 1963) which probably specifies foetal and adult cardiac myoglobin as well as skeletal muscle myoglobin. The grey seal genome also appears to contain a single myoglobin gene. This suggests that the elevated level of myoglobin in seal muscle is not the result of amplification of a myoglobin gene, but may be the result of an increased transcription rate from a single myoglobin gene. A comparison of the 5' regulatory sequences of a seal and terrestrial mammal myoglobin gene could prove interesting in elucidating the mechanism of this regulation.

4. ISOLATION OF A MYOGLOBIN GENE

The myoglobin cDNA insert of pSM178 was used to detect myoglobin gene sequences in a genomic library of Sau3A partial digests of grey seal DNA cloned into the BamHI replacement phage vector λ L47.1 (Loenen and Brammar, 1980). Two hybridising recombinants, λ SM.1 and λ SM.19 were isolated and shown to be overlapping isolates of a single region of the seal genome. The 3' end of the myoglobin gene was located by Southern blot analysis of these recombinants with cloned myoglobin cDNA. The restriction map around this region corresponds to the map of the seal myoglobin gene previously determined by Southern analysis of seal genomic DNA probed with myoglobin cDNA. Fragments from these two recombinants were subcloned into pAT153 and used for DNA sequencing.

The 3' exon of the seal myoglobin gene has been completely sequenced and is remarkably long comprising codons 106-153 plus 548 bp of 3'-untranslated sequence. The amino acid sequence deduced from the exon coding region corresponds exactly to the sequence reported for grey seal myoglobin from the harbour seal, Phoca vitulina (Bradshaw and Gurd, 1969). The sequence extends 5' into non coding DNA, presumably of an intervening sequence, for a further 308 bp. The position of this intervening sequence in the seal myoglobin gene (codons 105-106) is precisely homologous to the second intervening sequence in α - and β -globin genes (Efstratiadis et al., 1980). The exon-intron junction in the seal myoglobin gene conforms to the GT-AG rule of Breathnach and Chambon (1981) in that the final two bases of the intron are A G. The final 293 bp of the 3' untranslated region are identical to the sequence of the myoglobin cDNA clone, pSM178, fully confirming that this gene is the functional myoglobin gene in adult seal muscle.

Romero-Herrera et al. (1978) suggested that the 21 bases of the myoglobin mRNA specifying residues 133-139 of the protein may form a hairpin loop with the mRNA encoding residues 146-152. They postulate a sequence where 15 of the 21 base pairs are complementary and may form a stable structure. The sequence of the grey seal mRNA in this region is different to this postulated sequence and only 9 of the 21 base pairs are complementary. The proposed hairpin structure is therefore unlikely to form in the grey seal myoglobin mRNA sequence. There must be another explanation for the conservation of the seven invariant residues 133-139 probably involving functional and steric considerations.

The complete DNA sequence of the exons of the grey seal myoglobin gene has been completed by other workers and has now been published (Blanchetot et al., 1983). The coding region of the gene is interrupted by two very long introns of 4,800 and 3,400 bp. The three exon, two intron organisation is identical to that of all vertebrate α - and β -globin genes. This structure must therefore have been established before the myoglobin-haemoglobin divergence up to 800 million years ago. The soybean leghaemoglobin gene contains an additional intron within the central globin exon. Whether this extra intron was gained by insertion into a globin-like gene or whether a primordial gene ancestral to globin and leghaemoglobin contained three introns, one of which was eliminated early in animal evolution is unknown.

It remains to be seen whether this organisation of the grey seal myoglobin gene is peculiar to this species or whether it represents a common feature of all vertebrate myoglobin genes.

5. PROSPECTS

The grey seal myoglobin gene is the first myoglobin gene to be isolated. Gene sequence from this clone can be used to isolate the myoglobin genes of other species. It should be interesting to isolate the human myoglobin gene in order to compare its structure and organisation. The human globin gene family is well characterised at the molecular level. Information regarding the similarities and differences between human myoglobin genes and haemoglobin genes should prove of great evolutionary and functional interest.

The grey seal myoglobin gene has been used to isolate the human myoglobin gene from a human genomic library, (Weller et al., 1984). The amino acid sequence derived from the cloned human gene corresponds exactly

with the myoglobin sequence determined by Romero-Herrera and Lehmann (1974). The overall organisation of the human and grey seal myoglobin genes is similar. The human myoglobin gene is interrupted by two introns at codon 31 and between codons 105 and 106, that is, in identical positions to those in the seal myoglobin gene and all characterised vertebrate α - and β -globin genes (Blanchetot et al., 1983). These introns (5.8 kb and 3.6 kb) are much longer than their haemoglobin counter parts and slightly longer than the introns of the grey seal myoglobin gene. The 5'- and 3'-nontranslated regions of the human and grey seal myoglobin genes are of similar length and as a result the predicted mRNA length is similar. Analysis of the human myoglobin gene sequence reveals tandem repetitive sequence upstream of the gene with internal homologies. In addition a 33 base pair tandem repeat element is present in the first intron flanked by a direct repeat and sharing homology with other repetitive elements in the human genome.

Unique sequence DNA probes were isolated from the cloned human myoglobin gene and used to assign the gene to a specific human chromosome (Jeffreys et al., 1984). The myoglobin gene is located on chromosome 22 in the region 22q11 $\rightarrow$ 22q13. The myoglobin gene is therefore not linked to either the α - or β -globin gene clusters on chromosomes 16 or 11 respectively, and represents a third dispersed globin locus in the human genome.

As well as its use in isolating myoglobin genes from other species the seal myoglobin gene can be used as a probe to analyse the genomic DNA of many species in order to determine gene number and arrangement. In addition the nucleotide sequence comparison of myoglobin genes from many different species should allow more detailed molecular phylogenies to be constructed.

It should be interesting to study the organisation of myoglobin genes in diving and terrestrial mammals in order to understand the genetic basis for increased levels of the protein in the former. A study of promoter efficiency and messenger RNA stability could be carried out using myoglobin cloned gene sequences.

Isolated myoglobin gene sequences can be used to study the developmental expression of the myoglobin genes in different tissues. Myoglobin gene sequences can be cloned into expression vectors in order to study regulation of these genes in either prokaryotic or eukaryotic systems.

6. CONCLUDING REMARKS

This study was undertaken in order to isolate a mammalian myoglobin gene. From the results presented here it is clear that the experimental strategy adopted has proved successful in isolating the grey seal myoglobin gene. The use of this cloned DNA in the analysis of myoglobin genes from this and other species should enable a wealth of information to be obtained regarding myoglobin gene evolution and organisation. The analysis of myoglobin at the DNA level should complement the studies already carried out on myoglobin proteins. Myoglobin is an ancient and important member of the most extensively studied protein family. The analysis of myoglobin gene sequences should further enable us to elucidate the overall organisation and evolution of this multigene family.

The use of the techniques of recombinant DNA technology has revolutionised the study of molecular biology. These techniques have been used to isolate a mammalian myoglobin gene. Further technical advances should enable us to manipulate gene sequences in vitro and in vivo in order to gain a greater understanding of the control of gene structure and function.

Bibliography

- Aaij, C. and Borst, P. (1972).
The gel electrophoresis of DNA. *Biochim. Biophys. Acta.*, 269, 192-200.
- Adams, J.W., Kaufman, R.E., Kretschmer, P.J. and Nienhuis, A.W. (1980).
A family of long reiterated DNA sequences, one copy of which is next to the human β -globin gene. *Nucl. Acids Res.*, 8, 6113-6128.
- Affara, N.A., Robert, B., Jacquet, M., Buckingham, M.E. and Gros, E. (1980a).
Changes in gene expression during myogenic differentiation. I Regulation of mRNA 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 proteins encoded by myotube-specific cDNA sequences. *J. Mol. Biol.*, 140, 459-470.
- Andersen, M.E. and Gibson, Q.H. (1971).
Kinetic analysis of binding of oxygen and carbon monoxide to lamprey haemoglobin Petromyzon marinus and Petromyzon fluviatilis. *J. Biol. Chem.*, 246, 4790-4799.
- Appleby, C.A. (1974).
In 'The Biology of Nitrogen Fixation', Quispel, A. Ed., North-Holland Publishing Company, Amsterdam Oxford, pp.499-554.
- 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.
- Baralle, F.E., Shoulders, C.C., Goodbourn, S., Jeffreys, A. and Proudfoot, N.J. (1980a).
The 5' flanking region of human ξ -globin gene. *Nucl. Acids Res.*, 8, 4393-4404.
- Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980b).
The primary structure of the human ξ -globin gene. *Cell*, 21, 621-626.
- Barets, A. (1961).
Contribution à l'étude des systèmes moteurs "lent" et "rapide" du muscle latéral des téléostéens. *Arch. Anat. Micro. Morph. Exp.*, 50, 92-97.
- Barrie, P.A., Jeffreys, A.J. and Scott, A.F. (1981).
Evolution of the β -globin gene cluster in man and the primates. *J. Mol. Biol.*, 149, 319-336.

- Behlke, J. and Scheler, W. (1970).
Molecular properties of methaemoglobin of Chironomus plumosus L.
in solution. *Eur. J. Biochem.*, 3, 153-157.
- Bentley, D.L. and Rabbitts, T.H. (1980).
Human immunoglobulin variable region genes - DNA sequences of 2
VK genes and a pseudogene. *Nature*, 288, 730-733.
- Benton, W.D. and Davis, R.W. (1977).
Screening λ gt recombinant clones by hybridisation to single plaques
in situ. *Science*, 196, 180-182.
- Bernstein, S.I., Mogami, K., Donady, J.J. and Emerson, C.P. Jr. (1983).
Drosophila muscle myosin heavy chain encoded by a single gene in
a cluster of muscle mutations. *Nature*, 302, 393-397.
- Birnboim, H.C. and Doly, J. (1979).
A rapid alkaline procedure for screening recombinant plasmid DNA.
Nucl. Acids Res., 7, 1513-1523.
- Bishop, J.M. (1981).
Enemies within : the genesis of retrovirus oncogenes. *Cell*, 23, 5-6.
- Blake, C.C.F. (1978).
Do genes-in-pieces imply proteins-in-pieces? *Nature*, 273, 267.
- Blake, C.C.F. (1979).
Exons encode protein functional units. *Nature*, 277, 598.
- Blake, C.C.F. (1981).
Exons and the structure, function and evolution of haemoglobin.
Nature, 291, 616-618.
- Blanchetot, A., Wilson, V., Wood, D. and Jeffreys, A.J. (1983).
The seal myoglobin gene : an unusually long globin gene. *Nature*,
301, 732-734.
- Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K.,
Faber, H.E., Furlong, L., Grunwald, D.J., Kiefer, D.O., Moore, D.D.,
Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977).
Charon phages : safer derivatives of bacteriophage lambda for DNA
cloning. *Science*, 194, 161-169.
- Bonner, W.M. and Laskey, R.A. (1974).
A film detection method for tritium-labelled proteins and nucleic
acids in polyacrylamide gels. *Eur. J. Biochem.*, 46, 83-88.
- Borst, P. and Grivell, L.A. (1981).
One gene's intron is another gene's exon. *Nature*, 289, 439-440.
- Boulton, F.E., Huntsman, R.G., Lorkin, P.A. and Lehmann, H. (1969).
Abnormal human myoglobin; 53 (D4) Glutamic acid $\rightarrow$ lysine. *Nature*,
223, 832-833.

- Boulton, F.E., Huntsman, R.G., Lehmann, H., Lorkin, P.A. and Romero-Herrera, A.E. (1970).
Myoglobin variants. *Biochem. J.*, 119, 69P.
- Boulton, F.E., Huntsman, R.G., Romero-Herrera, A.E., Lorkin, P.A. and Lehmann, H. (1971a).
A human myoglobin variant 133(H10) lysine $\rightarrow$ asparagine. *Biochem. Biophys. Acta.*, 229, 871-876.
- Boulton, F.E., Huntsman, G.I., Yowson, A.E., Romero-Herrera, A.E., Lorkin, P.A. and Lehmann, H. (1971b).
The second variant of human myoglobin 138(H16) arginine $\rightarrow$ glutamine. *Brit. J. Haem.* 20, 69-73.
- Boulton, F.E., Huntsman, R.G., Romero-Herrera, A.E., Lorkin, P.A. and Lehmann, H. (1971c).
The third variant of human myoglobin showing an unusual amino acid substitution; 138(H16) arginine $\rightarrow$ tryptophan. *Biochim. Biophys. Acta.*, 229, 716-719.
- Boyer, H.W. and Roulland-Dussoix, D. (1969).
A complementation analysis of the restriction and modification of DNA in *E.coli*. *J. Mol. Biol.*, 41, 459-472.
- Boyer, S.H., Fainer, D.C. and Naughton, M.A. (1963).
Myoglobin : Inherited structural variation in man. *Science* 140, 1228-1231.
- Bradshaw, R.A. and Gurd, F.R.N. (1969).
Comparison of myoglobins from Harbour Seal, Porpoise and Sperm Whale. V. The complete amino acid sequences of Harbour Seal and Porpoise myoglobins. *J. Biol. Chem.*, 244, 2167-2181.
- Breathnach, R. and Chambon, P. (1981).
Organisation and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.*, 50, 349-383.
- Brenner, S. (1974).
New directions in molecular biology. *Nature*, 248, 785-787.
- Brisson, N. and Verma, D.P.S. (1982).
Soybean leghaemoglobin gene family : normal pseudo, and truncated genes. *Proc. Natl. Acad. Sci. USA.*, 79, 4055-4059.
- Buckingham, M.E. (1977).
Muscle protein synthesis and its control during the differentiation of skeletal muscle cells *in vitro*. In 'International Review of Biochemistry : Biochemistry of Cell Differentiation II', ed. Paul J. Univ. Park Press, Baltimore, vol 15, pp.269-332.
- Buell, G.N., Wickens, M.P., Carbon, J. and Schimke, R.T. (1979).
Isolation of recombinant plasmids bearing cDNA to hen ovomucoid and lysozyme mRNAs. *J. Biol. Chem.*, 254, 9277-9283.

- Chaconas, G. and Van de Sande, J.H. (1980).
5'-³²P labelling of RNA and DNA restriction fragments. *Methods in Enzymology*, 65, 75-85.
- Chan, S.J., Noyes, B.E., Agarwal, K.L. and Steiner, D.F. (1979).
Construction and selection of recombinant plasmids containing full length complementary DNAs corresponding to rat insulins I and II. *Proc. Natl. Acad. Sci. USA.*, 76, 5036-5040.
- Cleary, M.L., Haynes, J.R., Schon, E.A. and Lingrel, J.R. (1980).
Identification by nucleotide sequence analysis of a goat pseudo-globin gene. *Nucl. Acids. Res.*, 8, 4791-4802.
- Cleary, M.L., Schon, E.A. and Lingrel, J.B. (1981).
Related pseudogenes are the result of a gene duplication in the goat β -globin locus. *Cell*, 26, 181-190.
- Cleveland, D.W., Lopata, M.A., MacDonald, R.J., Cowan, N.J., Rutter, W.J. and Kirschner, M.W. (1980).
Number and evolutionary conservation of α - and β -tubulin and Cytoplasmic γ -actin genes using specific cloned cDNA probes. *Cell*, 20, 95-105.
- Coggins, L.W., Grindlay, G.J., Vass, K.J., Slater, A.A., Montague, P., Stinson, M.A. and Paul, J. (1980).
Repetitive DNA sequences near three human β -type globin genes. *Nucl. Acids. Res.*, 8, 3319-3333.
- Craik, C.S., Buchman, S.R. and Beychok, S. (1980).
Characterization of globin domains. Heme binding to the central exon product. *Proc. Natl. Acad. Sci. USA.*, 77, 1384-1388.
- Craik, C.S., Buchman, S.R. and Beychok, S. (1981).
O₂ binding properties of the product of the central exon of β -globin gene. *Nature*, 291, 87-90.
- Crick, F. (1979).
Split genes and RNA splicing. *Science*, 204, 264-271.
- Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M.L., Venta, P.J. and Tashian, R.E. (1982).
Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature*, 298, 297-300.
- Darnell, J.E. (1978).
Implications of RNA splicing in evolution of eukaryotic cells. *Science*, 202, 1257-1260.
- Davidson, E.H. and Posakony, J.W. (1982).
Repetitive sequence transcripts in development. *Nature*, 297, 633-635.
- Dayhoff, M.O. ed. (1972).
Atlas of Protein Sequence and Structure, vol. 5. Natn. Biomedical Res. Found. Silver Spring, Md.

- Deisseroth, A., Nienhuis, A., Lawrence, J., Giles, R., Turner, P. and Ruddle, F.H. (1978).
Chromosomal localisation of human β -globin gene on human chromosome 11 in somatic cell hybrids. *Proc. Natl. Acad. Sci. USA.*, 75, 1456-1460.
- Denhardt, D. (1966).
A membrane-filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.*, 23, 641-646.
- Devlin, R.B. and Emerson, C.P. Jr. (1978).
Coordinate regulation of contractile protein synthesis during myoblast differentiation. *Cell*, 13, 599-611.
- Dodgson, J.B., Strommer, J. and Engel, J.D. (1979).
Isolation of the chicken β -globin gene and a linked embryonic β -like globin gene from a chicken DNA recombinant library. *Cell*, 17, 879-887.
- Dolan, M., Sugarman, B.J., Dodgson, J.B. and Engel, J.D. (1981).
Chromosomal arrangement of the chicken β -type globin genes. *Cell*, 24, 669-677.
- Doolittle, W.F. (1978).
Genes in pieces : were they ever together? *Nature*, 272, 581-582.
- Durica, D.S., Schloss, J.A. and Crain, W.R. (1980).
Organisation of actin gene sequences in the sea urchin : molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin. *Proc. Natl. Acad. Sci. USA.*, 77, 5683-5687.
- 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.
- Engel, J.D. and Dodgson, J.B. (1980).
Analysis of the closely linked adult α -globin genes in recombinant DNAs. *Proc. Natl. Acad. Sci. USA.*, 77, 2596-2600.
- Engel, J.N., Gunning, P.W. and Kedes, L. (1981).
Isolation and characterisation of human actin genes. *Proc. Natl. Acad. Sci. USA.*, 78, 4674-4678.
- Englander, S.W. and Staley, R. (1969).
Measurement of the free and the H-bonded amides of myoglobin. *J. Mol. Biol.* 45, 277-295.
- Enquist, L. and Sternberg, N. (1979).
In vitro packaging of λ Dam vectors and their use in cloning DNA fragments. *Methods in Enzymology*, 68, 281-298.

- Firtel, R.A. (1981).
Multigene families encoding actin and tubulin. *Cell*, 24, 6-7.
- Firtel, R.A., Timm, R., Kimmel, A.R. and McKeown, M. (1979).
Unusual nucleotide sequences at the 5' end of actin genes in Dictyostelium discoidium. *Proc. Natl. Acad. Sci. USA.*, 76, 6206-6210.
- Fischman, D.A. and Zak, R. (1971).
Assembly of myofibrils in absence of protein synthesis. *J. Gen. Physiol.*, 57, 245-253.
- Flavell, R.A. (1982).
The mystery of the mouse α -globin pseudogene. *Nature*, 295, 370.
- Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982).
The complete nucleotide sequence of the chick α -actin gene and its evolutionary relationship to the actin gene family. *Nucl. Acids Res.*, 10, 3861-3876.
- Francki, R.I.B. and Randles, J.W. (1980).
Limited value of the thermal inactivation point, longevity in vitro and dilution endpoint as criteria for the characterization, identification and classification of plant virus. *Interviro*, 13, 91-98.
- Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980).
Molecular cloning and characterisation of the human β -like globin gene cluster. *Cell*, 19, 959-972.
- Froning, G.W., Daddario, J. and Hartung, T.E. (1968).
Color and myoglobin concentration in turkey meat as affected by age, sex and strain. *Poultry Sci.*, 47, 1827-1833.
- Fuchsman, W.H. and Appleby, C.A. (1979).
Separation and determination of the relative concentrations of the homogeneous components of soybean leghaemoglobin by isoelectric focussing. *Biochem. Biophys. Acta.*, 579, 314-324.
- Fyrberg, E.A., Bond, B., Hershey, N.D., Mixter, K.S. and Davidson, N. (1981).
The actin genes of Drosophila : protein coding regions are highly conserved but intron positions are not. *Cell*, 24, 107-116.
- Fyrberg, E.A., Kindle, K.L. and Davidson, N. (1980).
The actin genes of Drosophila : a dispersed multigene family. *Cell*, 19, 365-378.
- Fyrberg, E.A., Mahaffey, J.W., Bond, B.J. and Davidson, N. (1983).
Transcripts of the six Drosophila actin genes accumulate in a stage-and tissue-specific manner. *Cell*, 33, 115-123.

- Gallwitz, D. and Sures, I. (1980).
Structure of a split yeast gene : complete nucleotide sequence of the actin gene in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA., 77, 2546-2550.
- Geraci, G., Sada, A. and Cirotto, C. (1977).
Cooperative, low-molecular-weight dimeric myoglobins from the radular muscle of the gastropod mollusc Nassa mutabilis L. Eur. J. Biochem., 77, 555-560.
- Gilbert, W. (1978).
Why genes in pieces? Nature, 271, 501.
- Gilsin, V., Crkvenjakov, R. and Byns, C. (1974).
Ribonucleic acid isolated by caesium chloride centrifugation. Biochemistry, 13, 2633-2637.
- Gō, M. (1981).
Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature, 291, 90-92.
- Goff, S.P., Gilboa, E., Witte, O.N. and Baltimore, D. (1980).
Structure of the Abelson murine leukemia virus genome and the homologous cellular gene : studies with cloned viral DNA. Cell, 22, 777-785.
- Goodman, M., Moore, G.W. and Matsuda, G. (1975).
Darwinian evolution in the genealogy of haemoglobin. Nature, 253, 603-608.
- Grunstein M. and Hogness, D.S. (1975).
Colony hybridization : a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA., 72, 3961-3965.
- Gusella, J., Varsanyi-Breiner, A., Kao, F.-T., Jones, C., Puck, T.T., Keys, C., Orkin, S. and Housman, D. (1979).
Precise localisation of human β -globin gene complex on chromosome II. Proc. Natl. Acad. Sci. USA., 76, 5239-5243.
- Gwo-Shu, L., Lee, M.G.-S., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983).
Evolutionary history of a multigene family : an expressed human β -tubulin gene and three processed pseudogenes. Cell, 33, 477-487.
- Hamada, H., Petrino, M.G. and Kakunaga, T. (1982).
Molecular structure and evolutionary origin of human cardiac-muscle actin gene. Proc. Natl. Acad. Sci. USA., 79, 6465-6469.
- Hamer, D.H. and Leder, P. (1979).
Expression of the chromosomal mouse beta-major globin gene cloned in SV40. Nature, 281, 35-40.
- Hampe, A., Therwath, A., Soriano, P. and Galibert, F. (1981).
Nucleotide sequence analysis of a cloned duck beta-globin gene. Gene, 14, 11-21.

- Hardison, R.C., Butler, E.T., Lacy, E. and Maniatis, T. (1979).
The structure and transcription of four linked rabbit β -like globin genes. *Cell*, 18, 1285-1297.
- Hastings, K.E.M. and Emerson, C.P., Jr. (1982).
cDNA clone analysis of six coregulated mRNAs encoding skeletal muscle contractile proteins. *Proc. Natl. Acad. Sci. USA.*, 79, 1553-1557.
- Helfman, D.M., Feramisco, J.R., Fiddes, J.C., Thomas, G.P. and Hughes, S.H. (1983).
Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. *Proc. Natl. Acad. Sci. USA.*, 80, 31-35.
- Heywood, S.M. (1973).
Is there specificity in initiation of protein synthesis in eukaryotic cells? *Develop. Bio.*, 31, F1-F3.
- Heywood, S.M. and Kennedy, D.S. (1974).
Control of myoglobin synthesis during muscle development. *Developmental Biology*, 38, 390-393.
- Heywood, S.M., Kennedy, D.S. and Bester, A.J. (1974).
Separation of specific initiation factors involved in the translation of myosin and myoglobin mRNAs and isolation of a new RNA involved in transcription. *Proc. Natl. Acad. Sci. USA.*, 71, 2428-2431.
- Holmquist, R., Jukes, T.H., Moise, H., Goodman, M. and Moore, G.W. (1976).
The evolution of the globin family genes : concordance of stochastic and augmented maximum parsimony genetic distances for α -haemoglobin, β -haemoglobin and myoglobin phylogenies. *J. Mol. Biol.*, 105, 39-74.
- Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982).
Processed genes : a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. *Nature*, 296, 321-325.
- Hughes, S.E., Stubblefield, E., Payvar, F., Engel, J.D., Dodgson, J.B., Spector, D., Cordell, B., Schimke, R.T. and Varmus, H.E. (1979).
Gene localization by chromosome fractinnation : globin genes are on at least two chromosomes and three estrogen-inducible genes are on three chromosomes. *Proc. Natl. Acad. Sci. USA.*, 76, 1348-1352.
- Humphries, S.E., Whittall, R., Minty, A., Buckingham, M. and Williamson, R. (1981).
There are approximately 20 actin genes in the human genome. *Nucl. Acids Res.*, 9, 4895-4908.
- Hunt, T.L., Hurst-Calderone, S. and Dayhoff, M.O. (1978).
Atlas of Protein Sequence and Structure, 5, Suppl. 3. pp.229-251.
- Huxley, H.E. (1973).
Muscle contraction and cell motility. *Nature*, 244, 445-449.

- Hyldig-Nielsen, J.J., Jensen, E.Ø., Paludan, K., Paludan, K., Wiborg, O., Garrett, R., Jørgensen, P. and Marcker, K.A. (1982).
The primary structure of two leghaemoglobin genes from soybean.
Nucl. Acids Res., 10, 689-701.
- Jackson, D.A., Symons, R.H. and Berg, P. (1972).
Biochemical method for inserting new genetic information into DNA of SV40 : circular SV40 DNA molecules containing lambda phage genes and the galactose sperm of E.coli. Proc. Natl. Acad. Sci. USA., 69, 2904-2909.
- Jahn, C.L., Hutchinson, C.A., Phillips, S.J., Weaver, S., Haigwood, N.L., Voliva, C.F. and Edgell, M.H. (1980).
DNA sequence organisation of the β -globin complex in the Balb/c mouse. Cell, 21, 159-163.
- Jensen, E.Ø., Paludan, K., Hyldig-Nielsen, J.J., Jørgensen, P. and Marcker, K.A. (1981).
The structure of a chromosomal leghaemoglobin gene from soybean. Nature, 291, 677-679.
- Jeffreys, A.J. (1981).
Recent studies in gene evolution using recombinant DNA. In Genetic Engineering vol. 2. R. Williamson ed. Academic Press, London.
- Jeffreys, A.J. and Flavell, R.A. (1977).
A physical map of the DNA regions flanking the rabbit β -globin gene. Cell, 12, 429-439.
- Jeffreys, A.J., Craig, I.W. and Franke, U. (1979).
Localisation of the γ -, $A\gamma$ -, δ - and β -globin genes on the short arm of chromosome 11. Nature, 281, 606-608.
- 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., Barrie, P.A., Harris, S., Fawcett, D.H., Nugent, Z.J. and Boyd, A.C. (1982).
Isolation and sequence analysis of a hybrid δ -globin pseudogene from the brown lemur. J. Mol. Biol., 156, 487-503.
- Jeffreys, A.J., Wilson, V., Blanchetot, A., Weller, P., Van Kessel, A.G., Spurr, N., Solomon, E. and Goodfellow, P. (1984).
The human myoglobin gene : a third dispersed globin locus in the human genome. Nucl. Acids Research, 12, 3235-3243.
- Kagen, L.J. (1973).
In 'Myoglobin : biochemical, physiological and clinical aspects.' Columbia Series in Molecular Biology. Columbia University Press, New York.

- Kagen, L.J. and Christian, C.L. (1966).
Immunologic measurements of myoglobin in human adult and fetal skeletal muscle. *Amer. J. Physiol.* 211, 656-659.
- Kagen, L.J. and Linder, S. (1968).
Immunological studies of duck myoglobin. *Proc. Soc. Exp. Biol. Med.*, 128, 438-442.
- Karin, M. and Richards, R.I. (1982).
Human metallothionein genes - primary structure of the human metallothionein-III gene and a related processed gene. *Nature*, 299, 797-802.
- Katcoff, D., Nudel, U., Zevin-Sonkin, D., Carmon, Y., Shani, M., Lehrach, H., Frischauf, A.M. and Yaffe, D. (1980).
Construction of recombinant plasmids containing rat muscle actin and myosin light chain DNA sequences. *Proc. Natl. Acad. Sci. USA.*, 77, 960-964.
- Kates, J. (1970).
Transcription of the Vaccinia virus genome and the Occurrence of polyribonucleic acid sequences in messenger RNA. *Cold Spring Harb. Symp. Quant. Biol.*, 35, 743-752.
- Kendrew, J.C. (1963).
Myoglobin and the structure of proteins. *Science*, 139, 1259-1266.
- Kendrew, J.C., Dickerson, R.E., Strandberg, R.G., Hart, D.R., Davies, D.R., Phillips, D.C. and Shore, V.C. (1960).
Structure of myoglobin. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature*, 185, 422-427.
- Kendrew, J.C., Bodo, G., Dinzis, H.M., Parrish, R.G., Wyckoff, M. 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., Watson, B.E., Strandberg, R.E., Dickerson, D.C., 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.
- Kendrick-Jones, J., Szentkiralyi, E.M. and Szent-Gyorgyi, A.G. (1976).
Regulatory light chains in myosins. *J. Mol. Biol.*, 104, 747-775.
- Kindle, K.L. and Firtel, R.A. (1978).
Identification and analysis of *Dictyostelium* actin genes, a family of moderately repeated genes. *Cell*, 15, 763-778.
- King, M.C. and Wilson, A.C. (1975).
Evolution at two levels in humans and chimpanzees. *Science*, 188, 107-116.

- Kirby, K.S. (1957).
A new method for the isolation of DNAs : evidence on the nature of bonds between DNA and protein. *Biochem. J.*, 66, 495-504.
- Klein, R.D., Selsing, E. and Wells, R.D. (1980).
A rapid microscale technique for isolation of recombinant plasmid DNA suitable for restriction enzyme analysis. *Plasmid*, 3, 88-91.
- 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.
- Koppenheffer, T.L. and Read, K.R.H. (1969).
The adductor muscle myoglobins of the bivalve mollusc Mercenaria mercenaria L. *Comp. Biochem. Physiol.*, 31, 829-834.
- Lacy, E., Hardison, R.C., Quon, D. and Maniatis, T. (1979).
The linkage arrangement of four rabbit β -like globin genes. *Cell*, 18, 1273-1283.
- Lacy, E. and Maniatis, T. (1980).
The nucleotide sequence of a rabbit β -globin pseudogene. *Cell*, 21, 545-553.
- Laemmli, U.K. (1970).
Cleavage of structural proteins during the assembly of the head of bacteriophage T₄. *Nature*, 227, 680-685.
- Land, H., Crez, M., Hauser, H., Lindenmaier, W. and Schutz, G. (1983).
Synthesis of ds-cDNA involving addition of dCMP tails to allow cloning of 5'-terminal mRNA sequences. *Methods in Enzymology*, 100, 285-292.
- Lehmann, H. (1974).
In 'Man's Haemoglobins'. North Holland Publishing Co. Amsterdam. pp.89-90.
- Laki, K. (1971).
Regulatory proteins. In 'Contractile Proteins and muscle', edited by K. Laki. Marcel Dekker Inc., New York. pp.273-288.
- Lawrie, R.A. (1950).
Some observations on factors affecting myoglobin concentrations in muscle. *J. Agric. Sci.*, 40, 356-366.
- Lawrie, R.A. (1952).
Biochemical differences between red and white muscle. *Nature*, 170, 122-123.
- Lawrie, R.A. (1953a).
The activity of the cytochrome system in muscle and its relation to myoglobin. *Biochem. J.*, 55, 298-305.
- Lawrie, R.A. (1953b).
The relation of energy-rich phosphate in muscle to myoglobin and to cytochrome oxidase activity. *Biochem. J.*, 55, 305-309.

- Leder, A., Swan, D., Ruddle, F., D'Eustachio, P. and Leder, P. (1981).
Dispersion of α -like globin genes of the mouse to three different chromosomes. *Nature*, 293, 196-200.
- Leis, J.P. and Hurwitz, J. (1972).
Isolation and characterization of a protein that stimulates DNA synthesis from Avian Myeloblastosis Virus. *Proc. Natl. Acad. Sci. USA.*, 69, 2331-2335.
- Lemischka, I. and Sharp, P.A. (1982).
The sequences of an expressed rat α -tubulin gene and a pseudogene with an inserted repetitive element. *Nature*, 300, 330-335.
- Lauer, J., Shen, C-K.J. and Maniatis, T. (1980).
The chromosomal arrangement of human α -like globin genes : sequence homology and α -globin gene deletions. *Cell*, 20, 119-130.
- Lewis, M.S., Marumaya, K., Carroll, D.R., Kominz, D.R. and Laki, K. (1963).
Physical properties and polymerization reactions of native and inactivated G-actin. *Biochemistry*, 2, 34-39.
- Li, S.L. and Riggs, A. (1970).
The amino acid sequence of haemoglobin V from the lamprey, *Petromyzon marinus*. *J. Biol. Chem.*, 245, 6149-6169.
- Liebhauer, S.A., Goossens, M.J. and Kan, Y.W. (1980).
Cloning and complete nucleotide sequence of human 5'- α -globin gene. *Proc. Natl. Acad. Sci. USA.*, 77, 7054-7058.
- Lim, L. and Cannelakis, E.S. (1970).
Adenine rich polymer associated with rabbit reticulocyte mRNA. *Nature*, 227, 710-712.
- Lobban, P.E. and Kaiser, A.D. (1973).
Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.*, 78, 453-471.
- Loenen, A.A.M. and Brammar, W.J. (1980).
A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. *Gene*, 10, 249-259.
- Lueders, K., Leder, A., Leder, P. and Kuff, E. (1982).
Association between a transposed α -globin pseudogene and a retrovirus-like element in the Balb/c mouse genome. *Nature*, 295, 426-428.
- MacLeod, A.R. (1981).
Construction of bacterial plasmids containing sequences complementary to chicken α -tropomyosin mRNA. *Nucl. Acids Res.*, 9, 2675-2697.
- MacLeod, A.R. (1982).
Distinct α -tropomyosin mRNA sequences in chicken skeletal muscle. *Eur. J. Biochem.*, 126, 293-297.

- MacLeod, A.R., Karn, J. and Brenner, S. (1981).
Molecular analysis of the unc-54 myosin heavy-chain gene of Caenorhabditis elegans. Nature, 291, 386-390.
- Mandel, M. and Higa, A.J. (1970).
Calcium-dependent bacteriophage DNA infection. J. Mol. Biol., 53, 159-162.
- Manwell, C. (1963).
The chemistry and biology of haemoglobin in some marine clams. I. Distribution of the pigment and properties of the oxygen equilibrium. Comp. Biochem. Physiol., 8, 209-213.
- Maruyama, K. (1971).
Tropomyosin A. In 'Contractile Proteins and Muscle', edited by K. Laki. Marcel Dekker, New York, pp.289-301.
- Maxam, A.M. and Gilbert, W. (1980).
Sequencing end-labelled DNA with base-specific chemical cleavages. Meth. Enz., 65, 499-560.
- McKeown, M. and Firtel, R.A. (1981).
Differential expression and S' end-mapping of actin genes in Dictyostelium. Cell, 24, 799-807.
- McKeown, M., Taylor, W.C., Kindle, K.L., Firtel, R.A., Bender, W. and Davidson, N. (1978).
Multiple, heterogeneous actin genes in Dictyostelium. Cell, 15, 789-800.
- Medford, R.M., Wydro, R.M., Nguyen, H.T. and Nadal-Ginard, B. (1980).
Cytoplasmic processing of myosin heavy chain messenger RNA : evidence provided by using a recombinant DNA plasmid. Proc. Natl. Acad. Sci. USA., 10, 5749-5753.
- Millikan, G.A. (1939).
Muscle haemoglobin. Physiol. Rev. 19, 503-534.
- 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 characterization of a recombinant plasmid molecule containing a complementary DNA transcript of mouse α -actin mRNA. J. Biol. Chem., 256, 1008-1014.
- Morimoto, H., Lehmann, H. and Perutz, M.F. (1971).
Molecular pathology of human haemoglobin-stereochemical interpretation of abnormal oxygen affinities. Nature, 232, 408-413.
- Musin, B. (1969).
Myoglobin content in skeletal and heart muscles of dogs at different age periods. Byall. Eksper. Biol. Med., 67, 10-17.

- Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. and Ogata, K. (1982).
Molecular cloning and nucleotide sequences of the cDNAs to chicken skeletal muscle myosin two alkali lightchain mRNAs. *Nucl. Acids Res.*, 10, 6099-6110.
- 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.
- Nelson, J.A., Levy, J.A. and Leony, J.C. (1981).
Human placentas contain a specific inhibitor of RNA-directed DNA polymerase. *Proc. Natl. Acad. Sci. USA.*, 78, 2087-2104.
- Nevers, P. and Saedler, H. (1977).
Transposable genetic elements as agents of gene instability and chromosomal rearrangements. *Nature*, 268, 109-115.
- Ng, R. and Abelson, J. (1980).
Isolation and sequence of the gene for actin in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA.*, 77, 3912-3916.
- Nishioka, Y., Leder, A. and Leder, P. (1980).
Unusual α -globin-like gene that has cleanly lost both globin intervening sequences. *Proc. Natl. Acad. Sci.*, 77, 2806-2809.
- Nobbs, C.L., Watson, H.C. and Kendrew, J.C. (1966).
Structure of deoxymyoglobin : A crystallographic study. *Nature*, 209, 339-341.
- Nobrega, F.G. and Tzagoloff, A. (1980).
Assembly of the mitochondrial membrane systems - DNA sequence and organization of the cytochrome B gene in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, 255, 9828-9837.
- Nudel, U., Katcoff, D., Carmon, Y., Zevin-Sonkin, D., Levi, Z., Shaul, Y., Shani, M. and Yaffe, C. (1980).
Identification of recombinant phages containing sequences from different rat myosin heavy chain genes. *Nucl. Acids Res.*, 8, 2133-2146.
- Nudel, U., Katcoff, D., Zakut, R., Shani, M., Carmon, Y., Finer, M., Czosnek, H., Ginsburg, I. and Yaffe, D. (1982).
Isolation and characterization of rat skeletal muscle and cytoplasmic actin genes. *Proc. Natl. Acad. Sci. USA.*, 79, 2763-2767.
- Ohno, S. (1970).
Evolution by Gene Duplication. Springer-Verlag, Heidelberg.
- Ohno, S. and Morrison, M. (1966).
Multiple gene loci for monomeric haemoglobin of hagfish (*Eptatretus stoutii*). *Science*, 154, 1034-1036.

- Ordahl, C.P., Tilghman, F.M., Ovitt, C., Fornwald, J. and Largen, M.T. (1980).
Structure and developmental expression of the chick α -actin gene. *Nucl. Acids. Res.*, 8, 4948-5005.
- Orkin, S.H. (1978).
The duplicated human α -globin genes lie close together in cellular DNA. *Proc. Natl. Acad. Sci. USA.*, 75, 5950-5954.
- Palatnik, C.M., Storti, R.V. and Jacobson, A. (1979).
Fractionation and functional analysis of newly synthesised and decaying mRNAs from vegetative cells of Dictyostelium discoideum. *J. Mol. Biol.*, 128, 371-395.
- Patient, R.K., Elkington, J.A., Kay, R.M. and Williams, J.G. (1980).
Internal organization of the major adult α - and β -globin genes of Xenopus laevis. *Cell*, 21, 565-573.
- Perutz, M.F. (1965).
Structure and function of haemoglobin. I. A tentative atomic model of horse oxyhaemoglobin. *J. Mol. Biol.*, 13, 646-668.
- Perutz, M.F. (1970).
Stereochemistry of cooperative effects in haemoglobin. *Nature*, 228, 726-739.
- Perutz, M.F., Kendrew, J.C. and Watson, H.C. (1965).
Structure and function of haemoglobin. II. Some relations between polypeptide chain configuration and amino acid sequence. *J. Mol. Biol.*, 13, 669-678.
- Perutz, M.F., Muirhead, H., Cox, J.M. and Goaman, L.C.G. (1968).
Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution : The Atomic Model. *Nature*, 219, 131-139.
- Proudfoot, N.J. and Brownlee, G.G. (1976).
3' non-coding region sequences in eukaryotic mRNA. *Nature*, 263, 211-214.
- Proudfoot, N.J. and Maniatis, T. (1980).
The structure of a human α -globin pseudogene and its relationship to α -globin gene duplication. *Cell*, 21, 537-544.
- Ray, G.B. and Paff, G.H. (1930).
A spectrophotometric study of muscle haemoglobin. *Amer. J. Physiol.*, 94, 521-527.
- Read, K.R.H. (1967).
The radular muscle myoglobin of the gastropod mollusc Busycon caricum. Gmelin. *Comp. Biochem. Physiol.* 25, 81-96.
- Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977).
Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.*, 113, 237-251.

- Robert, B., Weydert, A., Caravatti, M., Minty, A., Cohen, P., Daubas, F., Gros, F. and Buckingham, M.E. (1982).
cDNA recombinant plasmid complementary to mRNAs for light chains 1 and 3 of mouse skeletal muscle myosin. *Proc. Natl. Acad. Sci. USA.*, 79, 2437-2441.
- Robinson, D. (1939).
The muscle haemoglobin of seals as an oxygen store in diving. *Science*, 90, 276-277.
- Romero-Herrera, A.E. and Lehmann, H. (1971).
Primary structure of human myoglobin. *Nature-New Biol.*, 232, 149-152.
- 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.
- Romero-Herrera, A.E., Lehmann, H., Joysey, K.A. and Friday, A.E. (1978).
On the evolution of myoglobin. *Proc. Roy. Soc. Ser. B.* 283, 61-163.
- Roninson, I.B. and Ingram, V.M. (1981).
cDNA sequence of a new chicken embryonic β -globin. *Proc. Natl. Acad. Sci. USA.*, 78, 4782-4785.
- Rossi-Fanelli, A., Antonini, E. and Povoledo, D. (1958).
Chemical and biochemical properties of a new type of myoglobin in molluscs. In 'Symposium on Protein Structure,' edited by Neuberger, A. Methuen and Co., London.
- Rourke, A.W. and Heywood, S.M. (1972).
Myosin synthesis and specificity of eukaryotic initiation factors. *Biochemistry*, 11, 2061-2066.
- Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979b).
Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature*, 280, 288-294.
- Sakano, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S. (1979a).
Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. *Nature*, 277, 627-633.
- Schaffner, W., Gross, K., Telford, J. and Birnstiel, M. (1976).
Molecular analysis of the histone gene cluster of Psammochinus miliaris : II. The arrangement of the five histone coding and spacer sequences. *Cell*, 8, 411-478.
- Schneiderman, L.J. (1962).
Myoglobin in the human foetus. *Nature*, 194, 191-192.
- Scholander, P.F. (1965).
Tension gradients accompanying accelerated oxygen transport in a membrane. *Science*, 149, 876-877.

- Schwartz, R.J., Haron, J.A., Katrina, N.R. and Dugaiczky, A. (1980).
Regulation of Muscle Differentiation : Cloning of sequences from
 α -actin messenger ribonucleic acid. *Biochemistry*, 19, 5883-5890.
- Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Givol, D., Katcoff, D.,
Carmon, Y., Reiter, J., Frischauff, A.N. and Yabbe, D. (1981).
Skeletal muscle actin mRNA. Characterization of the three prime
untranslated region. *Nucl. Acids. Res.*, 9, 579-589.
- Sidloil, U.R., Kleiman, L. and Schulman, H.M. (1978).
Biochemical evidence that leghaemoglobin genes are present in soybean
but not in Rhizobium genome. *Nature*, 273, 558-560.
- Smith, H.O. (1980).
Recovery of DNA from gels. *Methods in Enzymology*, 65, 371-380.
- Southern, E.M. (1975).
Detection of specific sequences among DNA fragments separated by
gel electrophoresis. *J. Mol. Biol.*, 98, 503-517.
- Storti, R.V., Horovitch, S.J., Scott, M.P., Rich, A. and Pardue, M.L.
(1978).
Myogenesis in primary cell cultures from Drosophila melanogaster :
protein synthesis and actin heterogeneity during development.
Cell, 13, 589-598.
- Sullivan, D., Brisson, N., Goodchild, B., Verma, D.P.S. and Thomas, D.Y.
(1981).
Molecular cloning and organization of two leghaemoglobin genomic
sequences of soybean. *Nature*, 289, 516-518.
- Sutcliffe, J.G. (1978).
pBR322 restriction map derived from the DNA sequence : accurate
DNA size markers up to 4361 nucleotide pairs long. *Nucl. Acids
Res.*, 5, 2721-2728.
- Takano, T. (1977).
Structure of myoglobin refined at 2.0Å⁰ resolution. *J. Mol. Biol.*,
110, 537-584.
- Tekeuchi, A. (1959).
Neuromuscular transmission of fish skeletal muscles investigated
with intracellular microelectrode. *J. Cell. Comp. Physiol.*, 54,
211-216.
- Terwilliger, R.C. and Read, K.R.H. (1971).
Oxygen equilibrium studies of the radular muscle myoglobins of the
gastropod molluscs, Buccinum undatum L. and Busycon canaliculatum
L. *Int. J. Biochem.*, 2, 253-258.
- Theorell, H. (1932).
Myoglobin I. Kristallisieren and Reinigung des Myoglobins sowie
vorläufige Mitteilung über sein Molekulargewicht. *Biochem. Z.*,
252, 1-10.

- Theorell, H. (1934).
Kristallinisches Myoglobin. V. Die Sauerstoffbindungskurve des Myoglobins. *Biochem. Z.*, 268, 73-81.
- Thompson, W.C., Buzash, E.A. and Heywood, S.M. (1973).
Translation of myoglobin messenger ribonucleic acid. *Biochemistry*, 12, 4559-4565.
- Tipler, T.D., Edwards, Y.H. and Hopkinson, D.A. (1978).
Developmental changes in protein profiles of human cardiac and skeletal muscle. *Annals. Human Genetics, London*, 41, 409-418.
- Tobin, S.L., Zulauf, E., Sanchez, F., Craig, E.A. and McCarthy, B.J. (1980).
Multiple actin related sequences in the Drosophila melanogaster genome. *Cell*, 19, 121-131.
- Twigg, A.J. and Sherrat, D. (1980).
Trans-complementable copy-number mutants of plasmid ColE1. *Nature*, 283, 216-218.
- Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J. and Goodman, H.M. (1977).
Rat insulin genes : Construction of plasmids containing coding sequences. *Science*, 196, 1312-1319.
- Umeda, P.K., Sinha, A.M., Jakovcic, S., Merten, S., Hsu, H-J., Subramanian, K.N., Zak, R. and Rabinowitz, M. (1981).
Molecular cloning of two fast myosin heavy chain cDNAs from chicken embryo skeletal muscle. *Proc. Natl. Acad. Sci. USA*, 78, 2843-2847.
- Van der Ploeg, L.H.T., Konings, A., Oort, M., Roos, D., Bernini, L., Flavell, R.A. (1980).
 γ - β -thalassaemia studies showing that deletion of the γ - and δ -genes influences β -globin gene expression in man. *Nature*, 283, 637-642.
- Vanin, E.F., Goldberg, G.I., Tucker, P.W. and Smithies, O. (1980).
A mouse β -globin-related pseudogene lacking intervening sequences. *Nature*, 286, 222-226.
- Villa-Komaraff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. (1978).
A bacterial clone synthesizing proinsulin. *Proc. Natl. Acad. Sci. USA*, 75, 3727-3731.
- Wahl, G.M., Stern, M. and Stark, G.R. (1976).
Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridisation by using dextran sulphate. *Proc. Natl. Acad. Sci. USA*, 76, 3683-3687.
- Weeds, A.G. and Lowey, S. (1971).
Structure of the myosin molecule II. The light chains of myosin. *J. Mol. Biol.*, 61, 701-725.

- Weller, P., Jeffreys, A.J., Wilson, V., Blanchetot, A. (1984).
Organisation of the human myoglobin gene. *EMBO J.* 3, 439-446.
- Whipple, G.H., Groth, A.H. and Robschey-Robbins, F.S. (1928).
Muscle haemoglobin concentration during growth as influenced by diet factors. *Amer. J. Physiol.*, 87, 185-188.
- White, J.M., Szur, L., Roberts, P., Lorkin, P.A. and Lehmann, H. (1973).
Heathrow- β G5 103 phenylalanine-leucine-new high affinity haemoglobin. *Brit. J. Haematology*, 25, 284-286.
- Wiborg, O., Hyldig-Nielsen, J.J., Jensen, E.Ø., Paludan, K. and Marcker, K.A. (1982).
The nucleotide sequence of two leghaemoglobin genes from soybean. *Nucl. Acids Res.*, 10, 3487-3494.
- Wilde, C.D., Growther, C.E., Cripe, T.P., Lee, M.G.S. and Cowan, N.J. (1982a).
Evidence that a human β -tubulin pseudogene is derived from its corresponding mRNA. *Nature*, 297, 83-84.
- Wilde, C.D., Crowther, C.E. and Cowan, N.J. (1982b).
Diverse mechanisms in the generation of human β -tubulin pseudogenes. *Science*, 217, 549-552.
- Williams, J.G. (1981).
The preparation and screening of a cDNA clone bank. In *Genetic Engineering* vol 1. R. Williamson ed. Academic Press, London.
- Williams, J.G. and Lloyd, M.M. (1979).
Changes in the abundance of polyA RNA during slime mould development measured using cloned molecular hybridization probes. *J. Mol. Biol.*, 129, 19-35.
- Wilson, A.C., Carlson, S.S. and White, T.J. (1977).
Biochemical evolution. *Ann. Rev. Biochem.*, 46, 573-639.
- Wittenberg, J.B. (1965).
Myoglobin-facilitated oxygen diffusion of oxygen. *J. Gen. Physiol.*, 49, 57-61.
- Wittenberg, J.B. (1970).
Myoglobin facilitated oxygen diffusion : Role of myoglobin in oxygen entry into muscle. *Physiol. Rev.*, 50, 559-636.
- Wittenberg, J.B., Appleby, C.A., Begersee, F.J. and Turner, G.L. (1975).
Leghaemoglobin - role of haemoglobin in nitrogen-fixing legume root module. *Ann. N.Y. Acad.*, 244, 28-34.
- Wood, K.O. and Lee, J.C. (1976).
Integration of synthetic globin genes into an *E.coli* plasmid. *Nucl. Acids Res.*, 3, 1961-1971.

- Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979).
Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus-2 fiber mRNA. *Cell*, 16, 851-861.
- Zambryski, P., Holsters, M., Kruger, K., Depicker, A., Schell, J., Van Montagu, M. and Goodman, H.M. (1980).
Tumour DNA structure in plant cells transformed by A. tumefaciens. *Science*, 209, 1385-1391.
- Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. and Wilson, A.C. (1980).
Rapid duplication and loss of genes coding for the α -chains of haemoglobin. *Proc. Natl. Acad. Sci. USA*, 77, 2158-2162.

ABSTRACT

THE ISOLATION OF A MAMMALIAN MYOGLOBIN GENE

D. WOOD

Myoglobin is the major haemoprotein of vertebrate muscle where it facilitates oxygen diffusion from the blood to the mitochondria and acts as an oxygen store. Elevated levels of myoglobin are found in diving mammals and birds.

Myoglobins and haemoglobins form part of the globin superfamily and are related evolutionarily. The ancestral globin gene underwent a duplication between 500 and 800 million years ago to give rise to monomeric myoglobins and the haemoglobin lines. The globin superfamily of proteins has been extensively characterised and sperm whale myoglobin was the first protein whose three-dimensional structure was determined. In addition the haemoglobin genes represent the best characterised multigene family. However the myoglobin gene has not been investigated. In order to understand the relationship of myoglobin genes in relation to other globin genes from an evolutionary and structural aspect, a mammalian myoglobin gene has been isolated.

After unsuccessful attempts to clone myoglobin messenger RNA from human muscle, grey seal muscle containing high levels of myoglobin was used to prepare polyadenylated mRNA. cDNA was prepared by reverse transcription of polyadenylated mRNA and cloned into a plasmid vector. 4% of cDNA recombinants contained myoglobin cDNA inserts. One clone, shown to contain only 3' non-translated mRNA sequences, was used to determine that grey seal myoglobin is coded by a single gene which is transcribed to give a 1400 nucleotide mRNA considerably longer than related haemoglobin mRNAs. This cDNA clone was also used to isolate a myoglobin genomic clone from grey seal DNA cloned into a λ bacteriophage vector. The 3' end of this gene has been completely sequenced and the gene is interrupted by an intervening sequence at codon 105, a position precisely homologous to that in α - and β -globin genes. The 3' exon of the grey seal myoglobin gene is remarkably long and comprises codons 106-153, which give an amino acid sequence identical to that of the harbour seal myoglobin, and 548 bp of untranslated sequence. This untranslated sequence contains a typical polyadenylation signal and is much longer than the corresponding sequences of α - and β -globin genes.